Chair Financière de la Cité - Toulouse School of Economics

Role and the Behaviour's Determinants of Long Term Investors

Activity Report, June 2009 – June 2010

Main goal of the Chair: The main goal of the research agenda is to analyze the social role of long term investors as well as their behavior's determinants. The determinants of the investors are studied by analyzing alternative preferences and changes in the fundamental state variables.

Main topics of the research agenda: The research agenda contains the following topics

- The role of long-term investors in the economy
- Portfolio choice when investors present a loss aversion
- Psychological determinants of investors' reactions to changing markets
- Volatility, markets uncertainty, and their impact on long term investors
- Inflation and assets' prices determination
- Ambiguity and risk aversion
- Payout policy and Ownership structure

Members of the TSE research team:

- Catherine Bobtcheff
- Frédéric Cherbonnier
- Patrick Fève
- Christian Gollier
- Alexander Guembel
- Astrid Hopfensitz
- Nour Meddahi
- Silvia Rossetto

Output of the first year of the Chair, June 2009- June 2010: The first year of the Chair was very productive. During this period, the members of the Toulouse team wrote eight academic papers and a book. This papers and book are presented below in connection to the different projects of the Chair. All the abstracts are given in Appendix I. The papers are in Appendix II while the book is given in Appendix III.

The role of long-term investors in the economy: The leader of the project, Christian Gollier, wrote three papers on the topic and a first draft of a book to be published by Princeton University. The titles of the papers and the book are given below:

- Gollier C., 2010, “Portfolio choices and asset prices: The comparative statics of ambiguity aversion”.

- Gierlinger J., Gollier C., 2010, “Socially efficient discounting under ambiguity aversion”.

- Gollier C., 2010, “Does flexibility enhance risk tolerance?”

- Gollier C., 2010, “Pricing the future: The economics of discounting and sustainable development”, to be published by Princeton University.

Christian Gollier presented part of his research in May 2010 to the clients of la Financière de la Cité. The talk was entitled

- “L’évaluation à court terme des stratégies d’investissement à long terme : L’allocation d’actifs au risque des nouvelles normes prudentielles”.

The slides of the presentation are given in Appendix IV.

Psychological determinants of investors’ reactions to changing markets:

- Hopfensitz A. and Wranik T., 2010, “How to adapt to changing markets: experience and personality in a repeated investment game”.

Inflation and assets’ prices determination:

- Fève P., Matheron J., and Sahuc J.-G., 2010, “Désinfation et chômage dans la zone euro : une analyse à l'aide d'un modèle VAR structurel”.

Ambiguity and risk aversion:

- Cherbonnier F. and Gollier C., 2010, “Decreasing aversion and prudence under ambiguity aversion”.

Payout policy and Ownership structure:

- Dhillon A. and Rossetto S., 2010, “Corporate control and multiple large shareholders”.

- Guembel A. and White L., 2010, “Good Cop, Bad Cop: Complementarities between debt and equity in disciplining management”

Appendix I

Abstracts of the papers

Gollier C., 2010, "Portfolio choices and asset prices: The comparative statics of ambiguity aversion"

Abstract: We investigate the comparative statics of "more ambiguity aversion" as defined by Klibanoff, Marinacci and Mukerji (2005) in the context of the static two-asset portfolio problem. It is not true in general that more ambiguity aversion reduces the demand for the uncertain asset. We exhibit some sufficient conditions to guarantee that, ceteris paribus, an increase in ambiguity aversion reduces the demand for the ambiguous asset, and raises the equity premium. For example, this is the case when the set of plausible distributions of returns can be ranked according to the monotone likelihood ratio order. We also show how ambiguity aversion distorts the price kernel in the alternative portfolio problem with complete markets for contingent claims.

Gierlinger J., Gollier C., 2010, "Socially efficient discounting under ambiguity aversion"

Abstract: In an economy with uncertain consumption growth and an ambiguity-averse representative agent, we provide sufficient conditions under which ambiguity aversion decreases the social discount rate. We identify two effects. The first is an ambiguity prudence effect, similar to the effect of prudence under expected utility. We show that it decreases the rate if and only if preferences satisfy decreasing ambiguity aversion. The second effect is observationnally equivalent to a deterioration of beliefs. This pessimism effect requires joint restrictions on the growth process and the agent's preferences to be signed. The calibration of the model suggests that the effect of ambiguity aversion on the way we should discount distant cash flows is potentially large.

Gollier C., 2010, "Does flexibility enhance risk tolerance?"

Abstract: The optimal decision under risk depends upon the ability of the decision makers to adapt their actions to the state of nature ex post. We examine two choice problems, one in which agents select their action ex post, and one in which they must commit on their action ex ante. Contrary to the intuition, it is not true in general that agents are more tolerant to risk in the flexible context than in the rigid one. We provide some sufficient conditions to guarantee that the optimal risk exposure is larger in the flexible context. We apply these results to examine various questions. In particular, we examine the effect of housing and labour markets rigidities and of rigid long-term saving plans on the demand for equity. We also compare the optimal portfolios in continuous and discrete time.

Hopfensitz A. and Wranik T., 2010, "How to adapt to changing markets: experience and personality in a repeated investment game"

Abstract: Investment behavior is traditionally investigated with the assumption that it is on average advantageous to invest. However, this may not always be the case. In this paper, we experimentally studied investment choices made by students and financial professionals facing alternately an advantageous and disadvantageous environment in a multi-round investment game. Expected returns from investment in the advantageous environment were higher than a safe alternative, while expected returns were lower in the disadvantageous environment. We investigate how experience and personality are related to choices. Investment behavior does not differ dependent on expected returns and professionals do not significantly differ from students. Personality predicts behavior and in particular we observe that openness to experience was an asset in unfavorable markets, leading to reduced risk taking.

Fève P., Matheron J., and Sahuc J.-G., 2010, "Désinfation et chômage dans la zone euro: une analyse à l'aide d'un modèle VAR structurel"

Abstract: The present paper investigates the dynamic effects of disinflation shocks for a number of real macroeconomic variables in the euro area. Using structural VARs, we identify disinflation shocks as the only shocks that can exert a long-run effect on inflation as well as other nominal variables cointegrating with inflation. These shocks are found to generate large recessionary effects, notably when it comes to investment, and triggers a persistent rise in unemployment and in the real interest rate. The analysis is complemented by computing inefficiency measures on goods and labor markets. We show that, after a disinflation shock, inefficiencies in the labor market seem to prevail. These conclusions are robust to modifications of our baseline identification scheme.

Cherbonnier F. and Gollier C., 2010, "Decreasing aversion and prudence under ambiguity aversion"

Abstract: We examine the conditions under which a risk-averse and ambiguity-averse agent exhibits some standard preference properties, as decreasing aversion and prudence. We explore two decision criteria (max-min and smooth ambiguity aversion), and two decision problems (acceptance of a lottery and the portfolio choice problem).

Dhillon A. and Rossetto S., 2010, "Corporate control and multiple large shareholders"

Abstract: Recent evidence on financial structures within forms suggests that concentrated ownership in the form of several large shareholders, rather than dispersed ownership, is the norm. This is puzzling given that often the stakes are too big for optimal diversification and too small to guarantee control. This paper attempts to provide an explanation. We consider a setting where multiple shareholders have endogenous conflicts of interest depending on the size of their stake. Such conflicts arise because larger shareholders tend to be less well diversified and would therefore prefer the firm to pursue more conservative investment policies, while dispersed shareholders prefer high risk, high return policies. A second blockholder (or more) can mitigate the conflict by shifting the voting outcome more towards

the dispersed shareholders' preferred investment policy and this raises the share price. Having a larger stake increases the ability to change decisions: however, it also changes the preferences of the shareholders, reducing the incentive to buy larger shares. The main contribution of the paper is to show the conditions under which blockholder equilibria exist when there are endogenous conflicts of interest. The model shows how different ownership structures affect firm value and IPO's underpricing.

Guembel A. and White L., 2010, “Good Cop, Bad Cop: Complementarities between debt and equity in disciplining management”

Abstract: In this paper we demonstrate an inherent conflict that can arise in a firm between inducing ex ante efficient monitoring and liquidation decisions by outside claimholders. This tension arises because the choice of liquidation decision when firm prospects are uncertain will influence incentives for monitoring to produce information about firm prospects. We show that when high levels of outside monitoring are desirable in order to induce managerial effort, it can be useful to follow an inefficient liquidation policy, because this will provide greater incentives for the monitor. This result in turn has implications for firm capital structure: the quantity of information generated about firm prospects - and hence firm value - can be improved by splitting a firm’s cash flow into a ‘safe’ claim (debt) and a ‘risky’ claim (equity) rather than selling a single claim, precisely because of the ex post conflicts of interest between claimholders that this creates. This generates a partial answer to the puzzle raised by Tirole (2001) as to why firms issue multiple securities when this leads to ex post conflicts of interest.

Appendix II

The papers

Portfolio choices and asset prices: The comparative statics of ambiguity aversion

Christian Gollier[1]
Toulouse School of Economics (LERNA and IDEI)

June 30, 2010

[1]The first version of this paper was entitled "Does ambiguity aversion reinforce risk aversion? Applications to portfolio choices and asset prices". The author thanks Markus Brunnermeier, Alain Chateauneuf, Ed Schlee, Jean-Marc Tallon, Thomas Mariotti, Jean Tirole, François Salanié, three anonymous referees and seminar participants at Paris 1, Georgia State, Helsinki, Princeton, Arizona State and Toulouse for helpful comments. This research was supported by the Chair of the Financière de la Cité at TSE, and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 230589.

Abstract

We investigate the comparative statics of "more ambiguity aversion" as defined by Klibanoff, Marinacci and Mukerji (2005) in the context of the static two-asset portfolio problem. It is not true in general that more ambiguity aversion reduces the demand for the uncertain asset. We exhibit some sufficient conditions to guarantee that, ceteris paribus, an increase in ambiguity aversion reduces the demand for the ambiguous asset, and raises the equity premium. For example, this is the case when the set of plausible distributions of returns can be ranked according to the monotone likelihood ratio order. We also show how ambiguity aversion distorts the price kernel in the alternative portfolio problem with complete markets for contingent claims.

Keywords: Smooth ambiguity aversion, monotone likelihood ratio, equity premium, portfolio choice, price kernel, central dominance.

1 Introduction

In this paper, we examine the standard static portfolio problem with one safe asset and one uncertain asset. The investor perceives ambiguity about what the true probability distribution for the excess return of the uncertain asset is. This ambiguity is expressed by a second order prior probability distribution over the set of plausible (first order) distributions of the excess return. Following Segal (1987) and Klibanoff, Marinacci and Mukerji (2005) (hereafter KMM), we introduce ambiguity attitude by relaxing reduction of first and second order probabilities. In words, the investor does not evaluate in the same way asset 1 yielding a return of 20% with probability p and asset 2 yielding a return of 20% with an unknown probability whose expectation is p, unlike in the standard Bayesian expected utility framework. Following KMM, we assume that the investor is ambiguity averse in the sense that he dislikes any mean-preserving spread in the space of first order probability distributions of excess returns. For example, he prefers asset 1 to asset 2. KMM proposes a nice decision criteria called "smooth ambiguity aversion" to characterize such preferences under uncertainty. For a given portfolio allocation, the ex ante welfare of the investor is measured by computing the (second order) expectation of a concave function ϕ of the (first order) expected utility u of final consumption conditional to each plausible distribution of the excess return. As usual, the concavity of the utility function u expresses risk aversion in the special case of risky acts, i.e. acts entailing consequences whose plausible probability distribution is unique. When ϕ is linear, we are back to the standard expected utility model in which the uncertain context can be reduced to a single compound probability distribution. When ϕ is concave, the investor is ambiguity averse, and the reduction of compound distributions does not hold.

Interestingly enough, KMM also define the comparative notion of "more ambiguity aversion". Consider two agents, respectively with function ϕ_1 and ϕ_2, who have the same beliefs expressed by the set of first and second order probability distributions. Suppose also that they have the same utility function u to evaluate risky acts. Agent ϕ_2 is more ambiguity averse than agent ϕ_1 if ϕ_1 prefers an uncertain act over a pure risky one whenever ϕ_2 does so. This is true if and only if function ϕ_2 is more concave than function ϕ_1, in the sense of Arrow-Pratt: $-\phi_2''/\phi_2'$ is uniformly larger than $-\phi_1''/\phi_1'$. In this paper, we examine the effect of such an increase in ambiguity aversion on the optimal demand for the uncertain asset. By doing so, we fix the first and second order beliefs, and the utility function u. We examine the effect of a concave transformation of function ϕ on the optimal portfolio.

The KMM model has two attractive features in comparison to other models of ambiguity such as the pioneering maxmin expected utility (or multiple-prior) model of Gilboa and Schmeidler (1989). First, it organizes a nice separation between ambiguity aversion and ambiguity, i.e., between tastes and beliefs. Without this feature, we could not perform the comparative statics of more ambiguity aversion. Second, the KMM model applies the expected utility machinery sequentially, first on order probability distributions, and then on the second order distribution. This allows us to exploit the huge armory of techniques amassed over the years to tackle questions involving risk under the expected utility framework to the analysis of problems involving non-expected utility involving ambiguity. This paper illustrates this point.

This question of the comparative statics of ambiguity aversion for the portfolio problem is parallel to the one of risk aversion. Since Arrow (1963) and Pratt (1964), it is well known that an increase in risk aversion reduces the demand for the risky asset. It is therefore quite surprising that, as we show in this paper, it is not true in general that more ambiguity aversion reduces the demand for the ambiguous asset. For cleverly chosen - but still not spurious - set of priors for the return of the risky asset, we show that the introduction of ambiguity aversion *increases* the investor's demand for the ambiguous asset. The intuition for why such counterexamples may exist can be explained as follows. The first-order condition of the portfolio problem with ambiguity aversion can be rewritten to take the form that it would take under expected utility, but with a distortion in the way the different first order probability distributions on excess return are compounded. Under expected utility, i.e., ambiguity neutrality, the compounding is made by using the true second order probability distribution. Under ambiguity aversion, this second order probability distribution is distorted by putting more weight on the plausible second order distributions yielding a smaller expected utility, as first observed by Taboga (2005). In spite of the fact that the ambiguity averse investor's beliefs cannot be reduced to a single compound probability distribution over excess returns, the introduction of ambiguity aversion is observationally equivalent to the effect of distorting the compound distribution used by the ambiguity-neutral investor. This distortion is pessimistic. It is well known from expected utility theory that pessimistic deteriorations in beliefs do not always reduce the demand for the risky asset.[1] As for Giffen goods in consumer theory, this deterioration in the terms of trade yields a wealth effect that may raise the asset demand.

The main objective of the paper is to characterize conditions under which more ambiguity aversion reduces the optimal exposure to uncertainty. This can be done by restricting either the set of utility functions and/or the set of possible priors. If we assume that the set of priors can be ranked according to the first-degree or second-degree stochastic dominance orders (FSD/SSD),

[1] See Rothschild and Stiglitz (1971), Fishburn and Porter (1976), Meyer and Ormiston (1985), Hadar and Seo (1990), Gollier (1995), Eeckhoudt and Gollier (1995), Abel (2002) and Athey (2002), and the bibliographical references in these papers.

we exhibit some simple sufficient conditions on the utility function to obtain an unambiguous comparative static property of the introduction of ambiguity aversion. These results are derived by using the following technique. It happens that any increase in ambiguity aversion deteriorates the observationally equivalent second order distribution in a very specific way. It transfers more weight on the worse priors in the sense of the monotone likelihood ratio (MLR) order. This puts a very specific structure to the notion of pessimism entailed by more ambiguity aversion. For example, if the plausible priors can be order according to FSD, their compound through the MLR deteriorated second order probability distribution generated by more ambiguity aversion entails a FSD deterioration of the behaviorally equivalent changes of beliefs under expected utility. This implies that, under this assumption, the following two problems are linked : 1) under the EU model, what are the conditions on the utility function u that guarantee that any FSD deterioration in the distribution of excess return of the risky asset reduces the demand for it; and 2) in the KMM model, what are the conditions that guarantee that any increase in ambiguity aversion reduces the demand for the ambiguous asset? The same property holds when replacing FSD by SSD. More sophisticated methods are required when considering other stochastic orders to rank plausible priors for the excess return. Let us just mention at this stage the result that is easiest to express: if the plausible priors can be ranked according to MLR (a special case of FSD), then it is always true that more ambiguity aversion reduces the demand for the ambiguous asset.

It is easy to translate these results about the effect of comparative ambiguity aversion on the demand for the ambiguous asset into its effect on the equity premium. Thus, our work is related to the recent developments about the effect of ambiguity aversion on the equity premium. Ju and Miao (2009) and Collard, Mukerji, Sheppard and Tallon (2009) examine a dynamic infinite-horizon portfolio problem in which the representative investor exhibits smooth ambiguity aversion and faces time-varying ambiguity about the second order distribution of the plausible probability distributions of consumption growth. These two papers consider different sets of risk-ambiguity attitudes (u, ϕ), and different stochastic processes for the first and second order probability distributions. They both use numerical analyses to solve the calibrated dynamic portfolio problem. They both conclude that ambiguity aversion raises the equity premium. Our paper demonstrates that these results are specific to the calibration under scrutiny. Other papers conclude in the same direction, but using other decision criterion for ambiguity aversion, either maxmin expected utility, Choquet expected utility, or robust control theory.[2]

The portfolio problem and two illustrations are presented in Section 2. Our main results are presented in Section 3 in which we derive sufficient conditions for the comparative statics of more ambiguity aversion. In Section 4, we examine a Lucas economy with a representative agent facing ambiguous state probabilities. We show how the ambiguity aversion of the representative agent affects the equity premium, the price kernel of the economy, and individual asset prices.

2 The smooth ambiguity model applied to the portfolio problem

Our model is static with two assets. The first asset is safe with a rate of return that is normalized to zero. The risky asset has a return x whose distribution is ambiguous in the sense that it is sensitive to some parameter θ whose true

[2]See for example Dow and Werlang (1992), Chen and Epstein (2002), Epstein and Wang (1994), Hansen, Sargent and Tallarini (1999), Maenhout (2004), Hansen and Sargent (2008).

value is unknown. The investor is initially endowed with wealth w_0. If he invests α in the risky asset, his final wealth will be $w_0 + \alpha x$ conditional to a realized return x of the risky asset.

The ambiguity of the uncertain asset is characterized by a set $\Pi = \{F_1, ..., F_n\}$ of plausible cumulative probability distributions for $\widetilde{x}$. Let $\widetilde{x}_\theta$ denote the random variable distributed as F_θ. We suppose that the support of all priors are bounded in $[x_-, x_+]$, with $x_- < 0 < x_+$. Based on his subjective information, the investor associates a second order probability distribution $(q_1, ..., q_n)$ over the set of priors Π, with $\Sigma_{\theta=1}^n q_\theta = 1$, where $q_\theta \geq 0$ is the probability that F_θ be the true probability distribution of excess returns. We hereafter denote $\widetilde{\theta}$ for the random variable $(1, q_1; 2; q_2; ...; n, q_n)$. Following Klibanoff, Marinacci and Mukerji (2005), we assume that the preferences of the investor exhibit smooth ambiguity aversion. For each plausible probability distribution F_θ, the investor computes the expected utility $U(\alpha, \theta) = Eu(w_0 + \alpha\widetilde{x}_\theta) = \int u(w_0 + \alpha x)dF_\theta(x)$ conditional to F_θ being the true distribution. We assume that u is increasing and concave, so that $U(., \theta)$ is concave in the investment α in the ambiguous asset, for all θ. Ex ante, for a given portfolio allocation α, the welfare of the agent is measured by $V(\alpha)$ with

$$V(\alpha) = \phi^{-1}\left(\sum_{\theta=1}^{n} q_\theta \phi(U(\alpha, \theta))\right) = \phi^{-1}\left(\sum_{\theta=1}^{n} q_\theta \phi\left(Eu(w_0 + \alpha\widetilde{x}_\theta)\right)\right),$$

$V(\alpha)$ can be interpreted as the certainty equivalent of the uncertain conditional expected utility $U(\alpha, \widetilde{\theta})$. The shape of ϕ describes the investor's attitude towards ambiguity. A linear ϕ means that the investor is neutral to ambiguity, and that he can reduce compound probability distributions to a single one $\Sigma_\theta q_\theta F_\theta$. On the contrary, a concave ϕ is synonymous of ambiguity aversion in the sense that the DM dislikes any mean-preserving spread of the conditional expected utility $U(\alpha, \widetilde{\theta})$.

An interesting particular case arises when the absolute ambiguity aversion $\eta(U) = -\phi''(U)/\phi'(U)$ is constant, so that $\phi(U) = -\eta^{-1}\exp(-\eta U)$. As proved by Klibanoff, Marinacci and Mukerji (2005), the ex ante welfare $V(\alpha)$ essentially exhibits a maxmin expected utility functional $V^{MEU}(\alpha) = \min_\theta Eu(w_0 + \alpha\widetilde{x}_\theta)$ à la Gilboa and Schmeidler (1989) when the degree η of absolute ambiguity aversion tends to infinity.

The optimal portfolio allocation α^* maximizes the ex ante welfare of the investor $V(\alpha)$. Because ϕ is increasing, α^* is the solution of the following program:

$$\alpha^* \in \arg\max_\alpha \quad \sum_{\theta=1}^{n} q_\theta \phi\left(Eu(w_0 + \alpha\widetilde{x}_\theta)\right). \tag{1}$$

If ϕ and u are strictly concave, the objective function is concave in α and the solution to program (1), when it exists, is unique. Let us observe that the demand for the ambiguous asset shares its sign with the equity premium $E\widetilde{x} = \Sigma_\theta q_\theta E\widetilde{x}_\theta$. All proofs are relegated to the Appendix.

Lemma 1 *The demand for the ambiguous asset is positive (zero/negative) if the equity premium is positive (zero/negative).*

This means that ambiguity aversion, as risk aversion, has a second order nature, as defined by Segal and Spivak (1990). As soon as the equity premium is positive, the demand for the ambiguous asset is positive, independent of the degree of ambiguity surrounding the distribution of returns. We hereafter assume that the equity premium is positive, so that α^* is positive.

In the remainder of this section, we examine two illustrations. Consider first the following special case in which an analytical solution can be found for α^*:

- Priors are normally distributed with the same variance σ^2, and with $E\widetilde{x}_\theta = \theta : \widetilde{x}_\theta \sim N(\theta, \sigma)$;[3]

- The ambiguity on the equity premium θ is itself normally distributed: $\widetilde{\theta} \sim N(\mu, \sigma_0)$;

- The investor's preferences exhibit constant absolute risk aversion: $u(z) = -A^{-1} \exp -Az \in \mathbb{R}_-$;

- The investor's preferences exhibit constant relative ambiguity aversion: $\phi(U) = -(-U)^{1+\gamma}/(1+\gamma)$. This function is increasing in $\mathbb{R}_-$ and is concave in this domain if γ is positive.

As is well-known, the normality of the priors and the constancy of absolute risk aversion implies that the Arrow-Pratt approximation is exact.[4] This implies in turn that

$$U(\alpha, \theta) = -A^{-1} \exp -A \left(w_0 + \alpha\theta - 0.5A\alpha^2\sigma^2 \right). \tag{2}$$

Because $\phi(U)$ is an exponential function of θ, and because $\widetilde{\theta}$ is normally distributed, the same trick can be used to compute $E\phi$. It yields

$$V(\alpha) = -A^{-1} \exp -A(w_0 + \alpha\mu - 0.5A\alpha^2(\sigma^2 + (1+\gamma)\sigma_0^2)). \tag{3}$$

The optimal demand for the risky asset is thus equal to

$$\alpha^* = \frac{\mu}{A(\sigma^2 + (1+\gamma)\sigma_0^2)}. \tag{4}$$

We see that under ambiguity ($\sigma_0^2 > 0$), the demand for the risky asset is decreasing in the relative degree γ of ambiguity aversion of the investor. This exponential-power specification for (u, ϕ) differs from the other three papers

[3] It is easy to extend this to the case of an ambiguous variance.

[4] For a simple proof, see for example Gollier (2001, page 57).

existing on this topic. Taboga (2005) examines an exponential-exponential specification. Ju and Miao (2009) used a power-power specification, whereas Collard, Mukerji, Sheppard and Tallon (2009) used a power-exponential specification. None of these three alternative problems can be solved analytically.

Consider alternatively the following counterexample such that

- $n = 2$, $\widetilde{x}_1 \sim (-1, 2/10; -0.25, 3/20; 0.75, 7/20; 1.25, 3/10)$ and $\widetilde{x}_2 \sim (-1, 1/5; 0, 1/5; 1, 3/5)$;
- $q_1 = 5\%$ and $q_2 = 95\%$;
- $u(z) = \min(z, 3 + 0.3(z - 3))$ and $w_0 = 2$;
- $\phi(U) = -\eta^{-1} \exp(-\eta U)$.

It is easy to check that $\widetilde{x}_1$ is riskier than $\widetilde{x}_2$ in the sense of Rothschild and Stiglitz (1970). We solved this problem numerically. Below a minimum threshold around 20 for the degree η of ambiguity aversion, the optimal holding of the ambiguous asset equals $\alpha^* = 1$. However, above this threshold, the introduction of ambiguity aversion *increases* α^* above the optimal investment of the ambiguity-neutral investor. For example, α^* equals 1.204 when $\eta = 50$. When η tends to infinity, the optimal investment in the risky asset tends to $\alpha^* = 4/3$, which is the optimal holding of the ambiguous asset for an ambiguity-neutral investor who believes that the distribution of excess return is $\widetilde{x}_1$ with certainty. In terms of portfolio allocation, it is observationally equivalent to increase absolute ambiguity aversion from zero to infinity or to replace beliefs $\widetilde{y}_1 \sim (\widetilde{x}_1, 5\%; \widetilde{x}_2, 95\%)$ by $\widetilde{y}_2 \sim (\widetilde{x}_1, 100\%)$ under expected utility. Notice that because $\widetilde{x}_1$ is riskier than $\widetilde{x}_2$, the extreme belief $\widetilde{y}_2$ is riskier than $\widetilde{y}_1$ in the sense of Rothschild and Stiglitz. This example illustrates the fact – first observed by Rothschild and Stiglitz (1971) – that it is not true in general that a riskier distribution of excess return reduces the demand for the risky asset in the expected utility model.

3 Effect of an increase of ambiguity aversion

The beliefs are represented by the set of priors $(\widetilde{x}_1, ..., \widetilde{x}_n)$ of the excess return of the risky asset, together with the second order distribution $(q_1, ..., q_n)$ on these priors. We compare two agents with the same beliefs and the same concave utility function u, but with different attitudes toward ambiguity represented by concave functions ϕ_1 and ϕ_2. The demand for the risky asset by agent ϕ_1 is expressed by α_1^* which must satisfy the following first-order condition:

$$\sum_{\theta=1}^{n} q_\theta \phi_1'(U(\alpha_1^*, \theta)) E\widetilde{x}_\theta u'(w_0 + \alpha_1^* \widetilde{x}_\theta) = 0. \tag{5}$$

Following Klibanoff, Marinacci and Mukerji (2005), we assume that the agent with function ϕ_2 is more ambiguity averse than agent ϕ_1 in the sense that there exists an increasing and concave transformation function k such that $\phi_2(U) = k(\phi_1(U))$ for all U in the relevant domain. We would like to characterize conditions under which the more ambiguity averse agent ϕ_2 has a smaller demand for the risky asset than agent ϕ_1: $\alpha_2^* \leq \alpha_1^*$. This would be the case if and only if

$$\sum_{\theta=1}^{n} q_\theta \phi_2'(U(\alpha_1^*, \theta)) E\widetilde{x}_\theta u'(w_0 + \alpha_1^* \widetilde{x}_\theta) \leq 0. \tag{6}$$

We would like to find conditions under which it is always true that (5) implies (6). Notice that this condition can be rewritten as

$$E\widetilde{y}_1 u'(w_0 + \alpha_1^* \widetilde{y}_1) = 0 \implies E\widetilde{y}_2 u'(w_0 + \alpha_1^* \widetilde{y}_2) \leq 0, \tag{7}$$

where $\widetilde{y}_i$ is a compound random variable which equals $\widetilde{x}_\theta$ with probability $\widehat{q}^i_\theta$, $\theta = 1, ..., n$, such that

$$\widehat{q}^i_\theta = \frac{q_\theta \phi'_i(U(\alpha^*_1, \theta))}{\sum_{t=1}^n q_t \phi'_i(U(\alpha^*_1, t))}. \tag{8}$$

Notice that the left equality in (7) can be interpreted as the first-order condition of the problem $\max_\alpha Eu(w_0 + \alpha\widetilde{y}_1)$ of an expected-utility-maximizing investor whose beliefs are represented by $\widetilde{y}_1 \sim (\widetilde{x}_1, \widehat{q}^1_1; ...; \widetilde{x}_n, \widehat{q}^1_n)$. Thus, the ambiguity averse agent ϕ_1 behaves as an EU-maximizing agent who would distort his second order beliefs from $(q_1, ..., q_n)$ to the "observationally equivalent probability distribution" $\widehat{q}^1 = (\widehat{q}^1_1, ..., \widehat{q}^1_n)$. The distortion factor $\phi'_1(U(\alpha^*_1, \theta))/\Sigma_t q_t \phi'_1(U(\alpha^*_1, t))$ is a Radon-Nikodym derivative, and the probability distribution $\widehat{q}^1$ is analogous to the risk-neutral probability distribution used in the theory of finance. Notice that the distortion functional described by equation (8) is endogenous, as it depends upon the portfolio allocation α^*_1 selected by the agent. The right inequality in (7) just means that shifting beliefs from $\widetilde{y}_1$ to $\widetilde{y}_2$ reduces the ambiguity-neutral investor's holding of the asset. These findings are summarized in the following lemma.

Lemma 2 *The change in preferences from (u, ϕ_1) to (u, ϕ_2) reduces the demand for the ambiguous asset if the EU agent with utility function u reduces his demand for the risky asset when his beliefs about the excess return shift from $\widetilde{y}_1 \sim (\widetilde{x}_1, \widehat{q}^1_1; ...; \widetilde{x}_n, \widehat{q}^1_n)$ to $\widetilde{y}_2 \sim (\widetilde{x}_1, \widehat{q}^2_1; ...; \widetilde{x}_n, \widehat{q}^2_n)$, where $\widehat{q}^i_\theta$ is defined by (8).*

This result was initially due to Taboga (2005). It precisely expresses the observational equivalence property that we already encountered in the counterexample presented in the previous section. It provides a test to determine whether more ambiguity aversion reduces demand. Observe that this test relies on two reduced probability distribution $\widetilde{y}_1$ and $\widetilde{y}_2$. However, it is not true that the ambiguity averse investor (u, ϕ_1) uses the corresponding reduced probability distribution $\widetilde{y}_1$ to evaluate the optimality of the different feasible portfolios. If he would do so, he would reevaluate the distribution of $\widetilde{y}_1$ for each portfolio, since vector $\widehat{q}_1$ is a function of α. In the smooth ambiguity aversion model, beliefs cannot be reduced to a single probability distribution on the state payoffs. But this lemma builds a bridge between the comparative statics of increased ambiguity aversion and the one of changes in risk in the classical EU model.

Let us examine how does changing function ϕ_1 into ϕ_2 modify the observationally equivalent probability distribution of the excess return. A first answer to this question is provided by the following lemma.

Lemma 3 *The following two conditions are equivalent:*

1. *Agent ϕ_2 is more ambiguity averse than agent ϕ_1;*
2. *Beliefs $\widehat{q}^2$ is dominated by $\widehat{q}^1$ in the sense of the monotone likelihood ratio order, i.e., $\widehat{q}^2_\theta / \widehat{q}^1_\theta$ is decreasing in θ, whenever $U(\alpha^*_1, 1) \leq U(\alpha^*_1, 2) \leq ... \leq U(\alpha^*_1, n)$.*

An increase in ambiguity aversion has an effect on demand that is observationally equivalent to a MLR-dominated shift in the prior beliefs. In other words, it distorts beliefs by favoring the worse priors in a very specific sense: if agent ϕ_1 prefers prior $\widetilde{x}_\theta$ over prior $\widetilde{x}_{\theta'}$, then, compared to agent ϕ_1, the more ambiguity averse agent ϕ_2 increases the distorted probability $\widehat{q}^2_{\theta'}$ relatively more than the probability $\widehat{q}^2_\theta$. Lemma 3 provides a justification to say that, in the case of the portfolio problem, more ambiguity aversion is observationally equivalent to more pessimism, i.e., to a MLR deterioration of beliefs. This result is central to prove our next proposition, in which we consider three dominance orders: first-degree stochastic dominance (FSD), second-degree stochastic dominance (SSD), and Rothschild and Stiglitz's increase in risk (IR).

Proposition 1 *Let D be one of the following three stochastic orders: FSD, SSD or IR. Suppose that $E\widetilde{x} > 0$, and that $(\widetilde{x}_1, ..., \widetilde{x}_n)$ can be ranked according to the stochastic order D . It implies that an increase in ambiguity aversion deteriorates the observationally equivalent probability distribution of the excess return in the sense of the stochastic order D: If $\exists k$ concave: $\phi_2 = k(\phi_1)$ and $\widetilde{x}_1 \preceq_D \widetilde{x}_2 \preceq_D ... \preceq_D \widetilde{x}_n$, then*

$$\widetilde{y}_2 \sim (\widetilde{x}_1, \widehat{q}_1^2; ...; \widetilde{x}_n, \widehat{q}_n^2) \preceq_D (\widetilde{x}_1, \widehat{q}_1^1; ...; \widetilde{x}_n, \widehat{q}_n^1) \sim \widetilde{y}_1.$$

Thus, if priors can be ranked according first-degree stochastic dominance, the increase in ambiguity aversion modifies the demand for the asset in the same direction as a FSD deterioration of the excess return in the expected utility model. The problem is that the comparative statics of an FSD deterioration in the excess return is in general ambiguous in the expected utility model. The intuition for this negative result is that a reduction in the return of an asset has a substitution effect and a wealth effect. As for the existence of Giffen goods in consumption theory, the wealth effect may induce an increase in the asset demand. Technically, it is not true in general that condition (7) holds when $\widetilde{y}_2 \preceq_{FSD} \widetilde{y}_1$. It is easy to see why: By definition of FSD, this would be true if and only if function $f(y) = yu'(w_0 + \alpha y)$ would be increasing, which is not true in general. As observed by Fishburn and Porter (1976), a sufficient condition for f to be increasing is that relative risk aversion $R(z) = -zu''(z)/u'(z)$ be smaller than unity.[5] It implies that this is also a sufficient condition for an increase of ambiguity aversion to reduce the demand for the ambiguous asset when priors can be ranked according to FSD. The same strategy can be used to examine the case when

[5]This is because $f'(y) = 1 - R(w_0 + \alpha y) + w_0 A(w_0 + \alpha y)$, with $A(z) = -u''(z)/u'(z)$.

priors can be ranked in the sense of Rothschild-Stiglitz increase in risk. In that case, the above proposition tells us that the observationally equivalent probability distribution $\widetilde{y}_2$ is an increase in risk compared to $\widetilde{y}_1$. Because f is not concave, this does not in general imply that condition (7) holds. As initially shown by Rothschild and Stiglitz (1971), it is not true in general that an increase in risk of the excess return of the risky asset reduces its demand. Hadar and Seo (1990) provided a sufficient condition, which guarantees that f is concave. This condition is that relative prudence is positive and less than 2, where relative prudence is defined by $P(z) = -zu'''(z)/u''(z)$ (Kimball (1990)). This proves the following result.

Proposition 2 *Suppose that* $u \in C^3$ *and* $E\widetilde{x} > 0$. *Any increase in ambiguity aversion reduces the demand for the risky asset if one of the following two conditions is satisfied:*

1. $(\widetilde{x}_1, ..., \widetilde{x}_n)$ *can be ranked according first-degree stochastic dominance, and* $R \leq 1$*;*

2. $(\widetilde{x}_1, ..., \widetilde{x}_n)$ *can be ranked according to the Rothschild and Stiglitz's riskiness order, and* $0 \leq P^r \leq 2$.

More generally, if the set of marginals can be ranked according to the SSD order, an increase in ambiguity aversion reduces the demand for the risky asset if relative risk aversion is less than unity, and relative prudence is positive and less than two. In the case of power utility function, relative prudence equals relative risk aversion plus one. This implies that when relative risk aversion is constant, and when priors can be ranked according to SSD, any increase in ambiguity aversion reduces the demand for the ambiguous asset if relative risk aversion is less than unity. This condition is not very convincing, since relative risk aversion is usually assumed to be larger than unity. Arguments have been provided based on introspection (Drèze (1981), Kandel and Stambaugh (1991), Gollier (2001)) or on the equity premium puzzle that can be solved in the canonical model only with a degree of relative risk aversion exceeding 40.

Rather than limiting the set of utility functions yielding an unambiguous effect, an alternative approach consists in restricting the set of priors. To do this, let us first introduce the following concepts, which rely on the location-weighted-probability function T_θ that is defined as follows:

$$T_\theta(x) = \int_{x_-}^{x} t dF_\theta(t). \tag{9}$$

Following Gollier (1995), we say that $\widetilde{x}_2$ is dominated by $\widetilde{x}_1$ in the sense of Central Dominance if there exists a nonnegative scalar m such that $T_2(x) \leq mT_1(x)$ for all $x \in [x_-, x_+]$.[6] Gollier (1995) showed that $\widetilde{x}_2 \preceq_{CD} \widetilde{x}_1$ is necessary and sufficient to guarantee that all risk-averse investors reduce their demand for the risky asset whose distribution of excess return goes from $\widetilde{x}_1$ to $\widetilde{x}_2$. SSD-dominance is not sufficient for CD-dominance. Notice that $\widetilde{x}_1$ and $\widetilde{x}_2$ in the counterexample of the previous section violate the CD condition. It implies that there exists a concave utility function such that the demand for the asset is increased when beliefs go from $(\widetilde{x}_1, 5\%; \widetilde{x}_2, 95\%)$ to the riskier $\widetilde{x}_1$.

Here is a partial list of stochastic orders that have been shown to belong to the wide set of CD:

[6] There is no simple interpretation of this stochastic order in the literature. However, observe that replacing $\widetilde{x}_1$ by $\widetilde{x}_2 \sim (\widetilde{x}_1, m; 0, 1-m)$ implies that $T_2 = mT_1$. This proportional probability transfer to the zero excess return has no effect on the risk-averse investor's demand for the risky asset. This explains the presence of the arbitrary scalar m in the definition of CD. Moreover a CD shift with $m = 1$ requires a reduction of the location-weighted-probability function T. For example, if one divides a probability mass p at some return $x = r < 0$ in two equal masses $(p/2, p/2)$, the transfer of mass to the left must be to the left of $2r$, which is a strong condition.

- Monotone Likelihood Ratio order (MLR) (Ormiston and Schlee (1993)). Notice that MLR is a subset of FSD.

- Strong Increase in Risk (Meyer and Ormiston (1985)): The excess return $\widetilde{x}_2$ is a strong increase in risk with respect to $\widetilde{x}_1$ if they have the same mean and if any probability mass taken out of some of the realizations of $\widetilde{x}_1$ is transferred out of the support of this random variable.

- Simple Increase in Risk (Dionne and Gollier (1992)): Random variable $\widetilde{x}_2$ is a simple increase in risk with respect to $\widetilde{x}_1$ if they have the same mean and $x(F_1(x) - F_2(x))$ is nonnegative for all x.

- Monotone Probability Ratio order (MPR) (Eeckhoudt and Gollier (1995), Athey (2002)): When the two random variables have the same support, we say that $\widetilde{x}_2$ is dominated by $\widetilde{x}_1$ in the sense of MPR if the cumulative probability ratio $F_2(x)/F_1(x)$ is nonincreasing. It can be shown that MPR is more general than MLR, but is still a subset of FSD.

The next result allows us to relax the conditions on u, but at the cost of restricting the set of priors.

Proposition 3 *Suppose that $E\widetilde{x} > 0$. Any increase in ambiguity aversion reduces the demand for the risky asset if the set of priors $(\widetilde{x}_1, ..., \widetilde{x}_n)$ can be ranked according to both SSD dominance and central dominance, that is, if $\widetilde{x}_\theta \preceq_{SSD} \widetilde{x}_{\theta+1}$ and $\widetilde{x}_\theta \preceq_{CD} \widetilde{x}_{\theta+1}$ for all $\theta = 1, ..., n-1$.*

To illustrate, because we know that MLR yields both first-degree stochastic dominance and central dominance, we directly obtain the following corollary.

Corollary 1 *Suppose that $E\widetilde{x} > 0$ and that $(\widetilde{x}_1, ..., \widetilde{x}_n)$ can be ranked according to the monotone likelihood ratio order. Then, any increase in ambiguity aversion reduces the demand for the risky asset.*

In this case, we conclude that ambiguity aversion and risk aversion goes into the same direction. A more general corollary holds where the MLR order is replaced by the more general MPR order.

It is noteworthy that the comparative statics of ambiguity aversion is much simpler when considering market participation. Of course, as observed in Lemma 1, our basic model is not well fitted to examine this question, since all agents should have a positive demand for equity as soon as the equity premium is positive (second order risk aversion).[7] Let us introduce a fixed cost C for market participation, so that the new model is with $U(\alpha,\theta) = Eu(w_0 - C + \alpha\widetilde{x}_\theta)$, $V(\alpha) = \phi^{-1}\left(\sum_\theta q_\theta \phi(U(\alpha,\theta))\right)$, $\overrightarrow{\alpha} = \arg\max V(\alpha)$, and $\alpha^* = \overrightarrow{\alpha}$ if $V(\overrightarrow{\alpha}) \geq u(w_0)$, and $\alpha^* = 0$ otherwise. Obviously, because V is the certainty equivalent of $U(\alpha,\widetilde{\theta})$ under function ϕ, an increase in the concavity of ϕ reduces $V(\alpha)$ for all α. Thus the condition for market participation $V(\overrightarrow{\alpha}) \geq u(w_0)$ is less likely to hold when ambiguity aversion is increased. This means that ambiguity aversion may explain the market participation puzzle (Haliassos and Bertaut (1995)).

4 Asset prices with complete markets

In this section, we extend the focus of our analysis to the effect of ambiguity aversion to the price of contingent claims. Consider a Lucas tree economy with a risk-averse and ambiguity averse representative agent whose preferences are characterized by increasing and concave functions (u, ϕ_i). Each agent is endowed with a tree producing an uncertainty quantity of fruits at the end of the period. There are S possible states of nature, with c_s denoting the number of fruits produced by trees in state s, $s = 1, ..., S$. The distribu-

[7]Thus, our story of the role ambiguity aversion to explain the market participation puzzle differs from the one by Dow and Werlang (1992) and Epstein and Schneider (2007), who consider a MEU model without participation cost.

tion of states is subject to some parameter uncertainty. Parameter θ can take value $1, ..., n$ with probabilities $(q_1, ..., q_n)$, and $p_{s|\theta}$ is the probability of state s conditional to θ. Let $p_s = \Sigma_\theta q_\theta p_{s|\theta}$ denote the unconditional probability of state s. Ex ante, there is a market for contingent claims. Agents trade claims of fruits contingent on the harvest. Assuming complete markets, the ambiguity averse and risk-averse agent whose preferences are given by the pair (u, ϕ_i) solves the following problem:

$$\max_{(x_1,...,x_S)} \sum_{\theta=1}^{n} q_\theta \phi_i \left(\sum_{s=1}^{S} p_{s|\theta} u(x_s) \right), \text{ s.t. } \sum_{s=1}^{S} \Pi_s (x_s - c_s) = 0, \tag{10}$$

where $x_s - c_s$ is the demand for the Arrow-Debreu security associated to state s, and Π_s is the price of that contingent claim. The first-order conditions for this program are written as

$$u'(x_s) \left[\sum_{\theta=1}^{n} q_\theta \phi_i' \left(\sum_{s'=1}^{S} p_{s'|\theta} u(x_{s'}) \right) p_{s|\theta} \right] = \lambda \Pi_s, \tag{11}$$

for all s, where λ is the Lagrange multiplier associated to the budget constraint. The market-clearing conditions impose that $x_s = c_s$ for all s, which implies the following equilibrium state prices :

$$\Pi_s^i = \widehat{p}_s^i u'(c_s), \tag{12}$$

for all s, where the distorted state probability $\widehat{p}_s^i$ is defined as follows:

$$\widehat{p}_s^i = \sum_{\theta=1}^{n} \widehat{q}_\theta^i p_{s|\theta} \text{ with } \widehat{q}_\theta^i = \frac{q_\theta \phi_i' \left(Eu(\widetilde{c}_\theta) \right)}{\sum_{t=1}^{n} q_t \phi_i' \left(Eu(\widetilde{c}_t) \right)}, \tag{13}$$

where $\widetilde{c}_\theta$ is distributed as $(c_1, p_{1|\theta}; ..., c_S, p_{S|\theta})$. Under ambiguity neutrality, we have that $\widehat{q}_\theta^i = q_\theta$, and $\widehat{p}_s^i$ is the true probability of state s computed from the compound first and second order probabilities. The aversion to ambiguity of the representative agent affects the equilibrium state prices in a way that is observationally equivalent to a distortion of beliefs in the EU model. This distortion takes the form of a transformation of the subjective prior distribution from $(q_1, ..., q_n)$ to $(\widehat{q}_1^i, ..., \widehat{q}_n^i)$ that is equivalent to the previous section with $\widetilde{c}_\theta = w_0 + \alpha_1^* \widetilde{x}_\theta$. Lemma 3 implies that $\widehat{q}^2$ is dominated by $\widehat{q}^1$ in the sense of MLR when ϕ_2 is more ambiguity averse than agent ϕ_1. The next proposition is a direct consequence of this observation.

Proposition 4 *Suppose that the set of priors $(\widetilde{c}_1, ..., \widetilde{c}_n)$ can be ranked according the stochastic order D (D =FSD, SSD or IR) . It implies that an increase in ambiguity aversion deteriorates the observationally equivalent probability distribution of consumption in the sense of the stochastic order D : If $\exists k$ concave: $\phi_2 = k(\phi_1)$ and $\widetilde{c}_1 \preceq_D \widetilde{c}_2 \preceq_D ... \preceq_D \widetilde{c}_n$, then $(\widetilde{c}_1, \widehat{q}_1^2; ...; \widetilde{c}_n, \widehat{q}_n^2) \preceq_D (\widetilde{c}_1, \widehat{q}_1^1; ...; \widetilde{c}_n, \widehat{q}_n^1)$.*

It is easy to reinterpret this result in terms of the impact of ambiguity aversion on the price kernel $\pi_s = \Pi_s/p_s$. Suppose that $c_s \neq c_{s'}$ for all (s, s'), so that we can substitute index $s = 1, ..., n$ by another index $s = c_1, ..., c_S$. In Figure 1, we have drawn the state price $\pi_s = \widehat{p}_s^i u'(c_s)/p_s$ as a function of c_s. Under ambiguity neutrality, this is a decreasing function, because u' is decreasing. The slope of the curve $\pi(c)$ describes the degree of risk aversion of the agent. From Proposition 4, ambiguity aversion tends to reinforce risk aversion. Indeed, if the priors can be ranked by FSD, an increase in ambiguity aversion has an effect on asset prices that is observationally equivalent to a FSD-deteriorating shift in beliefs, i.e., it tends to transfer the distorted probability mass $\widehat{p}$ from the good states to the bad ones. The corresponding shift in $\pi_s = \widehat{p}_s u'(c_s)/p_s$ is described in Figure 1a. If the priors can be ranked according to their riskiness, an increase in ambiguity aversion tends to transfer the distorted probability mass to the extreme states. This implies convexifying the price kernel, as depicted in Figure 1b.

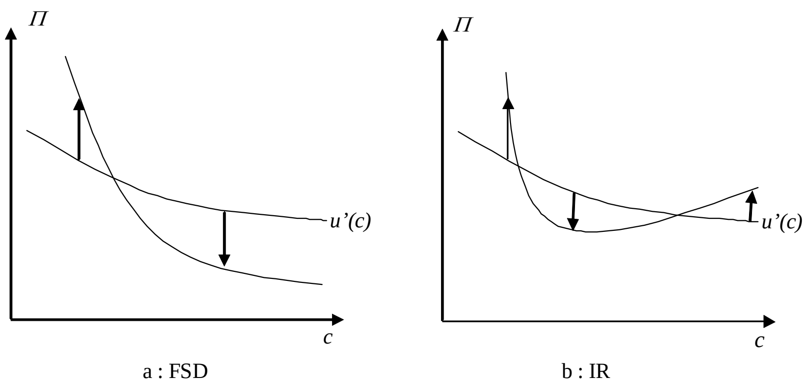


Figure 1: The effect of an increase in ambiguity aversion on the price kernel, when the priors can be ranked by the FSD order (a), or by the Rothschild-Stiglitz riskiness order (b).

The equilibrium price of trees equals $P_i = \Sigma_s \Pi_s^i c_s / \Sigma_s \Pi_s^i$. It is easily checked that this implies that

$$\sum_{\theta=1}^{n} q_\theta \phi_i'(Eu(\widetilde{c}_\theta)) E\left[(\widetilde{c}_\theta - P_i) u'(\widetilde{c}_\theta)\right] = 0.$$

It is intuitive that more ambiguity aversion should reduce the price of equity, thereby increasing the equity premium. We obtain that P_2 is smaller than P_1 if and only the following inequality holds:

$$\sum_{\theta=1}^{n} q_\theta \phi_2'(Eu(\widetilde{c}_\theta)) E\left[(\widetilde{c}_\theta - P_1) u'(\widetilde{c}_\theta)\right] \leq 0. \tag{14}$$

Technically, this condition is equivalent to condition (6) with $\widetilde{c}_\theta = w_0 + \alpha_1^* \widetilde{x}_\theta$, $\alpha_1^* = 1$ and $P_1 = w_0$. We conclude this section with the following proposition, which is a direct consequence of this observation together with the results presented in the previous section.

Proposition 5 *An increase in ambiguity aversion raises the equity premium if one of the following conditions is satisfied:*

1. $(\widetilde{c}_1, ..., \widetilde{c}_n)$ *can be ranked according first-degree stochastic dominance, and* $R \leq 1$*;*

2. $(\widetilde{c}_1, ..., \widetilde{c}_n)$ *can be ranked according to the Rothschild and Stiglitz's riskiness order, and* $0 \leq P^r \leq 2$*;*

3. $(\widetilde{c}_1 - P_1, ..., \widetilde{c}_n - P_1)$ *can be ranked according to both Second-degree Stochastic Dominance and Central Dominance, where* P_1 *is the initial price of equity.*

Property 3 implies for example that the equity premium is increasing in the degree of ambiguity aversion of the representative agent if the set of priors $(\widetilde{c}_1, ..., \widetilde{c}_n)$ can be ranked according to the MLR/MPR order.

5 Conclusion

In this paper, we explored the determinants of the demand for risky assets and of asset prices when investors are ambiguity averse. We have shown that, contrary to the intuition, ambiguity aversion may yield an increase in the demand for the risky and ambiguous asset, and a reduction in the demand for the safe one. In the same fashion, it is not true in general that ambiguity aversion raises the equity premium in the economy. We have first shown that the qualitative effect of an increase in ambiguity aversion in these settings is observationally equivalent to that of a shift in the beliefs of the investor in the standard EU model. If the set of plausible priors can be ranked according to the first degree stochastic dominance order, this shift is first degree stochastic deteriorating, whereas it is risk-increasing if these priors can be ranked according to the Rothschild-Stiglitz risk order. The problem originates from the observation already made by Rothschild and Stiglitz (1971) and Fishburn and Porter (1976) that a FSD/SSD deteriorating shift in the distribution of the return of the risky asset has an ambiguous effect on the demand for that asset in the EU framework. We heavily relied on the literature that emerged from this negative result to provide some sufficient conditions for any increase in ambiguity aversion to yield a reduction in the demand for the risky asset and an increase in the equity premium.

In most cases however, an increase in ambiguity aversion reduces the demand for the ambiguous asset, and it raises the equity premium. Two sets of findings confirm this view. First, the numerical analyses in the existing literature all go in that direction. We also showed that this is always true when the first and second order probability distributions are normal, and the pair (u, ϕ) are exponential-power functions. Second, some of our sufficient conditions cover a wide set of realistic situations. For example, if the set of priors can be ranked according to the well-known monotone likelihood ratio order, then it is always true that an increase in ambiguity aversion raises the equity premium. Our conclusion is that the potential existence of a counterintuitive effect of ambiguity aversion plays a role similar to the potential existence of Giffen goods in consumption theory. The observationally equivalent FSD deterioration of more ambiguity aversion has a wealth effect on the demand for the asset that may dominate the substitution effect. This is a rare event, but theoretical progress can rarely be made without understanding the mechanism that generates it. After all, the existence of Giffen goods is taught in Microeconomics 101.

References

Arrow, K.J., (1963), Liquidity preference, Lecture VI in "*Lecture Notes for Economics 285, The Economics of Uncertainty*", pp 33-53, undated, Stanford University.

Athey, S.C., (2002), Monotone Comparative Statics Under Uncertainty, *Quarterly Journal of Economics*, 117(1) : 187-223.

Abel, A.B., (2002), An exploration of the effects of optimism and doubt on asset returns, *Journal of Economic Dynamics and Control*, 26, 1075-1092.

Chen, Z;, and L. Epstein, (2002), Ambiguity, risk, and asset returns in continuous time, *Econometrica*, 70, 1403-1443.

Cecchetti, S.G., P.-S. Lam and N.C. Mark, (2000), Asset pricing with distorted beliefs: Are equity returns too good to be true?, *American Economic Review*, 90, 787-805.

Collard, F., S. Mukerji, K. Sheppard, and J.-M. Tallon, (2009), Ambiguity and historical equity premium, unpublished manuscript, Oxford University.

Dionne, G., and C. Gollier, (1992), Comparative Statics Under Multiple Sources of Risk with Applications to Insurance Demand, *The Geneva Papers on Risk and Insurance Theory*, 17.

Dow, J., and S. Werlang, (1992), Uncertainty aversion, risk aversion and the optimal choice of portfolio, *Econometrica,* 60, 197-204.

Drèze, J. H. (1981). "Inferring Risk Tolerance from Deductibles in Insurance Contracts", *The Geneva Papers* 6, 48-52.

Eeckhoudt, L., and C. Gollier, (1995), Demand for Risky Assets and the Monotone Probability Ratio Order, *Journal of Risk and Uncertainty*, 11, 113-122.

Epstein, L. G. and M. Schneider, (2007), Learning Under Ambiguity, *Review of Economic Studies*, 74, 1275-1303.

Epstein, L.G., and T. Wang, (1994), Intertemporal asset pricing under Knightian uncertainty, *Econometrica*, 62, 283-322.

Fishburn, P., and B. Porter, (1976), Optimal portfolios with one safe and one risky asset: Effects of changes in rate of return and risk, *Management Science*, 22, 1064-73.

Gilboa, I. and D. Schmeidler, (1989), Maximin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18, 141, 153.

Gollier, C., (1995), The Comparative Statics of Changes in Risk Revisited, *Journal of Economic Theory*, 66, 522-535.

Gollier, C., (2001), *The economics of risk and time*, MIT Press, Cambridge.

Hadar, J. and T.K. Seo, (1990), The Effects of Shifts in a Return Distribution on Optimal Portfolios, *International Economic Review*, 31, 721-736.

Haliassos, M., and C.C. Bertaut, (1995), Why do so few hold stocks?, *Economic Journal*, 105, 1110-1129.

Hansen, L.P., and T.J. Sargent, (2008), *Robustness*, Princeton University Press, Princeton and Oxford.

Hansen, L.P., T.J. Sargent, and T.D. Tallarini, (1999), Robust permanent income and pricing, *Review of Economic Studies*, 66.

Ju, N., and J. Miao, (2009), Ambiguity, learning and asset returns, unpublished manuscript, Boston University.

Kandel, S., and R.F. Stambaugh, (1991), Assets returns and intertemporal preferences, *Journal of Monetary Economics*, 27, 39-71.

Kimball, M.S., (1990), Precautionary savings in the small and in the large, *Econometrica*, 58, 53-73.

Klibanoff, P., M. Marinacci and S. Mukerji, (2005), A smooth model of decision making under ambiguity, *Econometrica* 73(6), 1849-1892.

Maenhout, P., (2004), Robust portfolio rules and asset pricing, *Review of Financial Studies*, 17, 951-983.

Meyer, J. and M.B. Ormiston, (1985), Strong Increase in Risk and Their Comparative Statics, *International Economic Review*, 26, 425-437.

Ormiston, M. and E. Schlee, (1993), Comparative Statics Under Uncertainty for a Class of Economic Agents, *Journal of Economic Theory,* 61, 412-422.

Rothschild, M. and J. Stiglitz, (1970), Increasing Risk: I. A Definition, *Journal of Economic Theory,* 2, 225-243.

Rothschild, M. and Stiglitz, J.E., (1971), Increasing Risk II: Its Economic Consequences, *Journal of Economic Theory*, 3, pp. 225-243.

Segal, U., (1987), The Ellsberg paradox and risk aversion: An anticipated utility approach, *International Economic Review*, 87, 175-202.

Segal, U. and A. Spivak, (1990), First order versus second order risk aversion, *Journal of Economic Theory*, 51, 111-125.

Taboga, M., (2005), Portfolio selection with two-stage preferences, *Finance Research Letters*, 2, 152-164.

Appendix: Proofs

Proof of Lemma 1

By concavity of the objective function in (1) with respect to α, we have that α^* is positive if the derivative of this objective function with respect to α evaluated at $\alpha = 0$ is positive. This derivative is written as

$$\sum_{\theta=1}^{n} q_\theta \phi'(u(w_0)) E\widetilde{x}_\theta u'(w_0) = u'(w_0)\phi'(u(w_0))E\widetilde{x},$$

where $E\widetilde{x} = \Sigma_\theta q_\theta E x_\theta$ is the equity premium. This concludes the proof. ■

Proof of Lemma 3

Because ϕ_1 and ϕ_2 are increasing in U, there exists an increasing function k such that $\phi_2(U) = k(\phi_1(U))$, or $\phi_2'(U) = k'(\phi_1(U))\phi_1'(U)$ for all U. Using definition (8), we obtain that

$$\frac{\widehat{q}_\theta^2}{\widehat{q}_\theta^1} = k'(\phi_1(U(\alpha_1^*, \theta))\frac{\sum_{t=1}^n q_t \phi_1'\left(U(\alpha_1^*, t)\right)}{\sum_{t=1}^n q_t \phi_2'\left(U(\alpha_1^*, t)\right)} \tag{15}$$

for all $\theta = 1, ..., n$. The Lemma is a direct consequence of (15), in the sense that the likelihood ratio $\widehat{q}_\theta^2/\widehat{q}_\theta^1$ is decreasing in θ if k' is decreasing in ϕ_1. ■

Proof of Proposition 1

Suppose that $\widetilde{x}_1 \preceq_D \widetilde{x}_2 \preceq_D ... \preceq_D \widetilde{x}_n$. It implies that $U(\alpha_1^*, 1) \leq U(\alpha_1^*, 2) \leq ... \leq U(\alpha_1^*, n)$. We have to prove that $(\widetilde{x}_1, \widehat{q}_1^1; ....; \widetilde{x}_n, \widehat{q}_n^1)$ is preferred to $(\widetilde{x}_1, \widehat{q}_1^2; ....; \widetilde{x}_n, \widehat{q}_n^2)$ by all utility functions v in C, that is

$$\sum_{\theta=1}^{n} \widehat{q}_\theta^2 Ev(\widetilde{x}_\theta) \leq \sum_{\theta=1}^{n} \widehat{q}_\theta^1 Ev(\widetilde{x}_\theta),$$

where C is the set of increasing functions if D=FSD, C is the set of increasing and concave functions if D=SSD, and C is the set of concave functions if D=IR. Combining the conditions that $\widetilde{x}_\theta \preceq_D \widetilde{x}_{\theta+1}$ and that $v \in C$ implies that $Ev(\widetilde{x}_\theta)$ is increasing in θ. The above inequality is obtained by combining this property with the fact that $\widehat{q}^2$ is dominated by $\widehat{q}^1$ in the sense of MLR (Lemma 3), a special case of FSD. ■

Proof of Proposition 3

The following lemma is useful to prove Proposition 3. Let K denote interval $[\min_\theta \alpha^*_\theta, \max_\theta \alpha^*_\theta]$, where α^*_θ is the maximand of $Eu(w_0 + \alpha\widetilde{x}_\theta)$.

Lemma 4 *Consider a specific set of priors $(\widetilde{x}_1, ..., \widetilde{x}_n)$ and a concave utility function u. They characterize function U defined by $U(\alpha, \theta) = Eu(w_0 + \alpha\widetilde{x}_\theta)$. Consider a specific scalar α^*_1 in K. The following two conditions are equivalent:*

1. *Any agent ϕ_2 that is more ambiguity averse than agent ϕ_1 with demand α^*_1 for the ambiguous asset will have a demand for it that is smaller than α^*_1;*
2. *There exists $\overline{\theta} \in \{1, ..., n\}$ such that*
$$U(\alpha^*_1, \theta)U_\alpha(\alpha^*_1, \theta) \geq U(\alpha^*_1, \overline{\theta})U_\alpha(\alpha^*_1, \theta) \tag{16}$$
for all $\theta \in \{1, ..., n\}$.

Proof: We first prove that condition 2 implies condition 1. Consider an agent $\phi_2 = k(\phi_1)$ that is more ambiguity averse than agent ϕ_1, so that the transformation function k is concave. The condition thus implies that

$$k'(\phi_1(U(\alpha^*_1, \theta)))U_\alpha(\alpha^*_1, \theta) \leq k'(\phi_1(U(\alpha^*_1, \overline{\theta})))U_\alpha(\alpha^*_1, \theta)$$

for all θ. Multiplying both side of this inequality by $q_\theta\phi'_1(U(\alpha^*_1, \theta)) \geq 0$ and summing up over all θ yields

$$\sum_{\theta=1}^{n} q_\theta\phi'_2(U(\alpha^*_1, \theta))U_\alpha(\alpha^*_1, \theta) \leq k'(\phi_1(U(\alpha^*_1, \overline{\theta}))) \sum_{\theta=1}^{n} q_\theta\phi'_1(U(\alpha^*_1, \theta))U_\alpha(\alpha^*_1, \theta) = 0.$$

The last equality comes from the assumption that agent ϕ_1 selects portfolio α_1^*. Thus, condition (6) is satisfied, thereby implying that α_2^* is less than α_1^*.

We then prove that condition 1 implies condition 2. Without loss of generality, rank the θs such that $U(\alpha_1^*, \theta)$ is increasing in θ. By contradiction, suppose that there exists a $\theta_0 < n$ such that $U_\alpha(\alpha_1^*, \theta_0) \geq 0$ and $U_\alpha(\alpha_1^*, \theta_0 + 1) \leq 0$. Select a prior distribution $(q_1, ..., q_n)$ so that $q_\theta = 0$ for all θ except for θ_0 and $\theta_0 + 1$. Select $q_{\theta_0} = q \in [0, 1]$ so that

$$q\phi_1'(U(\alpha_1^*, \theta_0))U_\alpha(\alpha_1^*, \theta_0) + (1 - q)\phi_1'(U(\alpha_1^*, \theta_0 + 1))U_\alpha(\alpha_1^*, \theta_0 + 1) = 0, \tag{17}$$

so that agent ϕ_1 selects portfolio α_1^*. Consider any concave transformation function k. It implies that

$$\begin{aligned} & \sum_{\theta=1}^{n} q_\theta \phi_2'(U(\alpha_1^*, \theta))U_\alpha(\alpha_1^*, \theta) \\ = \; & qk'(\phi_1(U(\alpha_1^*, \theta_0)))\phi_1'(U(\alpha_1^*, \theta_0))U_\alpha(\alpha_1^*, \theta_0) \\ & +(1 - q)k'(\phi_1(U(\alpha_1^*, \theta_0 + 1)))\phi_1'(U(\alpha_1^*, \theta_0 + 1))U_\alpha(\alpha_1^*, \theta_0 + 1). \end{aligned}$$

Because $U_\alpha(\alpha_1^*, \theta_0+1) \leq 0$ and $k'(\phi_1(U(\alpha_1^*, \theta_0+1))) \leq k'(\phi_1(U(\alpha_1^*, \theta_0)))$, this is larger than

$$k'(\phi_1(U(\alpha_1^*, \theta_0))) \left[q\phi_1'(U(\alpha_1^*, \theta_0))U_\alpha(\alpha_1^*, \theta_0) + (1 - q)\phi_1'(U(\alpha_1^*, \theta_0 + 1))U_\alpha(\alpha_1^*, \theta_0 + 1)\right] = 0.$$

It implies that condition (6) is violated, implying in turn that α_2^* is larger than α_1^*, a contradiction. ■

If we rank the θ in such a way that $U(\alpha_1^*, \theta)$ is monotone in θ, condition 2 is essentially a single-crossing property of function $U_\alpha(\alpha_1^*, \theta)$. To illustrate, suppose that $u(z) = -A^{-1}\exp(-Az)$ and $\widetilde{x}_\theta \sim N(\theta, \sigma^2)$, which implies that $U(\alpha, \theta)$ is increasing in θ and is given by equation (2). It implies that $U_\alpha(\alpha, \theta)$ has the same sign as $\theta - \alpha A\sigma^2$. It implies in turn that condition 2 in Lemma 4 is satisfied with $\overline{\theta} = \alpha A\sigma^2$. Our Lemma implies that ambiguity aversion reduces the demand for the risky asset in the exponential/normal case. This was shown in Section 2 in the special case of power ϕ functions.

We need to prove a second lemma in order to prepare for the proof of Proposition 3.

Lemma 5 *Suppose that* $\widetilde{x}_2$ *is centrally dominated by* $\widetilde{x}_1$. *Then,* $E\widetilde{x}_2 u'(w_0 + \alpha\widetilde{x}_2) \leq 0$ *for any* $\alpha \geq 0$ *such that* $E\widetilde{x}_1 u'(w_0 + \alpha\widetilde{x}_1) \leq 0$.

Proof: By assumption, there exists a positive scalar m such that $T_2(x) \leq mT_1(x)$. Integrating by part, we have that

$$\begin{aligned} E\widetilde{x}_2 u'(w_0 + \alpha\widetilde{x}_2) &= \int_{x_-}^{x_+} u'(w_0 + \alpha x) x dF_2(x) \\ &= u'(w_0 + \alpha x_+) T_2(x_+) - \alpha \int_{x_-}^{x_+} u''(w_0 + \alpha x) T_2(x) dx. \end{aligned}$$

This implies that

$$\begin{aligned} E\widetilde{x}_2 u'(w_0 + \alpha\widetilde{x}_2) &\leq m \left[u'(w_0 + \alpha x_+) T_1(x_+) - \alpha \int_{x_-}^{x_+} u''(w_0 + \alpha x) T_1(x) dx \right] \\ &= mE\widetilde{x}_1 u'(w_0 + \alpha\widetilde{x}_1). \end{aligned}$$

By assumption, this is nonpositive. ■

We can now prove Proposition 3. Condition $\widetilde{x}_\theta \preceq_{SSD} \widetilde{x}_{\theta+1}$ implies that $U(\alpha, \theta+1) \geq U(\alpha, \theta)$, whereas, by Lemma 5, condition $\widetilde{x}_\theta \preceq_{CD} \widetilde{x}_{\theta+1}$ implies that $U_\alpha(\alpha, \theta) \leq 0$ whenever $U_\alpha(\alpha, \theta+1) \leq 0$. This latter result implies that there exists a $\overline{\theta}$ such that $(\theta - \overline{\theta})U_\alpha(\alpha, \theta) \leq 0$ for all θ. This immediately yields condition 2 in Lemma 4, which is sufficient for our comparative static property. ■

Socially efficient discounting under ambiguity aversion

Johannes Gierlinger[1]
Toulouse School of Economics (LERNA)

Christian Gollier
Toulouse School of Economics (LERNA and IDEI)

June 30, 2010

[1]**Correspondence**: Toulouse School of Economics, Manufacture des Tabacs, Aile J.-J. Laffont, MF007, 21 allée de Brienne, 31000 Toulouse, France. Email: johannes.gierlinger@univ-tlse1.fr.
This research was supported by the Chair of the Financière de la Cité at TSE. Gierlinger acknowledges support through a FIR *Finance and Sustainability* research grant. We are thankful to Ivar Ekeland, Frederic Cherbonnier, and seminar participants at EARE 2008, ESEM 2008.

Abstract

In an economy with uncertain consumption growth and an ambiguity-averse representative agent, we provide sufficient conditions under which ambiguity aversion decreases the social discount rate. We identify two effects. The first is an ambiguity prudence effect, similar to the effect of prudence under expected utility. We show that it decreases the rate if and only if preferences satisfy decreasing ambiguity aversion. The second effect is observationnally equivalent to a deterioration of beliefs. This pessimism effect requires joint restrictions on the growth process and the agent's preferences to be signed. The calibration of the model suggests that the effect of ambiguity aversion on the way we should discount distant cash flows is potentially large.

Keywords: Decreasing ambiguity aversion, ambiguity prudence, Ramsey rule, sustainable development.

1 Introduction

The emergence of public policy problems associated with the sustainability of our development has raised considerable interest for the determination of a socially efficient discount rate. This debate has culminated in the publication of two reports about the evaluation of different public investments. The Copenhagen Consensus (Lomborg, 2004) put top priority on public programs yielding immediate benefits, like fighting AIDS and malnutrition and rejected measures to fight climate change as not being cost-effective. The Stern Review, on the other hand, (Stern, 2007) argues for decisive and immediate action against climate change.

Because global warming will really affect our economies in a relatively distant time horizon, the choice of the rate at which these costs are discounted plays a key role in reaching either conclusion. While Stern applies an implicit rate of 1.4% per year, the Copenhagen Consensus is based on a rate around 5%. For the sake of illustrating the power of discounting, consider a project which yields its benefits in t years time. For a horizon $t = 100$ the Copenhagen Consensus would require a rate-of-return 36 times higher than Stern.

As stated by the well-known Ramsey rule (Ramsey (1928)), the socially efficient discount rate (net of the rate of pure preference for the present) is equal to the product of relative risk aversion and the growth rate of consumption. With future generations likely being richer, the return on investment needs to be large enough to compensate for the increased intertemporal inequality that it generates. If we assume that the growth rate of wealth is 2% and relative risk aversion equals 2, this yields a discount rate of 4%.

However, this basic reasoning does not take into account the riskiness affecting the long-term growth of consumption. Hansen and Singleton (1983), Gollier (2002) and Weitzman (2007a), among others, have extended the Ramsey rule in this spirit. An exogenously given stochastic growth process adds a precautionary term to the Ramsey rule which reduces the discount rate under prudence: The willingness to save increases with risk since marginal utility is convex (Leland, 1968; Drèze and Modigliani, 1972).

The present paper goes one step further in recognizing the potential uncertainty on the growth process itself. Such parameter uncertainty on priors is typically referred to as statistical ambiguity or Knightian uncertainty. We believe that this assumption is realistic, especially for long-term forecasts.

Departing from the standard Subjective Expected Utility paradigm (SEU, Savage (1954)), we also assume that the representative agent is ambiguity-averse, i.e., that she dislikes mean-preserving spreads over prior beliefs. Indeed, starting with the pioneering work by Ellsberg (1961), ample evidence in favor of this hypothesis has been accrued.[1] All of which suggests that it is behaviorally

[1] The *Ellsberg Paradox* goes back to a thought experiment by Daniel Ellsberg (1961). Suppose two urns containing 90 balls of three colors. The first contains 30 red balls, with the remaining black and yellow in unknown proportion. The second one contains 30 balls of each color. Imagine a bet on drawing the color black. Ellsberg conjectured that a majority of bettors prefers to place it on the unambiguous urn. However, when asked to bet on the

meaningful to distinguish lotteries over prior distributions from lotteries over final outcomes. In what follows, we will consider a representative agent who displays "smooth ambiguity preferences", as recently proposed by Klibanoff, Marinacci and Mukerji (KMM, 2005, 2009). Accordingly, the agent computes the expected utility of future consumption conditional on each possible value of the uncertain parameter. She then evaluates her future felicity by computing the certainty equivalent of these conditional expected utilities, using an increasing and concave function ϕ. The concavity of this function implies that she dislikes any mean-preserving spread in the set of plausible beliefs, i.e. that she is ambiguity-averse. It can be shown that the smooth ambiguity family entails the well-known max-min criterion as a special case.

Intuitively, one might expect that ambiguity aversion should raise the agent's willingness to save in order to compensate for its adverse effect on the present value of future welfare. In this paper we show that this is not true in general: ambiguity aversion may increase the socially efficient discount rate. This is connected to two, possibly opposing, effects of ambiguity aversion on marginal utility. On the one hand, there is an ambiguity prudence effect, similar to the prudence effect in the expected utility framework. Independent of risk preferences, decreasing absolute ambiguity aversion (DAAA) has a positive effect on the willingness to save for the ambiguous future. This effect is related to future felicity being measured by the ϕ-certainty equivalent rather than the expectation of ϕ.

On the other hand ambiguity aversion acts like an implicit pessimistic shift in beliefs with respect to the expected utility benchmark. This has been observed by KMM (2005, 2007). It is as if probability weights were shifted towards unfavorable priors in the sense of the Monotone Likelihood Ratio order (MLR). However, pessimism does not in general imply a reduction of the interest rate. We derive pairs joint conditions on the risk attitude and the stochastic ordering of plausible distributions to guarantee that, under DAAA, the socially efficient discount rate will indeed be lower under ambiguity aversion.

This paper is related to Weitzman (2007a) and Gollier (2007b) in recognizing uncertainty as a determinant feature of the discounting problem. Weitzman (2007a) shows that uncertainty about the volatility of the growth process may yield a term structure of the discount rate that tends to minus infinity for very long time horizons. Gollier (2007b) provides a general typology for parameter uncertainty. He shows that the sign of the third or fourth derivative of the utility function are necessary to sign the effect on the efficient discount rate, depending upon its type. The present paper departs strongly from these works in acknowledging evidence in favor of ambiguity-averse preferences.

Jouini, Marin and Napp (2008) and Gollier (2007) consider the related question of how to aggregate diverging beliefs in a SEU framework. Jouini, Marin and Napp show that an aggregation bias might cause a rich evolution of the discount rate than in the representative agent models. In particular, the dis-

contrary (*not black*), they would also prefer to place the opposite bet on the unambiguous urn. This pattern would be incompatible with any stable belief in the SEU model. Experimental data confirms Ellsberg's "paradoxical" predictions (see e.g. Camerer and Weber, 1992).

count rate might be first increasing and only then approach its limit, namely the smallest individual rate.

The most active branch of the literature on uncertainty studies phenomena on financial markets. Methodologically, our paper is most closely related to Gollier (2006). He investigates the effect of ambiguity aversion on the demand for risky assets. Like in the present model, the pessimism effect on beliefs is shown to affect behavior differently than risk aversion.

Ju and Miao (2007) and Collard, Mukerji, Sheppard and Tallon (2008) investigate the quantitative effects of KMM preferences on asset prices. Using tractable functional forms, they are able to replicate several empirical phenomena, like low risk-free rates, which are difficult to explain within the standard expected utility model. The present paper shows, however, that the relation between the degree of ambiguity aversion and the risk-free rate needs not be negative in the general case.

The remainder of the paper is organized as follows. Section 2 introduces the basic model and presents the equilibrium pricing formula. In Section 3 an analytical example yields an adapted Ramsey-rule for the interest rate under ambiguity. We decompose the effect of ambiguity aversion into its two components in Section 4, whereas Sections 5 and 6 are devoted to respectively the ambiguity prudence effect and the pessimism effect. Section 7 investigates under which conditions our findings extend to any increase in ambiguity aversion. Finally, before concluding, we calibrate the model using two different specifications in Section 8.

2 The model

We consider an economy à la Lucas (1978). Each agent in the economy is endowed with a tree which produces $\widetilde{c}_t$ fruits at date t, $t = 0, 1, 2, \ldots$. There is a market for zero-coupon bonds at date 0 in which agents may exchange the delivery of one fruit today against the delivery of $e^{r_t t}$ fruits for sure at date t. Thus, the real interest rate associated to maturity t is r_t. The distribution of $\widetilde{c}_t$ is a function of a parameter θ, $\theta = 1, 2, \ldots, n$. This parametric uncertainty takes the form of a random variable $\widetilde{\theta}$ whose probability distribution is a vector $q = (q_1, \ldots, q_n)$, where q_θ is the probability that $\widetilde{\theta}$ takes value θ. Let the cumulative distribution function of $\widetilde{c}_t$ conditional on θ be denoted by $F_{t\theta}$. The crop conditional to θ is denoted $\widetilde{c}_{t\theta}$. An ambiguous environment for $\widetilde{c}_t$ is thus fully described by $\widetilde{c}_t \sim (\widetilde{c}_{t1}, q_1; \ldots; \widetilde{c}_{tn}, q_n)$. Conditional on θ, the expected utility of an agent who purchases α zero-coupon bonds with maturity t equals

$$U_t(\alpha, \theta) = Eu(\widetilde{c}_{t\theta} + \alpha e^{r_t t}) = \int u(c + \alpha e^{r_t t}) dF_{t\theta}(c).$$

We assume that u is three times differentiable, increasing and concave, so that $U(\cdot, \theta)$ is concave in the investment α, for all θ.

Following Klibanoff, Marinacci and Mukerji (2005) and its recursive generalization (Klibanoff, Marinacci and Mukerji, 2009), we assume that the preferences of the representative agent exhibit smooth ambiguity aversion. Ex ante, for a given investment α, her welfare is measured by $V_t(\alpha)$, which is the certainty equivalent of the conditional expected utilities:

$$\phi(V_t(\alpha)) = \sum_{\theta=1}^{n} q_\theta \phi(U_t(\alpha,\theta)) = \sum_{\theta=1}^{n} q_\theta \phi\left(Eu(\widetilde{c}_{t\theta} + \alpha e^{r_t t})\right). \tag{1}$$

Function ϕ describes the investor's attitude towards ambiguity (or parameter uncertainty). It is assumed to be three times differentiable, increasing and concave. A linear function ϕ means that the investor is neutral to ambiguity as her preferences simplify to the subjective expected utility functional $V_t^{SEU}(\alpha) = Eu(\widetilde{c}_t + \alpha e^{r_t t})$. In contrast, a concave ϕ is equivalent to ambiguity aversion. In other words, she dislikes mean-preserving spreads over candidate levels of $U_t(\alpha,\theta)$.

An interesting particular case arises when absolute ambiguity aversion $A(U) = -\phi''(U)/\phi'(U)$ is constant, so that $\phi(U) = -A^{-1}\exp(-AU)$. As proven by Klibanoff, Marinacci and Mukerji (2005), the ex-ante welfare $V_t(\alpha)$ tends to the max-min expected utility functional $V_t^{MEU}(\alpha) = \min_\theta Eu(\widetilde{c}_{t\theta} + \alpha e^{r_t t})$ when the degree of absolute ambiguity aversion ϕ tends to infinity. Thus, the max-min criterion à la Gilboa and Schmeidler (1989) is a special case of this model.

The optimal investment α^* maximizes the intertemporal welfare of the investor,

$$\alpha^* \in \arg\max_{\alpha} \qquad u(c_0 - \alpha) + e^{-\delta t} V_t(\alpha), \tag{2}$$

where parameter δ is the rate of pure preference for the present.

At this stage, it is important to point out that the basic assumptions underlying KMM models do not guarantee that the maximization problem (2) is convex. To see why, it suffices to recall that certainty equivalent functions need not be concave. Indeed, even if we imposed ϕ and u to be strictly concave, the solution to program (2), when it exists, need not be unique. To handle this problem we prove the following.

Proposition 1 *Suppose that ϕ has a concave absolute ambiguity tolerance, i.e., $-\phi'(U)/\phi''(U)$ is concave in U. This implies that V_t is concave in α.*

Proof. Relegated to the Appendix.

If the inverse of absolute ambiguity aversion increases at a linear or decreasing rate in U, then the KMM functional is concave in α. The above proposition includes the specifications which are most widely used in the literature, in particular the families of exponential and power functions.

Henceforward we will consider the following assumption satisfied.

Assumption 1 *The function ϕ exhibits a concave absolute ambiguity tolerance, i.e. $-\phi'(U)/\phi''(U)$ is concave in U everywhere.*

Thanks to Assumption 1, the necessary and sufficient condition to solve program (2) can be written as

$$u'(c_0 - \alpha^*) = e^{-\delta t} V_t'(\alpha^*).$$

Fully differentiating equation (1) with respect to α yields

$$V_t'(\alpha) = e^{r_t t} \frac{\sum_{\theta=1}^{n} q_\theta \phi' \left(Eu(\widetilde{c}_{t\theta} + \alpha e^{r_t t}) \right) Eu'(\widetilde{c}_{t\theta} + \alpha e^{r_t t})}{\phi'(V_t(\alpha))}.$$

Because we assume that all agents have the same preferences and the same stochastic endowment, the equilibrium condition on the market for the zero-coupon bond associated to maturity t is $\alpha^* = 0$. Combining the above two equations implies the following equilibrium condition:

$$r_t = \delta - \frac{1}{t} \ln \left[\frac{\sum_{\theta=1}^{n} q_\theta \phi' \left(Eu(\widetilde{c}_{t\theta}) \right) Eu'(\widetilde{c}_{t\theta})}{\phi'(V_t(0)) u'(c_0)} \right]. \tag{3}$$

This is also the socially efficient rate at which sure benefits and costs occurring at date t must be discounted in any cost-benefit analysis at date 0.

Under ambiguity-neutrality, the standard bond pricing formula $r_t = \delta - t^{-1} \ln \left[Eu'(\widetilde{c}_t)/u'(c_0) \right]$ obtains.[2] The riskiness of future consumption reduces the social discount rate if and only if $Eu'(\widetilde{c}_t)$ is larger than $u'(E\widetilde{c}_t)$,, that is to say, if and only if u' is convex and the agent displays *prudence* (see Leland, 1968; Drèze and Modigliani, 1972; or Kimball, 1990).

Our goal in this paper is to determine the conditions under which ambiguity aversion reduces the discount rate. An ambiguous environment $(\widetilde{c}_{t1}, q_1; ...; \widetilde{c}_{tn}, q_n)$ is said to be acceptable if the respective supports of the $\widetilde{c}_{t\theta}$ are in the domain of u, and if all $Eu'(\widetilde{c}_{t\theta})$ are in the domain of ϕ. The set of acceptable ambiguous environments is denoted Ψ.

3 An analytical solution

Let us consider the following specification:

- The plausible distributions of $\ln \widetilde{c}_{t\theta}$ are all normal with the same variance $\sigma^2 t$, and with mean $\ln c_0 + \theta t$.[3]
- The parameter θ is normally distributed with mean μ and variance σ_0^2.[4]
- The representative agent's preferences exhibit constant relative risk aversion $\gamma = -cu''(c)/u'(c)$, such that $u(c) = c^{1-\gamma}/(1-\gamma)$.

[2] See for example Cochrane (2001).

[3] In continuous time, this would mean that the consumption process is a geometric Brownian motion $d \ln c_t = \theta dt + \sigma dw$.

[4] We consider the natural continuous extension of our model with a discrete distribution for $\widetilde{\theta}$.

- The representative agent's preferences exhibit constant relative ambiguity aversion $\eta = -|u|\,\phi''(u)/\phi'(u) \geq 0$. This means that
$\phi(U) = k(kU)^{1-\eta k}/(1-\eta k)$, where $k = sign(1-\gamma)$ is the sign of u.

As is well-known, the Arrow-Pratt approximation is exact under CRRA and lognormally distributed consumption. Therefore, conditional to each θ, we have that

$$Eu(\widetilde{c}_{t\theta}) = (1-\gamma)^{-1}\exp(1-\gamma)(\ln c_0 + \theta t + 0.5(1-\gamma)\sigma^2 t).$$

We can again use the same trick to compute the ϕ-certainty equivalent V_t, since $\phi(Eu(\widetilde{c}_{t\theta}))$ is an exponential function and the random variable $\widetilde{\theta}$ is normal, which is another case where the Arrow-Pratt approximation is exact. It yields

$$V_t(0) = (1-\gamma)^{-1}\exp(1-\gamma)\Big(\ln c_0 + \mu t + 0.5(1-\gamma)\sigma^2 t + 0.5(1-\gamma)(1-k\eta)\sigma_0^2 t^2\Big).$$

However, in order to solve for the pricing rule (3) we are really interested in $V_t'(0)$. A convenient way to structure the algebra is to decompose $V_t'(0)$ in the following way: again exploiting the Arrow-Pratt approximation, we have on the one hand

$$\frac{E\phi'\left(Eu(\widetilde{c}_{t\theta})\right)}{\phi'(V_t(0))} = \exp\Big(\frac{1}{2}(1-\gamma)^2 k\eta\sigma_0^2 t^2\Big), \tag{4}$$

and on the other hand

$$\begin{aligned}\frac{E[\phi'\left(Eu(\widetilde{c}_{t\theta})\right)Eu'(\widetilde{c}_{t\theta})]}{E\phi'\left(Eu(\widetilde{c}_{t\theta})\right)} \quad = \quad \exp-\Big(\gamma(\ln c_0 + \mu t) \quad & -\frac{1}{2}\gamma^2(\sigma^2 t + \sigma_0^2 t^2) - \\ & -(\gamma(1-\gamma)k\eta)\sigma_0^2 t^2\Big).\end{aligned} \tag{5}$$

Finally, multiplying (4) by (5) and plugging the result into (3), yields the desired analytical expression:

$$r_t = \delta + \gamma\mu - \frac{1}{2}\gamma^2(\sigma^2 + \sigma_0^2 t) - \frac{1}{2}\eta\left|1-\gamma^2\right|\sigma_0^2 t. \tag{6}$$

Let g be the growth rate of expected consumption. It is easy to check that $g = \mu + 0.5(\sigma^2 + \sigma_0^2 t)$. It implies that the above equation can be rewritten as

$$r_t = \delta + \gamma g - \frac{1}{2}\gamma(\gamma+1)(\sigma^2 + \sigma_0^2 t) - \frac{1}{2}\eta\left|1-\gamma^2\right|\sigma_0^2 t. \tag{7}$$

The first two terms on the right-hand side of this equation correspond to the classical Ramsey rule. The interest rate is increasing in the growth rate of expected consumption g. When g is positive, decreasing marginal utility implies that the marginal utility of consumption is expected to be smaller in the future than it is today. This yields a positive interest rate. The third term expresses prudence. Because the riskiness of future consumption increases the expected marginal utility $Eu'(\widetilde{c}_t)$ under prudence, this has a negative impact on the discount rate.[5] Notice that the variance of consumption at date t equals $\sigma^2 t + \sigma_0^2 t^2$, so that it increases at an increasing rate with respect to the time horizon. Therefore, the precautionary effect has a relatively larger impact on the discount rate for longer horizons. This argument has been developed in Weitzman (2007a) and Gollier (2008) to justify a decreasing discount rate in an expected utility framework.

The final term reduces the discount rate under positive ambiguity aversion ($\eta > 0$). It is increasing with ambiguity aversion η, the degree of uncertainty σ_0, and with the time horizon t.

Thanks to the above specifications, absent ambiguity the term structure is flat. The mere presence of ambiguity (i.e. $\sigma_0^2 > 0$) causes the rates to decrease linearly over time. If in addition, the agent displays ambiguity aversion ($\eta > 0$), this decline steepens.

The following sections investigate whether it is true in general, that ambiguity aversion decreases the socially efficient discount rate for any maturity. Contrary to the example presented above, the next section reveals that ambiguity aversion might even decrease the willingness to save.

4 The two effects of ambiguity aversion

The usual bond-pricing formula under SEU yields a benchmark expression for the social discount rate

$$r_t = \delta - \frac{1}{t} \ln \left[\frac{Eu'(\widetilde{c}_t)}{u'(c_0)} \right], \tag{8}$$

where the random variable $\widetilde{c}_t$ describes future consumption, distributed as $(\widetilde{c}_{t1}, q_1; ...; \widetilde{c}_{tn}, q_n)$. Like in the analytical example from above, the effect of ambiguity-aversion on $V_t'(0)$ can be decomposes such that

$$r_t = \delta - \frac{1}{t} \ln \left[a \frac{Eu'(\overset{\circ}{\widetilde{c}_t})}{u'(c_0)} \right], \tag{9}$$

where the constant a is defined as

$$a = \frac{\sum_{\theta=1}^{n} q_\theta \phi' \left(Eu(\widetilde{c}_{t\theta}) \right)}{\phi'(V_t(0))}, \tag{10}$$

and where $\overset{\circ}{\widetilde{c}_t}$ is a distorted probability distribution $(\widetilde{c}_{t1}, \overset{\circ}{q}_1; ...; \widetilde{c}_{tn}, \overset{\circ}{q}_n)$ with the property that for any $\theta = 1, ..., n$,

$$\overset{\circ}{q}_\theta = \frac{q_\theta \phi' \left(Eu(\widetilde{c}_{t\theta}) \right)}{\sum_{\tau=1}^{n} q_\tau \phi' \left(Eu(\widetilde{c}_{t\tau}) \right)}. \tag{11}$$

[5] This precautionary effect is equivalent to reducing the growth rate of consumption g by the precautionary premium (Kimball (1990)) $0.5(\gamma+1)(\sigma^2+\sigma_0^2 t)$. Indeed, $\gamma+1 = -cu'''(c)/u''(c)$ is the index of relative prudence of the representative agent.

Notice the similarity between pricing formula (8) and (9). It implies that ambiguity aversion reduces the discount rate if

$$aEu'(\widetilde{c}_t^{\circ}) \geq Eu'(\widetilde{c}_t). \tag{12}$$

Moreover, Observe that this condition simplifies to $a \geq 1$ when the agent is risk neutral. Because we don't constrain the risk attitude in any way except risk aversion, condition $a \geq 1$ is necessary to guarantee that ambiguity aversion reduces the discount rate. For reasons that will be clarified in the next section, we will refer to $a \geq 1$ as the ambiguity prudence effect.

In the absence of an ambiguity prudence effect ($a = 1$), condition (12) becomes $Eu'(\widetilde{c}_t^{\circ}) \geq Eu'(\widetilde{c}_t)$, which is referred to as the pessimism effect. At this stage, it is enough to say that it comes from a distortion of the beliefs $(q_1, ..., q_n)$ on the likelihood of the different plausible probability distributions $(\widetilde{c}_1, ..., \widetilde{c}_n)$.

5 The ambiguity prudence effect

In this section, we focus on whether the constant a, defined by equation (10), is larger than unity. As stated above, this is necessary to guarantee that the discount rate is reduced and it becomes necessary and sufficient in the special case of risk-neutrality. Notice that in the latter case, a can be interpreted as the sensitiveness of the $\phi-$certainty equivalent of $\overline{c}_{\widetilde{\theta}} = E\left[\widetilde{c}_{t\widetilde{\theta}} \mid \widetilde{\theta}\right]$ with respect to an increase in saving.[6] We seek to determine whether one more unit saved yields an increase in the $\phi-$certainty equivalent future consumption. More generally, condition $a \geq 1$ can be rewritten as

$$\sum_{\theta} q_\theta \phi(u_\theta) = \phi(V_t) \quad \Longrightarrow \quad \sum_{\theta=1}^{n} q_\theta \phi'(u_\theta) \geq \phi'(V_t). \tag{13}$$

Similar questions have been raised in risk theory: Do expected-utility-preserving risks raise expected marginal utility? Along the same lines, we are able to conclude from expected utility theory that condition (13) requires ϕ to satisfy decreasing absolute ambiguity aversion (see e.g. Gollier, 2001, Section 2.5). Indeed, defining function ψ such that $\psi(\phi(U)) = \phi'(U)$ for all U, the above condition can be rewritten as

$$\sum_{\theta=1}^{n} q_\theta \psi(\phi_\theta) \geq \psi(\Sigma_\theta q_\theta \phi_\theta),$$

where $\phi_\theta = \phi(u_\theta)$ for all θ. This is true for all distributions of $(\phi_1, q_1; ...; \phi_n, q_n)$ if and only if ψ is convex. Because $\psi'(\phi(U)) = \phi''(U)/\phi'(U)$, this is true iff $A(U) = -\phi''(U)/\phi'(U)$ be non-increasing. This proves the following results.

Lemma 1 $a \geq 1$ *(resp.* $a \leq 1$*) for all acceptable ambiguous environments* $\widetilde{c} \in \Psi$ *if and only if absolute ambiguity aversion is non-increasing (resp. non-decreasing).*

[6] Define $V(s, \overline{c}_{\widetilde{\theta}})$ such that $\phi(s+V) = E\phi(s+\overline{c}_{\widetilde{\theta}})$. We have that $a = \partial V(s, \overline{c}_{\widetilde{\theta}})/\partial s$ at $s = 0$.

Proposition 2 *Suppose that the representative agent is risk-neutral. The socially efficient discount rate is smaller (resp. larger) than under ambiguity neutrality for all ambiguous environments $\widetilde{c}$ if and only if ϕ exhibits non-increasing (resp. non-decreasing) absolute ambiguity aversion.*

Rather than the extent of ambiguity aversion itself, what drives the result is how the degree of ambiguity aversion relates to conditional expected utility U. For instance, in the limit case with constant absolute ambiguity aversion, ambiguity has *no* effect on the equilibrium interest rate. The intuition for these results is easy to derive from the observation that the period-t felicity V_t is approximately equal to expected consumption minus the ambiguity premium. Moreover, the premium is itself proportional to ambiguity aversion A, which makes the willingness to save decreasing in A'. Thus, ambiguity aversion raises the willingness to save – therefore reducing the equilibrium interest rate – if absolute ambiguity aversion is decreasing.

Exactly as decreasing absolute risk aversion is unanimously accepted as a natural assumption for risk preferences, we believe that decreasing absolute ambiguity aversion (DAAA) is a reasonable property of uncertainty preferences. It means that a local mean-preserving spread in conditional expected utility has an impact on welfare that is decreasing in the level of utility where this spread is realized.

We call this the *ambiguity prudence effect* because it emerges as a consequence of the uncertainty of the future conditional expected utility. This raises the willingness to save exactly as the risk on future income raises savings in the standard expected utility model under "risk prudence". But contrary to risk prudence, which is characterized by $u''' \geq 0$, ambiguity prudence is described by decreasing absolute uncertainty aversion, which is weaker than $\phi''' \geq 0$. This is because, in the intertemporal KMM model, the future felicity is represented by the $\phi-$certainty equivalent of the conditional expected utilities, rather than by the expected $\phi-$valuation of the conditional expected utilities. Had we used this alternative model, ϕ' convex (concave) would have been the property to determine the sign of the effect.

However, once we allow for risk aversion, non-increasing ambiguity aversion is no longer sufficient to sign the impact of ambiguity on the discount rate, as shown by the following counterexample.

Counterexample 1. Let c_0 equal 2 and $\widetilde{c}_t$ shall take two olausible distributions $\widetilde{c}_{t1} \sim (1, 1/3; 4, 1/3; 7, 1/3)$ and $\widetilde{c}_{t2} \sim (3, 2/3; 4, 1/3)$, both equally plausible, i.e., $q_1 = q_2 = 1/2$. Further, preferences shall display constant relative risk aversion (CRRA) with $\gamma = 2$ such that $u(c) = -c^{-1}$, and $\delta = 0$. It is easy to check that absent ambiguity aversion the interest rate equals 9.24%. However, under constant absolute ambiguity aversion of $A = 2.11$, i.e., $\phi(U) = -\exp(-2.11U)$, tedious computations yield a discount rate of exactly zero: $r_t = 0$. ■

6 The pessimism effect

Counter-example 1 can be explained by the presence of a second effect, the pessimism effect. In the pricing formula (9), ambiguity has an effect which is equivalent to compute marginal expected utility using the distorted random variable $\widetilde{c}_t^{\circ}$ instead of $\widetilde{c}_t$. The distortion of these implicit beliefs depends upon the degree of ambiguity aversion and is governed by rule (11). This section is devoted to characterize how the distortion affects the discount rate. If we find that it is pessimistic in the sense of FSD, then we are able to unambiguously sign the effect on the discount rate.

To examine this question we begin by comparing of the distorted probabilities $\overset{\circ}{q} = (\overset{\circ}{q}_1, ..., \overset{\circ}{q}_n)$ to the original probabilities $q = (q_1, ..., q_n)$.

Let the priors θ be ordered such that $Eu(\widetilde{c}_{t1}) \leq Eu(\widetilde{c}_{t2}) \leq ... \leq Eu(\widetilde{c}_{tn})$ and the agent prefers θ to be large. We hereafter show that ambiguity aversion is equivalent to a distortion of the prior beliefs on parameter $\widetilde{\theta}$ in the sense of the Monotone Likelihood Ratio Order (MLR). By definition, a shift of beliefs from q to $\overset{\circ}{q}$ entails a deterioration in the sense of the *monotone likelihood ratio ordering* (MLR) if $q^{\circ}_{\widetilde{\theta}}/q_{\widetilde{\theta}}$ and $\tilde{\theta}$ are anti-comonotonic. Observe from (11) that $q^{\circ}_{\theta}/q_{\theta}$ is proportional to $\phi'(Eu(\widetilde{c}_{t\theta}))$. Thus, since ϕ' is decreasing, we know that $q^{\circ}_{\widetilde{\theta}}/q_{\widetilde{\theta}}$ and $E\left[u(\widetilde{c}_t) \mid \widetilde{\theta}\right]$ are anti-comonotonic, establishing the following property.

Lemma 2 *The subsequent conditions are equivalent:*

1. *Beliefs q° are dominated by q in the sense of the monotone likelihood ratio order for any set of marginals $(\tilde{c}_{t1}, ..., \tilde{c}_{tn})$ such that $Eu(\tilde{c}_{t1}) \leq ... \leq Eu(\tilde{c}_{tn})$.*
2. *ϕ is concave.*

The intuitive interpretation is that ambiguity aversion is characterized by an MLR-dominated shift in the prior beliefs. In other words, it biases beliefs by favoring the worse marginals in a very specific sense: if the agent prefers marginal $\widetilde{c}_{t\theta}$ to marginal $\widetilde{c}_{t\theta'}$, then, the ambiguity-averse representative agent increases the implicit prior probability q°_{θ} relatively more than the implicit prior probability $q^{\circ}_{\theta'}$. This result gives some flesh to our pessimism terminology. It also generalizes – and builds a bridge to – the max-min case where all the weight is transferred to the worst θ.[7]

However, the MLR deterioration of the distribution $\widetilde{\theta}$ of priors is not enough to ensure a negative pessimism effect on the discount rate, as shown by Counterexample 1. Instead, the crucial requirement would be that the distortion overweights scenarios which yield larger conditional expected *marginal* utility. The above lemma says something different, namely that the distortion overweights scenarios which yield larger conditional expected utility. Therefore, to

[7] This equivalence between concave transformation functions and stochastic orderings is well known (see Lehmann, 1955). Economic applications notably include rank-dependent utility models (see Quiggin, 1995).

obtain the desired result, we need to find conditions such that u and $-u'$ agree on the ranking of scenarios.

Lemma 3 *The following two conditions are equivalent:*

1. *The pessimism effect reduces the discount rate, i.e.,* $Eu'(\overset{\circ}{\widetilde{c}_t}) \geq Eu'(\widetilde{c}_t)$, *for all* ϕ *increasing and concave;*

2. $E\left[u(\widetilde{c}_t) \mid \widetilde{\theta}\right]$ *and* $E\left[u'(\widetilde{c}_{t\theta}) \mid \widetilde{\theta}\right]$ *are anti-comonotonic.*

Proof: To prove that $2 \Rightarrow 1$, suppose that $E\left[u(\widetilde{c}_t) \mid \widetilde{\theta}\right]$ and $E\left[u'(\widetilde{c}_{t\theta}) \mid \widetilde{\theta}\right]$ be anti-comonotonic. Since ϕ' is decreasing, our assumption implies that $\phi'(E\left[u(\widetilde{c}_t) \mid \widetilde{\theta}\right])$ and $E\left[u'(\widetilde{c}_{t\theta}) \mid \widetilde{\theta}\right]$ are comonotonic. By the covariance rule, it implies that

$$\begin{aligned} Eu'(\overset{\circ}{\widetilde{c}_t}) &= \frac{\sum_{\theta=1}^{n} q_\theta \phi'\left(Eu(\widetilde{c}_{t\theta})\right) Eu'(\widetilde{c}_{t\theta})}{\sum_{\theta=1}^{n} q_\theta \phi'\left(Eu(\widetilde{c}_{t\theta})\right)} \\ &\geq \frac{\left[\sum_{\theta=1}^{n} q_\theta \phi'\left(Eu(\widetilde{c}_{t\theta})\right)\right]\left[\sum_{\theta=1}^{n} q_\theta Eu'(\widetilde{c}_{t\theta})\right]}{\sum_{\theta=1}^{n} q_\theta \phi'\left(Eu(\widetilde{c}_{t\theta})\right)} \\ &= \sum_{\theta=1}^{n} q_\theta Eu'(\widetilde{c}_{t\theta}) = Eu'(\widetilde{c}_t). \end{aligned}$$

In order to prove that $1 \implies 2$, suppose by contradiction that $Eu(\widetilde{c}_{t1}) < Eu(\widetilde{c}_{t2}) < ... < Eu(\widetilde{c}_{tn})$, but there exists $\theta \in [1, n-1]$ such that $Eu'(\widetilde{c}_{t\theta}) \leq Eu'(\widetilde{c}_{t\theta+1})$. Then, consider any increasing and concave ϕ that is locally linear for all $U \leq Eu(\widetilde{c}_{t\theta})$ and for all $U \geq Eu(\widetilde{c}_{t\theta+1})$, and has a strictly negative derivative in between these bounds. For any such function ϕ, we have that $\phi'(E\left[u(\widetilde{c}_t) \mid \widetilde{\theta}\right])$ and $E\left[u'(\widetilde{c}_{t\theta}) \mid \widetilde{\theta}\right]$ are anti-comonotonic. Using the covariance rule as above, that implies that $Eu'(\overset{\circ}{\widetilde{c}_t}) < Eu'(\widetilde{c}_t)$, a contradiction.■

6.1 The CARA case

By consequence of Lemma 3, in order to sign the pessimism effect, we need to look for conditions such that u and $-u'$ indeed "agree" on a ranking of lotteries $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$. Consider first an agent who satisfies constant absolute risk aversion (CARA), i.e.,

$$u(c) = -\frac{1}{A}\exp(-Ac)$$

Since $u'(c) = \exp(-Ac)$, functions u and $-u'$ represent the same preferences over priors. Hence the following result.

Proposition 3 *Under CARA preferences, the pessimism effect always reduces the socially efficient discount rate.*

6.2 The general case

One might conjecture that the previous result extends to any concave u. That is, if both u and $-u'$ are increasing, it may seem natural that their expectations agree on the ranking of lotteries. Yet, the theory of stochastic dominance tells us that the solution is not that simple. Indeed, a necessary and sufficient condition for any two increasing utility functions on agree on the ranking of two lotteries is that they be ranked along first-degree stochastic dominance (FSD).

However, ranking the priors according to FSD is rather restrictive. It would be desirable to extend this result to a weaker stochastic order. For instance, consider the second-degree stochastic dominance order (SSD). It guarantees that $Ef(\widetilde{c}_{t\theta})$ is increasing in θ for all increasing and concave functions f. Indeed, prudence means precisely that $f = u$ and $f = -u'$ are increasing and concave. Using these conditions, we are able to once more apply Lemma 3 to obtain the following.

Proposition 4 *The pessimism effect reduces the socially efficient discount rate if*

1. *The set of marginals $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ can be ranked according to FSD.*
2. *The set of marginals $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ can be ranked according to SSD and u exhibits prudence.*

Essentially, the previous result exploits results on how changes in risk affect savings decisions under ambiguity-neutrality. Indeed, being risk-averse (prudent) means that an SEU agent would like to save more if an unfair (zero-mean) risk is added to her wealth (Leland, 1968; Drèze and Modigliani, 1972).

In the following proposition, we put forward a third pair of sufficient conditions. Compared to the SSD/prudence requirement, we impose a stronger restriction on utility functions as we replace prudence by the stronger DARA condition. In return we are able to relax SSD to the weaker stochastic order introduced by Jewitt (1989).

Definition 1 *We say that $\widetilde{c}_{\theta'}$ dominates $\widetilde{c}_{\theta}$ in the sense of Jewitt if the following condition is satisfied: for all increasing and concave u, if agent u prefers $\widetilde{c}_{\theta'}$ to $\widetilde{c}_{\theta}$, then all agents more risk-averse than u also prefer $\widetilde{c}_{\theta'}$ to $\widetilde{c}_{\theta}$.*

Of course, from the definition itself, if $\widetilde{c}_{\theta'}$ dominates $\widetilde{c}_{\theta}$ in the sense of SSD, this preference order also holds in the sense of Jewitt, thereby showing that this order is weaker than SSD. Jewitt (1989) shows that distribution function $F_{t\theta'}$ dominates $F_{t\theta}$ in this sense if and only if there exists some w in their joint support $[a, b]$, such that

$$\begin{aligned}
&\int_a^x (F_{t\theta'}(z) - F_{t\theta}(z))dz \geq 0 \qquad \text{for all} \qquad x \in [a, w], \\
&\int_a^w (F_{t\theta'}(z) - F_{t\theta}(z))dz = 0, \\
&\int_a^x (F_{t\theta'}(z) - F_{t\theta}(z))dz \qquad \text{is non-increasing} \quad \text{on } [w, b].
\end{aligned}$$

Hence the result.[8]

Proposition 5 *The pessimism effect reduces the socially efficient discount rate if the set of marginals* $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ *can be ranked according to Jewitt's stochastic order and* u *exhibits decreasing absolute risk aversion.*

Proof: Decreasing absolute risk aversion means that $v = -u'$ is more concave than u in the sense of Arrow-Pratt. By definition of Jewitt's stochastic order, it implies that $Eu(\widetilde{c}_{t\theta'}) \geq Eu(\widetilde{c}_{t\theta})$ implies that $Ev(\widetilde{c}_{t\theta'}) \geq Ev(\widetilde{c}_{t\theta})$, or equivalently, that $Eu'(\widetilde{c}_{t\theta'}) \leq Eu'(\widetilde{c}_{t\theta})$. Using Lemma 3 concludes the proof. ■

Finally, combining Lemma 1 with Propositions 3, 4 and 5 yields our main result.

Proposition 6 *Suppose that the representative agent exhibits non increasing absolute ambiguity aversion (DAAA). Then, ambiguity aversion reduces the socially efficient discount rate if one of the following conditions holds:*

1. *The set of marginals* $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ *can be ranked according to FSD and* u *is increasing and concave.*
2. *The set of marginals* $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ *can be ranked according to SSD and* u *is increasing, concave, and exhibits prudence.*
3. *The set of marginals* $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ *can be ranked according to Jewitt (1989) and* u *is increasing and concave, and exhibits DARA.*
4. u *exhibits constant absolute risk aversion.*

Observe that the result in our analytical example in Section 3 fits condition *1*: A mere translation in the distribution constitutes a first-degree stochastic dominance. Yet, in many circumstances, the degrees of riskiness also differ across the plausible distributions, usually implying that the plausible prior distributions cannot be ranked according to FSD. Condition *2* provides a sufficient condition on risk attitudes if marginals can only be ranked according to second-degree stochastic dominance, which contains Rothschild-Stiglitz's increases in risk as a particular case. It turns out that in this case, in addition to risk-aversion, the representative agent should also be prudent. Note that even the weaker Jewitt-ordering from condition *3* only requires decreasing absolute risk aversion. This property is widely accepted in the economic literature. It is in particular compatible with the observation that more wealthy individuals tend to take more portfolio risk.[9]

[8] Two random variables fulfill Definition 1 if there exists a consumption level w in their support such that, conditional on the outcome being lower than w, $F_{t\theta'}$ dominates $F_{t\theta}$ in the sense of SSD, whereas conditional on the outcome being higher than w, $F_{t\theta'}$ dominates $F_{t\theta}$ in the sense of FSD.

[9] In counterexample 1, the two random variables $\widetilde{c}_{t1}$ and $\widetilde{c}_{t2}$ cannot be ranked according to SSD. This is why we obtain that ambiguity aversion raises the interest rate in spite of the fact that $u'(c) = c^{-2}$ is convex.

7 The comparative statics of an increase in ambiguity aversion

Our results up to now characterize the effect of smooth ambiguity aversion on the equilibrium interest rate, starting from the ambiguity-neutral benchmark. A natural question to ask is whether our results hold for any increase in ambiguity aversion.

For this purpose, consider two economies, $i = 1, 2$, identical up to the level of ambiguity aversion, with the agent in $i = 2$ more ambiguity-averse. That is to say that $\phi_2(U) = \psi(\phi_1(U))$ for all U, with the property that ψ is increasing and concave. According to the adjusted pricing formula in (9) an increase in ambiguity aversion decreases the social discount rate if and only if

$$a_2 Eu'(\tilde{c}_t^2) \geq a_1 Eu'(\tilde{c}_t^1), \tag{14}$$

where a_i is defined as in (10) with ϕ being replaced by ϕ_i, and where $\tilde{c}_t^i$ is random future consumption distorted by weights q_θ^i, as in (11). Naturally, taking ϕ_1 linear, we retrieve condition (12) from the SEU benchmark.

At the outset, we are able to generalize our findings about the pessimism effect to any increase in ambiguity aversion.

Lemma 4 *The following two conditions are equivalent:*

1. *Beliefs q^2 are dominated by q^1 in the sense of the monotone likelihood ratio order for any set of marginals $(\widetilde{c}_{t1}, ..., \widetilde{c}_{tn})$ such that $Eu(\widetilde{c}_{t1}) \leq ... \leq Eu(\widetilde{c}_{tn})$.*

2. *Agent $i = 2$ is more ambiguity-averse than $i = 1$.*

Proof: Note that we need to find that $q_{\tilde{\theta}}^2 / q_{\tilde{\theta}}^1$ and $\tilde{\theta}$ are anti-comonotonic. Using (11), we can rewrite the ratio as

$$\frac{q_\theta^2}{q_\theta^1} = \psi'\left(\phi_1(Eu(\widetilde{c}_{t\theta}))\right) \frac{\sum_{\tau=1}^n q_\tau \phi_1'\left(Eu(\widetilde{c}_{t\tau})\right)}{\sum_{\tau=1}^n q_\tau \phi_2'\left(Eu(\widetilde{c}_{t\tau})\right)}.$$

The fraction on the RHS does not change with θ. Furthermore, ψ' is decreasing in its argument. Finally, since the argument $\phi_1(Eu(\widetilde{c}_{t\theta}))$ is itself increasing with θ by assumption, we get the desired result.■

Hence, under the stochastic order conditions from Proposition 6, more ambiguity aversion reinforces the pessimism effect, which makes saving more attractive. However, it is clear from section 5 that an increase of ambiguity aversion need not reinforce the ambiguity *prudence* effect a.[10] For small amounts of ambiguity, the relation between ambiguity aversion to ambiguity on a can be approximated in the following way.

[10] For instance, introducing *increasing* absolute ambiguity aversion will in fact raise the interest rate if the representative agent is risk neutral.

Lemma 5 *Consider a family of ambiguous environments parametrized by $k \in R$ and a vector $(u_1, ..., u_n) \in R^n$ such that $Eu(\widetilde{c}_{t\theta}(k)) = u_0 + ku_\theta$ for all θ. Let us define $a(k) = \Sigma_\theta q_\theta \phi'(Eu(\widetilde{c}_{t\theta}(k)))/\phi'(V(k))$, where $\phi(V(k)) = \Sigma_\theta q_\theta \phi(Eu(\widetilde{c}_{t\theta}(k)))$. We have that*

$$a(k) = 1 - \frac{1}{2} Var(ku_{\widetilde{\theta}}) \frac{\partial}{\partial u_0} \left(\frac{-\phi''(u_0)}{\phi'(u_0)} \right) + o(k^2), \tag{15}$$

where $\lim_{k \to 0} o(k^2)/k^2 = 0$.

Proof: Observe first that $V(0) = u_0$, $V'(0) = Eu_{\widetilde{\theta}}$, and $V''(0) = Var(u_{\widetilde{\theta}})\phi''(u_0)/\phi'(u_0)$. Notice also that $a(0) = 1$. We have in turn that

$$a'(k) = \frac{E\left[u_{\widetilde{\theta}}\phi''(u_0 + ku_{\widetilde{\theta}})\right]\phi'(V(k)) - E\left[\phi'(u_0 + ku_{\widetilde{\theta}})\right]\phi''(V(k))V'(k)}{\phi'(V(k))^2}.$$

It implies that $a'(0) = 0$. Differentiating again the above equality at $k = 0$ yields

$$\begin{aligned} {\phi_0'}^2 a''(0) &= E\left[u_{\widetilde{\theta}}^2\right]\phi_0'\phi_0''' + \left(Eu_{\widetilde{\theta}}\right)^2 \phi_0''^2 - \left(Eu_{\widetilde{\theta}}\right)^2 \phi_0''^2 - \\ &\quad - \left(Eu_{\widetilde{\theta}}\right)^2 \phi_0'\phi_0''' - \phi_0'\phi_0'' V''(0) \\ &= \left(E\left[u_{\widetilde{\theta}}^2\right] - \left(Eu_{\widetilde{\theta}}\right)^2\right)\left(\phi_0'\phi_0''' - \phi_0''^2\right), \end{aligned}$$

where $\phi_0^{(i)} = \phi^{(i)}(u_0)$. This implies that

$$a''(0) = -Var(u_{\widetilde{\theta}}) \frac{\partial}{\partial u_0} \left(\frac{-\phi''(u_0)}{\phi'(u_0)} \right).$$

The Taylor expansion of a yields $a(k) = a(0) + ka'(0) + 0.5k^2 a''(0) + o(k^2)$. Collecting the successive derivatives of a concludes the proof.■

Accordingly, for small degrees of ambiguity, a_2 is larger than a_1 if and only if

$$\frac{\partial}{\partial u_0} \left(\frac{-\phi_2''(u_0)}{\phi_2'(u_0)} \right) \geq \frac{\partial}{\partial u_0} \left(\frac{-\phi_1''(u_0)}{\phi_1'(u_0)} \right). \tag{16}$$

In others words, if locally, at the ambiguity-free expected utility level u_0, absolute ambiguity aversion decreases more rapidly under ϕ_2 than under ϕ_1 the ambiguity prudence effect is more negative in economy 2.

Unfortunately, as the following example shows, even if condition (16) holds for all u_0, this is not sufficient to guarantee $a_2 \geq a_1$.

> **Counterexample 2.** Let $\phi(U) = U^{1-\eta}/(1-\eta)$ be defined on R^+. Observe that $-\phi''(U)/\phi'(U) = \eta/U$ is positive and decreasing in its domain. Moreover, an increase in η raises both ambiguity aversion, and the speed at which absolute ambiguity aversion decreases with

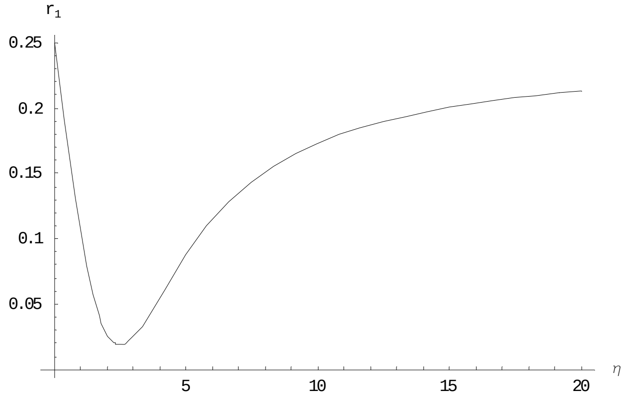


Figure 1: The discount rate as a function of relative ambiguity aversion. We assume that $\phi(U) = U^{1-\eta}/(1-\eta)$, $u(c) = c$, $\delta = 0.25$, $\overline{c}_1 = 0.5$, $\overline{c}_2 = 1.5$ and $p = 0.5$.

U. Proposition 5, yields that a is increasing in η when the risk on U is small. To show that this is not true for large degrees of ambiguity suppose risk-neutrality $u(c) = c$, $n = 2$ equally likely plausible probability distributions with $\overline{c}_1 = 0.5$ and $\overline{c}_2 = 1.5$, and let $\delta = 0.25$. In Figure 1, we draw the socially efficient discount rate r_t for $t = 1$ as a function of the degree of relative ambiguity aversion η. As stated in Proposition 2, we see that the discount rate $r_1(\eta)$ under ambiguity aversion is always smaller than under ambiguity neutrality $(r(0))$. However, the relationship between the discount rate and the degree of ambiguity aversion is not monotone. For example, increasing relative ambiguity aversion from $\eta = 3$ to any larger level *raises* the discount rate. ■

With a counter-example based on the most common family of utility functions $\phi(U) = U^{1-\eta}/(1-\eta)$, there is no hope for convincing sufficient conditions to guarantee an increase in savings. To summarize, we are left with three special cases where signing the effect on a is possible:

i) The degree of ambiguity aversion is small and condition (16) is satisfied;

ii) The initial degree of ambiguity aversion is small, so that Proposition 2 can be used as an approximation;

iii) The initial ϕ_1 function exhibits non decreasing ambiguity aversion, whereas the final ϕ_2 function exhibits non increasing ambiguity aversion. This implies that $a_1 \leq 1 \leq a_2$.

Combining any of these conditions with any of the three conditions from Proposition 6 is sufficient to guarantee that a marginal increase in ambiguity aversion reduces the socially efficient discount rate.

8 Numerical illustrations

8.1 The power-power normal-normal case

As observed in Section 3, we can solve analytically for the socially efficient discount rate by taking a "power-power" specification. That is, CRRA risk preferences and CRAA ambiguity preferences allow for an exact solution if both ambiguity and the logarithm of consumption are normally distributed. For our quantitative analysis, we parametrize the model according to the *quartet of Twos*, as put forward by Weitzman (2007b). We assume a rate of pure preference for the present $\delta = 2\%$, a degree of relative risk aversion $\gamma = 2$, a mean growth rate of consumption $g = 2\%$, and standard deviation of growth $\sigma = 2\%$. We introduce ambiguity by assuming that the growth-trend has a normal distribution with standard deviation $\sigma_0 = 1\%$. In other words, consumers believe that with a 95% probability, the growth trend lies between 0% and 4%. The Ramsey rule (7) implies

$$r_t = 5.88\% - 3\sigma_0^2 t(1 + \eta/2). \tag{17}$$

As usual, in the absence of ambiguity, the Ramsey rule prescribes a flat discount rate of 5.88%. This is no longer true under ambiguity, even for SEU agents ($\eta = 0$), as shown by Weitzman (2007a) and Gollier (2007b).

This is because ambiguity creates fatter tails in the distribution of future consumption. Indeed, ambiguity increases the volatility of log-consumption at date t by $\sigma_0^2 t^2$. Accordingly, the prudent agent wants to save more for the remote future, and the interest rate should fall with the time-horizon. If in addition, the agent exhibits ambiguity aversion, the social discount rate decreases more quickly, as seen in equation (17).

In order to calibrate the model, one needs to evaluate the degree of relative ambiguity aversion η. Consider therefore the following thought experiment.[11] Suppose that the growth rate of the economy over the next 10 years is either 20% – with probability π–, or 0%. Further, suppose that the true value of π is unknown. Rather, it is uniformly distributed on $[0, 1]$, as in the Ellsberg game in which the player has no information on the proportion of black and white balls in the urn.

Let us define the certainty equivalent growth rate $CE(\eta)$ as the sure growth rate of the economy that yields the same welfare as the ambiguous environment described above. It is implicitly defined by the following condition:

$$\left(k\frac{(1+CE)^{1-\gamma}}{1-\gamma}\right)^{1-k\eta} = \int_0^1 \left(k\left(\pi\frac{1.2^{1-\gamma}}{1-\gamma} + (1-\pi)\frac{1^{1-\gamma}}{1-\gamma}\right)\right)^{1-k\eta} d\pi,$$

where γ is set at $\gamma = 2$. In Figure 2, we plot the certainty equivalent as a function of the degree of relative ambiguity aversion. In the absence of ambiguity aversion (or if π is known to be equal to 50%), the certainty equivalent growth

[11] This is based on a 10-year version of the calibration exercise performed by Collard, Mukerji, Sheppard and Tallon (2008), who considered a power-exponential specification.

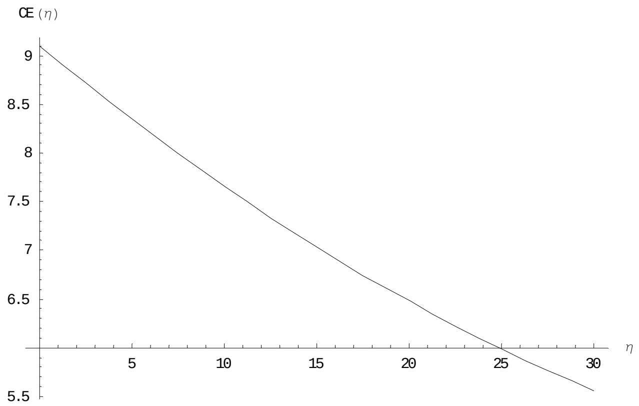


Figure 2: The certainty equivalent growth rate CE (in %) as a function of relative ambiguity aversion η. We assume that the growth rate is either 20% or 0% respectively with probability π and $1-\pi$, with $\pi \sim U(0,1)$. Relative risk aversion equals $\gamma = 2$.

rate equals $CE(0) = 9.1\%$. Surveying experimental studies, Camerer (1999) reports ambiguity premia $CE(0) - CE(\eta)$ in the order of magnitude of 10% of the expected value for such an Ellsberg-style uncertainty. This environment yields a reasonable ambiguity premium of 10%, i.e., a 1% reduction in the growth rate. Thus, ambiguity aversion should reduce the certainty equivalent from 9.1% to around 8%. From Figure 2, this is compatible with a degree of relative ambiguity aversion between $\eta = 5$ and $\eta = 10$.

Table 1 reports the values of efficient rates for projects with maturity 10 and 30 respectively.

Table 1: The social discount rate at the benchmark "quartet of twos", with $\sigma_0 = 1\%$.

t	$\boldsymbol{\eta = 0}$	$\boldsymbol{\eta = 5}$	$\boldsymbol{\eta = 10}$
10	5.58%	4.83%	4.08%
30	4.98%	2.73%	0.48%

While ambiguity aversion has no effect on the short term interest rate, its effect on the long rate is important. The discount rate for a cash flow occurring in 30 years is reduced from 4.98% to 2.73% when relative ambiguity aversion goes from $\eta = 0$ to $\eta = 5$.

The discrepancies between the settings call for an empirical separation between standard risk and ambiguity in an economy. While the former shifts the level of the yield curve, the latter determines its slope. A negative slope increases the relative importance of long-term costs and benefits.

8.2 An AR(1) process for log consumption with an ambiguous long-term trend

Clearly, in our benchmark economy, we abstract from rich consumption dynamics, notably any serial correlation. It is thus not surprising that our predictions do not fare well when confronted with the term structure of interest rates observed on financial markets. Thus, we will relax the assumption of uncorrelated growth rates and allow for persistence of shocks, as in Collard, Mukerji, Sheppard and Tallon (2008) and Gollier (2008). We hereafter show that this model can produce the desired non-linear term structure in the short run and the medium run. While, in the limit, it generates a linearly decreasing term structure in the long run.

Consider first an auto-regressive consumption process of order 1 à la Vasicek (1977), but in which the long-term growth μ of log consumption around which the actual growth mean-reverts is uncertain:

$$\begin{aligned}
\ln c_{t+1} &= \ln c_t + x_t \\
x_t &= \xi x_{t-1} + (1-\xi)\mu + \varepsilon_t \\
\varepsilon_t &\sim N(0,\sigma^2),\ \varepsilon_t \perp \varepsilon_{t'} \\
\mu &\sim N(\mu_0,\sigma_0^2),
\end{aligned} \tag{18}$$

where $0 \leq \xi \leq 1$. That is, system (18) describes an AR(1) consumption process with unknown trend. The polar case without persistence ($\xi = 0$), amounts to the discrete time equivalent of the geometric Brownian motion considered in Section 3 and calibrated here above. In contrast, $\xi = 1$ describes shocks on the growth of log consumption that are fully persistent. Using the same techniques which led us to equation (6), we obtain the following generalization:

$$r_t = \delta + \gamma\frac{EX_t}{t} - \frac{1}{2}\gamma^2\frac{Var\left[X_t \mid \mu\right] + Var\left[E[X_t \mid \mu]\right]}{t} - \frac{1}{2}\eta\left|1-\gamma^2\right|\frac{Var\left[E[X_t \mid \mu]\right]}{t}, \tag{19}$$

where X_t is defined as

$$X_t = \ln c_t - \ln c_0 = \mu t + (x_{-1} - \mu)\frac{\xi(1-\xi^t)}{1-\xi} + \sum_{\tau=1}^{t}\frac{1-\xi^\tau}{1-\xi}\varepsilon_{t-\tau}.$$

It yields

$$\frac{EX_t}{t} = \mu_0 + (x_{-1} - \mu_0)\frac{\xi(1-\xi^t)}{t(1-\xi)},$$

$$\frac{Var\left[X_t \mid \mu\right]}{t} = \frac{\sigma^2}{(1-\xi)^2} + \sigma^2\frac{\xi(1-\xi^t)}{t(1-\xi)^3}\left[\frac{\xi(1+\xi^t)}{1+\xi} - 2\right],$$

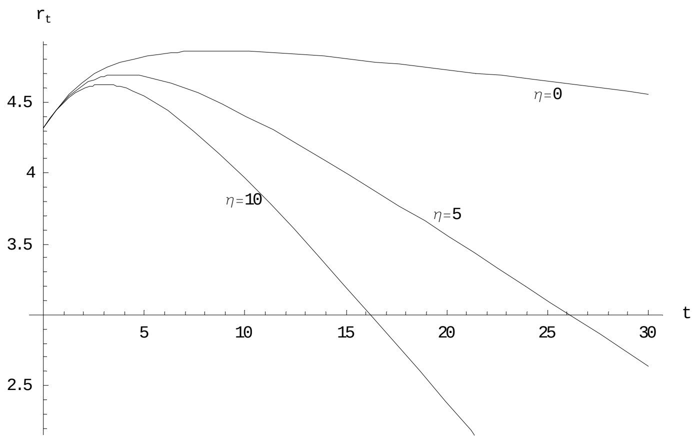


Figure 3: The term structure of discount rates in the case of an AR(1) process with ambiguous long-term trend, $\delta = 2\%$, $\gamma = 2$, $\mu_0 = 2\%$, $\sigma = 2\%$, $\sigma_0 = 1\%$, $x_{-1} = 1\%$, and $\xi = 0.7$.

and

$$\frac{Var\left[E[X_t \mid \mu]\right]}{t} = \frac{\sigma_0^2}{t}\left(t - \frac{\xi(1-\xi^t)}{1-\xi}\right)^2.$$

To illustrate, suppose that $\delta = 2\%$, $\gamma = 2$, $\mu_0 = 2\%$, $\sigma = 2\%$, $\sigma_0 = 1\%$,and $x_{-1} = 1\%$. Following Backus, Foresi and Telmer (1998), suppose also that $\xi = 0.7$ year^{-1}, such that a shock has a half-life of 3.2 years. In Figure 3, we have drawn the term structure of discount rates for 3 different degrees of ambiguity aversion: $\eta = 0$, 5, and 10. We can see that, as in the absence of persistence, the role of ambiguity aversion is to force a downward slope on the yield curve for long time horizons. This is confirmed by the following observation:

$$\lim_{t\to\infty} \frac{\partial r_t}{\partial t} = -\frac{1}{2}\eta\left|1-\gamma^2\right|\sigma_0^2.$$

8.3 An AR(1) process for log consumption with an ambiguous degree of mean reversion

Consider alternatively an auto-regressive consumption process of order 1 with a known long-term trend, but in which the coefficient of mean reversion is unkown:

$$\begin{aligned} \ln c_{t+1} &= \ln c_t + x_t \\ x_t &= \xi x_{t-1} + (1-\xi)\mu + \varepsilon_t \\ \varepsilon_t &\sim N(0,\sigma^2),\ \varepsilon_t \perp \varepsilon_{t'} \\ \xi &\sim U(\underline{\xi},\overline{\xi}). \end{aligned}$$

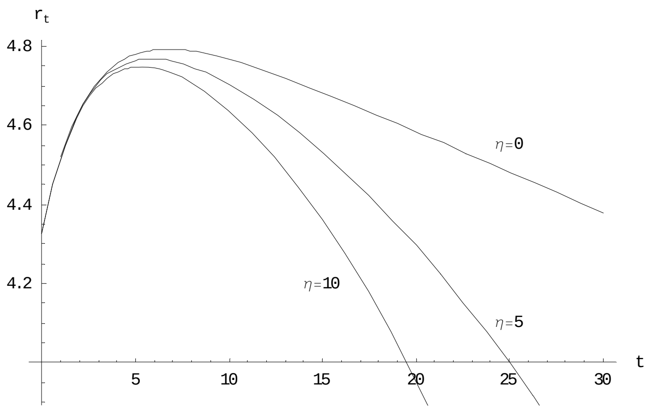


Figure 4: The term structure of discount rates in the case of an AR(1) with an ambiguous mean reversion coefficient, with $\delta = 2\%$, $\gamma = 2$, $\mu = 2\%$, $\sigma = 2\%$, $x_{-1} = 1\%$, and $\xi \sim U(0.5, 0.9)$.

There is no analytical solution for the discount rate, which must be computed numerically by estimating the following two terms, deduced from equation (3) (we normalized $c_0 = 1$):

$$\frac{E\phi'\left(Eu\right)Eu'}{u'(c_0)} = b(E\exp(G))$$

and

$$\phi'(V_t(0)) = b\left(E\exp\left(H\right)\right)^{\frac{-k\eta}{1-k\eta}},$$

with

$$\begin{aligned} G &= -(\gamma + k\eta(1-\gamma))E\left[X_t \mid \xi\right] + \frac{1}{2}\left(\gamma^2 - k\eta(1-\gamma)^2\right)Var\left[X_t \mid \xi\right], \\ H &= (1-k\eta)(1-\gamma)E[X_t \mid \xi] + \frac{1}{2}(1-k\eta)(1-\gamma)^2 Var[X_t \mid \xi]. \end{aligned}$$

In Figure 4 we draw the term structure of the discount rate with the same parameter values as in the previous section, except that $\mu = 2\%$ and $\xi \sim U(0.5, 0.9)$. As before, longer time horizons yields more ambiguity in the set of plausible distributions of consumption, which implies that ambiguity aversion has a stronger negative impact on the discount rates associated to these longer durations.

9 Conclusion

The present paper has shown how ambiguity-aversion affects the efficient rate to discount future costs and benefits of investment projects. In line with recent literature, our analysis suggests that parameter uncertainty might be decisive for long-term policy appraisals. We found that, in general, it is not true that ambiguity aversion decreases the discount rate. However, we identified moderate requirements on risk-attitudes and the statistical relation among prior distributions, such that decreasing ambiguity aversion should induce us to use a smaller discount rate. Our numerical illustrations suggest the effect of ambiguity aversion on the discount rate to be large, in particular for longer time horizons.

References

Backus D., Foresi S. and Telmer C. (1998), Discrete-time models of bond pricing, NBER working paper.

Camerer, C., (1999), Ambiguity-aversion and non-additive probability: Experimental evidence, models and applications, in Uncertain decisions: Bridging theory and experiments, ed. by L. Luini, pp. 53-80, Kluwer Academic Publishers.

Camerer, C. and M. Weber, (1992), Recent developments in modeling preferences: uncertainty and ambiguity, *Journal of Risk and Uncertainty*, 5, 325–370.

Cochrane, J., (2001), *Asset Pricing*, Princeton University Press.

Collard, F., S. Mukerji, K. Sheppard and J.-M. Tallon, (2008), Ambiguity and the historical equity premium, mimeo, Toulouse School of Economics.

Drèze, J.H. and F. Modigliani, (1972), Consumption decisions under uncertainty, *Journal of Economic Theory*, 5, 308–335.

Ellsberg, D., (1961), Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75, 643-669.

Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18, 141–153.

Gollier, C., and M.S. Kimball, (1996), Toward a systematic approach to the economic effects of uncertainty: characterizing utility functions, Discussion paper, University of Michigan.

Gollier, C., (2001), *The Economics of Risk and Time*, MIT Press, Cambridge.

—— (2002), Time horizon and the discount rate, *Journal of Economic Theory*, 107, 463-473.

—— (2006), Does ambiguity aversion reinforce risk aversion? Applications to portfolio choices and asset pricing, IDEI Working Paper, n. 357.

—— (2007), Whom should we believe? Aggregation of heterogeneous beliefs, *Journal of Risk and Uncertainty*, 35, 107-127.

—— (2008), Discounting with fat-tailed economic growth, *Journal of Risk and Uncertainty*, 37, 171–186.

Groom, B., P. Koundouri, E. Panopoulou and T. Pantelidis, (2004), Model selection for estimating certainty equivalent discount, mimeo, UCL, London.

Hansen, L.P. and K. Singleton, (1983), Stochastic consumption, risk aversion and the temporal behavior of assets returns, *Journal of Political Economy*, 91, 249-268.

Jewitt, I., (1989), Choosing between risky prospects: the characterization of comparative statics results, and location independent risk, *Management Science*, 35, 60–70.

Jouini, E., J.-M. Marin, and C. Napp (2008), Discounting and divergence of opinion, mimeo.

Ju, N. and J. Miao (2007). Ambiguity, learning, and asset returns, AFA 2009 San Francisco meetings paper.

Kimball, M.S., (1990), Precautionary savings in the small and in the large, *Econometrica*, 58, 53–73.

Klibanoff, P., M. Marinacci, and S. Mukerji, (2005), A smooth model of decision making under ambiguity. *Econometrica*, 73(6), 1849–1892

—— (2009), Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144, 930-976.

Lehmann, E., (1955), Ordered families of distributions, *Annals of Mathematical Statistics*, 26, 399–419.

Leland, H., (1968), Savings and uncertainty: The precautionary demand for savings, *Quarterly Journal of Economics*, 45, 621–36.

Lomborg, B., (2004, ed.), *Global Crises, Global Solutions. Copenhagen Consensus Challenge.* Cambridge University Press.

Lucas, R., (1978), Asset prices in an exchange economy, *Econometrica*, 46, 1429–1446.

Quiggin, J., (1995), Economic choice in generalized utility theory, *Theory and decision*, 38, 153-171.

Ramsey, F.P., (1928), A mathematical theory of savings, *The Economic Journal*, 38, 543-59.

Rothschild, M. and J. Stiglitz, (1970), Increasing risk: I. A definition, *Journal of Economic Theory*, 2, 225–243.

Savage, L.J., (1954), *The foundations of statistics*, New York: Wiley. Revised and Enlarged Edition, New York: Dover (1972).

Stern, N., (2007), *The Economics of climate change: The Stern review.* Cambridge University Press, Cambridge.

Vasicek, 0., (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-188.

Weitzman, M.L., (2007a), Subjective expectations and asset-return puzzles, *American Economic Review*, 97, 1102–1130.

—— (2007b), A review of the Stern Review on the economics of climate change, *Journal of Economic Literature*, 45, 703–724.

Appendix

Proof of Proposition 1. In order to prove this result, we need the following Lemma, which is Theorem 106 in Hardy, Littlewood and Polya (1934), Proposition 1 in Polak (1996), and Lemma 8 in Gollier (2001).

Lemma 6 *Consider a function $\phi : \mathbb{R} \to \mathbb{R}$, twice differentiable, increasing and concave. Consider a vector $(q_1, ..., q_n) \in \mathbb{R}^n_+$ with $\sum_{j=1}^n q_j = 1$, and a function f from $\mathbb{R}^n$ to $\mathbb{R}$, defined as*

$$f(U_1, ..., U_n) = \phi^{-1}(\sum_{\theta=1}^{n} q_\theta \phi(U_\theta)).$$

Define function T such that $T(U) = -\frac{\phi'(U)}{\phi''(U)}$. Function f is concave in $\mathbb{R}^n$ if and only if T is weakly concave in $\mathbb{R}$.

Having established the above, consider two scalars α_1 and α_2 and let us denote $U_{i\theta} = Eu(\widetilde{c}_{t\theta} + \alpha_i e^{r_t t})$. Using the notation introduced in the Lemma, it implies that $V_t(\alpha_i) = f(U_{i1}, ..., U_{in})$. Because u is concave, we have that, for any (λ_1, λ_2) such that $\lambda_i \geq 0$ and $\lambda_1 + \lambda_2 = 1$,

$$\begin{aligned} \lambda U_{1\theta} + \lambda_2 U_{2\theta} &= E\left[\lambda_1 u(\widetilde{c}_{t\theta} + \alpha_1 e^{r_t t}) + \lambda_2 u(\widetilde{c}_{t\theta} + \alpha_2 e^{r_t t})\right] \\ &\leq Eu(\widetilde{c}_{t\theta} + \alpha_\lambda e^{r_t t}) =_{def} U_{\lambda\theta}, \end{aligned}$$

for all θ, where $\alpha_\lambda = \lambda_1\alpha_1 + \lambda_2\alpha_2$. Because f is increasing in R^n, this inequality implies that

$$\begin{aligned} V_t(\lambda_1\alpha_1 + \lambda_2\alpha_2) &= f(U_{\lambda 1}, ..., U_{\lambda n}) \\ &\geq f(\lambda_1 U_{11} + \lambda_2 U_{21}, ..., \lambda_1 U_{1n} + \lambda_2 U_{2n}). \end{aligned} \tag{20}$$

Suppose that $-\phi'/\phi''$ be concave. By the Lemma, it implies that

$$\begin{aligned} f(\lambda_1 U_{11} + \lambda_2 U_{21}, ..., \lambda_1 U_{1n} + \lambda_2 U_{2n}) &\geq \lambda_1 f(U_{11}, ..., U_{1n}) + \lambda_2 f(U_{21}, ..., U_{2n}) \\ &= \lambda_1 V_t(\alpha_1) + \lambda_2 V_t(\alpha_2). \end{aligned} \tag{21}$$

Combining equations (20) and (21) yields $V_t(\lambda_1\alpha_1 + \lambda_2\alpha_2) \geq \lambda_1 V_t(\alpha_1) + \lambda_2 V_t(\alpha_2)$, i.e., V_t is concave in α. ■

Does flexibility enhance risk tolerance?

Christian Gollier[1]
University of Toulouse

June 30, 2010

[1]This research was supported by the Chair of the Financière de la Cité at TSE.

Abstract

The optimal decision under risk depends upon the ability of the decision makers to adapt their actions to the state of nature ex post. We examine two choice problems, one in which agents select their action ex post, and one in which they must commit on their action ex ante. Contrary to the intuition, it is not true in general that agents are more tolerant to risk in the flexible context than in the rigid one. We provide some sufficient conditions to guarantee that the optimal risk exposure is larger in the flexible context. We apply these results to examine various questions. In particular, we examine the effect of housing and labour markets rigidities and of rigid long-term saving plans on the demand for equity. We also compare the optimal portfolios in continuous and discrete time.

Keywords: dynamic portfolio choice, risk sharing, continuous versus discrete time

1 Introduction

Arrow (1963) and Pratt (1964) were the first to relate the optimal decision under risk to the shape of the decision marker's utility function on consumption. They assumed that the only decision is about which lottery to accept, and that the outcome of the lottery is immediately consumed. Drèze and Modigliani (1966, 1972), Mossin (1969) and Spence and Zeckhauser (1972) noticed that this is rarely a realistic assumption. Most agents will react to the outcome of the lottery by making additional decisions. If they find it optimal ex post, households holding stocks will react to a crash on financial markets by moving to a smaller house, by working more, by reducing their stock holding, or by reducing their saving. In other words, they will rarely compensate their immediate financial loss by a corresponding reduction in their consumption. The expectation of these ex post actions affects the optimal attitude towards risk ex ante. Mossin (1968), Merton (1969) and Samuelson (1969) used backward induction to determine the optimal portfolio when investors are fully flexible both on their saving decision and on their portfolio rebalancement strategy.

In this paper, we examine a quite general dynamic choice problem to determine the effect of flexibility on the optimal risk attitude. We compare the choices under risk in two different contexts. In the flexible context, the agent first chooses a lottery in a choice set, and then takes an action x after observing the outcome of the lottery. In the rigid context, the agent must commit on an action before observing the state of nature. The intuition suggests that the agent should be more risk-prone in the flexible context than in the rigid one.[1] Spence and Zeckhauser (1972) showed this in a numerical example in which the Cobb-Douglass agent's ex post choice problem is a standard consumption choice with two goods. Using the same framework in continuous time, Bodie, Merton and Samuelson (1992) and Chetty and Szeidl (2003) showed that flexibility enhances risk tolerance for a subset of the parameter values of the Cobb-Douglass specification. Postelwaite, Samuelson and Silverman (2005) and Flavin and Nakagawa (2005) obtained the same kind of result for CES utility functions. The same result holds when the agent's utility function is log linear rather than Cobb-Douglass. Thus,

[1]This point is related to the irreversibility effect first developed by Henry (1974) and Arrow and Fischer (1974). However, the irreversibility effect is about the strategic preservation of more flexibility for the future.

rigidities on housing and labor markets can potentially explain the equity premium puzzle that arises when we assume full flexibility.

However, Machina (1982) used an exponential-quadratic utility function to show numerically that it is not true in general that flexibility induces a more risk-prone behavior. It may thus be possible that the loss of flexibility on housing and labor decisions raises the demand for risky assets. Our contribution to this literature is twofold. We first derive some sufficient conditions that yield an unambiguous effect on the risky choice. In the second part of the paper, we discuss various applications of these general results.

As explained by Drèze and Modigliani (1972), the intuition that flexibility enhances risk tolerance is based on the well-known result that, in the flexible EU context, the value of information is always non-negative. It implies that the certainty equivalent of any lottery is larger in the flexible context than in the rigid one. If the choice problem is to choose between a risky prospect or a risk free prospect, any lottery that is acceptable in the rigid context is also acceptable in the flexible one. However, as for example explained in Gollier (2001, chapter 6), the enlargement of the lottery acceptance set due to flexibility does not mean that risk aversion is globally reduced. For example, it does not imply that the demand for risky asset is reduced. The enlargement of acceptable lottery just implies that flexibility reduces risk aversion locally around the initial safe wealth. This is because the direction of the adaptation of the ex post action always reduces the sensitiveness of the marginal utility to change in wealth. This direct effect of flexibility may however be dominated by an "action" effect. It comes from the fact that the action selected ex ante in the rigid context may yield a low aversion to risk on wealth.

We are mostly interested in determining the effect of various rigidities on the optimal one-risky-one-risk-free-asset portfolio. More generally, we examine a class of ex ante decision problems under uncertainty in which the payoff is linear: $\widetilde{z} = z_0 + \alpha\widetilde{r}$, as is the case in the portfolio problem, the coinsurance problem, or the capacity problem of the firm under price uncertainty. For this class of ex ante choice problems, if the objective function is supermodular or submodular, the optimal risk exposure α in the flexible context is larger than in a rigid context in which the action is selected as if wealth would equal z_0 with certainty. Thus, when the option to invest in stocks has a negligible effect on the optimal action in the rigid context, super/submodularity is sufficient to imply that flexibility enhances risk tolerance. Because we also show that the option to invest is stocks only yields a second-order effect on the optimal rigid action, this result is useful in many applications.

The most obvious applications of this theory are when the second stage choice problem is a standard consumption allocation problem under certainty. After observing their portfolio return, consumers decide how to allocate their wealth into consumption for various goods. The existence of durable goods like housing, or the rigidities existing on the labor market, have an effect on the consumers' optimal portfolio allocation. Shroyen (2010) shows that the rigidity always raises the local risk aversion when the exogenous level of consumption of the rigid good happens to be optimal at the wealth level under consideration. When preferences are Cobb-Douglass, the optimal action in the rigid context is independent of the option to invest in the risky asset. If the parameters of the Cobb-Douglass utility function yield a value function that is less risk-averse than the log, the utility function is supermodular. Therefore, we can use the general result summarized above to claim that flexibility always raises the demand for stocks under these conditions, as shown by Bodie, Merton and Samuelson (1992). We also show that when the parameters of the Cobb-Douglass preferences yield a value function that is more risk-averse than the log, it is always possible to find an initial wealth and a distribution of excess returns such that flexibility reduces the demand for the risky asset.

Following Gabaix and Laibson (2002), we also examine the effect of reducing the frequency at which the consumption plan is reoptimized given the observation of the portfolio return since the previous reset date. There are several potential reasons for why consumers are unable to adapt their consumption level to shocks on their wealth in continuous time, as the existence of transaction costs or the willingness to commit on a specific saving plan to solve a time inconsistency problem. We examine a dynamic choice problem in which the second stage problem is a consumption/saving problem under certainty. Time-additive intertemporal utility functionals are supermodular. We show that the reduction of the frequency of reoptimizations of the consumption plan yields a reduction in the demand for stocks when the utility function exhibits harmonic absolute risk aversion (HARA), a class of preferences that includes power, exponential and logarithmic functions. The same model can be used to determine the effect of the households' inability to internally share risk efficiently on the demand for stocks.

We also apply our general model to the analysis of the reduction of the frequency at which the portfolio is rebalanced. We show that it reduces the demand for stocks when relative risk aversion is constant and less than unity. When relative risk aversion is larger than unity, we can always find a distribution of excess returns such that the reduction of rebalancement frequency raises the demand for stocks. This analysis sheds some light on the relationship between discrete time and continuous time in the modern theory of finance.

2 Some general results

The utility U of the decision-maker depends upon wealth $z \in \mathbb{R}$ and a decision x belonging to a decision set A. Function $U : \mathbb{R} \times \mathbb{A} \to \mathbb{R}$ is increasing in z and concave in x. We consider two dynamic choice problems. In the flexible framework, the agent first selects a risk-taking decision. Once the outcome z of the risk exposure is observed, the agent selects the $x \in A$ that maximizes U :

$$v(z) = \max_{x \in A} U(z, x). \tag{1}$$

Because U is concave in x, the solution $x^*(z)$ of the above program is unique. In this flexible framework, the attitude on risk in the first stage of the decision problem is fully characterized by the indirect utility function v. In other words, the first-stage choice among various lotteries can be rationalized by an expected utility functional

$$V^*(\widetilde{z}) = Ev(\widetilde{z}),$$

where $\widetilde{z}$ is the random variable describing the distribution of wealth. Because the operator max is convex, the value function v is not necessarily concave.

The alternative framework is rigid in the sense that the choice of $x \in A$ cannot be made sensitive to the realization of $\widetilde{z}$. In this rigid framework, the choice of x must be made before observing the realization of z. The choice among various lotteries is rationalized in the rigid framework by the following functional:

$$\widehat{V}(\widetilde{z}) = \max_{x \in A} EU(\widetilde{z}, x). \tag{2}$$

Because concavity is preserved by summation, the solution $\widehat{x}(\widetilde{z})$ of program (2) is unique. As first stated by Mossin (1969), Drèze and Modigliani (1972) and Spence and Zeckhauser (1972), $\widehat{V}$ is not an expected utility functional. In particular, $\widehat{V}$ violates the independence axiom. It implies that contrary to what we have in the flexible framework, the risk attitude in the rigid world cannot be expressed by the concavity of a von Neumann-Morgenstern utility function. This is an important source of complexity of the analysis.

The main objective of the paper is to determine the conditions under which flexibility enhances risk tolerance. To be more precise, consider a specific first-stage choice problem under uncertainty that is specified by a risk opportunity set S, a subset of real-valued random variables. Let $\widehat{\widetilde{z}} \in S$ and $\widehat{x} = \widehat{x}(\widehat{\widetilde{z}}) \in A$ be respectively the optimal risk on wealth and the optimal choice in the rigid context. The problem is to determine whether the optimal risk in the flexible framework is riskier than $\widehat{\widetilde{z}}$? To answer this question, one should compare the concavity of functions $v(.)$ and $U(., \widehat{x})$. Notice that it is essential here to specify not only function U, but also the risk opportunity set S. A change in S potentially modifies the optimal $\widehat{x}$ and the reference function $U(., \widehat{x})$. In the remainder of this section, we consider two first-stage choice problems: the acceptance of a lottery, and a portfolio allocation problem.

2.1 The first stage problem is a binary choice

In the following Proposition, we consider the simplest risk opportunity set, which is limited to two random variables, one of which being degenerated. We show that flexibility always enlarges the set of acceptable risks. It is a direct consequence of the well-known property that the value of information is nonnegative. This value of information raises the willingness to accept risk.

Proposition 1 *For any pair* $(z_0, \widetilde{z})$, z_0 *being degenerated,* $\widehat{V}(\widetilde{z}) \geq \widehat{V}(z_0)$ *implies* $V^*(\widetilde{z}) \geq V^*(z_0)$.

Proof: Suppose that $\widehat{V}(\widetilde{z}) \geq \widehat{V}(z_0)$. The proof is directly obtained by the following sequence of relations:

$$\begin{aligned} V^*(\widetilde{z}) &= E \max_{x \in A} U(\widetilde{z}, x) \\ &\geq EU(\widetilde{z}, \widehat{x}(\widetilde{z})) = \widehat{V}(\widetilde{z}) \\ &\geq \widehat{V}(z_0). \qquad \blacksquare \end{aligned}$$

In this binary riskfree-risky choice problem, we can evaluate the effect of flexibility by estimating the difference in the risk premia. Let us consider a risk-free wealth z_0 and the introduction of a small risk $k\widetilde{\varepsilon}$ with $E\widetilde{\varepsilon} = 0$. In the flexible context, the maximum premium $\pi^*(k)$ that the decision-maker is ready to pay to eliminate risk $k\widetilde{\varepsilon}$ satisfies the following condition:

$$Ev(z_0 + k\widetilde{\varepsilon}) = v(z_0 - \pi^*(k)). \tag{3}$$

In the rigid context, the risk premium $\widehat{\pi}(k)$ associated to risk $k\widetilde{\varepsilon}$ is defined by

$$\max_{x \in A} EU(z_0 + k\widetilde{\varepsilon}, x) = \max_{x \in A} U(z_0 - \widehat{\pi}(k), x). \tag{4}$$

In Proposition 2, we show that the aversion to small risk is larger in the rigid context than in the flexible one.

Proposition 2 *Suppose that U is twice differentiable and that $A = \mathbb{R}$. The risk premia associated to zero-mean risk $k\widetilde{\varepsilon}$ respectively in the rigid and flexible contexts satisfy the following properties:*

$$\widehat{\pi}(k) = \frac{1}{2}k^2\sigma_\epsilon^2 \left[\frac{-U_{zz}(z_0, x^*(z_0))}{U_z(z_0, x^*(z_0))}\right] + o(k^2) \tag{5}$$

$$\pi^*(k) = \frac{1}{2}k^2\sigma_\epsilon^2 \left[-\frac{v''(z_0)}{v'(z_0)}\right] + o(k^2), \tag{6}$$

with

$$-\frac{v''(z)}{v'(z)} = -\frac{U_{zz}(z, x^*(z))}{U_z(z, x^*(z))} + \frac{U_{zx}^2(z, x^*(z))}{U_z(z, x^*(z))U_{xx}(z, x^*(z))}. \tag{7}$$

Consequently, in the case of small risk, flexibility always reduces risk aversion.

Proof: See the Appendix. ■

In the rigid context, we see that the risk premium for small risk is not different from what one would obtain with a completely exogenous x that would be set equal to $x^*(z_0)$, the optimal action when $z = z_0$ with certainty. Indeed, equation (5) tells us that the premium associated to small risk $k\widetilde{\varepsilon}$ is the one that a von Neumann-Morgenstern agent with utility $\widehat{v}_0(.) = U(., x^*(z_0))$ would be ready to pay. The fact that the choice of action x is affected by the existence of risk $k\widetilde{\varepsilon}$ has only a second-order effect on $\widehat{\pi}$. This is because, as noticed in the proof, $d\widehat{x}/dk = 0$ when evaluated at $k = 0$: the introduction of a small risk does not affect the optimal action at the margin. In consequence, as clearly explained by Machina (1984), the expected utility approximation (5) holds in the rigid context. The preference functional $\widehat{V}$ is Fréchet differentiable at z_0 with local utility function $U(., x^*(z_0))$.

The picture is completely different in the flexible context. The risk aversion on wealth is determined by the Arrow-Pratt index $-v''/v'$ of the indirect utility function v . As shown by Machina (1984), the absolute aversion to temporal risk is the sum of two terms that are expressed in the right-hand side of (7). Evaluated at $z = z_0$, the first term is the absolute risk aversion of $\widehat{v}_0(.) = U(., x^*(z_0))$, as in the rigid context. The second term in the right-hand side of (7) originates from the flexibility of action $x^*(z)$. It is always negative. Thus, flexibility always reduces local risk aversion.

The above two propositions suggest that flexibility always leads to a globally less risk-averse behavior. However, as noticed by Machina (1982), this is not true in general. To see this, notice that $v(.)$ is less concave than $U(., \widehat{x})$ in the sense of Arrow-Pratt if

$$\frac{-zU_{zz}(z, x^*(z))}{U_z(z, x^*(z))}\,[1 - F(z)] \leq \frac{-zU_{zz}(z, \widehat{x})}{U_z(z, \widehat{x})} \tag{8}$$

for all z, where $F(z)$ is an index of "flexibility tolerance" that is defined by:

$$F(z) = \frac{-U_{zx}^2(z, x^*(z))}{U_{zz}(z, x^*(z))U_{xx}(z, x^*(z))} \geq 0. \tag{9}$$

We see that flexibility has two effects on risk aversion. The direct effect of flexibility is expressed by the multiplicative term $(1 - F) < 1$ in the left-hand side of (8), as explained above. We qualify this effect to be "direct" because it directly comes from the fact that the action x^* is sensitive to the outcome z. It is useful to see why the direct flexibility effect always reduces risk aversion. Suppose that U_{zx} is positive. It implies that x^* is increasing in z. It implies that an increase in z raises x^*, which in turn tends to raise U_z in the flexible context. Similarly, if U_{zx} is negative, an increase in z reduces x^* and this effect raises U_z. In both cases, the direct flexibility effect on the marginal utility of wealth is to make it less sensitive to wealth. In short, it reduces risk aversion.

The second effect comes from the fact that the absolute risk aversion of U at z is evaluated at $x^*(z)$ in the flexible context, and at $\widehat{x}$ in the rigid one. Let us refer to this second effect as the "action effect". It can be either positive or negative. In Figure 1, these two effects can easily be isolated. Considering a specific z, the action effect is obtained by comparing the concavity indexes of the two plain curves corresponding to $U(., \widehat{x})$ and $U(., x^*(z))$. The direct flexibility effect is obtained by comparing the concavity indexes of the plain curve $U(., x^*(z))$ and the dashed curve $v(.)$.

Flexibility would raise risk aversion if the unambiguously risk-prone direct effect of flexibility is more than compensated by the action effect. Whereas Proposition 2 states that this cannot be the case for small risks because $\widehat{x} = x^*(z)$, we cannot exclude this possibility for larger risks.

2.2 The first stage problem is a portfolio choice

If v is not globally less risk-averse than $U(., \widehat{x})$, we cannot sign the effect of flexibility on the optimal portfolio allocation. We illustrate this point by the following example. We consider a two-stage decision problem under uncertainty. In the first stage, the agent is endowed with wealth $z_0 = 1$. He invests in a portfolio of two assets, one of which is risk-free with a zero return. The other asset has a return distributed as $\widetilde{r} \sim (-1, 1/10; 10, 9/10)$. In the second stage, after observing the outcome of this investment, the agent can either accept or reject a lottery $\widetilde{y} \sim (-0.2, 1/2; 1, 1/2)$. The agent has constant relative risk aversion $\gamma = 4$.[2] It can be shown that the optimal strategy in the first stage consists in investing $\alpha = 14.98\%$ of z_0 in the risky asset. In the second stage, the agent accepts lottery $\widetilde{y}$ only if the risky asset return is positive. Suppose alternatively that the two decisions must be made

[2]Notice that this model is characterized by an objective function $U(z, x) = E\left[z + x\widetilde{y}\right]^{1-\gamma}/(1-\gamma)$ and a choice set $A = \{0, 1\}$.

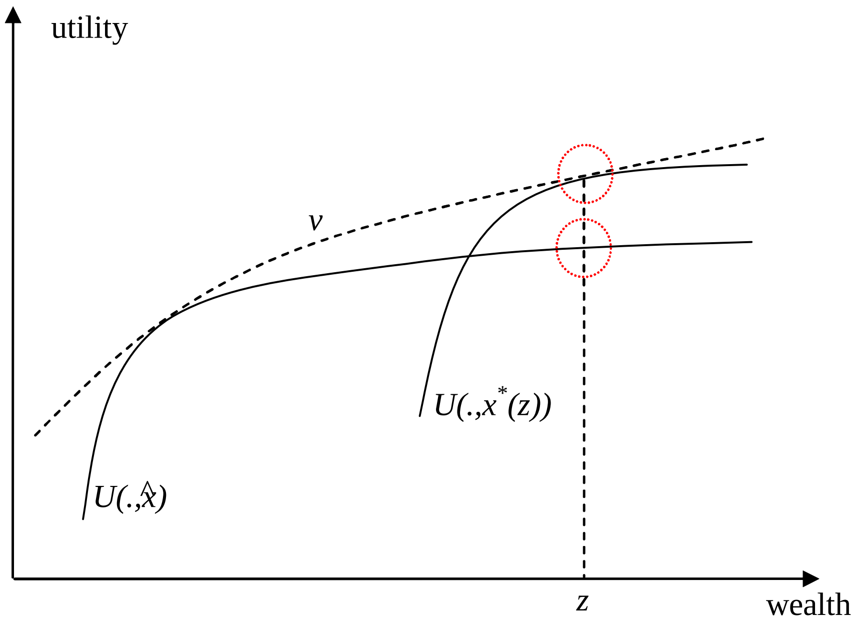


Figure 1: The value function and utility functions conditional to choice $x^*(z)$ and $\widehat{x}$.

simultaneously, i.e., that the lottery decision cannot be made after observing the portfolio return. In this rigid context, it can be shown that the optimal decision is to reject the lottery and to invest a share $\alpha = 15.90\%$ of z_0 in the risky asset. From the observation of the demand for the risky asset, it appears that the investor is more averse towards portfolio risk in the flexible context than in the rigid one. Flexibility does not enhance risk tolerance.

As in the above example, we will hereafter focus much of our attention on the one-risk-free-one-risky-asset choice problem in the first stage. In the flexible context, this first stage problem can be written as

$$\alpha^* = \arg\max_{\alpha} Ev(z_0 R + \alpha \widetilde{r}), \tag{10}$$

where z_0 is initial wealth, $\widetilde{r}$ is the excess return of the risky asset, R is the gross risk-free rate and α is the demand for this asset.[3] We hereafter suppose without loss of generality that the $E\widetilde{r} > 0$, which implies that α^* is positive. In the following Lemma, we show that the demand for the risky asset in the flexible context is larger than when the action x is exogenously fixed at $x^*(z_0)$.

Proposition 3 *Consider the first stage portfolio problem described by program (10). Suppose that U is concave in its first argument. If U is either supermodular or submodular, then α^* is larger than $\alpha_0 = \arg\max_\alpha EU(z_0R+ \alpha\widetilde{r}, x^*(z_0))$, i.e., the demand for the risky asset in the flexible context is larger than when action x is exogenously fixed at $x^*(z_0)$, its optimal level without the risky asset.*

Proof: Suppose that U is supermodular (submodular). By Topkis' theorem (Topkis (1978)), $x^*(z)$ is nondecreasing (nonincreasing) in z. Because U_z is nondecreasing (nonincreasing) in x, it implies that

$$rU_z(z_0R + \alpha^* r, x^*(z_0)) \le rU_z(z_0R + \alpha^* r, x^*(z_0 + \alpha^* r)) = rv'(z_0R + \alpha^* r)$$

for all r. Taking the expectation, we obtain that

$$E\widetilde{r}U_z(z_0R + \alpha^*\widetilde{r}, x^*(z_0)) \le E\widetilde{r}v'(z_0R + \alpha^*\widetilde{r}) = 0.$$

[3]Many other economic problems can be written in that way, as the coinsurance problem, the capacity problem of the risk-averse firm under price uncertainty, or various risk prevention problems.

Because U is concave in z, this inequality implies that the optimal demand for risky asset is less than α^* when x is exogenously fixed at $x^*(z_0)$. ■

This Proposition sheds some light on the above numerical example. It can be shown that the optimal choice when $z = z_0$ with certainty is to reject lottery $\widetilde{y}$, i.e., $a_0 = 0$. The optimal decision to invest $\alpha = 15.90\%$ of initial wealth in the risky asset does not affect the decision to reject this lottery. The fact that $U(z, x) = E\left[z + x\widetilde{y}\right]^{1-\gamma}/(1-\gamma)$ is neither supermodular nor submodular explains why the demand $\alpha^* = 14.98\%$ for the risky asset is not larger than in the rigid context.

Proposition 3 has some interest on its own, but it does solve our problem only when the optimal action in the rigid context is close to $x^*(z_0)$. But in general, the option to invest in the risky asset also has an effect on the optimal rigid action: $\widehat{x}$ will in general differ from $x^*(z_0)$. As we will see in the applications, the action effect may be stronger than the direct flexibility effect, inducing the demand for the risky asset to be reduced by flexibility. This can be the case only for large portfolio risks, i.e., when the equity premium is large relative to the return volatility.

3 Applications

3.1 Cobb-Douglass: Labor, health and housing

We start the list of applications of the above theory with a second stage choice represented by the standard consumption problem under certainty, whereas the first stage choice is the portfolio problem examined in the previous section. Wealth z is allocated to the consumption of $n+1$ goods indexed from $i = 0$ to $i = n$.[4] The price of good $i > 0$ relative to the numeraire good $i = 0$ is denoted p_i. The utility of bundle $(x_0, x_1, ..., x_n)$ is measured by $u(x_0, x_1, ..., x_n)$. The choice set is $A = R^n_+$ and the objective function is

[4]Strictly speaking, this model is a particular case of the above general model only when $n = 1$. In the case of Cobb-Douglass, it is easy to show that the choice problem with $n > 1$ is equivalent to the one with $n = 1$ for a composite bundle of the goods $(x_1, ..., x_n)$ with shares $(\gamma_1/p_1, ..., \gamma_n/p_n)$.

written as

$$U(z, x_1, ..., x_n) = u\left(z - \sum_{i=1}^{n} p_i x_i, x_1, ..., x_n\right). \tag{11a}$$

Following for example Bodie, Merton and Samuelson (1992), let us consider the following Cobb-Douglass specification:

$$u(x_0, x_1, ..., x_n) = \gamma_0^{-1} \prod_{i=0}^{n} x_i^{\gamma_i}, \tag{12}$$

where the elements in vector $(\gamma_0, ..., \gamma_n)$ are restricted to have the same sign and where $\Gamma = \sum_{i=0}^{n} \gamma_i$ is less than unity. In the rigid context, the consumption of the goods other than good $i = 0$ must be chosen before observing the outcome of the risk on z, so that the only way to adapt to a shock on wealth is to modify the consumption of the numeraire good x_0 accordingly.

Using Proposition 2, we first examine the case of small risks. Under specification (11a) and (12), the relative risk aversion locally at z_0 respectively in the rigid and flexible contexts are measured by

$$-\frac{z_0 U_{zz}(z_0, \widehat{x}_1, ..., \widehat{x}_n)}{U_z(z_0, \widehat{x}_1, ..., \widehat{x}_n)} = \Gamma(\gamma_0^{-1} - 1) \text{ and } -\frac{z_0 v''(z_0)}{v'(z_0)} = 1 - \Gamma. \tag{13}$$

As stated in Proposition 2, the local risk aversion is reduced by flexibility. The effect is particularly powerful when the γ_is are positive and Γ is close to unity, yielding a risk behavior close to risk neutrality in the flexible context. For example, when $\gamma_0 = \Sigma_{i=1}^{n} \gamma_i = 0.5$, the share of wealth invested in equity is infinite in the flexible case, whereas it approximately equals $E\widetilde{r}/\sigma_{\widetilde{r}}^2$ in the rigid context.

We then show that the optimal bundle $(\widehat{x}_1, \widehat{x}_2, ..., \widehat{x}_n)$ in this rigid context is not affected by the option to invest in the risky asset.

Lemma 1 *Suppose that the decision maker must determine her demand for consumption goods $i = 1, ..., n$ before observing the return of her portfolio. Suppose also that the utility function $u : R_+^{n+1} \to R$ satisfies (12). It implies that the optimal bundle of these consumption goods is not affected by the option to invest in equity.*

Proof: Consider an arbitrary bundle $(x_1, ..., x_n) \in R_+^{n+1}$ with $z_0 - \Sigma_{i=1}^n p_i x_i > 0$. It is easy to check that the portfolio α that is optimal conditional to this bundle is characterized by

$$\alpha(x_1, ..., x_n) = b\left(z_0 R - \sum_{i=1}^{n} p_i x_i\right)$$

where scalar b is implicitly defined by $E\widetilde{r}(1 + b\widetilde{r})^{\gamma_0 - 1} = 0$. It implies that the optimal bundle in the rigid context must maximize the following objective function

$$\gamma_0^{-1} E(1 + b\widetilde{r})^{\gamma_0}\left(z_0 R - \sum_{i=1}^{n} p_i x_i\right)^{\gamma_0} \prod_{i=1}^{n} x_i^{\gamma_i}.$$

The optimal solution of this program is independent of $\widetilde{r}$. ■

It implies that the optimal demand for the n goods in the rigid framework is $\widehat{x}_i = \gamma_i z_0 R / p_i \Gamma$. Because the optimal second stage choice is independent of the investment opportunity, we can use Proposition 3 to determine the effect of flexibility in this framework. The objective function (11a) cannot be submodular. It is easy to check that it is supermodular if and only if the γs are positive. The following Corollary is thus a direct consequence of the combination of Proposition 3 and Lemma 1.

Corollary 1 *Consider the two stage problem with a portfolio choice and a budget allocation with multiple goods, assuming a Cobb-Douglass utility function (12) with $\gamma_i > 0$, $i = 0, ..., n$. Under this assumption, the demand for equity is larger in the flexible context than in the rigid one.*

The assumption the γ_is are positive is equivalent to the flexible indirect utility function being less risk-averse than the log. In fact, this assumption implies that the indirect utility function $U(., \widehat{x}_1, ..., \widehat{x}_n)$ in the rigid context is globally more concave than the indirect utility function v in the flexible one. To check this, we measure the relative risk aversion at an arbitrary wealth level z by

$$-\frac{zU_{zz}(z, \widehat{x}_1, ..., \widehat{x}_n)}{U_z(z, \widehat{x}_1, ..., \widehat{x}_n)} = \frac{(1 - \gamma_0)\Gamma z}{z\gamma_0 + (z - z_0)(\Gamma - \gamma_0)} \tag{14}$$

in the rigid context, and

$$-\frac{zv''(z)}{v'(z)} = 1 - \Gamma \tag{15}$$

in the flexible one. In fact, the flexibility index $F(z)$ defined by (9) is constant and equal to $F = (\Gamma - \gamma_0)/\Gamma(1 - \gamma_0)$. Risk aversion in the rigid context is uniformly larger than risk aversion in the flexible context if and only if $z\Gamma$ is uniformly larger than $z_0(\Gamma - 1)$. This is the case only when the γ_is are positive. It is the case only locally around $z = z_0$ when the γ_is are negative. In this latter case, the local aversion to risk at large wealth levels is larger in the flexible context! In the following Proposition, we prove that the effect of flexibility is intrinsically ambiguous for this set of utility functions.

Proposition 4 *Consider the two stage problem with a portfolio choice and a budget allocation with multiple goods, assuming a Cobb-Douglass utility function (12) with $\gamma_i < 0$, $i = 0, ..., n$. We can always find an initial wealth level z_0 and a distribution of the equity return $\widetilde{r}$ such that the demand for equity is smaller in the flexible context than in the rigid one.*

Proof: See the Appendix. ■

To illustrate, suppose that $n = 1$, $p_1 = 1$, $\gamma_0 = \gamma_1 = -2$, $z_0 R = 1$ and $\widetilde{r} \sim (-0.2, \pi; 5, 1 - \pi)$ with $\pi = 0.00404$. In the rigid context, it is optimal to take $\widehat{x}_1 = 0.5$ and $\widehat{\alpha} = 100\%$. The demand for the risky asset in the flexible context is reduced to $\alpha^* = 77\%$. Observe that this counterexample is based on a very skewed distribution of $\widetilde{r}$ in order to take advantage of the larger local risk aversion of $v(.)$ compared to $U(., \widehat{x})$ at large wealth levels.

Bodie, Merton and Samuelson (1992) examined the consequences of the flexibility of labor supply on the optimal portfolio. Relying on some specifications of the consumption/leisure utility function as the one presented in Corollary 1, they claimed that the ability to vary labor supply as a function of the portfolio return should induce young individuals to raise their demand for risky assets. All their numerical simulations yield the same conclusion. We have shown above that this result does not hold in general, except for small portfolio risks. We have also shown above that other variables like health and housing can be included in the utility function without changing the structure of the results. The presence of high transaction costs and taxes in the housing market together with a rigid health insurance system in many countries of continental Europe may also explain the relative resistance of European households to invest in stocks.

3.2 Additive model: Saving and risk-sharing

We now examine a simple portfolio-saving problem under certainty. At the beginning of period $t = 0$, individuals initially endowed with wealth z_0 determine their α for the risky asset with return $\widetilde{r}$. Wealth z_0 includes the human capital, which is the net present value of the flow of future labor incomes. In the flexible context, they observe their portfolio return before determining their consumption plan $(x_0, x_1, ..., x_n)$ over the remaining $n+1$ dates. In the rigid context, the future consumption plan $(x_1, ..., x_n)$ must be determined before observing the portfolio return in the first period, i.e., the portfolio risk must be entirely allocated to the consumption in the first period. Assuming a time-additive utility function, the objective function for this problem can be written as

$$U(z, x_1, ..., x_n) = u_0\left(z - \sum_{i=1}^{n} p_i x_i\right) + \sum_{i=1}^{n} u_i(x_i), \tag{16}$$

where $z = z_0 + \alpha\widetilde{r}$ is the wealth accumulated in the first period, and p_i is the price of consumption at date i relative to consumption at date 0. In the special case of a flat yield curve, $p_i = R^{-i}$. The utility functions u_i on consumption at $i = 0, ..., n$ are assumed to be increasing and concave.

During the last three decades, tax incentives have been developed for long-term saving schemes, as the 401k in the US, or the PEL or PERP in France. Following Laibson (1997), these tax incentives can be justified on the basis that individual have difficulties to commit themselves to save for their retirement. This may be due for example because of a time-inconsistency problem due to the hyperbolic structure of their time preferences. However, there is a clear cost associated to the system, since it does not allow individuals to use their savings as a buffer stock to smooth shocks on their incomes. We hereafter focus on another indirect cost of rigid saving schemes. When the contributions to the long-term saving plan are fixed ex-ante, households cannot time-diversify their portfolio risks. This induces them to be less tolerant to portfolio risks, thereby reducing the benefit that they can extract from the positive equity premium. We examine the condition under which the demand for equity is reduced by the rigidity of the saving plan.

Function U in (16) is neither supermodular nor submodular. However, a rewriting of this choice problem using backward induction allows us to use Proposition 3. To see this, let us define the value function J such that

$$J(x) = \max_{x_1,\ldots,x_n} \sum_{i=1}^{n} u_i(x_i) \text{ s.t. } \sum_{i=1}^{n} p_i x_i = x,$$

where x is the amount saved from z in the first period. Using this definition, we can rewrite the objective function (16) as

$$U^{\circ}(z,x) = u_0(z-x) + J(x).$$

Because U° is supermodular in (z,x), we know from Proposition 3 that the demand α^* for the risky asset in the flexible context is larger than the demand α_0 that is optimal when saving is exogenously fixed at $x^*(z_0)$, the optimal saving when the individual has no access to the equity market. However, except in the case where u_0 would be logarithmic, we know that the access to the equity market affects the optimal saving $\widehat{x}$ in the rigid context. We would be done if we would have that $\widehat{\alpha}$ be smaller than α_0. Because the optimal portfolio risk $\widehat{\alpha}$ in this context must maximize $Eu_0(z_0 - \widehat{x} + \alpha\widetilde{r})$, whereas α_0 maximizes $Eu_0(z_0 - x^*(z_0) + \alpha\widetilde{r})$, we would have $\widehat{\alpha} \leq \alpha_0$ if u_0 exhibits decreasing absolute risk aversion and $\widehat{x} > x_0$. To sum up, the question boils down to determining the conditions under which the access to the equity market raises the optimal saving. This is the case if

$$Eu_0'(w_0 + \alpha_0\widetilde{r}) \leq u_0'(w_0), \tag{17}$$

with $w_0 = z_0 - x^*(z_0)$, and where α_0 must satisfy condition $E\widetilde{r}u_0'(w_0 + \alpha_0\widetilde{r}) = 0$. Gollier and Kimball (1996) and Gollier (2001, Proposition 75) solved this question. Using the concept of risk tolerance $T_0(z) = -u_0'(z)/u_0''(z)$, they showed that the necessary and sufficient condition to guarantee that inequality (17) holds for all z_0 and all acceptable distributions of $\widetilde{r}$ is that the derivative of risk tolerance T_0 be uniformly larger than unity: $T_0'(z) \geq 1$ for all z. This condition is stronger than decreasing absolute risk aversion, which is equivalent to $T_0'(z) \geq 0$ for all z. This proves the first part of the following proposition.

Proposition 5 *Consider the two stage problem with a portfolio choice and a saving decision. The demand for equity is larger when the saving plan is* *flexible than when it must be fixed before observing the portfolio return if one of the two conditions is satisfied:*

1. *The derivative of risk tolerance T_0 is uniformly larger than unity;*
2. *The derivative of risk tolerance T_0 is uniformly negative.*

Proof: It just remains to prove the sufficiency of condition 2, which is equivalent to increasing absolute risk aversion. Following Gollier and Kimball (1996), this condition implies that inequality (17) is reversed, yielding $\widehat{x} \leq x_0$. Increasing absolute risk aversion implies in turn that the maximum of $Eu_0(z_0 - \widehat{x} + \alpha\widetilde{r})$ is smaller than the one of $Eu_0(z_0 - x^*(z_0) + \alpha\widetilde{r})$, or $\widehat{\alpha} \leq \alpha_0$. Because U is supermodular, we also know that $\alpha_0 \leq \alpha^*$. This concludes the proof of the sufficiency of condition 2. ■

Both conditions 1 and 2 imply that $\widehat{\alpha}$ be smaller than α_0, which we know is smaller than α^*. In the intermediate case of weakly decreasing absolute risk aversion, i.e., when absolute prudence is in between one and two time the absolute risk aversion, $\widehat{\alpha}$ is larger than $\alpha_0 \leq \alpha^*$, and the Gollier and Kimball's result does not allow us to conclude whether $\widehat{\alpha}$ is larger or smaller than α^*.

We can use an alternative approach by explicitly measuring the risk tolerance indexes of $v(.)$ and $U(., x_1, ..., x_n)$. Fully differentiating the first-order condition $v'(z) = u_0'(z - x^*(z)) = p_i^{-1}u_i'(x_i^*(z))$ yields

$$T_v(z) = T_0(z - x^*(z)) + \sum_{i=1}^{n} p_i T_i(x_i^*(z)), \tag{18}$$

where $T_v(z) = -v'(z)/v''(z)$ and $T_i(z) = -u_i'(z)/u_i''(z)$ are the indexes of absolute risk tolerance of v and u_i respectively. Suppose that functions u_0 and u_i are proportional to each other ($u_i = \beta_i u_0$), so that $T_0 \equiv T_i \triangleq T$. Under the additional assumption that T is convex, we obtain that

$$\begin{aligned} T_v(z) &= (1+P)\left[\frac{1}{1+P}T(z - x^*(z)) + \sum_{i=1}^{n} \frac{p_i}{1+P}T(x_i^*(z))\right] \\ &\geq (1+P)T\left(\frac{z - x^*(z) + \sum_{i=1}^{n} p_i x_i^*(z)}{1+P}\right) = (1+P)T\left(\frac{z}{1+P}\right), \end{aligned}$$

with $P = \Sigma_{i=1}^{n} p_i$. Suppose moreover that the absolute tolerance of risk on consumption is subhomogeneous, which implies that the right-hand side of the above inequality is larger than $T(z)$.[5] Under decreasing absolute risk aversion, this is in turn larger than $T(z-x)$ for all $x \geq 0$. This proves the following proposition.

Proposition 6 *Consider the two stage problem with a portfolio choice and a saving decision. Suppose that $u_i(.) = \beta_i u_0(.)$. The demand for equity is larger when the saving plan is flexible than when it is fixed ex-ante at an arbitrary level $x \geq 0$ if the absolute risk tolerance is increasing, convex and subhomogeneous.*

Under this condition, the indirect utility function $v(.)$ is less concave than $U(., x)$ for all x. This condition is satisfied for the important set of HARA utility functions

$$u(z) = \zeta(\eta + \frac{z}{\gamma})^{1-\gamma} \tag{19}$$

with $\zeta(1-\gamma)\gamma^{-1} > 0$ to ensure monotonicity and concavity. The corresponding absolute risk tolerance $T(z) = \eta + z/\gamma$ is increasing, convex and subhomogeneous when η and γ are both nonnegative. This covers the classical cases of constant relative risk aversion ($\eta = 0$) and of constant absolute risk aversion ($\gamma \to \infty$). Chiappori (1999) and Bodie, Merton and Samuelson (1992) have examined in a consumption/leisure context the special case of log linear preferences where $\eta = 0$ and $\gamma \to 1$, implying $u(z) = \log z$.

Lynch (1996) and Gabaix and Laibson (2002) examined a model of delayed adjustments in consumption. Investors can rebalance their portfolio every period, but they adapt their consumption to change in portfolio wealth only every D periods. They showed that the reduced flexibility of consumption plans due to these delays can potentially explain the equity premium puzzle. Because they assume constant relative risk aversion, we know from Proposition 6 that this loss of flexibility reduces the demand for equity, thereby increasing the equity premium. Their decision problem can be written as

$$J(z) = \max_{\substack{(c_0,\ldots,c_{D-1}) \\ (\alpha_0,\ldots,\alpha_{D-1})}} \sum_{t=0}^{D-1} \beta^t u(c_t) + \beta^D EJ\left(\left(z - \sum_{t=0}^{D-1} \frac{c_t}{R^t}\right) R^D + \sum_{t=0}^{D-1} \alpha_t \widetilde{r}_t\right), \tag{20}$$

[5] A function T is subhomogeneous if and only if $kT(z) \geq T(kz)$ for all z and all $k \geq 1$.

where the consumption plan $(c_0, ..., c_{D-1})$ between two reset dates must be chosen at the previous reset date, whereas the portfolio choices $(\alpha_0, ..., \alpha_{D-1})$ can depend upon the history of returns since that date. We assume that the excess returns $(\widetilde{r}_0, ..., \widetilde{r}_{D-1})$ are independent and identically distributed. When $D = 1$, this model is the standard portfolio-saving model of Merton (1969) with full flexibility. Assuming that $u(z) = z^{1-\gamma}/(1-\gamma)$, it is easy to check that the optimal demand for equity at each reset date is proportional to z, the total wealth of the investor at that date. The optimal share of wealth invested in equity equals

$$\frac{\alpha(z; D)}{z} = b\left(R\beta E(1 + b\widetilde{r})^{-\gamma}\right)^{D/\gamma}, \tag{21}$$

where b solves $E\widetilde{r}(1 + b\widetilde{r})^{-\gamma} = 0$. Because $R\beta E(1 + b\widetilde{r})^{-\gamma}$ must be less than unity for the solution to be bounded, we obtain that the share of wealth invested in equity is exponentially decreasing with the length D of time between two reset dates. This illustrates Proposition 6. If the equity premium $E\widetilde{r}$ is small compared to the volatility $\sigma^2_{\widetilde{r}}$, we obtain the following approximation for the portfolio allocation rule:

$$\frac{\alpha(z; D)}{z} \simeq b \exp - \left[\frac{\delta - \mu}{\gamma} + \frac{\gamma - 1}{\gamma}\frac{(E\widetilde{r})^2}{2\gamma\sigma^2_{\widetilde{r}}}\right] D, \tag{22}$$

where $\delta = -\log\beta$ is the rate of pure preference for the present and $\mu = \log R$ is the risk-free rate. To illustrate, suppose that investors rebalance their portfolio every year, but they adapt their consumption to their portfolio wealth only every D years. For $\gamma = 4$, $\delta - r = 2\%$, $E\widetilde{r} = 7\%$ and $\sigma_{\widetilde{r}} = 17\%$, we obtain that the share of wealth invested in equity is approximately equal to $0.61 \exp -0.021D$. When there is no delay in the adjustment of consumption ($D = 1$, flexible context), the optimal share invested in equity is 59.3%. It goes down to 58.1% (49.1%) if consumption is planned every two (ten) years.[6]

Going back to the original problem expressed by (16), observe that it can be reinterpreted as a risk-sharing problem between $n + 1$ agents with utility functions u_i, $i = 0, ..., n$. The transfer of one unit of the consumption good

[6]Gabaix and Laibson (2002) get an effect of the low frequency of consumption adjustments on the equity premium that is much stronger than what this analysis suggests. This is because a (rotating) small fraction $1/D$ of investors participates to the sharing of the market risk in each period.

from agent i yields p_i units to the reference agent 0. The intuition suggests that syndicates that are able to internally share risks efficiently should accept more portfolio risk, compared to syndicates in which individual 0 bears the entire portfolio risk, whereas agents $i \neq 0$ get a fixed income x_i. Propositions 5 and 6 provides sufficient conditions for this property to hold. An immediate extension of these results are obtained by combining risk sharing with one's future self and risk sharing with another individual.

3.3 Dynamic portfolio management

In the previous section, we examined a portfolio-saving model in which we introduced some rigidities in the way consumption can be adjusted to financial shocks. In this section, we alternatively examine the effect of introducing some rigidities in the way the portfolio allocation can be adjusted to these shocks. We consider the pure investment problem in which consumption takes place only at the end of the investment period. There are three dates. At date $t = 0$, the agent is endowed with wealth z_0. This wealth can be invested in two assets, one of which is risk-free. At date $t = 1$, after observing the portfolio return, the investor can rebalance her portfolio in the flexible context. At date $t = 2$, the portfolio is liquidated and the accumulated wealth is consumed. The gross risk-free rate in each subperiod is denoted R. The return of the risky asset in the first and second subperiods are denoted respectively $\widetilde{r}$ and $\widetilde{r}_1$. We assume that $\widetilde{r}$ and $\widetilde{r}_1$ are independent. This application is a special case of our general model with

$$U(z, x) = Eu(zR + x\widetilde{r}_1), \tag{23}$$

where z is the wealth accumulated at the intermediary date, and x is the euro investment in the risky asset in the second subperiod. The optimal investment in the risky asset in this flexible context at date $t = 0$ solves the following backward induction problems:

$$\alpha^* = \max_{\alpha} Ev(z_0 R + \alpha\widetilde{r}) \quad \text{with} \quad v(z) = \max_{x} Eu(zR + x\widetilde{r}_1). \tag{24}$$

In the rigid context, the demand for the risky asset in the two subperiods must be determined at date $t = 0$:

$$(\widehat{\alpha}, \widehat{x}) = \arg\max_{\alpha, x} Eu(z_0 R^2 + \alpha\widetilde{r}R + x\widetilde{r}_1). \tag{25}$$

When $\widetilde{r}$ and $\widetilde{r}_1$ are identically distributed, it must be that $\widehat{x} = \widehat{\alpha}R$. This corresponds to a "buy-and-hold" strategy in which the excess return of the risky portfolio during the first subperiod is automatically reinvested in the risk-free asset during the second subperiod.

The objective function U defined in (23) is neither supermodular nor submodular, so that the only general result that we can use from section 2 is that the ability of the investor to rebalance her portfolio induces her to raise her demand for the risky asset at date $t = 0$ ($\alpha^* > \widehat{\alpha}$) when the expected excess return $E\widetilde{r}$ in the first subperiod is small compared to its standard deviation. In order to compare α^* and $\widehat{\alpha}$ for larger portfolio risk in the first subperiod, let us assume that the utility function u exhibits constant relative risk aversion γ. It is well-known since Mossin (1968) that $v(z) = ku(Rz)$ for all z, which means that myopia is optimal. In the flexible context described by (24), the demand α^* at $t = 0$ is independent of the distribution of the return of the risky asset in the second subperiod. It implies that

$$\alpha^* = \max_{\alpha} Eu(z_0R^2 + \alpha\widetilde{r}R). \tag{26}$$

Thus, in the CRRA case, the question can be restated in a completely static framework. It boils down to determining the effect on the demand for the risky asset $\widetilde{r}$ of the introduction of another risky asset $\widetilde{r}_1$. Gollier (2001, Proposition 36) showed that independent static risks $\widetilde{r}$ and r_1 are substitutes, i.e., $\widehat{\alpha} < \alpha^*$, if absolute risk aversion is decreasing and absolute prudence is decreasing and larger than twice the absolute risk aversion. In the CRRA case, it is easy to check that these conditions hold when relative risk aversion is less than unity. Gollier, Lindsey and Zeckhauser (1997) showed that this condition is necessary to guarantee the result when returns are identically distributed.

Proposition 7 *Consider the two-period investment problem with constant relative risk aversion and independent and identically distributed returns. If relative risk aversion is less than unity, the initial demand for the risky asset is larger when the portfolio can be flexibly rebalanced at the end of the first period than when the investor follows a rigid buy-and-hold strategy. When relative risk aversion is larger than unity, it is always possible to find a distribution of excess returns so that the initial demand is smaller in the flexible context than in the more rigid one.*

To understand the necessity of relative risk aversion less than unity, consider the case of utility function $u(z) = -z^{-3}$, which exhibits constant relative risk aversion $\gamma = 4$. Normalize the risk free rate to zero. If the yearly excess return of the risky asset is distributed as $(-0.1, 2/3, 10, 1/3)$, the optimal investment in the risky asset in the flexible context equals 16.16% of initial wealth, whereas it goes up to 16.27% in the rigid context for a time horizon of two years. Observe that this counterexample is built on a very skewed distribution of excess returns. For more symmetric distributions, the standard intuition that an increase of the frequency at which the portfolio can be rebalanced raises the demand for the risky asset. To illustrate, consider the more realistic assumption that the distribution of excess return is normally distributed with mean 4.6% per year and a standard deviation of 14.2%.[7] In this case, the optimal portfolio allocation is to invest 56.35% of the investor's wealth if the investor rebalances her portfolio every year. This share of wealth invested in the risky asset goes sligthly down to 56.01% when the investor rebalances her portfolio only every two years.[8]

4 Conclusion

Because the value of information is nonnegative in the expected utility model, flexibility always raises the expected utility of the decision maker. In this paper, we examine the comparative statics consequence of flexibility on the attitude towards risk ex ante. At the exception of Machina (1982), the general tone of the literature is that flexibility enhances risk tolerance. Workers who can adjust their labor supply to their financial wealth should take more risk on financial markets. Spouses should take more risk on their job if these risks are efficiently shared within their household. Investors should invest more in risky assets if they can rebalance their portfolio more frequently. Consumers who have access to the credit market should take more risk than those who cannot adjust their saving/credit to the shocks on their incomes.

[7] We calibrate our model on the real excess return of SP500 over the period 1963-1995.

[8] The demand for the risky asset goes down to 55.66%, 55.28% and 54.87% when the frequency of portfolio rebalancement is respectively every 3, 4 and 5 years. We solve the problem numerically by using a discrete approximation for the normal distribution of $\log \widetilde{r}$ with $n = 20$ equally distant points between $\mu - 5\sigma$ and $\mu + 5\sigma$. An increase in n over 20 has no effect on the first four digits of α.

When the decision under risk is a choice between a risky prospect or a safe one, or when it entails only small risks, the intuition that flexibility enhances risk tolerance is correct. For more general choice problems under uncertainty, additional restrictions on preferences or on the characteristics of the available ex post adjustments are necessary to guarantee the comparative statics result. When the choice under uncertainty is a portfolio allocation, we have shown that the submodularity or the supermodularity of the objective function is useful to derive sufficient conditions in various applications.

References

Arrow, K.J., (1963), Liquidity preference, Lecture VI in "*Lecture Notes for Economics 285, The Economics of Uncertainty*", pp 33-53, undated, Stanford University.

Arrow, K.J. and A.C. Fischer, (1974), Environmental preservation, uncertainty and irreversibility, *Quarterly Journal of Economics*, 88, 312-319.

Bodie, Z., R.C. Merton and W.F. Samuelson, (1992), Labor supply flexibility and portfolio choice in a life cycle model, *Journal of Economic Dynamics and Control*, 16, 427-449.

Chetty, R., and A. Szeidl, (2003), Consumption commitments and asset prices, mimeo, Berkeley.

Chiappori, P.-A., (1999), Labor supply and efficient risk sharing, mimeo, University of Chicago.

Drèze, J.H. and F. Modigliani, (1972), Consumption decisions under uncertainty, *Journal of Economic Theory*, 5, 308-335.

Flavin, M., and S. Nakagawa, (2005), A model of housing in the presence of adjusment costs: A structural interpretation of habit persistence, unpublished paper.

Gabaix, X., and D. Laibson, (2002), The 6D Bias and the Equity Premium Puzzle, *NBER Macroeconomics Annual*, MIT Press, Ben Bernanke and Ken Rogoff ed., pp. 257-312.

Gollier, C., (2001), *The economics of risk and time*, MIT Press, Cambridge MA.

Gollier, C., and M.S. Kimball, (1996), Toward a systematic approach to the economic effects of uncertainty: characterizing utility functions, Discussion paper, University of Michigan.

Gollier, C., J. Lindsey, and R.J. Zeckhauser, (1997), Investment Flexibility and the Acceptance of Risk, *Journal of Economic Theory*, 76, 219-42.

Henry, C., (1974) , Investment decisions under uncertainty: the irreversibility effect, *American Economic Review*, 64, 1006-1012.

Laibson, D.I., (1997), Golden eggs and hyperbolic discounting, *Quarterly Journal of Economics*, 62, 443-479.

Lynch, A., (1996), Decision frequency and synchronization across agents: Implications for aggregate consumption and equity return, *Journal of Finance*, 51, 1479-1497.

Machina, M.J., (1982), Flexibility and the demand for risky assets, *Economics Letters*, 10, 71-76.

Machina, M.J., (1984), Temporal risk and the nature of induced preferences, *Journal of Economic Theory*, 33, 199-231.

Merton, R.C., (1969), Lifetime portfolio selection under uncertainty: The continuous- time case, *Review of Economics and Statistics*, 51, pp. 247-257.

Mossin, J., (1968), Optimal multiperiod portfolio policies, *Journal of Business*, 215-229.

Mossin, J., (1969), A note on uncertainty and preferences in a temporal context, *American Economic Review*, 59, 172-174.

Postlewaite, A., L. Samuelson and D. Silverman, (2005), Consumption commitments and preferences for risk, unpublished paper, University of Pennsylvania.

Pratt, J., (1964), Risk aversion in the small and in the large, *Econometrica*, 32, 122-136.

Samuelson, P.A., (1969), Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics*, 51, 239-246.

Shroyen, F., (2010), Income risk aversion with quantity constraints, Norwegian School of Economics and Business Administration.

Spence, M., and R. Zeckhauser, (1972), The effect of the timing of consumption decisions and the resolution of lotteries on the choice of lotteries, *Econometrica*, 40, 401-403.

Topkis, D., (1978), Minimizing a submodular function on a lattice, *Operations Research*, 26, 305-21.

Appendix: Proof of Proposition 2

We first examine the flexible context. The first-order condition of program 1 is written as

$$U_a(z, x^*(z)) = 0. \tag{27}$$

Fully differentiating this condition yields

$$x^{*\prime}(z) = -\frac{U_{zx}(z, x^*(z))}{U_{xx}(z, x^*(z))}. \tag{28}$$

Using the envelop theorem, we have that

$$v'(z) = U_z(z, x^*(z)). \tag{29}$$

Fully differentiating this condition implies that

$$v''(z) = U_{zz}(z, x^*(z)) + x^{*\prime}(z)U_{zx}(z, x^*(z)) \tag{30}$$

Combining equations (28), (29) and (30) yields property (7). Condition (6) is in Pratt (1964).

We now turn to the analysis of the rigid context. It is obvious that $\widehat{\pi}(0) = 0$. Using the envelop theorem, fully differentiating condition (4) yields

$$E\widetilde{\varepsilon}U_z(z_0 + k\widetilde{\varepsilon}, \widehat{x}(k)) = -\widehat{\pi}'(k)U_z(z_0 - \widehat{\pi}(k), \overline{x}(k)), \tag{31}$$

where functions $\widehat{x}(k)$ and $\overline{x}(k)$ satisfy the following conditions:

$$EU_x(z_0 + k\widetilde{\varepsilon}, \widehat{x}(k)) = 0 \tag{32}$$

$$U_x(z_0 - \widehat{\pi}(k), \overline{x}(k)) = 0. \tag{33}$$

We have that $\widehat{x}(0) = \overline{x}(0) = x^*(z_0)$ and $\widehat{\pi}'(0) = 0$. Fully differentiating equations (31), (32) and (33) shows that $\widehat{x}'(0) = \overline{x}'(0) = 0$ and

$$\begin{aligned} & E\widetilde{\varepsilon}^2U_{zz}(z_0 + k\widetilde{\varepsilon}, \widehat{x}(k)) + \widehat{x}'(k)E\widetilde{\varepsilon}U_{zx}(z_0 + k\widetilde{\varepsilon}, \widehat{x}(k)) \\ = \; & (\widehat{\pi}'(k))^2U_{zz}(z_0 - \widehat{\pi}(k), \overline{x}(k)) - \widehat{\pi}''(k)U_z(z_0 - \widehat{\pi}(k), \overline{x}(k)) + \overline{x}'(k)U_{zx}(z_0 - \widehat{\pi}(k), \overline{x}(k)). \end{aligned}$$

Evaluating this at $k = 0$ yields

$$\widehat{\pi}''(0) = \sigma_\epsilon^2 \frac{-U_{zz}(z_0, x^*(z_0))}{U_z(z_0, x^*(z_0))}.$$

It implies that

$$\begin{aligned}\widehat{\pi}(k) &= \widehat{\pi}(0) + k\widehat{\pi}'(0) + \frac{1}{2}k^2\widehat{\pi}''(0) + o(k^2) \\ &= \frac{1}{2}k^2\sigma_\epsilon^2 \frac{-U_{zz}(z_0, x^*(z_0))}{U_z(z_0, x^*(z_0))} + o(k^2).\end{aligned}$$

This concludes the proof.■

Proof of Proposition 4

In the rigid context, the optimal bundle is characterized by $\widehat{x}_i = \gamma_i z_0 R/p_i\Gamma$, $i = 1, ..., n$, yielding a total expenditure equaling $z_0 R(1 - (\gamma_0/\Gamma))$. It implies that the portfolio choice problem in this context can be written as

$$\widehat{\alpha} = \arg\max_{\alpha} E\widehat{u}(z_0 R + \alpha\widetilde{r}),$$

where $\widehat{u}(z) == \gamma_0^{-1}\left(z - z_0\left(1 - \gamma_0\Gamma^{-1}\right)\right)^{\gamma_0}$. In the flexible context, the optimal portfolio solves

$$\alpha^* = \arg\max_{\alpha} Ev(z_0 R + \alpha\widetilde{r}),$$

with $v(z) = k\ z^\Gamma$. Gollier and Kimball (1996)[9] have shown that the necessary and sufficient condition for α^* to be larger than $\widehat{\alpha}$ for all acceptable distributions of the equity return $\widetilde{r}$ is that v be "centrally less risk-averse" than $\widehat{u}$ around z_0, i.e., that

$$h(r; z_0) = r\left(\frac{v'(z_0 R + r)}{v'(z_0 R)} - \frac{\widehat{u}'(z_0 R + r)}{\widehat{u}'(z_0 R)}\right) \geq 0 \qquad (34)$$

[9]See also Gollier (2001), section 6.3.3.

for all r. For r positive, this condition can be rewritten as

$$\left(1+\frac{r}{z_0 R}\right)^{\Gamma-1} > \left(1+\frac{\Gamma}{\gamma_0}\frac{r}{z_0 R}\right)^{\gamma_0-1}.$$

When r tends to infinity, the two sides of this inequality tend to zero. However, because $\Gamma-1<\gamma_0-1<0$, the left-hand side converges quicker to zero than the right-hand side. This implies that the necessary condition (34) is violated for large r. ■

HOW TO ADAPT TO CHANGING MARKETS:

EXPERIENCE AND PERSONALITY IN A REPEATED INVESTMENT GAME

Astrid Hopfensitz
Toulouse School of Economics
University of Toulouse 1
France
Astrid.Hopfensitz@univ-tlse1.fr

Tanja Wranik
Department of Psychology
University of Geneva
Switzerland
Tanja.Wranik@unige.ch

WORK IN PROGRESS AND INCOMPLETE - PLEASE DO NOT CIRCULATE

2010

ABSTRACT:

Investment behavior is traditionally investigated with the assumption that it is on average advantageous to invest. However, this may not always be the case. In this paper, we experimentally studied investment choices made by students and financial professionals facing alternately an advantageous and disadvantageous environment in a multi-round investment game. Expected returns from investment in the advantageous environment were higher than a safe alternative, while expected returns were lower in the disadvantageous environment.

We investigate how experience and personality are related to choices. Investment behavior does not differ dependent on expected returns and professionals do not significantly differ from students. Personality predicts behavior and in particular we observe that openness to experience was an asset in unfavorable markets, leading to reduced risk taking.

JEL: D14, D53, D81, G11, C91, C93

* We would like to thank Michel Odehnal and Stephane Peysson for help with programming the experimental interface and Marina Bianchi-Galanti and Cristina Rivera for the excellent help running the experiments. Financial support through a NETSPAR research grant awarded to Astrid Hopfensitz and Tanja Wranik is gratefully acknowledged. We would also like to thank the chair 'Financière de la Cité' at TSE. This is an interdisciplinary project between economics and psychology, and both authors contributed equally to the research.

Introduction

A large body of research in economics and finance has focused on understanding risk taking behaviors in financial markets. Much of this research has been motivated by the observation that stocks have significantly higher average long term returns that bonds (Mehra and Prescott, 1985)[1] and that investment in bonds is higher than reasonable levels of risk aversion would predict. Thus, a major research question is to understand why investment in stocks is not higher. Accordingly, most experimental studies in this area have been constructed such that investment decisions are made between "risky" and "safe" projects, where returns from the risky project(s) are on average higher than returns from the safe project(s).

In real life, returns from stocks and returns from bonds are long term averages, and obviously investment in stocks might be more or less advantageous during certain periods compared to others. Unfortunately, there has so far been little research to examine how investors react during unfavorable or changing market periods. Since generally people are observed to be risk averse (Holt and Laury, 2002) almost no investment would be predicted in an environment where expected returns from a risky option are lower than from a safe option. Consequently unfavorable investment tasks have so far been little studied by economists. Meanwhile psychologists have developed the "Iowa Gambling Task", which represents a gambling situation where a risky option gives high returns but has a lower expected value than a safe option. This task has been used with emotionally impaired brain patients to confirm that lack of emotional competence can

[1] Specifically: "Over the ninety-year period 1889-1978 the average real annual yield on the Standard and Poor 500 Index was seven percent, while the average yield on short-term debt was less than one percent".

lead to long term disadvantageous investment decisions (Bechara et al., 1997)[2]. Normal controls are able to distinguish between the two tasks. While this result indicates that people are able to select situations with higher returns in expected terms it does not tell us how people react if all available options are disadvantageous. The aim of our study is to understand how investor reactions and behaviors differ in advantageous versus disadvantageous market conditions and how these are related to both experience and personality.

We present results from an economic experiment investigating behaviors when market conditions change. Specifically, we compare behaviors of naive investors (i.e. students) with experienced investors (i.e. professionals of the financial industry). Previous research has shown that professionals often show the same biases as naive investors, and also that professionals can sometimes have stronger biases than non-professionals (Haigh and List, 2005). Clearly, there are many reasons (training, experience, etc.) why professionals might react differently from students (e.g. Burns, 1985, Potters and van Winden, 2000). Another important reason might be self-selection of people with specific personality profiles into certain professions. Such self selection has been observed for entrepreneurs (Brandstaetter, 1997) and financial traders (Lo et al., 2005). When comparing behavior between experienced and inexperienced investors, we therefore also took into account their specific personality profile. Consequently, we will explore the relationship between experience, personality, and investment behavior in two market conditions: one of which is 'advantageous' the other 'disadvantageous'.

[2] However, such patients also invest more than controls in situations where investment is advantageous (Shiv et al, 2005a/2005b). Thus a lack of emotions generally leads to more risk taking and gambling, independent of whether the participant is in a good or a bad market condition.

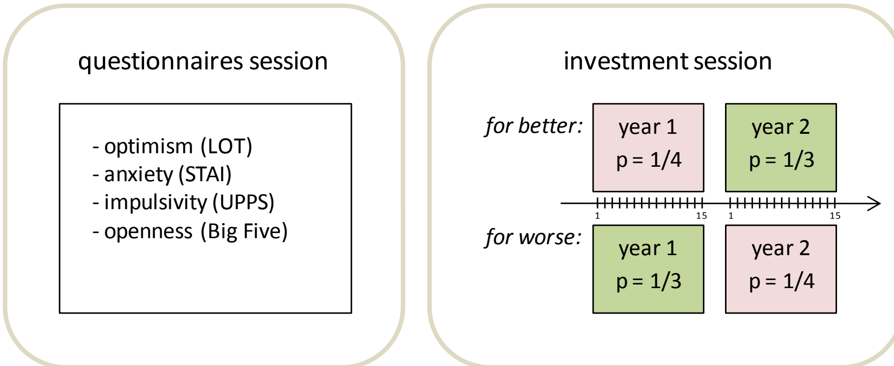


Figure 1: Sessions of the experiment. Note that for students the questionnaire session took place about one week before the investment session. For professionals the questionnaire session followed immediately the investment session.

1 Methods and Predictions

In this paper, we present an experiment in which students and finance professionals participated in a repeated investment game. The game consisted of two "years" of 15 rounds each. In one "year" investment was advantageous, and in the second "year" investment was disadvantageous. In addition to investment behavior, we also report a number of personality traits that were measured in a separate session (see Figure 1).

The investment task is based on the type of repeated investment task generally used to study myopic loss aversion (Gneezy and Potters, 1997; Thaler et al., 1997; Bellmare et al., 2005). In this task, participants repeatedly make decisions concerning the allocation of points into two projects. In our case, participants received 100 points each period, which they could use for investment. Note that previous earnings could not be used for investment, and that therefore the available budget stayed constant throughout the task. Participants could chose to invest their budget into a “safe” project, in which each point invested was simply added to the cumulative earnings, or a “risky” project. The risky project was a project in which the participant had a probability p of receiving the invested amount multiplied by 2.5 plus the initial investment, and a probability (1-p) that the amount invested into the project during the round would be lost. In past studies the probability of winning was set to p = 1/3, which meant that investment was on average advantageous (Expected Value = 1.17 > 1) compared to the safe project.

In our study, participants had to make investment decisions in two different market conditions. The risky option in the first market had a probability of winning of p = 1/3, and in the second market a probability of winning of p = 1/4. We will call the year with a probability of p = 1/3 the "good year" since investment was on average advantageous (Expected Value = 1.17). In contrast, investment in the year with p = 1/4 was in expected terms disadvantageous (Expected Value = 0.875), and we will refer to it as a "bad year". The first treatment variable was the order in which the participant entered into the good versus the bad year (see Figure 1: 'for better' and 'for worse').

To control for personality differences, participants also filled out a number of standardized personality scales during a second session. Variables of interest included trait optimism, trait anxiety, impulsivity, and openness to experience. Trait optimism was measured with the 10-item LOT-R (Carver and Scheier, 2001), which includes four filler items, three positively-worded items, and three reverse-coded items. Respondents indicate their degree of agreement with statements such as, "In uncertain times, I usually expect the best," using a five-point response scale ranging from "strongly disagree" to "strongly agree." Negatively-worded items are reversed, and a single score is obtained indicating the degree of optimism. Anxiety was measured with the 20-item STAI-Trait questionnaire (Spielberger, 1972). This instrument assesses the relative frequency of general nervousness or anxiousness in different contexts, and participants rate the relative frequency with which they engage in the described behavior on a four point scale

(1 = almost never; 4 = almost always). General personality was measured using the Big Five Personality Inventory (BFI) developed by John and Srivastava (1999). The "Big Five" are broad categories of personality traits thought to be the most parsimonious in describing inter-individual variation in behavioral propensities. The BFI includes items pertaining to Extraversion (e.g., talkative, energetic), Agreeableness (e.g., kind, warm), Conscientiousness (e.g., efficient, organized), Neuroticism (e.g., moody, touchy), and Openness to Experience (e.g., imaginative, complex). The 44-items are presented as a series of affirmations, and participants are asked to indicate the extent to which they agree or disagree with them, using a 1 (disagree strongly) to 5 (agree strongly) response format. For this study we only examined Openness to Experience. Finally, we measured impulsivity using the UPPS Impulsive Behavior scale (Whiteside and Lynam, 2001). This instrument measures four distinct traits related to impulsivity: (a) lack of premeditation; (b) urgency; (c) sensation-seeking; and (d) lack of perseverance. The 45 items are presented as a series of affirmations, and participants are asked to indicate the extent to which they agree or disagree with them, using a 1 (agree strongly) to 4 (disagree strongly) response format.

1.1 Procedures

The experiment was conducted in Spring 2009 at the University of Geneva, Switzerland. Student participants were recruited by announcements promising a monetary reward and were asked to sign up for two one-hour sessions. The first was a questionnaire session; the second was the experimental session in which participants completed the investment task. Professionals were invited by their human resources (HR) manager per e-mail (see Appendix B). For them, monetary rewards were not explicitly mentioned in the invitation, and they accepted in order to help the research of the HR manager who invited them. For practical reasons, professionals were only required to come to the University laboratory once, and to complete both sessions on the same day. To avoid carry over effects from the personality questionnaires, professionals first participated in the experimental session and then filled out the personality questionnaires. Both students and professionals were paid their earnings from the investment task at the end. Average earnings for students were around 31.3 CHF (approx. 27 USD) and for professionals 58.9 CHF (approx. 52 USD). In total, 31 professionals (22 men, 9 women; mean age 43.9, std. dev. 9.25) and 46 students (25 men, 21 women; mean age 27.0, std. dev. 8.02) participated in the study.

Professionals came from a small private bank in Switzerland. The bank employs around 100 people in four different locations in Switzerland. The areas of expertise are: private banking, institutional asset management, fund administration and services for independent asset managers. The aim of the bank is to "apply advanced financial techniques to client service, to protect their assets from the hazards of speculation, and to ensure regular returns over the medium and long term". The size of the bank requires small teams to work in close liaison with the asset managers. The proximity also allows them to share information with respect to financial markets and sectors of particular interest. The 31 professionals participating in the experiment came from various sites of the bank. They also represent a variety of nationalities: the majority was born in Switzerland, but also three Italians, one Spanish, one German, one Swedish, two Japanese, and one Canadian participated. The professional’s education background included: practical banking training completed by theoretical courses (5 participants), commercial diploma (8 participants), and university graduates in political economy, economics, mathematics, engineering and econometrics. The average length of service in the same bank was 11 years for women and 9 years for men. Table I summarizes areas of responsibility of participants. The majority of participants are asset and relationship managers. The 20 professionals that fall into this category manage assets from 50 to 250 millions CHF. Dependent on the agreement with the client this work can consist in either managing a portfolio under a management mandate (i.e. for a given 'risk level') or deciding on investment together with the client when no mandate is given.[3] A further 9 participants are financial analysts that have no direct client contact however are responsible for providing recommendations and analysis to the asset managers.

Table I: Area of responsibility of professionals

		Female	Male	Total
Asset and relationship managers	senior	5	13	18
	junior	2		2
Financial analyst and investment funds managers		2	7	9
Trader		1		1
Chief financial officer		1		1

Even though the order of the two sessions differed for professionals and students, the same protocol was applied for each of the two sessions. At the beginning of the questionnaire session, participants were informed that they would have to fill out a number of questionnaires concerning their personality. It took participants between 40 to 60 minutes to answer all the questionnaires. At the beginning of the investment task, participants were informed that they would participate in an investment game in which they could earn points that would be converted to real money at a specified exchange rate at the end of the session. Students received 30 CHF[4] and professionals 60 CHF (equivalent to 3000 points) as initial capital and were handed the money in envelopes[5].

[3] According to the banks' annual report in 2008 around 815 million CHF were held in funds under own management and about 1462 million CHF were under portfolio management mandate.
[4] At time of study 30 CHF equaled approximately $25.
[5] Doubling the earnings for finance professionals was based on discussions with our contact from human resources from the implied bank. Note that in a similar study professional traders earned about $30 for participation (Haigh and List, 2005).

This money was the capital that could be used for investment in the two years of 15 rounds each. In each round, participants made decisions concerning 100 points from their initial capital. Points had to be distributed between two projects: a safe project and a risky project. In one of the two years the risky project had an expected value higher than the safe project in the other year the value was lower. Specifically in 'good' years investment in the risky project resulted in gains of 2.5 times the investment with probability p=1/3 and in bad years the probability of gains was p=1/4. Probabilities and returns for the first year were described in the instructions and it was made clear that more information would follow after year one. After finishing the fifteen rounds of the first year a short note informed participants about the new probability of gains from the risky project. It was made clear that besides this nothing had changed in the game. A short questionnaire after the initial instructions verified that participants did understand the information given. To control for order effects the order of 'good' and 'bad' years was counterbalanced across participants.

After participants had read the instructions, they answered a number of control questions and were invited to address any remaining questions to the experimenter. To keep feedback comparable across treatment, outcomes from investment were predetermined by random sequences that were equally distributed across treatments.

1.2 Predictions and Hypothesis

Our first question will focus on whether the probability of the risky project will indeed influence investment amounts. Given that the expected value of the risky project is larger than the safe project for good markets, and lower than the safe project for bad markets, we expect a risk neutral decision maker to invest fully in good markets and not to invest in bad markets. Risk aversion might lead to intermediate investment for good markets, but certainly investment should be higher for good markets than for bad markets. Our first hypothesis is therefore:

HYPOTHESIS 1: Investment is on average and for each subject higher for good market conditions than for bad market conditions.

Predictions could differ if risk aversion depends on how earnings are evaluated with respect to some reference point (Kahneman and Tversky, 1979). In such cases, a within-subject comparison for investment in different markets might be problematic, since risk taking behavior will always be influenced by previous outcomes and probabilities. Past studies have already confirmed that previous outcomes will influence future choices. Examples are the "hot hand effect" and "gamblers fallacy" (Tversky and Kahneman, 1971; Croson and Sundali, 2005). However, it is not known whether these effects are influenced by the specific probabilities of winning. We will thus investigate the presence of these effects and their relative strength given the two probability conditions. If probabilities are reflected in the gamblers fallacy (which has been labeled as the "law of small numbers"), we expect this effect to be influenced by market conditions and to be stronger for bad markets. Specifically if the probability of loosing is high, observing a gain will in case of the gamblers fallacy lead to the belief that the next round will more likely be a loosing round.

HYPOTHESIS 2: If the gambler’s fallacy occurs, the effect should be stronger for bad markets than for good markets.

Our third question concerns whether professionals show significantly different trading strategies than students. If we observe differences, we will explore whether these are due to different personality profiles or due to experience and training specific to professionals. As previous research has suggested (Sjöberg and Engelberg, 2009), professionals in the financial industry tend to be a self-selected sample characterized by certain personality traits enabling them to take high risks. Thus, we expect professionals to be less anxious and more optimistic than our student population.

HYPOTHESIS 3: Professionals show a personality profile different from students. Namely they are characterized by less anxiety and more optimism.

We further predict that personality will influence investment behavior for the two different markets. Specifically optimism will be related to more risk taking, leading to higher investment for both market conditions. Trait anxiety can be related to the anticipation of anxiety in which case economic models of anticipatory emotions predict less risk taking (Wu, 1999; Caplin and Leahy, 2001). However studies on gamblers have shown a link between high trait anxiety and an urge to gamble (Zangeneh et al., 2008). Thus, we might expect anxiety to have either a negative or positive impact on risk taking. Impulsivity implies more impulsive reactions to outcomes. This can be either a tendency to give in to risk aversion once small gains are made or a tendency to continue gambling once losses are encountered (Lynam and Miller, 2004). Impulsivity might therefore show a very different impact for good or for bad markets. For good markets it might decrease risk taking while for bad markets it might increase risk taken. Finally, we also include Openness to Experience, to allow for a personality trait related to curiosity and novel approaches to new situations. Generally research focuses on why stock investment is 'not high enough' given the long term observation that returns from stocks are on average higher than returns from bonds (e.g. Mehra and Prescott, 1985). This leads to a general tendency to consider risk taking as advantageous, and especially so if it is presented as an investment setting. Consequently, unfavorable probabilities might be ignored. We predict that participants high in openness will be more likely to adapt to a novel situation and be more likely to react appropriately to the low probability of winning in the bad market.

HYPOTHESIS 4: We expect investment behavior to be influenced by personality. The influence on risk taking will differ between good and bad market years.

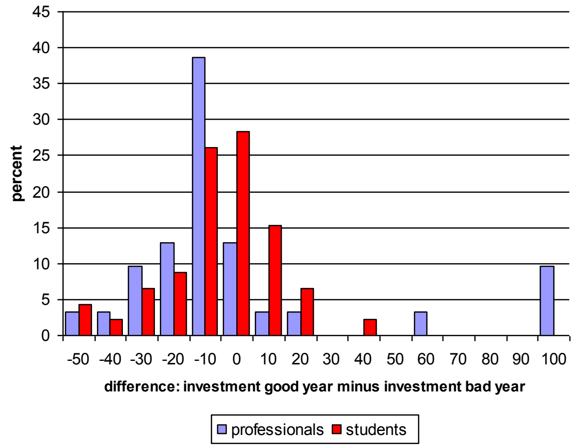


Figure 2: Histogram of investment difference between good and bad years for professionals and students.

2 Results

We now come to a presentation of the results. In the next section, we will give general descriptive statistics concerning investment behavior across the different treatments. In sections 2.2 and 2.3, we will then investigate the impact of markets, experience, and personality on investment choices.

2.1 General descriptive statistics

Overall investment over treatments and rounds is at 42.1%, with average investment in good years at 42.9% and in bad years at 40.6%. Investment in good years is thus slightly higher, however this difference is not significant (Wilcoxon sign-rank test, p=0.978). Moreover, this difference is mainly due to the professional's behavior. While professionals invested 45.4% in good markets and 39.6% in bad markets (Wilcoxon sign-rank test, p= 0.616), students had an

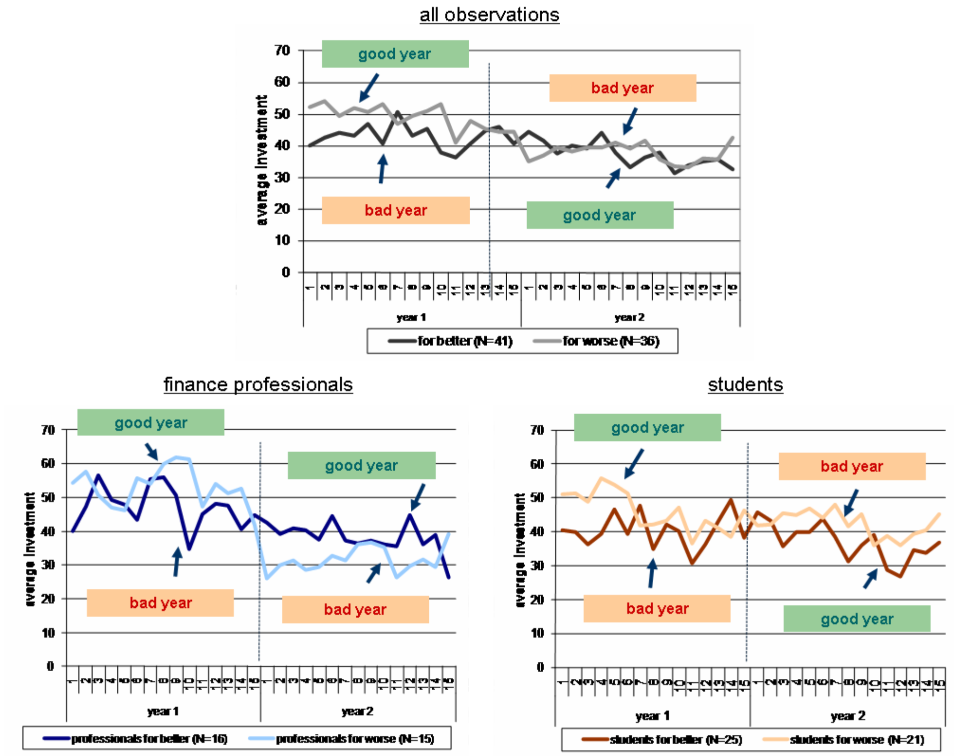


Figure 3: Investment over rounds and years. Top panel: all observations. Bottom left: professionals. Bottom right: students.

average investment of 41% in both market conditions. The difference for professionals is mainly due to four out of 31 participants (i.e. 13 %) who invest on average at least 60 units more in good years than in bad years (see also Figure 2). Ignoring these few observations, we observe for professionals a slight bias to lower investment for good than bad years.[6] Overall, our results show that investment is largely unaffected by probabilities of winning.

This result is further confirmed by investment over rounds in each year. In Figure 3, we present the investment timeline for investors who first face a good and then a bad year ("for worse") and for investors who first face a bad and then a good year ("for better"). It is striking that we do not observe a strong difference for the treatment order. Specifically, although we observe slightly

[6] Excluding the four participants investing at least 60 units more in good markets we observe for professionals a mean of -6.54 while students show a mean of -0.09.

higher investments during the first years for investors in a good market, investments are very similar in the second year. This tendency is stronger for professionals (bottom, left panel) than for students.[7] Students show almost no difference in investment for the two years. For professionals, we observe that the order of the years matters. Overall, we observe a clearly negative time trend over years. Investment in the first year is significantly higher than investment in the second year, independent of whether a good year was followed by a bad year or the other way round. Specifically, investment decreased from 43.0% to 37.4% when markets changed for better, and investment decreased from 49.1% to 37.9% when markets changed for worse. In both cases this difference is significant (Wilcoxon signed-rank test: "for better" $p = 0.027$; "for worse" $p = 0.064$).

Further support for this result is obtained from a probit regression (
Table II, column 1). We regress the probability of having on average higher investment in the good market year compared to the bad market year[8] on being a student, female and the order of years. We observe that students invest slightly more in good years and that mostly the order of the two years influences the difference. We conclude with the following first result:

RESULT 1: Independent of training, investment is similar for good and for bad market conditions.

[7] The clear difference for finance professionals is due to the fact that the previously identified 4 participants that invested at least 60 points more in good versus bad markets were in the treatment 'for worse'.

[8] The explained variable is 1 if investment was on average larger in the good market year compared to the bad market year and is 0 otherwise. The results presented in table

Table II are robust to restricting the data set by excluding players that are equal in both markets (value = 0) or ignoring the previously observed outliers with values larger than 60.

Table II: Probit regression of higher investment in good versus bad market year

	higher investment in good year	higher investment in good year
student (dummy)	0.553	0.542
	(1.76)*	(1.45)
female (dummy)	0.164	-0.246
	(0.53)	(0.69)
bad year after good year (dummy)	0.735	0.750
	(2.45)**	(2.27)**
optimism (LOT)		0.890
		(2.21)**
anxiety (STAI)		0.949
		(1.72)*
impulsivity (UPPS)		-0.810
		(1.05)
openness (Big Five)		1.015
		(2.64)***
Constant	-1.637	-8.542
	(3.06)***	(2.85)***
Observations	77	77
Pseudo R^2	0.089	0.241

Absolute value of z statistics in parentheses

** significant at 10%; ** significant at 5%; *** significant at 1%*

2.2 Investment dynamics

To investigate investment strategies based on previous outcomes we will examine behavioral dynamics over rounds and years. In Table III (columns (1) and (3)) we present results from a random effects tobit regression of investment at time t on previous outcomes and relative earnings[9] as well as dummies for student, year (and period), and gender. We show separate regression results for investment made under bad market conditions (column 1) and good market

[9] Relative earnings at time t are the aggregate earnings of all previous time periods. Counting a positive outcome for an investment of level x as a gain of $2.5*x$ and a negative outcome as a loss of x. Thus relative earnings can be seen as a deviation of the current budget level from a references point. The reference point is taken as possible earnings if no investment would have been made so far.

conditions (column 3). As earlier results have shown (cf. Hopfensitz, 2009), both previous gains (i.e. having won in the previous rounds) and relative earnings (i.e. deviations from a reference point) have a significant impact on investment.

We observe evidence of the gambler's fallacy, namely that investment is reduced after a lucky round, as well as a general reduction in investment for higher levels of earnings. These results are also confirmed by answers to an open-ended question concerning the participants' investment strategy (see Appendix C). A large amount of participants refers in this to the idea of 'avoiding' or to 'equaling out' losses and to reactive strategies based on outcomes from earlier rounds. As predicted the gamblers fallacy is more pronounced for bad than for good market conditions. Further as observed in section 3.1 investment in the second year is generally lower. However across the periods of each year we observe a general positive time trend. We observe no significant difference between students and finance professionals, however a clear gender effect. Women take less risk in both market environments, a result consistent with the literature (Croson and Gneezy, 2009). Note that all of these effects are quite similar for good and bad market conditions.

RESULT 2: Results from a random effects tobit regression show that winning and relative earnings have a negative impact on investment for both market conditions. We observe no difference for students and professionals however a clear gender effect.

Table III: Random effects tobit regression of investment for good and bad market year

	"bad" year		"good" year	
	(1)	(2)	(3)	(4)
	investment	investment	investment	investment
win previous round (dummy)	-6.951	-7.462	-4.571	-3.866
	(2.54)**	(2.75)***	(1.73)*	(1.49)
total relative earnings at time t	-0.070	-0.060	-0.095	-0.103
	(6.42)***	(5.51)***	(9.51)***	(11.05)***
period (1 to 15)	0.717	0.755	1.113	1.208
	(2.78)***	(2.96)***	(3.73)***	(4.14)***
student (dummy)	-0.663	15.475	-4.170	9.266
	(0.15)	(3.53)***	(0.78)	(2.34)**
year (1 or 2)	-8.766	-14.617	-29.179	-14.576
	(2.35)**	(4.20)***	(6.04)***	(4.40)***
female (dummy)	-15.996	-6.220	-17.180	-21.298
	(3.96)***	(1.48)	(3.28)***	(5.58)***
optimism (LOT)		23.563		7.343
		(5.92)***		(1.70)*
anxiety (STAI)		22.084		18.687
		(4.21)***		(3.04)***
impulsivity (UPPS)		47.264		-15.193
		(5.42)***		(1.67)*
openness (Big Five)		-9.529		8.800
		(2.31)**		(2.34)**
Constant	69.848	-147.768	123.983	30.156
	(7.08)***	(5.55)***	(12.36)***	(0.95)
Observations	1155	1155	1155	1155
Number of id	77	77	77	77
Wald Chi 2	100.62	186.72	178.87	199.33

Absolute value of z statistics in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

2.3 Personality and investment

In addition to reactions to outcomes we expect a large heterogeneity of risk taking across participants. To understand which individual factors influence investment decisions we will now turn to an analysis of the personality profile of participants. Since professionals in the financial sector might be self-selected based on their personality we will first compare personality traits between students and finance professionals. We will then use personality characteristics to explain investment behavior in the two markets.

Table IV: Overview of personality variables for students and professionals.

	professional (N=31)		student (N=46)		
	Mean	Std. Dev.	Mean	Std. Dev.	Wilcoxon rank sum test
optimism (LOT)	3.75	0.46	3.46	0.62	p = 0.052
anxiety (STAI)	1.73	0.33	2.21	0.49	p = 0.000
impulsivity (UPPS)	2.22	0.23	2.29	0.21	p = 0.271
openness (Big Five)	3.72	0.45	3.72	0.53	p = 0.859

We first compared personality scores for professionals and students. Table IV gives an overview of mean ratings for optimism, anxiety, impulsivity and openness. Professionals score higher on optimism and lower on anxiety compared to students (Wilcoxon rank sum test, $p < 0.052$). We observe no other significant differences. The results concerning optimism and anxiety reflect the common perception of personality characteristics of finance professionals. It might seem surprising that we do not find any difference concerning impulsivity, however it should be noted that professionals score significantly lower on one of the impulsivity subscales: namely "lack of perseverance" (professionals: 1.82; students: 2.08; p=0.008).

Given these differences, we add personality measures to the tobit regressions (

Table III, columns (2) and (4)).[10] The effect of winning previously, relative earnings, and period in the year remain mostly unchanged. However, the student dummy now shows a positive coefficient. In addition, the gender effects is no longer significant for bad markets. Thus, the effect of being a student and female seem to be related to the included personality variables.

Overall, we observe that personality variables have a significant impact on investment behavior. Optimism and anxiety are significant for both market conditions. Optimism generally leads to higher investment, and especially so under bad market conditions. Given that participants lose money (on average) under bad market conditions, optimism leads to a stronger effect in bad than in good markets. In good markets optimism is less influential because investment is in itself advantageous (on average). Anxiety has a similar effect. What might be surprising is that anxiety is positively related with investment. Thus, trait anxiety is related to taking more risk, as predicted by the literature on gambling (Zangeneh et al, 2008) and not to reduced risk taking as has been suggested by economic models (Wu, 1999; Caplin and Leahy, 2001)[11].

Impulsivity shows a differential effect on behavior for good and bad markets. While impulsivity leads (on average) to more investment for bad markets, it leads (on average) to less investment for good markets. Since investment is favorable in good markets and unfavorable in bad markets, this implies that high levels of impulsivity will lead to unfavorable investment decisions and therefore losses in both market conditions. That impulsivity might have different effects given the

[10] To control for the robustness of our results we also ran regressions excluding the four finance professionals that were close to maximizing the expected value in both market conditions. Our qualitative results concerning optimism, anxiety and openness are unchanged (solely openness is no longer significant in good markets). However the fact that these participants were able to stay with a predetermined strategy and not to react to outcomes already indicates that they were not very impulsive. Consequently the significant coefficient for impulsivity disappears when these participants are excluded.

[11] Note that also earlier economic experiments did fail to observe reduced risk taking for participants high on trait anxiety (Hopfensitz and van Winden, 2008).

market conditions is important given the previous observation that sensation seeking will increase trading in general (Grinblatt and Keloharju, 2009).

Finally, we observe that openness reduces investment in bad markets.[12] People high in 'openness to experience' have a stronger preference for novelty, variety, and complexity (McCrae, 1996) and to be less conventional and think more deeply about new information than those low in 'openness to experience' (McCrae, 1987). Therefore, those high in this trait may have been more attentive to the probabilities presented and considered the implications of these probabilities. This result is further supported by an extension of the regression presented in section 3.1 (

Table II, column 2). When including personality traits in the regression of the probability of investing more in the good market year we observe a strongly significant positive effect for openness. We conclude the Openness to Experience might favor investment behaviors in changing markets, because new information is more likely to be integrated and used in decision-making.

RESULT 3: We observe significant differences in the personality profile of professionals and students. Moreover, personality variables significantly influence investment behavior and the specific impact depends on the market condition.

3 Summary and conclusions

Real markets are variable, and risk taking and investment in stock will be more advantageous in some periods rather than others. We wanted to determine if and how investors would adapt to

[12] This result is further supported by a significant correlation between mean investment in the bad market case and openness (corr. coef = -0.306, p = 0.008).

changing market conditions, and to examine if professional experience or personality could help predict the capacity to adapt. Surprisingly, professional investors did not show more adaptive responses to changing markets, than students.

Personality plays a role in risk taking and adaptation. Openness to experience can help investors to take into account new information to challenge dominant responses to risk taking. Being open to variable market conditions and alternative investment strategies can be an asset, and is a capacity that can be both selected and trained. Impulsivity, on the other hand, is unfavorable in all accounts. This was also found in an examination of long term investment strategies for clients in regards to their retirement investment (Ameriks et al., 2009). Thus, impulsivity and sensation seeking, which often characterize trader personality (Sjöberg and Engelberg, 2009), may have to be reconsidered. Finally, both optimism and anxiety have more complex relationships to risk taking and adaptation than previously thought, and the widely held belief that optimism is a positive and anxiety a negative trait for investment may prove to be false. Moderated levels of anxiety have indeed been shown to be an asset for long term investment decision-making (Ameriks et al., 2009). Given the current financial crisis, and the repeated demonstration that many financial institutions collectively take unreasonable levels of risk, a discussion on how professionals are selected and trained may warrant further exploration and discussions.

References:

Ameriks, J., T. Wranik and P. Salovey (2009). Emotional intelligence, personality, impulsivity, and investor behavior. *CFA Institute Research Foundation Publications.*

Bechara, A., H. Damasio, D. Tranel and A. Damasio (1997). Deciding advantageously before knowing the advantageous strategy. *Science*. 275:1293–1294.

Bellemare, C., M. Krause, S. Kroeger and C. Zhang (2005). Myopic loss aversion: information feedback vs. investment flexibility. *Economics Letters*. 87(3):319-324.

Brandstaetter, H. (1997). Becoming an entrepreneur -- A question of personality structure? *Journal of Economic Psychology*. 18(2-3): 157-177.

Burns, P. (1985). Experience in decision making: A comparison of students and businessmen in a simulated progressive auction, in V. L. Smith, ed. *Research in Experimental Economics* (JAI Press, Greenwich).

Caplin, A., and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55-79.

Carver, C. and M. Scheier (2001). Optimism, pessimism, and self-regulation. In Chang and Edward (eds.). *Optimism & pessimism: Implications for theory, research, and practice*. Washington: American Psychological Association.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*. 47(2): 448-474.

Croson, R. and J. Sundali (2005). The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos. *The Journal of Risk and Uncertainty*. 30:195-209.

Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*. 112(2):631-645.

Grinblatt, M. and M. Keloharju (2009). Sensation Seeking, Overconfidence, and Trading Activity. *Journal of Finance*. 64(2): 549-578.

Haigh, M. and J. List, 2005, Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance*, 60(1):523-534.

Holt, C., and S. K. Laury, 2002, Risk Aversion and Incentive Effect, *The American Economic Review*, 92(5):1644-1655.

Hopfensitz, A. (2009). Previous outcomes and reference dependence: A meta study of repeated investment tasks with and without restricted feedback. *MPRA working paper nr. 16096*

Hopfensitz, A. and van Winden, F. (2008). Dynamic Choice, Independence, and Emotions. *Theory and Decision,* 64: 249-300.

Hopfensitz, A. and T. Wranik (2008). Psychological and Environmental Determinants of Myopic Loss Aversion. *NETSPAR discussion paper.* nr. 2008-013

John, O. P. and S. Srivastava (1999). The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In L.A. Pervin and O.P. John (eds.), *Handbook of Personality: Theory and Research*. New York: Guilford Press.

Kahneman, D., and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica*. 47:263-291.

Lo, A.W, D.V. Repin and B. N. Steenbarger (2005). Fear and Greed in Financial Markets: A Clinical Study of Day-Traders. *The American Economic Review*. 95(2): 352-359.

Lynam, D. R. and J. D. Miller (2004). Personality pathways to impulsive behavior and their relation to deviance: Results from three samples. *Journal of Quantitative Criminology,* 20: 319-341.

McCrae, R. R. (1987). Creativity, divergent thinking, and Openness to Experience. *Journal of Personality and Social Psychology*, 52: 1258-1265.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin,* 120: 323-337.

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15:145-161.

Potters, J. and F. van Winden (2000). Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization*, 43(4): 499-522.

Shiv, B., G. Loewenstein and A. Bechara (2005a). The dark side of emotion in decision-making: When individuals with decreased emotional reactions make more advantageous decisions. *Cognitive Brain Research*. 23:85-92.

Shiv, B., G. Loewenstein, A. Bechara, H. Damasio and A. Damasio (2005b). Investment Behavior and the Negative Side of Emotions. *Psychological Science*. 16(6):435-439.

Sjöberg, L. and E. Engelberg (2009). Attitudes to economic risk taking, sensation seeking and values of business students specializing in finance. *Journal of Behavioral Finance*, 10: 33-43.

Spielberger, C. D. (1972). Anxiety as an emotional state. In *Anxiety: Current trends in theory and research, 1*. New York: Academic Press.

Thaler, R., A. Tversky, D. Kahneman and A. Schwartz (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics*. 112(2):647-661.

Tversky, A. and D. Kahneman (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Whiteside, S. P. and D.R. Lynam (2001). The Five Factor Model and Impulsivity: Using a Structural Model of Personality to Understand Impulsivity. *Personality and Individual Differences*, 30: 669-689.

Wu, G., (1999). Anxiety and Decision Making with Delayed Resolution of Uncertainty, *Theory and Decision*, 46(2):159-199.

Zangeneh, M., A. Grunfeld and S. Koenig (2008). Individual factors in the development and maintenance of problem gambling. In M. Zangeneh, A. Blaszczynski and E. Turner (eds.). *In the pursuit of winning: Problem gambling theory, research and treatment*. New York: Springer Science + Business Media.

Appendix A: Instructions

General instructions [for year 1 - good market; values for bad market in brackets; values for finance professionals in parenthesis]

Welcome: you are about to take part in a decision making experiment, in which you can earn real money. The experiment has 2 parts, which we will call years. The amount of money you can earn will depend on the decisions you make.

Dependent on your decisions, you can earn a significant amount of money.

During the experiment your earnings will be calculated in Unige Francs (UGF). At the end of the two years, these UGF will be converted into CHF and your earnings will be paid out to you in CHF using the following exchange rate:

100 UGF = 1 CHF **(2 CHF)**

At the beginning of the experiment you will receive from us 30 CHF **(60 CHF)**, which = 3000 UGF. This is your capital stock. You will have access to 1500 UGF of your capital stock at the beginning of each year. You can decide to either keep these UGF or to invest them in the experiment and try and earn more money. The details of this investment procedure will be explained to you below. At the end of the experiment we will pay you any earnings that you accumulated from the two years in addition to your 3000 UGF capital stock. If you lose money during the experiment, you will have to pay us back the losses from your capital stock at the end of the experiment.

During the experiment we will also ask you to answer a number of questions. These questions concern what you think and how you feel.

There are no right or wrong answers. You need to follow the decision strategy that feels right to you and to make those choices that come natural and that seem like the best choices for you. In addition, you should report those evaluations and emotions that are closest to your real thoughts and feelings. All answers are completely anonymous and confidential.

General instructions

During this experiment you will have to make investment decisions for 15 rounds in two investment years. This means that you will be making decisions for 30 rounds in total. In each of these rounds, you can invest 100 UGF from your capital stock of 3000 UGF.

We will now explain to you your options in year 1. After the 15 rounds of year 1 we will explain to you the situation in the second year.

Instructions for year 1

In year 1 you will have to make investment decisions for 15 rounds. In each of these rounds, you can invest 100 UGF from your capital stock. Each round you have to decide how you want to split these 100 UGF over two investment options.

We will call the two options: option A and option B.

Option A: In this option you will neither gain nor lose money. In other words, will always keep the number of UGF you put into option A.

Option B: The outcome from this option will be determined at the end of each round. In particular, we will pick one random number between 1 and 100. This is equivalent to picking a ball from an urn.

Imagine an urn with 100 balls in it, 33 [25] of these balls are orange, 67 [75] of these balls are blue.

- If the ball that is picked is orange (that is in 1/3 [1/4] of the cases) the UGF you placed in this option will be multiplied by 2.5. You will then receive 2.5 times the number of UGF you put into option B, in addition to the number of UGF you originally place into this option.
- If the ball that is picked is blue (that is in 2/3 [3/4] of the cases) you will lose the number of UGF you put into option B.

Note: For all fifteen rounds of year 1 we will always use the same urn. The number of orange and blue balls in this urn represents the market conditions of year 1.

Example

Imagine that in one round, you decide to split your 100 UGF by placing 50 UGF into option A and 50 UGF into option B.

If the randomly picked ball is orange (i.e. if the random number is smaller or equal to 33 [25]), you will receive 2.5*50 = 125, in addition to your 100 UGF for that round. Your capital stock will therefore increase by 125 UGF.

If the randomly picked ball is blue (i.e. if the random number is larger than 33 [25]), you will lose the 50 UGF you put in option B. Your capital stock will therefore decrease by 50 UGF.

Summary for year 1

- Year 1 is made up of 15 investment rounds.
- In each round, you can decide how to split 100 UGF of your capital stock between two options.
- At the end of each round we will pick a random number between 1 and 100.
- If the randomly picked ball is orange (i.e. if the random number is smaller or equal to 33 [25]) your earnings will be: 100 UGF + 2.5 * the number of UGF you placed into option B
- If the randomly picked ball is blue (i.e. if the random number is larger than 33 [25]) your earnings will be: 100 UGF – the number of UGF you placed into option B

Appendix B: Recruitment letter for professionals

Dear all,

Today I need your collaboration to help me with the research that is part of my master thesis in human resources, which I have been following by now for 18 months.

Do you know 'behavioral finance'? This is the study of investors' behavior in financial markets from a 'psychological' perspective or as stated by the American economist Richard Thaler, following an open minded approach.

Individual differences (personality, training …) play an important role in everyday decisions and studies have shown that they are also important in financial investment decisions.

[…] I would like to investigate with an experimental study the importance of personality differences for financial decisions. To conduct this research your professional help will be indispensable. You just have to give me 2 and a half hour of your time […].

The study will be conducted in small groups: each of you will work individually and completely anonymous (not only will you be protected by our professional vow of silence but we will also sign an individual declaration with each of you promising complete confidentiality). It is neither an exam nor an evaluation and our bank will only be informed about the aggregate results from this study, thus this research will be fully anonymous.

I am sure that this study will interest you … and it will even hold some small and nice surprises. […]

Appendix C: self reported investment strategies (translated from the French)

finance professionals

invest by year		invest by prob.		
1	**2**	**good**	**bad**	**self reported strategy**
0,00	0,00	0,00	0,00	no gambles
0,00	0,00	0,00	0,00	Calculate the expected gains and choose the option where it is higher (but, thinking about it, I fear to have miscalculated for the first part…)
11,33	0,60	0,60	11,33	
15,20	2,73	2,73	15,20	
15,00	7,00	15,00	7,00	conservation of capital
32,00	17,33	17,33	32,00	I started very prudent thus 2/3 option A and 1/3 option B, if during one parte the first draws were blue balls I increased option B, if however the first draws were positive I decreased the gamble in the latter and sometimes I put 100% in option A. Also note that in the second year I increased a bit the bets due to the number of orange balls with respect to blue balls.
20,00	20,00	20,00	20,00	conservation of capital, by placing 80% on the 100% safe, and by taking risk with 20%.
20,67	2,00	20,67	2,00	The probability of winning was very much to my disfavor, thus to win you should not gamble.
26,67	20,67	20,67	26,67	
25,33	18,80	25,33	18,80	All depends on the profile, the characteristics of the person, the global assets etc.
31,67	26,00	26,00	31,67	Do such that you take not too much risk concerning the investment and when the markets become too volatile try to keep as much capital as possible.
25,33	30,00	30,00	25,33	Distribution of risk, protection of capital and growth of portfolio.
63,33	33,53	33,53	63,33	After multiple identical tries 50/50 or 67/33, I believed that there would necessarily be a winning ball. Sometimes luck made things well.
80,00	34,00	34,00	80,00	During the second year I tried to bet on each option as a function of its expected gains.
53,33	35,60	35,60	53,33	
48,33	48,33	48,33	48,33	
43,33	48,67	48,67	43,33	Conservation of capital
54,00	54,00	54,00	54,00	Invest at the beginning an amount a bit higher or equal in option A. Then, in the case of gain, reduce option B in the favor of option A, in case of a loss, increase option B by reducing option A.
54,67	77,33	54,67	77,33	
55,33	34,67	55,33	34,67	Intermediate amount of risk.
58,67	90,00	58,67	90,00	
60,00	61,33	60,00	61,33	
63,33	60,00	60,00	63,33	
62,67	90,67	62,67	90,67	
74,67	12,00	74,67	12,00	Year 1: start by a small amount invested and double that amount until I win. Year 2: Worry of keeping the capital, less gambling, aiming better and less big, not to get under amount initially invested in year 1. I never play more than what I have already gained
76,00	80,00	76,00	80,00	I took a maximum of risk.
86,67	80,00	80,00	86,67	
93,33	0,00	93,33	0,00	A =0 B=100 in experiment 1; A=100 B=0 in experiment 2
100,00	0,00	100,00	0,00	The proposition was preferable in the long term, thus I bet all onto B. 2. The proposition was on the long term unfavorable, thus I bet all on A.
100,00	0,00	100,00	0,00	100% risk at the beginning then 0%, because probability changed from 33% to 25%.
100,00	100,00	100,00	100,00	

students				
invest by year		**invest by prob.**		
1	**2**	**good**	**bad**	**self reported strategy**
0,00	0,00	0,00	0,00	Not invest anything, thus no risk of losing money.
0,00	0,00	0,00	0,00	
0,00	0,00	0,00	0,00	
0,00	3,33	3,33	0,00	Not invest to keep my initial capital but I gave in to the temptation in one round
36,33	5,20	5,20	36,33	
11,33	5,87	5,87	11,33	
15,00	9,33	9,33	15,00	
10,67	10,00	10,67	10,00	It is worth to invest always a bit (10%) in the 29 rounds, but to keep a cold head. If you do not invest you also do not gain. You only keep your assets. If you invest all, you risk to loose all. The importance is to play adequately even though you cannot control all the parameters (here: in this test the programming of winning and losing rounds).
40,00	10,67	10,67	40,00	I tried to choose the logic that seemed to derive from the first year.
11,33	11,33	11,33	11,33	
40,33	12,33	12,33	40,33	Not to lose my initial capital. Do better in year 2 than in year 1.
9,33	13,33	13,33	9,33	Do not risk too much and do not go below a certain amount.
14,47	17,33	17,33	14,47	Sort of a hesitant luck… sometimes I tried, but if I lost I did not invest during the following round. Fear of losing too much….
18,00	9,87	18,00	9,87	The probabilities and the expectancies of gains for each round
24,33	14,33	24,33	14,33	Not much risk taking but by putting a bit hope on gaining a little more!
29,67	25,33	25,33	29,67	prudence
23,33	26,67	26,67	23,33	Place more when the probabilities where advantageous. Place more if I just won something, less when I just lost.
29,33	25,33	29,33	25,33	
33,53	13,67	33,53	13,67	
45,47	35,67	35,67	45,47	1 out of 4 cases the victory is possible (on average) - Gamble a large amount (from time to time) to try to win the most at the beginning and then once the victory was obtained, just put 10 more or less since a gain is won.
36,33	21,99	36,33	21,99	In year 2 I put more about 1 out of 4
86,67	36,67	36,67	86,67	
21,67	38,33	38,33	21,67	With respect to the number of the round. The 5 is my lucky number (normally) thus I waited for round nr. 5. Besides at the end I put large amounts because any ways I had lost a lot thus why not try to increase.
34,20	38,87	38,87	34,20	Calculate the probabilities and 6th sense.
43,00	61,33	43,00	61,33	Take the maximum amount of risk by limiting the losses.
46,00	45,33	46,00	45,33	
50,00	26,67	50,00	26,67	At the beginning I played all, then I tried to do half half, and finally I just wanted to conserve my gains.
50,00	60,00	50,00	60,00	I tried different combinations
70,00	53,67	53,67	70,00	
54,33	29,33	54,33	29,33	Count the number of times that I could win and the likely order in which the computer could present the good/bad investments.
35,33	55,00	55,00	35,33	Equalizing the losses encountered during one round by increasing the investment in the following round.
46,67	55,00	55,00	46,67	
54,40	56,00	56,00	54,40	From the second year on, taking the probabilities into account: if the probability is 1/3 and I invest 70 each round, after three rounds I lose 140 but I gain 175, thus gain of 35. However the probability of 1/3 rarely came true.
58,33	46,67	58,33	46,67	

invest by year		invest by prob.		
1	**2**	**good**	**bad**	**self reported strategy**
60,00	48,33	60,00	48,33	trust and calmness
60,00	100,00	60,00	100,00	
60,33	63,67	63,67	60,33	Very carefully, during the second year, since I had the right for three trials, I put once 20 in option A and 80 in option B, I checked how that went. Then, I put 80 in A and 20 in B. Like that I kept a basis, then I put for example 100 in B, if I won it was good too bad when I lost.
64,33	72,33	64,33	72,33	You only won once in a part, thus once I won I didn't invest much.
66,67	46,67	66,67	46,67	During the first part, when I lost twice in a row 50, I tried to double how much I invested in option B. During the second part, I was more reserved, because the chances of winning decreased. Thus I split my investment between A and B. When I lost too much in the second part, I sometimes continued dividing between A and B, because else there is no interest in the game.
73,33	46,67	73,33	46,67	
82,00	75,33	75,33	82,00	
80,00	100,00	80,00	100,00	
100,00	86,67	86,67	100,00	Try to calculate the mathematic gain, positive for the first experiment and negative for the second. Then, do as Closewitz, do not change your strategy until the numbers make sense… which was not the case here.
100,00	98,67	98,67	100,00	
50,00	100,00	100,00	50,00	
100,00	100,00	100,00	100,00	

Désinflation et chômage dans la zone euro : une analyse à l'aide d'un modèle VAR structurel *

Patrick FEVE †, Julien MATHERON ‡, Jean-Guillaume SAHUC §

14 octobre 2010

A paraître dans *Annales d'Economie et de Statistique*

*Nous tenons à remercier deux rapporteurs anonymes et le rédacteur en chef (S. Roux) ainsi que F. Collard, M. Dupaigne, C. Pfister et F. Portier pour leurs précieuses remarques et suggestions. Nous remerçions également la *Financière de la Cité* pour son soutien. Nous restons seuls responsables des erreurs et imprécisions rémanentes. Les vues exprimées dans cet article sont nôtres et n'engagent en aucune façon la Banque de France.

†Toulouse School of Economics (Université de Toulouse, GREMAQ, IDEI et IUF) et Banque de France.

‡Banque de France (DEMS-SEPS) et SDFi – Université de Paris-Dauphine.

§Banque de France (DEMS-SEPS) et Audencia Ecole de Management.

Résumé :
Cet article étudie les effets dynamiques des chocs de désinflation sur un ensemble de variables réelles et nominales de la zone Euro. A l'aide d'un modèle VAR structurel, nous identifions le choc de désinflation comme étant le seul choc ayant un effet permanent sur l'inflation à long terme ainsi que sur les variables nominales qui co–intègrent avec elle. Nous montrons que ce choc a des effets récessionnistes importants, notamment sur l'investissement, et provoque une hausse persistante du taux de chômage et du taux d'intérêt réel. Nous complétons cette analyse en calculant des mesures d'inefficience sur les marchés des biens et du travail et nous montrons que ce sont principalement ces dernières qui ont prévalues. Ces conclusions sont robustes à différentes mesures et schémas d'identification des chocs de désinflation.

Mots–Clés : modèle VAR Structurel, restriction de long terme, désinflation

Classification JEL : C32, E31, E52

Abstract :
The present paper investigates the dynamic effects of disinflation shocks for a number of real macroeconomic variables in the euro area. Using structural VARs, we identify disinflation shocks as the only shocks that can exert a long–run effect on inflation as well as other nominal variables cointegrating with inflation. These shocks are found to generate large recessionary effects, notably when it comes to investment, and triggers a persistent rise in unemployment and in the real interest rate. The analysis is complemented by computing inefficiency measures on goods and labor markets. We show that, after a disinflation shock, inefficiencies in the labor market seem to prevail. These conclusions are robust to modifications of our baseline identification scheme.

Keywords : SVAR, long–run restrictions, disinflation

JEL Classification : C32, E31, E52

1 Introduction

Le taux d'inflation dans la zone Euro a fortement baissé durant les dernières décennies, passant de 12% au début des années 1980 à 4% au début des années 1990. En dépit de certaines disparités, les inflations des différents pays de la zone présentent toutes une tendance marquée à la baisse sur cette période [1]. Durant cette phase de désinflation, le taux de chômage a augmenté de façon marquée et l'activité économique, notamment l'investissement, a sensiblement chuté dans la zone Euro. La figure 1 illustre ce fait pour la période 1980(1)–1990(1). Bien que l'évolution des variables réelles et nominales au cours de cette période puisse résulter d'autres facteurs et/ou chocs, un changement permanent dans la politique monétaire visant à réduire le taux d'inflation semble être un candidat explicatif naturel [2]. L'objectif de cet article est précisément d'évaluer l'importance quantitative de cette thèse.

Le lien entre activité réelle et inflation a fait l'objet d'une littérature empirique abondante. Une première approche utilise des régressions en coupe pour étudier les effets de l'inflation sur la croissance (voir par exemple Andres et Hernando, 1999, Kormendi et Meguire, 1985, McCandless et Weber, 1995). Cette approche se concentre exclusivement sur des relations de long terme et fait abstraction des dynamiques d'ajustement à court terme. Une seconde approche étudie ces dynamiques de court terme à l'aide de modèles VAR Structurels (VARS) récursifs (voir Christiano *et alii.*, 1999, pour un panorama exhaustif). Ces modèles supposent en général que ces chocs sont transitoires et ne représentent que des déviations de court terme vis à vis d'un comportement régulier de la politique monétaire. Ainsi, les modifications permanentes de l'inflation, telles qu'observées dans la zone Euro durant les trente dernières années (et surtout durant les années quatre–vingts), ne peuvent pas être appréhendées et analysées à l'aide de ce cadre empirique.

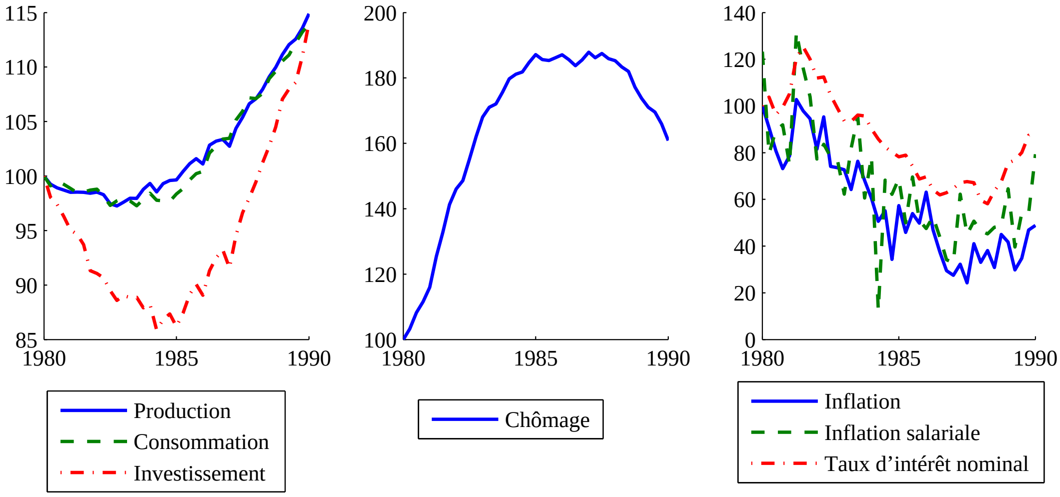
L'objectif de cet article se situe au carrefour de ces deux approches empiriques. Plus particulièrement, nous cherchons à étudier les réponses de court et moyen termes de l'économie faisant suite à des modifications permanentes de la cible d'inflation. La difficulté principale réside alors dans l'identification du choc de désinflation. A cette fin, nous abandonnons la représentation récursive discutée par Christiano *et alii.* [1999] et suivons la méthodologie VARS introduite par Bullard et Keating [1995] [3]. Cette approche adapte au cas des chocs de désinflation la technique d'identification des chocs d'offre proposée par Blanchard et Quah [1989]. Plus précisément, le choc de désinflation est identifié comme le seul choc

1. Voir Blanchard et Muet [1993] et Cling et Meunier [1986] pour des analyses sur données françaises.

2. Fitoussi et Passet [2000] montrent que la politique monétaire a eu une grande influence sur l'évolution du chômage.

3. Voir aussi Andres *et alii.* [2002], Coenen et Vega [2001], Dolado *et alii.* [2000] et Vlaar [2005].

FIGURE 1 – Désinflation des années quatre–vingt



Notes : toutes les variables sont normalisées à 100 en 1980 :1.

ayant un effet permanent sur le taux d'inflation ainsi que sur les variables nominales qui cointègrent avec elle (inflation salariale, taux d'intérêt nominaux, taux de croissance de la monnaie dans notre cas). Notre spécification de référence impose de plus la neutralité de long terme du choc de désinflation, interprété ici comme un choc permanent de politique monétaire, suivant ainsi une version faible de la doctrine monétariste (l'inflation est toujours un phénomène monétaire dans le long terme). Une originalité de notre démarche est de combiner l'approche VARS avec le calcul de différentes mesures d'inefficience sur les marchés des biens et du travail basées sur un exercice de comptabilité du cycle à la GALÍ *et alii.* [2003, 2007]. Cette analyse nous permet d'évaluer les conséquences d'un choc de désinflation et de mesurer quelles inefficiences ont joué un rôle dans la propagation de ses effets.

Comme l'a suggéré BLANCHARD [2003], deux mécanismes concourent à expliquer les effets récessionnistes d'un choc de désinflation. En premier lieu, un tel choc peut entraîner un accroissement persistant du taux d'intérêt réel, se traduisant par une baisse de l'investissement. En second lieu, si les salaires nominaux sont plus visqueux que les prix, en raison de rigidités nominales ou réelles, un choc de désinflation peut entraîner une hausse du salaire réel et un accroissement du chômage. Une contribution du modèle VARS envisagé ici est d'intégrer différentes variables permettant d'étudier ces mécanismes de propagation. Plus précisément, nous prêtons une attention particulière au taux de chômage, au salaire réel, à l'investissement et au taux d'intérêt réel.

A partir de notre schéma d'identification, nous trouvons que le choc de désinflation a un effet persistant et important sur l'activité réelle. Il induit une baisse persistante du produit réel par tête, se traduisant par un ratio de sacrifice d'environ 5% à l'horizon de cinq ans, où ce nombre est mesuré traditionnellement comme la baisse cumulée de la production rapportée à la variation de l'inflation annualisée. En d'autres termes, la politique de désinflation aurait engendré un manque à gagner correspondant à 5% de la production potentielle en l'absence de cette politique. La réponse de l'investissement présente une baisse encore plus marquée, reflétant la réponse positive et persistante du taux d'intérêt réel (à court et à long termes). Enfin, la réponse dynamique du chômage suggère un arbitrage persistant (au moins dans le moyen terme) avec l'inflation, puisque le taux de chômage augmente de façon persistante pour atteindre un pic positif après deux ans. En revanche, le salaire réel présente une réponse positive dans le très court terme mais décroît rapidement et s'ajuste à la baisse. Nous complétons ces résultats à l'aide du calcul de la réponse des marges sur les prix et les salaires. Cette dernière augmente de façon persistante et présente un profil en cloche, détériorant ainsi l'efficacité globale de l'économie. Nous évaluons finalement la robustesse de nos résultats à des spécifications alternatives du modèle VARS.

L'article est organisé comme suit. Dans une première section, nous présentons l'identification d'un choc de désinflation à partir d'un modèle VAR Structurel et nous discutons des résultats pour une représentation de référence. Une seconde section reporte différentes analyses de robustesse (prise en compte des chocs pétroliers, définitions alternatives de l'inflation, relâchement de certaines contraintes identifiantes). Une dernière section conclut.

2 Caractérisation des chocs de désinflation dans la zone euro

Dans cette section, nous présentons dans un premier temps la procédure d'identification du choc de désinflation à l'aide d'un modèle VARS. Nous exposons ensuite les effets dynamiques de ce choc sur un ensemble de variables d'intérêt. Enfin, nous menons un exercice de comptabilité du cycle à l'aide du modèle VARS et de restrictions théoriques additionnelles afin de caractériser les conséquences d'un choc de désinflation sur les marges de prix et de salaires.

2.1 Identification à partir d'un modèle VAR structurel

Nous étudions des données trimestrielles relatives à la zone euro, sur la période 1980(1)–2004(4). Nous nous intéressons au vecteur Z_t défini comme

$$Z_t = (\Delta\pi_t, \hat{y}_t, c_t - y_t, x_t - y_t, \hat{u}_t, \hat{w}_t, i_t^l - \pi_t, i_t^l - i_t^s, \gamma_t - \pi_t)', \tag{1}$$

où $\Delta\pi_t$ est la différence première du taux d'inflation π_t, $\hat{y}_t$ est la composante cyclique du logarithme du PIB, $c_t - y_t$ est le logarithme du rapport consommation sur PIB, $x_t - y_t$ est le logarithme du rapport investissement sur PIB, $\hat{u}_t$ est la composante cyclique du taux de chômage, $\hat{w}_t$ est la composante cyclique du logarithme du salaire réel, i_t^l est le taux d'intérêt nominal de long terme, i_t^s est le taux d'intérêt nominal de court terme et γ_t est le taux de croissance de M3[4]. Ce dernier est corrigé de la réunification allemande (1990(3)) et de l'entrée de la Grèce dans la zone Euro (2001(1)). Plus précisément, γ_t est obtenu comme le résidu de la régression de la série d'origine sur des variables indicatrices correspondant à ces deux dates.

Le choix des variables dans Z_t est en partie inspiré de Blanchard [2003]. Ce dernier identifie deux conséquences possibles d'une désinflation : en premier lieu, un accroissement du taux d'intérêt réel, se traduisant par une hausse du prix du capital et une baisse de l'investissement ; en second lieu, un accroissement des salaires réels, dû à des rigidités nominales ou réelles, et se traduisant par une hausse du chômage. Les variables incluses dans Z_t permettent en principe de capter ces mécanismes. Par ailleurs, le taux court est inclus dans l'analyse pour deux raisons. En premier lieu, nous voulons étudier la dynamique du taux court, vu comme instrument de la politique monétaire (en complément de M3). En second lieu, en considérant à la fois le taux court et le taux long, nous pouvons quantifier le niveau de crédibilité de la politique de désinflation. En particulier, une réaction du taux long plus prononcée que celle du taux court peut être interprétée comme un signe de crédibilité.

Les composantes cycliques du PIB, du salaire réel et du chômage sont obtenues à l'aide du filtre à bande passante de Christiano et Fitzgerald [2003], en éliminant les mouvements de période supérieure à 60 trimestres. Ce faisant, nous suivons la définition du moyen terme adoptée par Blanchard [1997] et par Staiger *et alii.* [2002][5]. Un intérêt du filtre de Christiano et Fitzgerald [2003] est qu'il permet d'éliminer la présence d'une racine unitaire dans la série originale. De fait, si les tests usuels de racine unitaire rejettent sans ambiguïté l'hypothèse de racine unitaire pour les ratios consommation–produit et investissement–produit, ils donnent des résultats plus ambigus pour le produit, le taux de chômage et le salaire réel, comme le suggèrent les résultats du tableau 1.

Pour compléter la discussion de la spécification de Z_t, il est aussi important d'évaluer si l'inflation et les autres variables nominales incluses dans le vecteur Z_t peuvent être caractérisées par un processus intégré d'ordre un dans notre échantillon. A cette fin, nous effectuons les mêmes tests de racine unitaire que ci-dessus. Les résultats sont reportés dans le tableau 2.

4. Voir l'annexe A pour une définition détaillée des données et de leurs sources. Les données brutes sont reportées sur le graphique A.1.

5. Galí *et alii.* [2003] adoptent la même procédure d'extraction de la composante cyclique.

Tableau 1. Tests de racine unitaire sur les variables réelles

	ADF	PP
u_t	-2.61	-3.15
	[9.43]	[2.63]
y_t	-2.78	-3.26
	[20.88]	[7.91]
$c_t - y_t$	-2.76	-2.76
	[6.70]	[6.70]
$i_t - y_t$	-3.00	-4.07
	[3.85]	[0.17]
w_t	-0.72	-0.72
	[96.87]	[96.87]

Notes : ADF : test de Dickey-Fuller Augmenté. Les retards sont selectionnés à l'aide du critère Hannan-Quinn. PP : tests de Phillips-Perron, avec spécification AR spectrale estimée par MCO, où les retards sont là aussi séléctionnés à l'aide du critère Hannan-Quinn. Pour u, $c-y$ et $i-y$, les tests autorisent seulement une constante tandis que pour y et w ils autorisent à la fois une constante et une tendance linéaire. Dans tous les cas, l'hypothèse nulle est la racine unitaire. Les seuils des statistiques du test sont entre crochets et en pourcentage.

De manière générale, les tests suggèrent que les différentes variables nominales incluses dans Z_t, c'est à dire l'inflation, les taux d'intérêt nominaux de court et long termes et le taux de croissance de la monnaie, présentent toutes une racine unitaire. Nous obtenons de plus que ces variables en différence première sont toutes stationnaires. A partir de ces propriétés univariées, nous étudions dans un second temps s'il existe une relation de long terme entre ces variables. Nous testons alors si le taux d'intérêt réel de long terme ex–post $(i_t^l - \pi_t)$, l'écart taux long–taux court $(i_t^l - i_t^s)$ et le taux de croissance des encaisses réelles $(\gamma_t - \pi_t)$ sont stationnaires. De façon générale, les résultats sont plutôt favorables à la stationarité de ces variables, ce qui suggère qu'elles partagent la même tendance stochastique de long terme. Nous imposons donc cette relation de cointégration dans le modèle VAR Structurel.

Nous supposons que le vecteur Z_t évolue selon le processus VAR

$$Z_t = A_0 + \sum_{i=1}^{\ell} A_i Z_{t-i} + e_t \tag{2}$$

où ℓ est le nombre de retards et e_t est un bruit blanc de matrice de variance–covariance Σ. Le nombre de retards ℓ est sélectionné à l'aide de critères d'information standards. Dans nos différentes expériences, nous obtenons $\ell = 1$ ou $\ell = 2$.

Notre hypothèse d'identification des chocs de désinflation consiste à imposer que seuls ces

Tableau 2. Tests de racine unitaire sur les variables nominales

	Niveau		Différence première	
	ADF	PP	ADF	PP
π_t	-1.99	-2.67	-9.66	-14.23
	[29.09]	[8.30]	[0.00]	[0.00]
i_t^l	-0.78	-0.46	-5.34	-9.67
	[81.93]	[89.34]	[0.00]	[0.00]
i_t^s	-1.19	-1.00	-6.17	-6.17
	[67.70]	[74.32]	[0.00]	[0.00]
γ_t	-2.27	-6.11	-7.83	-27.84
	[18.44]	[0.00]	[0.00]	[0.00]
$i_t^l - \pi_t$	-3.30	-4.26		
	[1.74]	[0.09]		
$i_t^l - i_t^s$	-3.22	-3.07		
	[2.19]	[3.13]		
$\gamma_t - \pi_t$	-5.79	-5.73		
	[0.00]	[0.00]		

Notes : ADF : test de Dickey-Fuller Augmenté. Les retards sont selectionnés à l'aide du critère Hannan-Quinn. PP : tests de Phillips-Perron, avec spécification AR spectrale estimée par MCO, où les retards sont là aussi séléctionnés à l'aide du critère Hannan-Quinn. Nous incluons une constante, mais pas de tendance déterministe. Dans tous les cas, l'hypothèse nulle est la racine unitaire. Les seuils des statistiques du test sont entre crochets et en pourcentage.

derniers peuvent affecter le niveau de l'inflation à long terme[6]. Cette restriction d'identification est analogue à celle initialement proposée par BLANCHARD et QUAH [1989] dans le cadre des chocs d'offre. Formellement, nous supposons que les innovations canoniques e_t sont une combinaison linéaire des chocs structurels η_t, soit $e_t = S\eta_t$, où S est une matrice de passage. Dans un premier temps, nous supposons que les chocs structurels sont orthogonaux et normalisons ces derniers de telle sorte que $\mathrm{E}\{\eta_t\eta_t'\} = I_m$, où I_m est la matrice identité et

6. Il faut noter que le schéma d'identification adopté dans ce papier implique que le choc de désinflation peut avoir lieu à toutes les périodes de l'échantillon. Cette stratégie empirique diffère d'une approche narrative (voir ROMER et ROMER, 1989 et 1994) qui sélectionne un petit nombre d'épisodes associés à des politiques actives de désinflation. Comme l'ont noté CHRISTIANO *et alii.*. [1999], un avantage de l'approche narrative réside dans le fait que l'économètre n'a pas à spécifier explicitement une règle monétaire ou à imposer un schéma particulier d'identification afin d'identifier les réponses de l'économie. Une telle approche n'est pas possible dans notre cas dans la mesure où il n'existe pas d'analogues des minutes du FOMC ni pour les banques centrales nationales de la zone euro ni pour la BCE.

m la dimension de Z_t. Posons alors $C(L) = B(L)S$, où

$$B(L) = (I_m - A_1 L - \cdots - A_\ell L^\ell)^{-1}.$$

Dans un second temps, étant donné l'ordre des variables dans Z_t dans l'équation (1), nous imposons que seul le premier élément de la première ligne de $C(1)$ soit a priori non nul. En pratique, nous sélectionnons S de telle sorte que $C(1)$ soit la matrice triangulaire inférieure obtenue par la décomposition de Cholesky de $B(1)\Sigma B(1)'$. Les chocs de désinflation ainsi identifiés sont invariants à toute transformation orthonormale de la matrice extraite de S en éliminant la première ligne et la première colonne de cette dernière.

Compte tenu de la spécification retenue, le choc de désinflation aura un effet permanent non seulement sur l'inflation mais aussi sur les autres variables nominales incluses dans Z_t qui cointègrent avec elle, c'est–à–dire les taux d'intérêt nominaux de court et long termes et le taux de croissance de M3. De plus, en spécifiant la composante cyclique du produit, du salaire réel et du taux de chômage dans Z_t et en considérant les rapports consommation–produit et investissement–produit, nous imposons implicitement que les chocs de désinflation ne peuvent pas avoir d'effets à long terme sur les variables réelles du modèle VAR Structurel. En d'autres termes, les chocs de désinflation peuvent s'interpréter comme des chocs permanents sur la politique monétaire qui sont neutres dans le long terme.

2.2 Les effets agrégés d'un choc de désinflation

Avant de présenter les réponses des variables agrégées, nous étudions le comportement temporel du choc de désinflation. Suivant certaines études empiriques (Stock et Watson, 2007, et Cogley et Sargent, 2007) ainsi que des spécifications plus structurelles (Ireland, 2007, Smets et Wouters, 2005, et de Walque, Smets et Wouters, 2006), nous supposons que la cible implicite d'inflation de la banque centrale évolue selon une marche aléatoire

$$\pi_t^\star = \pi_{t-1}^\star + \sigma_\pi \eta_t^\pi \tag{3}$$

où η_t^π est le choc de désinflation identifié à partir du modèle (2) selon notre schéma d'identification et σ_π est l'élément 1×1 de S. En procédant de la sorte, nous redimensionnons les chocs de désinflation pour les rendre de même "taille" que les données. La valeur initiale de $\pi_t^\star$ est fixée à la moyenne de l'inflation observée sur la période 1975(1)–1979(4). Cette représentation est évidemment une simplification, introduite à des fins illustratives. En effet, le modèle VARS nous permet d'identifier l'innovation structurelle de la cible implicite d'inflation. En revanche, le modèle VARS reste silencieux quant au comportement du banquier central et ne peut donc pas être utilisé pour identifier structurellement le processus suivi par cette cible implicite.

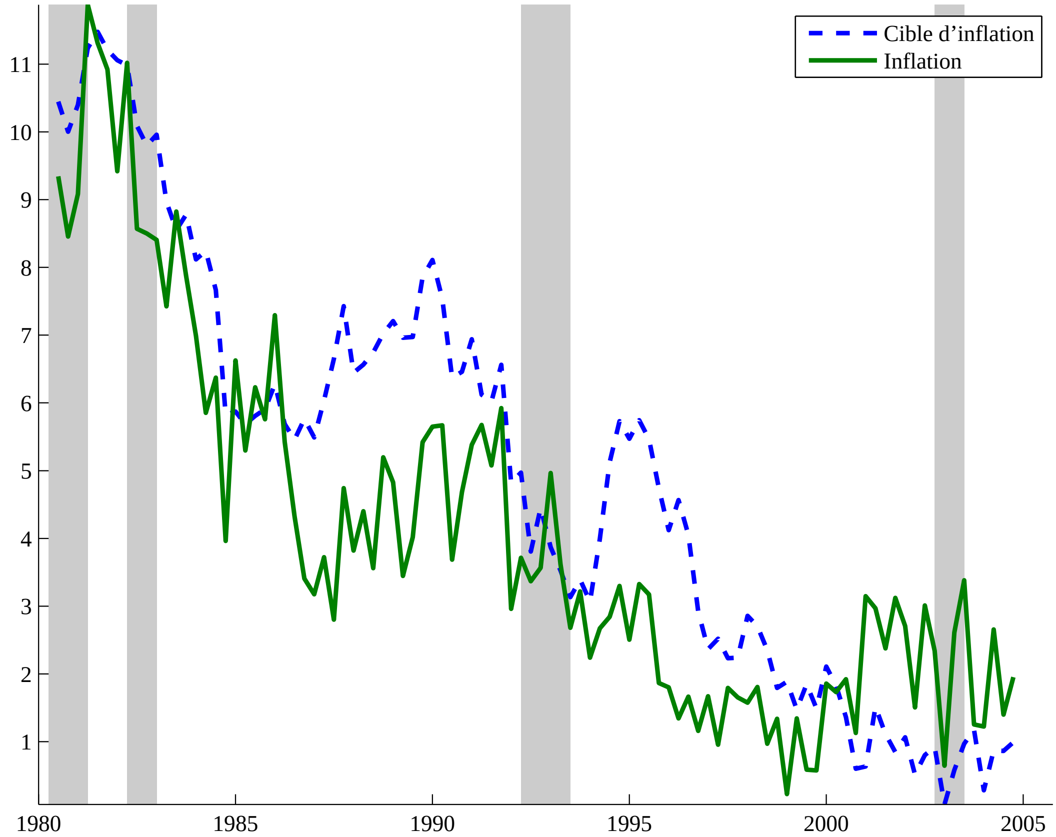
Il convient de remarquer que, pour l'essentiel de la période d'estimation considérée dans cette étude (1980(1)-2004(4)), les pays constituant la zone euro avaient chacun une banque centrale et ne mettaient pas en œuvre une politique monétaire unique. Le choix de modélisation retenu ici, d'ailleurs adopté dans la majeure partie de la littérature (en particulier dans la littérature SVAR, comme chez Peersman et Smets, 2001 ou DSGE, comme chez Smets et Wouters, 2003), est de traiter la banque centrale unique sous–jacente à l'équation (3) comme une banque centrale virtuelle, moyenne pondérée des 12 banques centrales nationales de l'époque. Cette approche a le défaut de gommer l'hétérogénéité des politiques monétaires précédant l'introduction de l'euro. Toutefois, cette approximation est cohérente avec la rapide adoption par de nombreux pays d'objectifs de politique monétaire convergents, en particulier, la réduction du niveau de l'inflation, comme le suggère le graphique A.2 qui reporte les trajectoires d'inflation des 12 pays membres avant l'adoption de l'euro. Ce graphique illustre clairement que les pays membres ont quasiment tous procédé à des politiques de désinflation en même temps ou, à tout le moins, sur la même période. Les épisodes de désinflation identifiés par Ball [1994] et Anderson et Wascher [1999] pour les pays membres les plus importants confirment cette observation. Par exemple, ces études identifient les épisodes 1981–1986 pour la France, 1980–1986 pour l'Allemagne et 1980–1987 pour l'Italie. En outre, vers la fin de l'échantillon, du fait des critères de convergence imposés par le traité de Maastricht, les taux d'inflation convergent quasi mécaniquement vers un niveau commun. De ce fait, il ne semble pas y avoir de changement de régime entre les périodes pre et post 1999. C'est en tous les cas l'hypothèse de travail retenue dans cet article.

La figure 2 reporte cette cible. Les aires grisées correspondent aux différents épisodes de récession, obtenus en utilisant une méthode similaire à celle de Watson [1994] [7]. Ces épisodes de récession sont 1980(1)–1981(1), 1982(1)–1982(4), 1992(1)–1993(2), et 2002(3)–2003(2). Ces dates coïncident à peu près avec celles identifiées par le "Euro Area Business Cycle Dating Committee". La seule exception est la récession du début des années quatre–vingt qui est séparée en deux récessions avec notre procédure. La différence provient du fait que nous utilisons un algorithme mécanique de sélection à partir du PIB par tête, alors que le comité adopte une procédure de jugement.

Cette figure met en évidence la grande similitude d'évolution du taux d'inflation de la zone euro et de la cible implicite d'inflation poursuivie par la politique monétaire. Il faut rappeler

7. Plus précisément, les dates de récession sont identifiées en utilisant l'algorithme de Bry–Boschan sur le logarithme du niveau du PIB réel par tête.

FIGURE 2 – Cible d'inflation identifiée par le modèle VARS

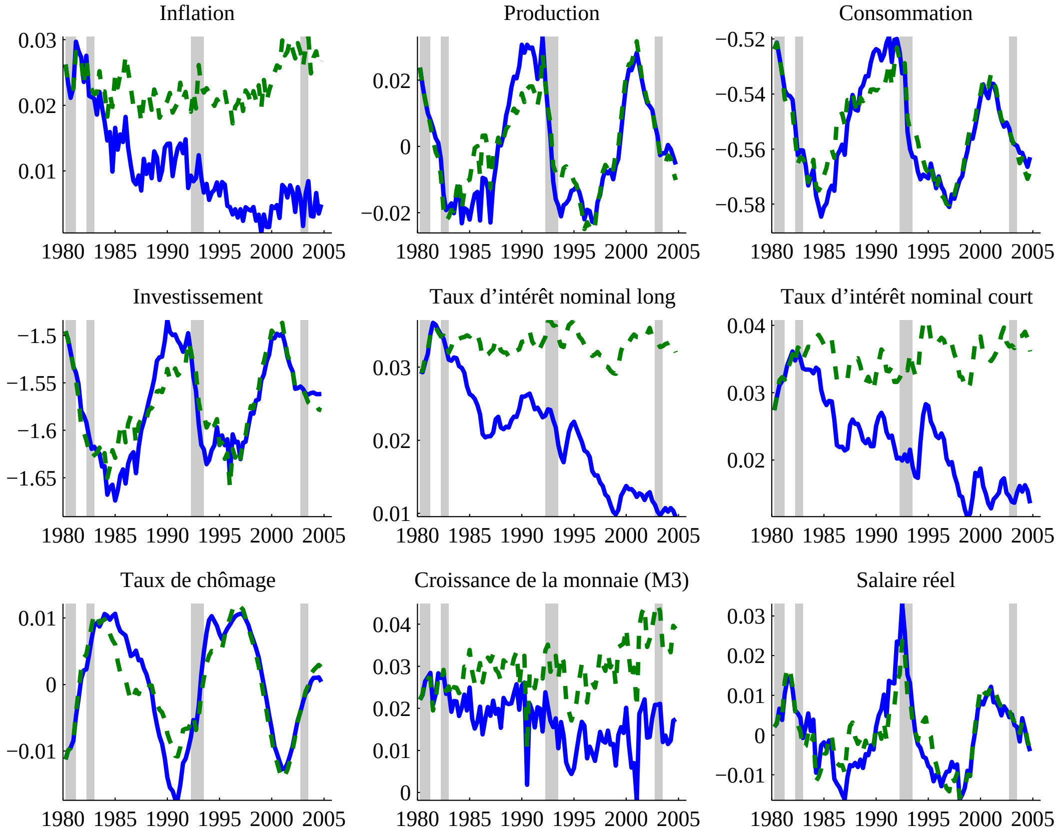


Notes : la cible d'inflation est identifiée à partir du modèle (2) et l'équation (3), en utilisant le schéma d'identification discuté dans le texte. La zone grisée correspond aux dates de récession identifiées à l'aide de l'algorithme de Bry–Boschan appliqué au logarithme du PIB par tête en niveau.

que notre schéma d'identification suppose une origine monétaire de la composante non–stationnaire de l'inflation. Ces deux variables tendent toutes les deux à baisser fortement dans les premières moitiés des années quatre–vingts et quatre–vingt–dix. La politique de désinflation ainsi identifiée par le modèle VARS peut être associée à deux grandes récessions de la zone Euro, c'est à dire celles de 1982(1)–1982(4) et de 1992(1)–1993(2).

Dans l'esprit d'IRELAND [2007], nous complétons cette analyse en simulant le modèle VAR sous l'hypothèse contre–factuelle que les chocs de désinflation ont été identiquement nuls sur l'échantillon. Nous reconstruisons alors les trajectoires des variables d'intérêt. Les résultats sont reportés sur la figure 3 pour les composantes cycliques du produit, de la consommation, de l'investissement et du chômage ainsi que pour l'inflation, le taux de croissance de la monnaie et les taux d'intérêt nominaux de court et de long termes. Le graphique illustre que sur la période 1982–1988, les composantes cycliques du produit, de la consommation et de l'investissement auraient été plus élevées sans chocs de désinflation, en restant toutefois sous leur niveau tendanciel. Concernant le chômage, l'effet de la désinflation se fait essentiellement

FIGURE 3 – Trajectoires des variables avec ou sans chocs de désinflation



Notes : les lignes continues correspondent au cas avec choc de désinflation ; les lignes brisées correspondent au cas où le choc de désinflation est identiquement nul sur l'échantillon.

sentir dans la première moitié des années 1980. Ces différences semblent moins marquées sur la fin de l'échantillon, après 1995. En ce qui concerne les variables nominales, la différence est, par construction, beaucoup plus criante. Il est cependant intéressant de remarquer qu'en l'absence de chocs de désinflation, l'inflation et le taux de croissance de la monnaie se seraient fortement accrus passé 2000. Les chocs de désinflation ont servi à contrebalancer les effets inflationnistes des autres chocs.

Pour mieux comprendre les résultats précédents, nous présentons maintenant les fonctions de réponse à un choc de désinflation dans notre spécification (1) de Z_t. Pour chacune des réponses, nous calculons les intervalles de confiance à l'aide d'une technique de bootstrap [8].

8. Nous calculons $N = 1000$ simulations bootstrap du modèle VAR structurel en construisant N nouvelles séries temporelles des résidus canoniques, notées $\{e_t(i)\}_{t=1}^T$, $i = 1, ..., N$, où le $t^{\text{ème}}$ élément de $\{e_t(i)\}_{t=1}^T$ est tiré avec remise dans $\{e_t\}_{t=1}^T$. En utilisant les paramètres estimés du modèle VAR et les valeurs historiques observées comme valeurs initiales, nous construisons alors N séries temporelles de Z_t, $\{Z_t(i)\}_{t=1}^T$. Pour chaque simulation, le modèle VAR spécifié dans l'équation (2) est estimé et les fonctions de réponse sont calculées à partir de la matrice de passage estimée à chaque simulation.

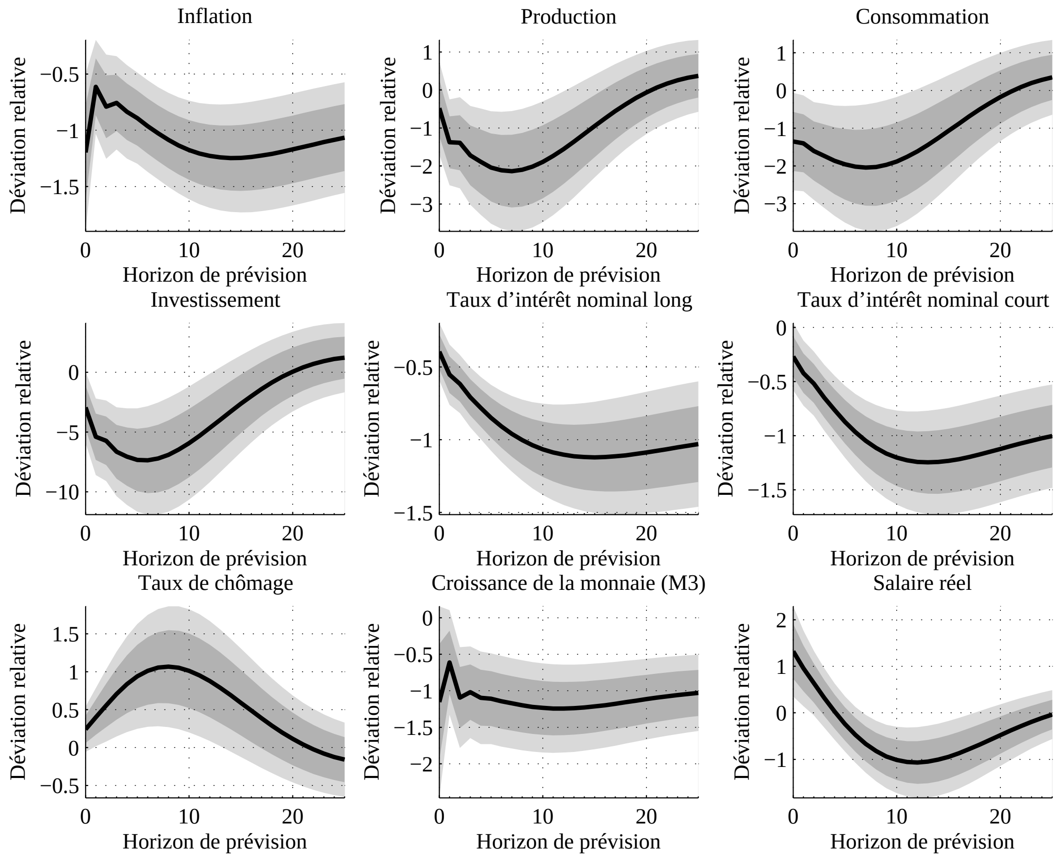
Afin de simplifier l'interprétation, nous reportons les réponses de l'inflation, du taux de croissance de la monnaie et des taux d'intérêt nominaux ainsi que les réponses des composantes cycliques de la consommation, de l'investissement [9]. Ces réponses sont obtenues à partir de celles des différents éléments du vecteur Z_t. Ainsi, la réponse de l'inflation est obtenue en cumulant la réponse du premier élément de Z_t. De même, celle de la consommation s'obtient en additionnant la réponse de la deuxième et de la troisième variable de Z_t.

La figure 4 reporte les réponses dynamiques des neuf variables incluses dans Z_t à un choc permanent de désinflation. Elles sont toutes normalisées de sorte que l'inflation trimestrielle est réduite de 1% à long terme. Nous considérons dans un premier temps les réponses des variables nominales. L'inflation et le taux de croissance de la monnaie présentent des profils très similaires. Toutes les deux commencent par diminuer fortement, puis fluctuent avant de converger lentement vers leur nouvelle valeur de long terme. Le fait que l'inflation et le taux de croissance de la monnaie présentent dans le très court terme des réponses aussi proches conforte l'idée que les chocs de désinflation peuvent être interprétés comme des chocs permanents sur la politique monétaire. Dans le même temps, le taux d'intérêt nominal de court terme ne répond pratiquement pas à l'impact, puis décroît régulièrement et sur–ajuste légèrement sa nouvelle valeur de long terme. De manière contrastée, le taux nominal de long terme baisse significativement à l'impact, ce qui implique que l'écart de taux diminue significativement à court terme. Ce résultat suggère que les agents ont en partie intégré la baisse de l'inflation engendrée par la politique de désinflation.

Concernant les variables réelles, nous obtenons les résultats suivants. Tout d'abord, le choc de désinflation a un effet négatif, persistant et significatif sur le produit par tête. Plus précisément, une baisse de un pour cent de l'inflation à long terme induit une baisse de plus de deux pour cent du produit après deux ans. Au bout de cinq ans, la réponse du produit s'annule. Le ratio de sacrifice associé est important. Ce dernier est défini de façon usuelle dans la littérature comme la réponse cumulée du produit divisée par la baisse de l'inflation annuelle (ici 4%). Après cinq ans, le ratio de sacrifice estimé est égal à 4.9% du produit cumulée avec un écart–type de 2.29. Le ratio de sacrifice augmente régulièrement avec l'horizon de la réponse pour se stabiliser approximativement douze ans après le choc de désinflation. Il faut noter que le produit et l'inflation varient dans le même sens, ce qui permet d'interpréter le choc de désinflation comme un choc de demande. La consommation baisse suite au choc, mais de façon un peu moins marquée que le produit, reflétant ainsi une phénomène de lissage. De façon contrastée, l'investissement présente une baisse deux fois plus forte que celle du produit.

9. Notre spécification de Z_t implique que la consommation, l'investissement et le produit partagent la même composante permanente.

Figure 4 – Réponses dynamiques à un choc de désinflation

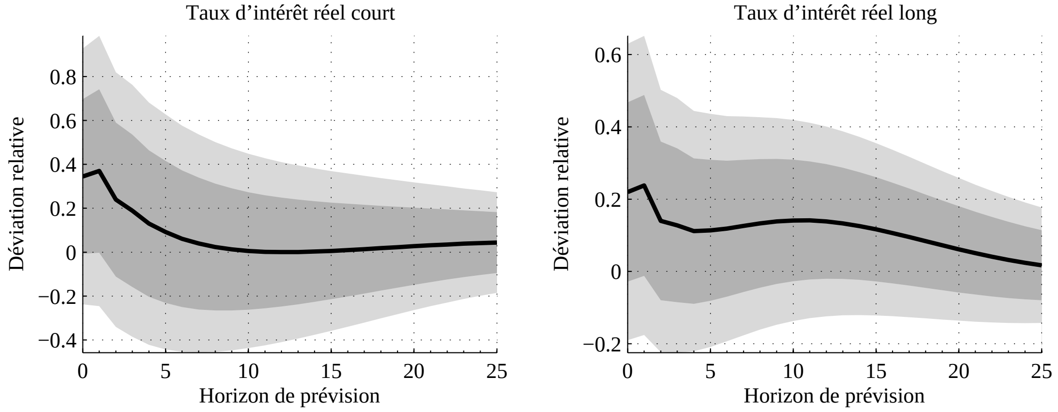


Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%.

Afin de mieux comprendre cette baisse de l'investissement, il est instructif de regarder les réponses des taux d'intérêt réels ex–ante (à court et à long termes). Celles-ci sont reportées dans la figure 5. Pour les deux cas, les taux ex–ante sont obtenus comme la différence entre le taux nominal et l'inflation anticipée à un pas, cette dernière étant calculée à l'aide de la réponse avancée de l'inflation dans le modèle VARS. Les deux taux réels répondent positivement, leur réponse à l'impact étant approximativement égale à la moitié (en valeur absolue) de la réponse de long terme de l'inflation. De plus, ces réponses positives apparaissent persistantes. Ceci suggère que les politiques de désinflation ont augmenté de façon persistante le coût du capital, entraînant ainsi une baisse importante de l'investissement. Il convient toutefois de remarquer que ces réponses ne sont pas précisément estimées.

Le taux de chômage ne réagit pas à l'impact mais présente dans les périodes suivantes un profil en cloche. Cette réponse du taux de chômage suggère un coût réel important de la politique de désinflation, puisque la composante cyclique du chômage s'accroît d'un point après deux ans. Cette dynamique suggère l'existence d'un arbitrage inflation–chômage per-

Figure 5 – Réponses dynamiques des taux d'intérêt réels à un choc de désinflation



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%.

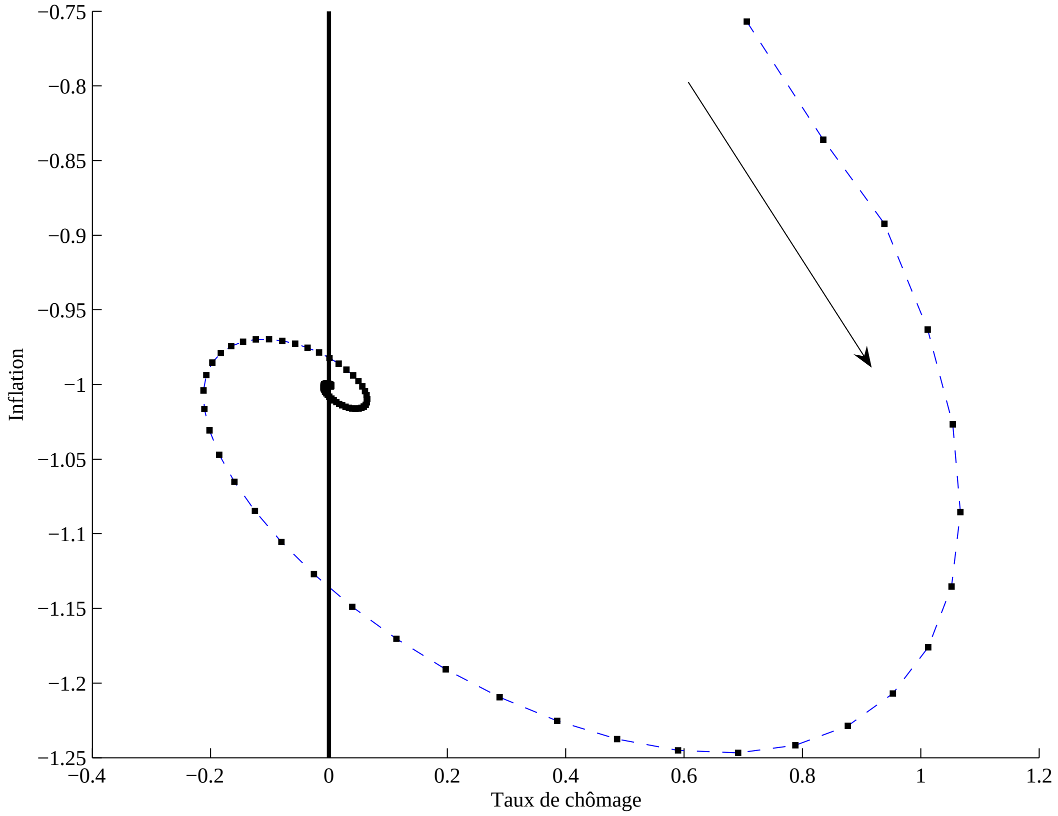
sistant. Ce point est mis en évidence dans la figure 6 qui reporte la dynamique d'ajustement impliquée par le modèle VARS autour de la courbe de Phillips de long terme.

Dans un premier temps, l'inflation baisse tandis que le chômage augmente. Au cours de cet épisode, en effet, les salaires réels augmentent instantanément (voir figure 4). Après quelques périodes, le taux de chômage revient à sa valeur de long terme tandis que l'inflation rejoint sa nouvelle cible. Le retour du chômage à son niveau initial est permis par la baisse du salaire réel qui joue ici le rôle de la force de rappel. Ce graphique illustre bien que même s'il y a neutralité à long terme de la politique de désinflation, la dynamique d'ajustement du chômage est extrêmement lente.

Comme l'a suggéré Blanchard [2003], les effets négatifs de la politique de désinflation sur l'emploi peuvent provenir de deux sources. Tout d'abord, la désinflation peut provoquer une hausse du salaire réel, lorsque les salaires nominaux présentent une forte inertie à court terme. Ensuite, en augmentant le coût du capital (le taux d'intérêt réel de long terme), une telle politique induit une baisse persistante de l'accumulation de capital physique. Dans la mesure où le travail et le capital sont des facteurs de production complémentaires, la baisse de l'investissement se traduit par une hausse du chômage. Ces deux effets sont présents dans le modèle VARS, puisque à la fois le salaire réel (voir la figure 4) et le taux d'intérêt réel (voir la figure 5) augmentent à court terme. Toutefois, seule la réponse du taux d'intérêt réel est persistante.

Nous complétons ces résultats en calculant la contribution du choc de désinflation à la variance des neuf variables incluses dans Z_t. Le tableau 3 reporte cette décomposition de variance pour différents horizons.

FIGURE 6 – Dynamique d'ajustement autour de la courbe de Phillips de long terme



Notes : les réponses correspondant à la première année consécutive au choc de désinflation sont omises du graphique afin d'en faciliter la lecture. La flèche indique le sens d'enroulement de la courbe de Phillips de court terme.

Concernant les variables nominales, le choc de désinflation est le choc dominant de leur fluctuations et ce, même dans le court terme (avec une exception pour le taux nominal de court terme). Par exemple, ce choc représente 69% de l'inflation, 67% du taux d'intérêt nominal de court terme, 86% du taux d'intérêt nominal de long terme et 41% du taux de croissance de la monnaie après trois ans. Au bout de cinq ans, les chocs de désinflation représentent entre 50% et 90% des fluctuations de ces variables nominales et par construction cette contribution augmente avec l'horizon. Bien que notre schéma d'identification impose la neutralité à long terme des chocs sur la politique de désinflation, ceux–ci ont un effet non négligeable sur les variables réelles. Ils représentent 28% de la variance de la composante cyclique du produit et 38% de celle de l'investissement après trois ans. Cette contribution apparaît de même ampleur pour la consommation et le taux de chômage (entre 25% et 30%) pour le même horizon. Pour des horizons plus longs, par exemple après dix ans, cette contribution se situe autour de 25% pour les variables réelles. Ce résultat reflète la réponse graduelle du secteur productif aux chocs sur la politique monétaire.

Tableau 3. Décomposition de la variance

	Horizon de prévision				
	0	12	20	40	∞
π_t	36.37	68.76	78.68	85.69	100.00
	[1.16, 59.73]	[23.06, 85.17]	[41.56, 90.16]	[63.43, 94.73]	[–]
$\hat{y}_t$	1.82	28.34	27.14	25.72	25.71
	[0.03, 23.30]	[1.85, 58.59]	[2.35, 58.80]	[2.65, 58.29]	[2.64, 58.28]
$\hat{c}_t$	13.47	26.38	25.59	25.20	25.19
	[0.14, 41.30]	[2.59, 59.97]	[3.20, 59.96]	[3.69, 58.62]	[3.65, 59.31]
$\hat{x}_t$	9.22	38.17	35.06	32.28	32.23
	[0.05, 30.16]	[3.44, 61.63]	[3.95, 61.00]	[4.30, 60.16]	[4.31, 60.17]
i_t^l	29.81	86.48	91.55	95.39	100.00
	[4.92, 55.65]	[55.56, 91.43]	[68.41, 94.74]	[83.52, 97.35]	[–]
i_t^s	8.95	66.99	76.00	84.60	100.00
	[0.25, 46.23]	[35.49, 87.22]	[50.41, 91.90]	[69.84, 95.57]	[–]
$\hat{u}_t$	6.47	30.01	28.76	26.82	26.79
	[0.03, 26.07]	[1.57, 62.07]	[2.02, 62.10]	[2.30, 61.03]	[2.33, 60.90]
γ_t	8.71	40.76	51.56	63.31	100.00
	[0.07, 29.84]	[8.96, 63.69]	[17.94, 73.58]	[37.09, 84.38]	[–]
$\hat{w}_t$	20.26	21.30	25.05	24.09	24.05
	[0.22, 39.44]	[2.25, 38.29]	[2.70, 45.37]	[3.08, 45.60]	[3.17, 45.87]

Notes : les chiffres entre crochets sont les intervalles de confiance obtenus à l'aide d'une technique de bootstrap. Un — indique que l'intervalle de confiance est dégénéré.

2.3 Désinflation et mesures d'inefficience

Les réponses obtenues à partir du modèle (2) peuvent être utilisées afin de mener un exercice de comptabilité du cycle et de fournir des mesures d'inefficience dans la zone euro. Les éléments permettant de telles mesures sont présents dans le modèle (2) puisque le vecteur Z_t inclut différentes variables relatives à ces mesures d'inefficience. Malheureusement, le modèle VAR ne permet pas de dériver formellement et directement de telles mesures. Pour ce faire, nous devons coupler certaines restrictions issues de la théorie avec les réponses des différentes variables contenues dans Z_t. A cette fin, nous suivons Galí *et alii.* [2003, 2007] et nous définissons la mesure d'inefficience (gap_t) comme

$$gap_t = mrs_t - mpn_t$$

où mrs_t et mpn_t sont respectivement le logarithme du taux marginal de substitution entre consommation et loisir et le logarithme de la productivité marginale du travail. Comme l'ont montré Galí *et alii.* [2007], la mesure gap_t permet de quantifier simplement les coins sur les marchés des biens et du travail. En effet, la mesure d'inefficience peut être définie comme la somme de deux marges,

$$gap_t = -(\mu_t^p + \mu_t^w)$$

où μ_t^p et μ_t^w représentent respectivement les logarithmes de la marge sur les prix et de la marge sur les salaires. La marge moyenne sur les prix dans l'économie est donnée par la différence entre la productivité marginale du travail et le salaire réel

$$\mu_t^p = mpn_t - w_t,$$

tandis que la marge moyenne sur les salaires est donnée par la différence entre le salaire réel et le taux marginal de substitution entre loisir et consommation

$$\mu_t^w = w_t - mrs_t,$$

où w_t est le salaire réel. En faisant les hypothèses que les entreprises sont preneuses de prix sur le marché du travail et en l'absence de coût d'ajustement sur le facteur travail, la marge sur le prix μ_t^p est égale à la différence entre le prix des biens et le coût marginal nominal du travail. De façon symétrique, la marge sur les salaires est définie comme la différence entre le salaire réel et la désutilité du travail, les deux étant exprimées en unité de consommation. La marge sur les salaires peut inclure de nombreux coins, tels que les taxes sur le revenus du travail ou bien des salaires d'efficience (voir Chari *et alii.*, 2007). Cette décomposition permet d'une part d'évaluer comment la mesure globale d'inefficience a évolué suite à un choc de désinflation, mais aussi de mesurer le rôle respectif des marges sur les marchés des biens et du travail.

A ce stade, il est nécessaire de faire des hypothèses plus spécifiques sur les préférences et la technologie. Concernant les préférences, nous supposons que la fonction d'utilité consommation–loisir est de la forme fonctionnelle suivante

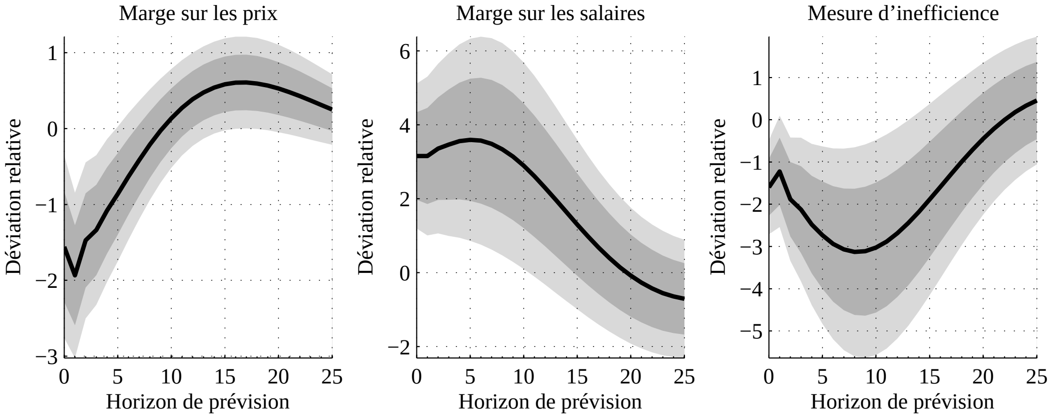
$$u(C_t, N_t) = \log(C_t) - \Xi_t \frac{N_t^{1+\phi}}{1+\phi}$$

de sorte que le taux marginal de substitution est donné par

$$mrs_t = c_t + \phi n_t + \bar{\xi}_t$$

où c_t représente le logarithme de la consommation par tête et n_t le logarithme des heures par tête. Cette dernière est définie comme le produit des heures moyennes travaillées d'un

FIGURE 7 – Marges sur les prix et salaires et mesure d'inefficience



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%. Les calculs sont effectués sous l'hypothèse $\phi = 2$.

individu et du nombre de personnes employées, le tout rapporté à la population en âge de travailler. Le paramètre $\phi \geq 0$ est l'inverse de l'élasticité Frishienne de l'offre de travail. La variable $\bar{\xi}_t \equiv \log(\bar{\Xi}_t)$ représente les mouvements longs qui affectent l'offre de travail. Ces mouvements peuvent représenter des changements institutionnels ou démographiques qui sont supposés non pertinents à la fréquence du cycle.

A ce niveau, la spécification du taux marginal de substitution mrs_t pose deux problèmes de mesure. Le premier point concerne la valeur retenue pour le paramètre ϕ. Dans notre expérience de référence, nous fixons $\phi = 2$, une valeur proche de celle estimée par SMETS et WOUTERS [2003] pour la zone euro[10]. Le second point concerne les propriétés dynamiques de $\bar{\xi}_t$. Nous supposons ici que ces chocs sur l'offre de travail sont non corrélés avec les chocs de désinflation.

Concernant la technologie, nous faisons l'hypothèse d'une fonction de production avec une élasticité constante par rapport au facteur travail. Il s'ensuit que, à un terme constant près, la productivité marginale du travail est donnée par

$$mpn_t = y_t - n_t,$$

où y_t représente le logarithme du produit par tête.

A partir des définitions de mpn_t et de mrs_t et à l'aide des hypothèses ci-dessus, nous pouvons définir les composantes cycliques des mesures d'inefficience μ_t^p, μ_t^w et gap_t, notées $\hat{\mu}_t^p$, $\hat{\mu}_t^w$ et $\widehat{gap}_t$, respectivement. Ces composantes cycliques sont obtenues en substituant $\hat{w}_t$, $\hat{y}_t$, $\hat{c}_t$ et $\hat{n}_t$

10. Nous étudions la sensibilité de nos résultats à la valeur de ϕ.

Tableau 4. Décomposition de la variances des mesures d'inefficience

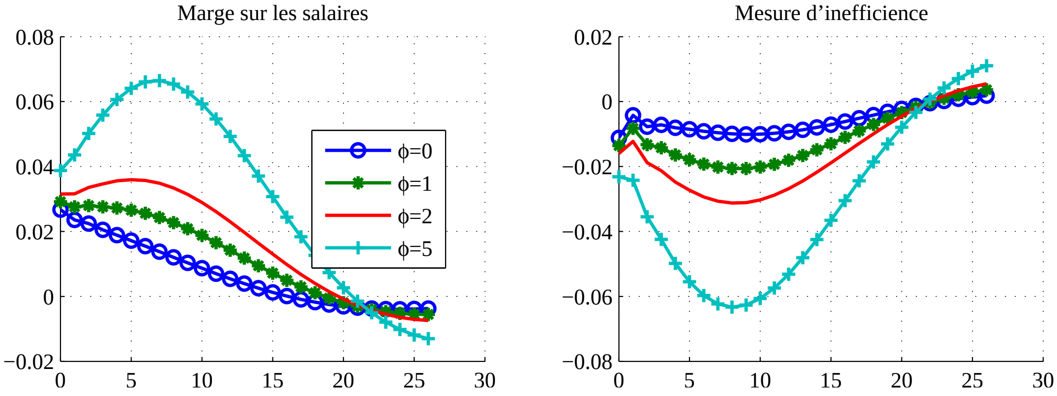
	Horizon de prévision				
	0	12	20	40	∞
$\hat{\mu}_t^p$	20.68	25.38	25.81	25.64	25.64
	$[0.25, 43.69]$	$[2.74, 48.81]$	$[2.95, 48.75]$	$[2.99, 49.39]$	$[2.99, 49.49]$
$\hat{\mu}_t^w$	32.00	30.68	28.95	28.14	28.14
	$[0.80, 55.07]$	$[2.55, 62.96]$	$[2.77, 62.52]$	$[3.38, 61.64]$	$[3.45, 61.67]$
$\widehat{gap}_t$	16.79	27.95	27.21	26.21	26.19
	$[0.46, 34.26]$	$[2.20, 59.96]$	$[3.01, 61.36]$	$[3.32, 59.91]$	$[3.33, 60.25]$

Notes : les chiffres entre crochets sont les intervalles de confiance obtenus à l'aide d'une technique de bootstrap. Les calculs sont effectués sous l'hypothèse $\phi = 2$.

au salaire réel, au produit, à la consommation et au heures travaillées dans les expressions précédentes. Comme le filtre à bande passante est linéaire, la composante cyclique de la somme est la somme des composantes cycliques. Cependant, le vecteur Z_t dans le modèle (2) comprend la composante cyclique du taux de chômage, et non pas celle des heures travaillées par tête. Toutefois, en supposant qu'à court terme la participation sur le marché du travail et les heures de travail par individu ne changent pas trop après le choc de désinflation, la composante cyclique du taux de chômage satisfait $\hat{u}_t \simeq -\hat{n}_t$, c'est à dire que sa déviation à la référence de long terme est approximativement égale (au signe moins près) à celle des heures travaillées. A l'aide du modèle VARS, nous sommes alors en mesure d'identifier les réponses de $\hat{\mu}_t^p$, $\hat{\mu}_t^w$ et $\widehat{gap}_t$ au choc de désinflation que nous obtenons à partir des réponses des composantes cycliques du produit, de la consommation, du chômage et du salaire réel. Ces réponses sont reportées dans la figure 7. Il convient de remarquer que, étant données nos définitions de l'inefficience sur les marchés des biens et du travail, une réponse négative de $\widehat{gap}_t$ implique un accroissement global de l'inefficience (c'est à dire sur les deux marchés simultanément). De façon symétrique, une hausse de $\hat{\mu}_t^p$ (respectivement de $\hat{\mu}_t^w$) implique une baisse d'efficacité sur le marché des biens (respectivement sur le marché du travail).

La réponse négative et persistante de $\hat{\mu}_t^p$ indique que l'inefficience sur le marché des biens a diminué suite à la politique de désinflation. En d'autres termes, il semblerait que les marges des entreprises sur le prix des biens aient été sensiblement réduites. La marge sur les prix suit de façon très proche l'évolution du produit, révélant ainsi une forte pro–cyclicité (du moins suite à ce choc) dans la zone Euro. Concernant la marge sur les salaires, les résultats sont forts différents. Après un choc de désinflation, cette marge augmente instantanément et présente un profil en cloche pour les périodes suivantes. Ces mouvements contra–cycliques de la marge sur les salaires en presence de chocs de désinflation sont en accord avec ceux

FIGURE 8 – Sensibilité à ϕ



Notes : la figure ne reporte pas la réponse de la marge $\hat{\mu}_t^p$ sur le prix puisque celui–ci ne dépend pas de l'élasticité de l'offre de travail. La figure inclut la marge $\hat{\mu}_t^w$ sur le salaire ainsi que la mesure d'inefficience $\widehat{gap}_t$ pour les valeurs $\phi = \{0, 1, 2, 5\}$. Afin de simplifier la comparaison, nous ne reportons pas les intervalles de confiance.

obtenus sur des données de la zone Euro par GALÍ *et alii.* [2003]. De plus la réponse positive de la marge sur les salaires excède (en valeur absolue) celle de la marge sur les prix. Il en découle que la mesure d'inefficience $\widehat{gap}_t$ augmente de façon persistante après le choc de désinflation, suggérant ainsi que les distortions sur le marché du travail dans la zone Euro ont progressé suite à cette politique.

Nous calculons également la contribution du choc de désinflation à la variance des marges et de la mesure globale d'inefficience. Le tableau 4 reporte ces décompositions à différents horizons. Le choc de désinflation a un effet important sur ces mesures, ce qui est cohérent avec les effets réels de ce choc (voir le tableau 3). Il représente 25% de la variance de la marge sur les prix après trois ans et près de 31% de la marge sur les salaires. Dans le cas de la mesure globale, cette contribution est du même ordre quoique un peu plus faible (à peu près 28%), puisque la réponse de la marge sur les prix vient pour partie contrebalancer celle de la marge sur les salaires. Pour des horizons plus importants, par exemple après dix ans, cette contribution est de l'ordre de 25%. Bien qu'il ne soit pas une source dominante de fluctuations dans les mesures d'inefficience, le choc de désinflation représente une fraction non négligeable de leurs mouvements cycliques.

Nous évaluons la robustesse de ces résultats à différents choix du paramètres ϕ. Les résultats avec différentes valeurs de ϕ comprises entre 0 et 5 sont reportés dans la figure 8. Par construction, la marge sur les prix présente la même réponse que dans la figure 7. La réponse de la marge sur les salaires présente le même signe mais l'ampleur de la réponse diffère selon les valeurs de ϕ. Quand ϕ prend des valeurs importantes, l'emploi répond peu de sorte que la marge sur les salaires augmente plus fortement. On obtient alors que la mesure d'inefficience croît plus fortement suite à un choc de désinflation.

3 Robustesse des résultats

Nous évaluons maintenant la robustesse de ces résultats à différentes modifications. Nous étudions successivement le rôle des variables étrangères (notamment le prix du pétrole et le taux de change réel), la définition des prix, les restrictions de long terme, la spécification du produit et du chômage et les origines non–monétaires des chocs de désinflation [11].

3.1 Prise en compte de variables étrangères

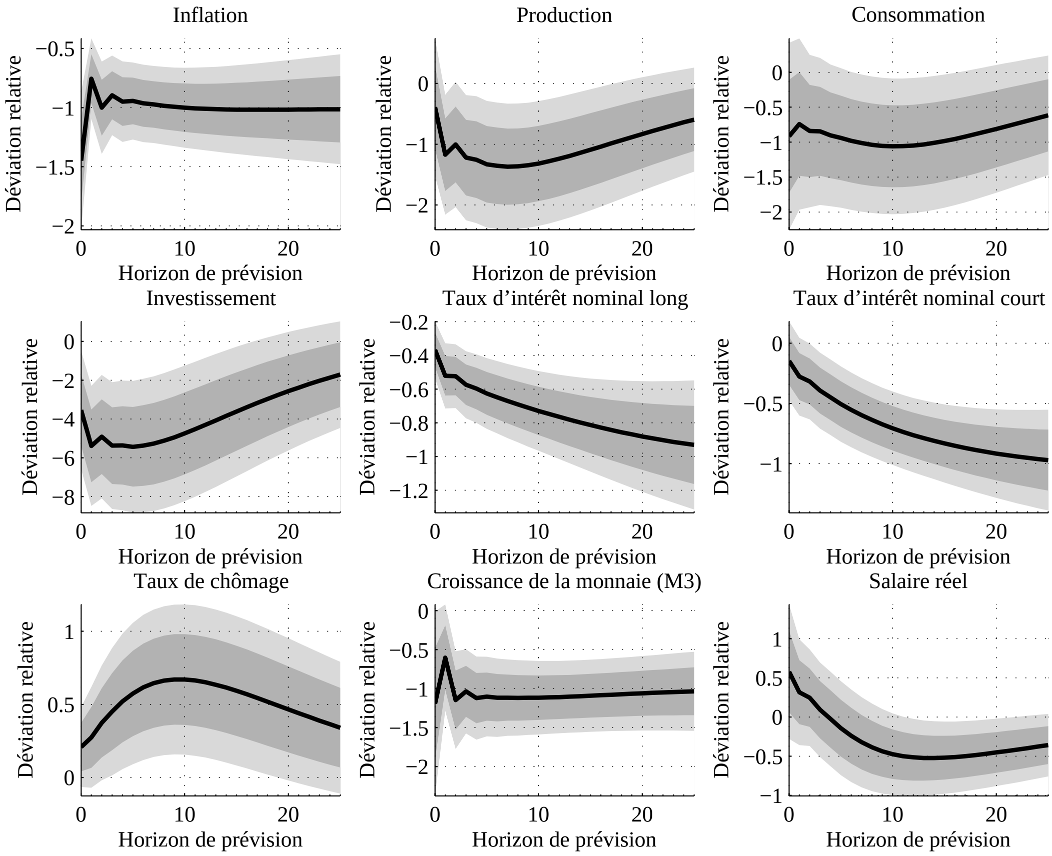
Dans cet article, nous avons implicitement traité la zone euro comme une économie fermée, au sens où aucune variable extérieure à la zone n'a été prise en compte dans notre spécification de référence. Cette stratégie peut constituer un point de départ intéressant, étant donné la faible ouverture extérieure de la zone. Toutefois, il est important d'examiner la robustesse de nos résultats à la prise en compte de ces variables extérieures. Nous suivons Peersman et Smets [2002] et incluons un vecteur de variables exogènes X_t dans le modèle VAR. Comme ces auteurs, nous incluons dans X_t le PIB américain en écart à une tendance linéaire qui nous sert d'approximation pour la demande mondiale, le taux de croissance des prix du pétrole et le taux de change réel \$–euro [12]. Le prix du pétrole est ici calculé comme le produit du West Texas Crude Oil Price et du taux de change nominal Euro–\$. En pratique, nous incluons la valeur contemporaine et un retard pour X_t. Les résultats sont reportés sur le graphique 9.

Nous obtenons que la dynamique des variables nominales n'est pas affectée par la prise en compte de X_t. En revanche, nous trouvons que les effets sur les variables réelles sont près de deux fois moins amples mais beaucoup plus persistants. Au total, la valeur du ratio de sacrifice, qui se monte à 4.95 à cinq ans dans l'intervalle $[0.3636, 8.2315]$, est pratiquement inchangée par rapport à notre spécification de référence. De ce point de vue, il est intéressant de remarquer que les effets dynamiques du choc de désinflation sur l'inflation ne doivent pas

11. Nous avons également évalué la sensibilité des résultats à une modification de la période d'estimation, par exemple en terminant l'échantillon au milieu des années quatre–vingt–dix. Les résultats obtenus sur cet échantillon plus court sont les mêmes pour les réponses, mais la précision des estimations s'avère fortement réduite. Ce résultat est assez standard dans la littérature VAR Structurel (voir Sims, 1998) qui montre que de tels modèles dynamiques peuvent produire des réponses relativement erratiques sur de petits échantillons. Nous avons aussi estimé un modèle de plus petite taille, comprenant la composante cyclique du produit, l'inflation, le taux de croissance de la monnaie et le taux de change réel. Les résultats sont similaires pour les variables nominales. En revanche, la réponse du produit est à la fois plus ample et moins persistante.

12. L'hypothèse de stationarité du PIB américain autour d'une tendance déterministe n'est pas rejetée par les tests de racine unitaire ADF et PP. Nos résultats sont préservés si nous substituons la composante cyclique du PIB américain obtenue par application du filtre de Christiano et Fitzgerald [2003].

Figure 9 – Robustesse à la prise en compte de variables étrangères exogènes



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à −1%.

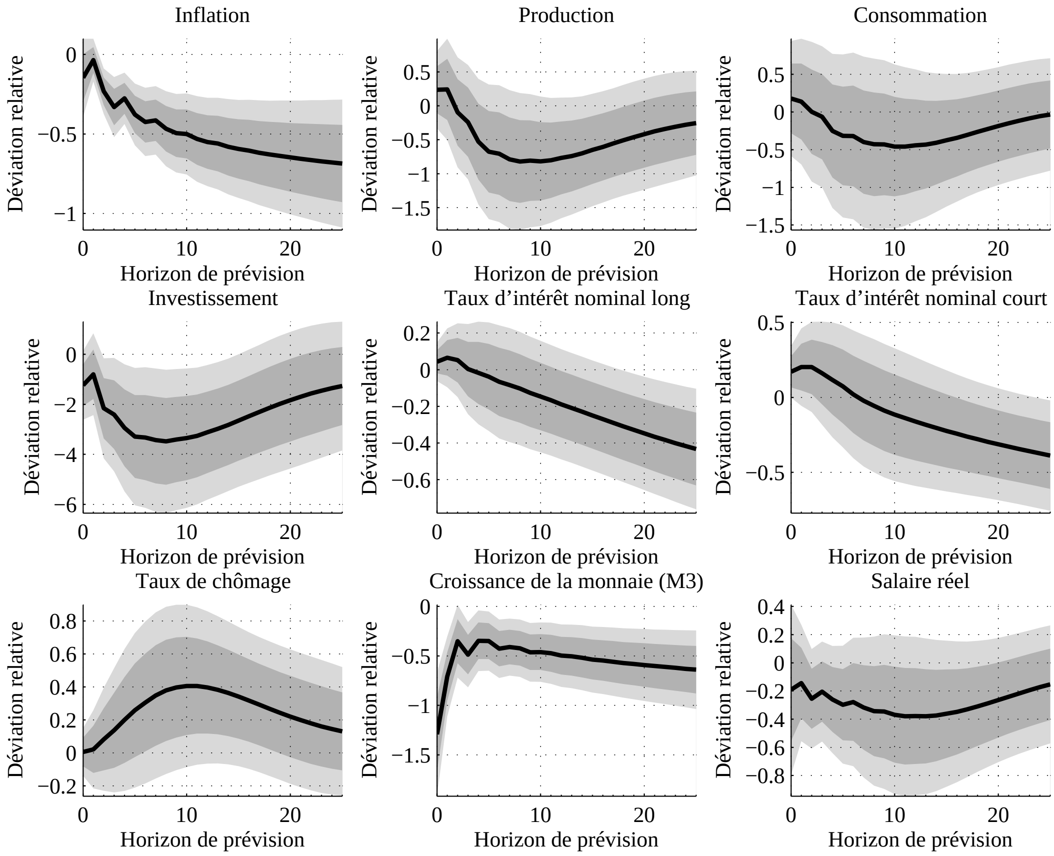
être confondus avec les effets du contre–choc pétrolier de la deuxième moitié des années 1980 [13].

3.2 Spécification des prix

Dans notre spécification de référence, le taux d'inflation est obtenu à partir de l'indice de prix du PIB, comme dans la grande majorité des modèles VARS monétaires. Cependant, nous

13. Nous avons aussi inspecté les liens de causalité entre le choc de désinflation estimé dans notre spécification de référence du modèle (2) et le taux de croissance du prix du pétrole. Le test de causalité au sens de Granger est mené en régressant le choc de désinflation estimé sur une constante et les valeurs retardées du taux de croissance du prix de pétrole (de un à quatre retards). La statistique de Fisher donne un seuil du test de 73.75% tandis que celui de la statistique de Wald est égal à 71.69%. Ces tests conduisent donc à rejeter cette causalité. Ces résultats confirment que le choc de désinflation identifié dans la spécification de référence n'est pas confondu avec les effets de l'importante baisse du prix du pétrole du milieu des années quatre–vingts.

FIGURE 10 – Robustesse à la définition de l'inflation



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%.

cherchons à évaluer si les réponses estimées sont sensibles à d'autres définitions de l'inflation dans la zone Euro. A cette fin, nous considérons l'indice de prix à la consommation (IPCH). En effet, cet indice de prix est connu comme étant celui suivi par la BCE (voir BCE, 2004) et comme le prix ayant servi de référence à la participation à la zone. Il apparaît donc légitime de l'utiliser afin d'identifier les effets d'un choc de désinflation.

Comme le montre la figure 10, cette nouvelle définition de l'inflation ne modifie que très faiblement nos précédentes conclusions. Néanmoins, quelques résultats intéressants apparaissent avec cette nouvelle définition. Tout d'abord, l'inflation présente un profil plus lisse d'ajustement vers sa nouvelle valeur de long terme. Ensuite, les effets de la politique de désinflation sont plus persistants. Par exemple, le produit atteint sa baisse maximale après dix trimestres au lieu de 8 dans notre spécification de référence. De même, la hausse maximale du taux de chômage apparaît après 10 trimestres au lieu de 8. Finalement, cette propriété empirique est partagée par presque toutes les variables. Enfin, dans la plupart des cas, nous obtenons des intervalles de confiance plus larges. Par exemple, la réponse du chômage n'est pas précisément estimée.

3.3 Restrictions sur la relation de long terme entre variables nominales

Nous évaluons maintenant les effets des restrictions cointégrantes entre les variables nominales. Contrairement à notre spécification de référence, nous n'imposons plus la restriction que les variables nominales incluses dans le vecteur Z_t partagent la même tendance stochastique à long terme. Ainsi, le vecteur Z_t est maintenant spécifié de la manière suivante

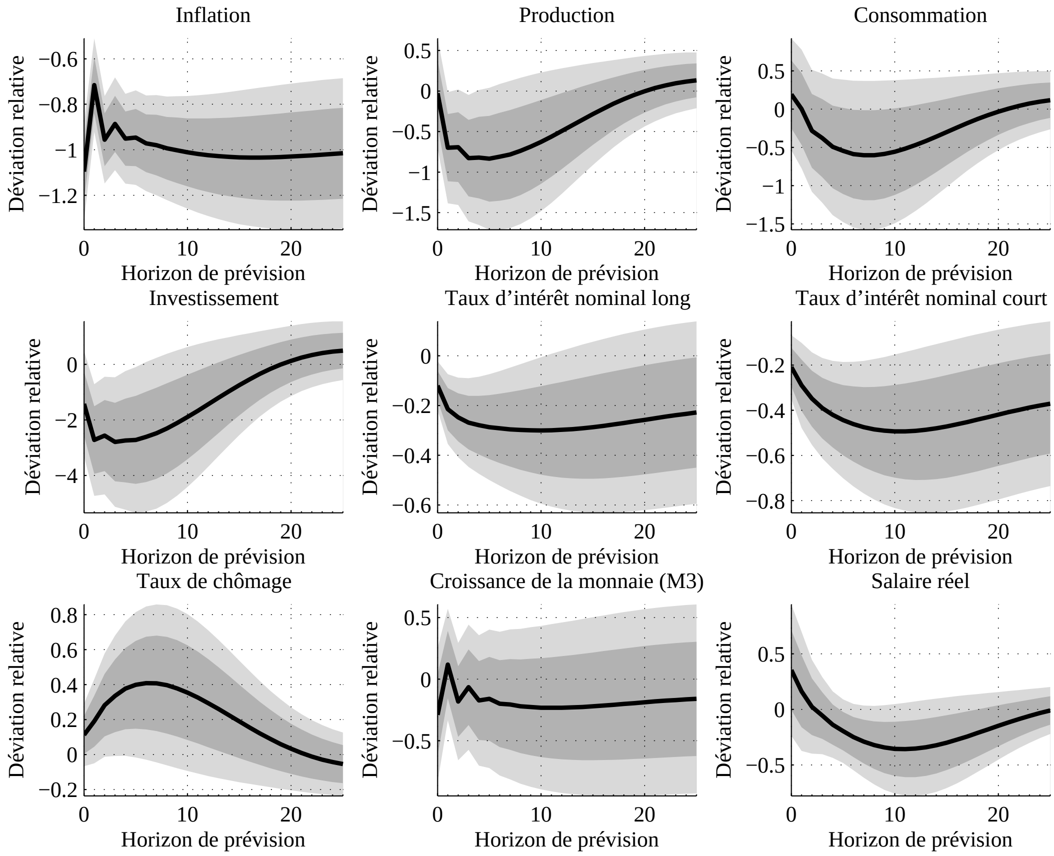
$$Z_t = (\Delta\pi_t, \hat{y}_t, c_t - y_t, x_t - y_t, \hat{u}_t, \hat{w}_t, \Delta i_t^s, \Delta i_t^l \Delta\gamma_t)'$$

Nous maintenons la même spécification des composantes cycliques du produit, du chômage et du salaire réel. De plus, nous utilisons le même schéma d'identification du choc de désinflation. La figure 11 reporte les réponses dynamiques des différentes variables incluses dans Z_t. La comparaison avec la figure 4 montre que les restrictions cointégrantes sur les variables nominales n'ont que peu d'effet sur les réponses des variables réelles et nominales : *(i)* l'inflation, les taux d'intérêt de court et de long termes et le taux de croissance de la monnaie baissent suite au choc de désinflation, *(ii)* le produit, la consommation et l'investissement diminuent et *(iii)* le salaire réel et le chômage augmentent. La différence notable ne concerne pas le profil des réponses, mais leurs intervalles de confiance. Lorsque l'on n'impose pas ces relations de cointégration, la précision des réponses estimées diminue fortement. Cette absence de précision concerne non seulement les variables nominales mais aussi l'ensemble des variables réelles. Dans ce cas, le modèle VARS est tellement imprécis qu'il n'est plus possible d'obtenir des résultats clairs sur les effets agrégés des chocs de désinflation, à l'exception de la réponse de l'inflation.

3.4 Spécification du chômage et du produit

Dans notre spécification de référence, les mouvements cycliques du produit et du chômage sont obtenus par application du filtre de Christiano et Fitzgerald [2003]. Etant donné le schéma d'identification, les chocs de désinflation ne peuvent donc pas avoir d'effets permanents sur le produit (et donc sur la consommation et l'investissement puisqu'ils apparaissent sous forme de ratios) et le taux de chômage. Suivant la littérature sur l'hystérèse, il est légitime d'évaluer si ces chocs peuvent avoir des effets permanents sur les variables réelles (voir Blanchard et Summers, 1986 et Buiter et Miller, 1985). A cette fin, le produit

FIGURE 11 – Robustesse aux restrictions de cointégration



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à −1%.

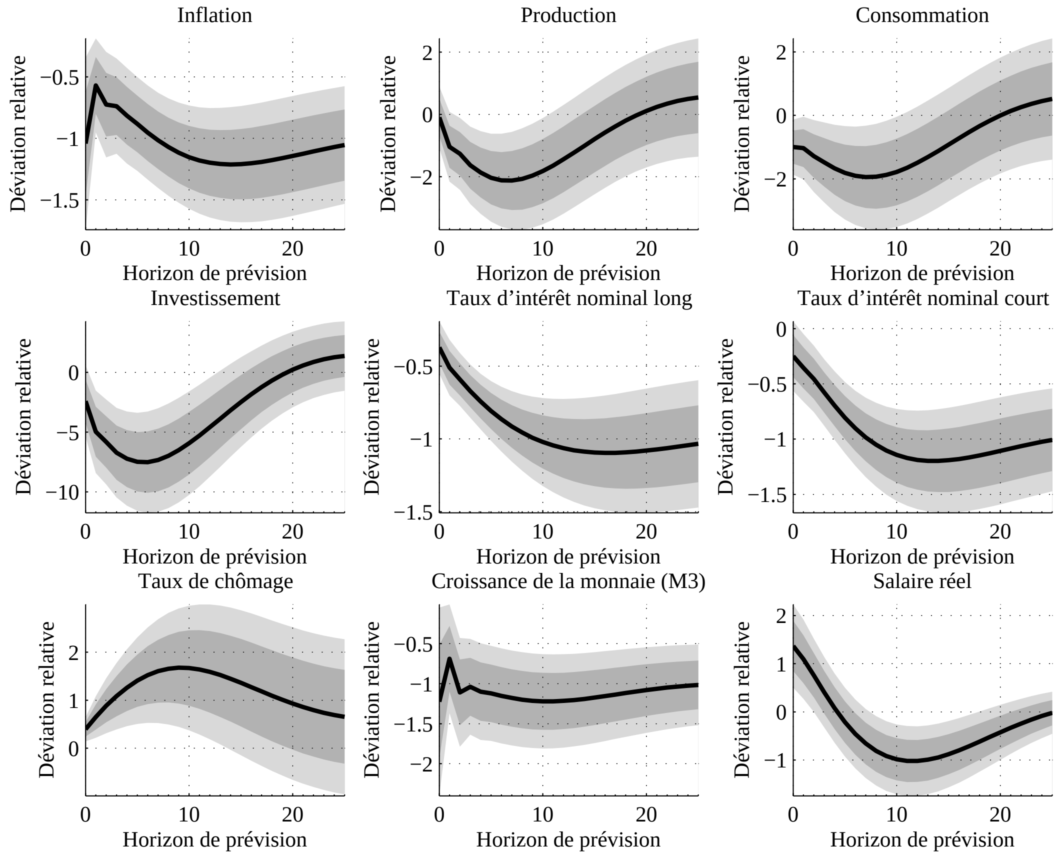
et le chômage sont maintenant spécifiés en différence dans le vecteur Z_t

$$Z_t = (\Delta\pi_t, \Delta y_t, c_t - y_t, x_t - y_t, \Delta u_t, \hat{w}_t, i_t^l - \pi_t, i_t^l - i_t^s, \gamma_t - \pi_t)',$$

les autres variables étant spécifiées comme dans (1). Avec cette nouvelle spécification, la politique monétaire n'est plus contrainte à être neutre à long terme. Les résultats sont reportés dans le figure 12.

Comme le montre la figure, toutes les variables, excepté le chômage, présentent les mêmes fonctions de réponse que dans la spécification de référence. En revanche, les intervalles de confiance sont plus larges. L'effet à long terme du choc de désinflation sur le produit est négatif, mais faible. Plus précisément, le choc de désinflation conduit à réduire asymptotiquement le produit par tête de 0.29% avec un intervalle de confiance de $[-1.9158\%, 1.8420\%]$. Si le produit contient une tendance stochastique, un choc de désinflation ne semble donc pas en expliquer une grande part. Ce résultat conforte donc le point de vue de la neutralité à long terme de ces chocs.

Figure 12 – Robustesse à la spécification du produit et du chômage



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%.

Concernant le taux de chômage, le profil des réponses est fortement modifié. Tout d'abord, la réponse maximale est maintenant deux fois plus forte que dans le cas de référence. Ensuite, la réponse à long terme ne présente pas de retour marqué et reste positive après 20 trimestres. Ce choc conduit à une hausse de 0.81 point du taux de chômage avec un intervalle de confiance de $[-0.8346\%, 1.9145\%]$. L'effet permanent du choc de désinflation sur le taux de chômage reste ainsi très imprécis. En dépit de ces différences, il apparaît que la spécification de référence et celle autorisant un effet réel permanent délivrent un message assez proche concernant les effets récessionistes persistants des politiques de désinflation à court et à moyen termes.

3.5 Relâchement de la doctrine monétariste

Pour l'instant, nous avons imposé que seuls les chocs de désinflation peuvent avoir un effet permanent sur les variables nominales. Cette restriction est cohérente avec la doctrine monétariste (du moins à long terme) selon lequel que “l'inflation est toujours un phénomène monétaire” (FRIEDMAN, 1968). Cette vision en écarte une autre, appelée “l'approche opportuniste” de la désinflation (voir BOMFIM et RUDEBUSCH, 2000 et ORPHANIDES et WILCOX, 2002). Selon cette dernière, la banque centrale peut être tentée d'exploiter une expansion engendrée par un choc d'offre positif afin de mettre en œuvre une politique de désinflation. Si l'approche opportuniste s'avère la bonne, on devrait alors s'attendre à ce que différents chocs d'offre puissent avoir des effets à long terme sur l'inflation. Nous modifions alors notre schéma d'identification afin d'évaluer cette dernière approche.

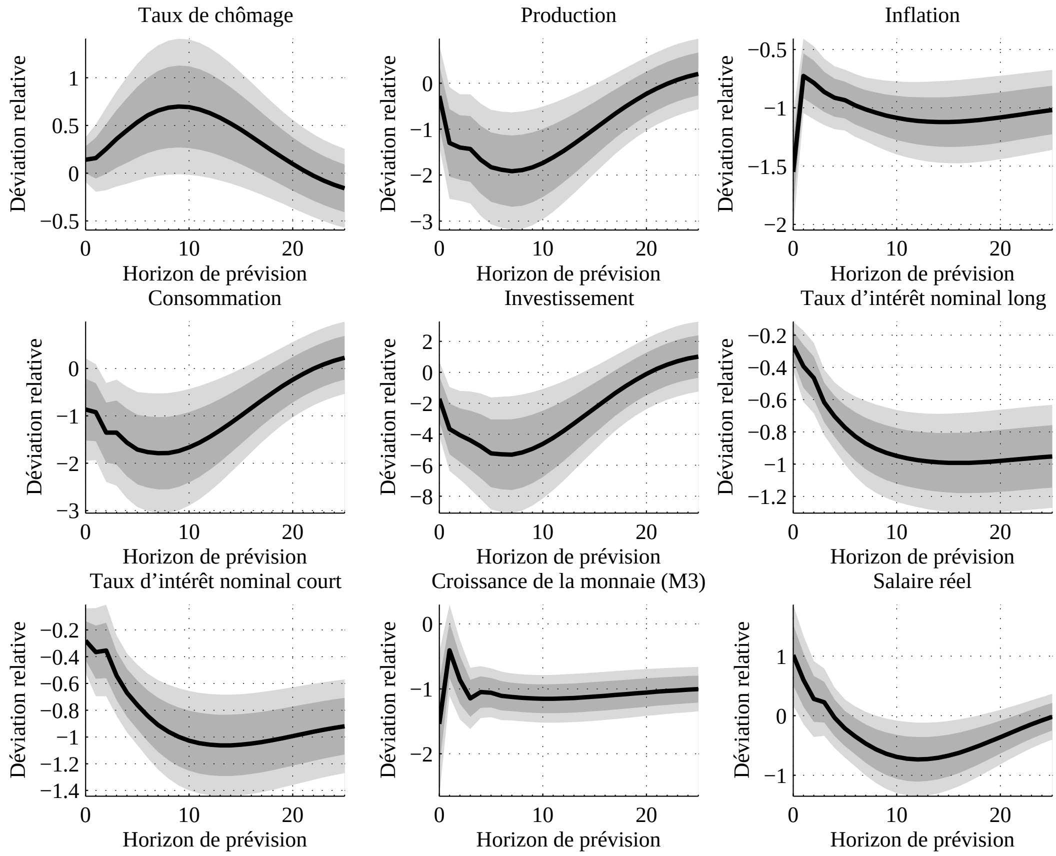
A cette fin, tout en maintenant le principe d'identification de long terme, nous modifions le vecteur Z_t de la façon suivante

$$Z_t = (\Delta u_t, \Delta y_t, \Delta \pi_t, c_t - y_t, x_t - y_t, \hat{w}_t, i^l_t - \pi_t, i^l_t - i^s_t, \gamma_t - \pi_t)'.$$

En utilisant cette définition de Z_t dans le modèle (2) et en utilisant la stratégie d'identification de BLANCHARD et QUAH [1989], nous imposons les restrictions à long terme suivantes : *i)* seul le choc de chômage, noté η^u_t, peut avoir un effet à long terme sur le taux de chômage ; il peut aussi affecter asymptotiquement la production et l'inflation, *ii)* le choc de production, noté η^y_t, peut affecter de façon permanente le produit et l'inflation et *iii)* le choc de désinflation η^π_t peut expliquer une partie de la dynamique de l'inflation à long terme, mais il n'a aucun effet à long terme sur le produit et le chômage. Etant donnée la spécification de Z_t, le choc de désinflation est contraint à nouveau à n'avoir aucun effet réel à long terme. Le choc de chômage peut s'interpréter comme un choc sur les coins du marché du travail qui modifie de façon permanente le facteur travail. On peut penser par exemple à des chocs permanents sur la taxation des revenus du travail ou bien encore un choc de préférence permanent sur l'offre de travail (voir CHANG *et alii.*, 2007, DUPAIGNE *et alii.*, 2007). Le choc de production pourrait s'interpréter, suivant BLANCHARD et QUAH [1989], comme un choc technologique permanent. Cependant, comme l'a noté GALÍ [1999], des chocs modifiant de façon permanente le facteur travail (comme notre choc de chômage) peuvent affecter le produit à long terme. Notre choc de produit est donc un possible mélange de ces deux types de chocs. Le choc de désinflation reçoit quant à lui une interprétation identique à celle de notre spécification de référence, n'était qu'il n'explique plus à lui seul le niveau de long terme de l'inflation.

Les réponses au choc de désinflation sous cette représentation alternative sont reportées dans la figure 13. Nous obtenons un profil très similaire pour toutes les variables. La seule différence concerne l'amplitude et la persistance de l'effet récesionniste de la désinflation, qui apparaissent plus faibles mais se résorbent plus lentement. Ce résultat se reflète dans le ratio de sacrifice, qui est maintenant égal à 6.07% après cinq ans. Il faut noter que celui–ci

Figure 13 – Robustesse au relâchement de la doctrine monétariste



Notes : les aires grisées correspondent à l'intervalle de confiance à 90% (clair) et à 68% (foncé), obtenus à l'aide d'une technique de Bootstrap. Afin de faciliter l'interprétation, la taille du choc de désinflation est normalisée de sorte que la réponse asymptotique de l'inflation soit égale à -1%.

n'est pas précisément estimé, faisant écho aux intervalles de confiance plus larges dans cette configuration.

Cet exercice permet également d'évaluer si les chocs "non–monétaires" contribuent à la dynamique de l'inflation à long terme. Les effets quantitatifs de ces chocs peuvent être directement obtenus à partir de la troisème ligne de la matrice de passage S. Les deux premières colonnes de cette ligne nous donnent les effets permanents sur l'inflation de η_t^u et η_t^y, tandis que la troisième permet de mesurer la contribution du choc autonome de désinflation. En notant $IRF_\infty(\pi)$ la réponse globale de l'inflation à long terme, on peut décomposer cette dernière comme suit

$$IRF_\infty(\pi) = -\underset{(0.0012)}{0.0002}\eta_t^u - \underset{(0.0004)}{0.0005}\eta_t^y + \underset{(0.0004)}{0.0040}\eta_t^\pi.$$

Dans cette dernière équation les chiffres entre parenthèses représentent les écart–types des effets à long terme des différents chocs. Afin d'interpréter plus structurellement cette équation, il faut d'abord s'assurer du statut des autres chocs. Par exemple, le choc η_t^y a un effet positif sur le produit et un effet négatif sur l'inflation. Il est donc possible de le labéliser comme un choc d'offre. Le choc η_t^u devrait quant à lui recevoir le statut de choc de demande puisqu'il fait baisser le produit et l'inflation. Cependant, en présence d'une approche opportuniste de la désinflation, la distinction choc d'offre–choc de demande devient délicate puisque, ici, la cible d'inflation et le niveau de la production covarient positivement.

Notre estimation montre que l'effet de η_t^y est négatif, ce qui signifie, dans le cas d'une politique de désinflation, que la politique monétaire a cherché à tirer partie de ces chocs pour réduire l'inflation. Cet effet est ainsi en accord avec l'approche opportuniste [14]. Cependant, ce résultat doit être utilisé avec prudence car les écart–types des coefficients suggèrent que ces mécanismes ne sont pas significatifs. Nous obtenons une conclusion similaire en ce qui concerne η_t^u. Ces résultats sont d'ailleurs confirmés par la contribution asymptotique des deux chocs η_t^u et η_t^y à la dynamique de l'inflation. Ils représentent 16.61% de la variance de l'inflation avec un intervalle de confiance à 95% égal à $[0.67\%, 80.64\%]$. Ainsi, même si nous autorisons des contributions non–monétaires de l'inflation, nous trouvons que la doctrine monétatiste est apparamment une dimension robuste de l'inflation dans la zone Euro.

Pour compléter cette analyse, nous proposons finalement de réinterpréter la cible implicite d'inflation dans le cadre de l'approche opportuniste. Pour ce faire, nous amendons l'équation (3) en y autorisant la présence des innovations η_t^u et η_t^y. Formellement, l'équation de la cible implicite s'écrit

$$\pi_t^\star = \pi_{t-1}^\star + \sigma_\pi \eta_t^\pi + \sigma_u \eta_t^u + \sigma_y \eta_t^y. \tag{4}$$

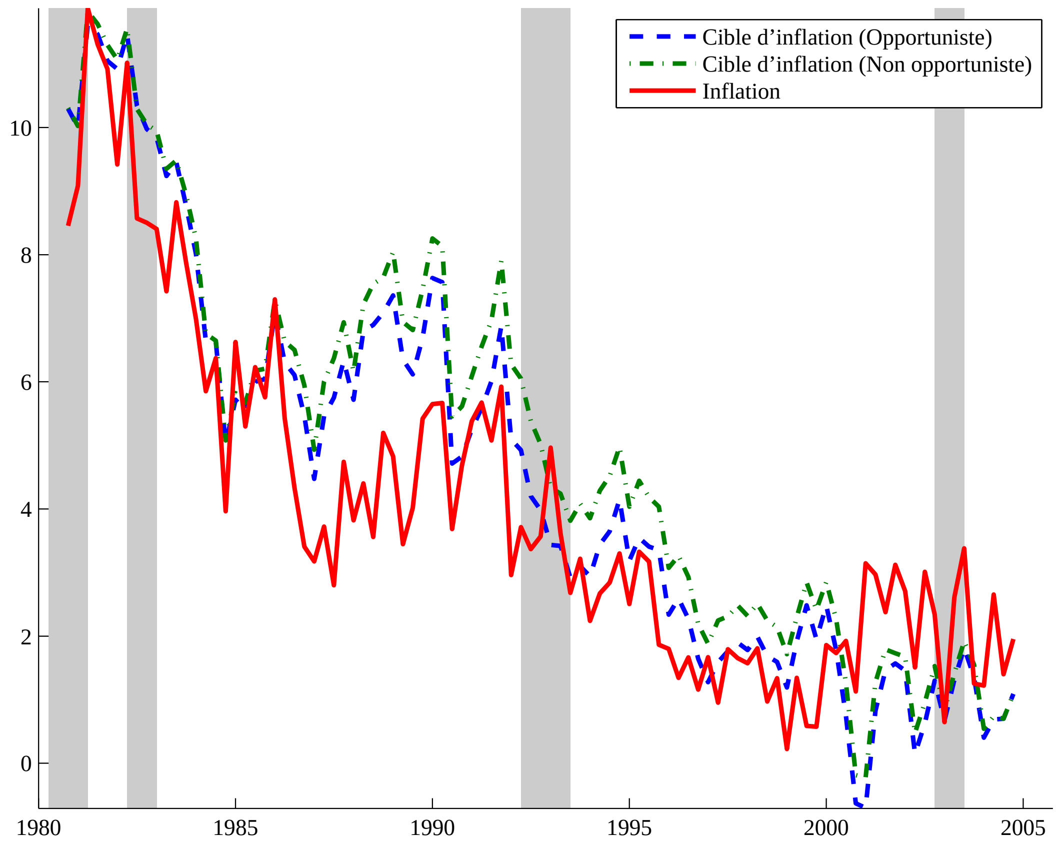
Dans cette équation, σ_u est l'élément 3×1 de la matrice de passage S, σ_y est l'élément 3×2 de S et σ_π est l'élément 3×3 de S. A des fins de comparaison, nous calculons une cible alternative sous l'hypothèse contre–factuelle que $\sigma_u = \sigma_y = 0$. Les trajectoires ainsi obtenues sont reportées dans le graphique 14. Comme le laissaient entrevoir les effets de long terme des autres chocs structurels, nous n'obtenons pas de différence marquée entre les deux séries de cible d'inflation. Même s'il existe un effet opportuniste, les mouvements de basse fréquence de l'inflation semblent essentiellement d'origine monétaire.

4 Conclusion

Dans cet article, nous utilisons la méthodologie VAR Structurel afin d'identifier les effets agrégés de chocs de désinflation dans la zone Euro. A l'aide de cette stratégie empirique, nous obtenons qu'un choc monétaire permanent de désinflation engendre des effets récessionnistes importants. Le produit par tête, la consommation et l'investissement baissent

14. Ireland [2007] obtient des résultats assez similaires pour les Etats-Unis à l'aide d'un modèle DSGE estimé par maximum de vraisemblance.

FIGURE 14 – Cible d'inflation et approche opportuniste



Notes : la cible d'inflation est identifiée à partir du modèle (4) en utilisant le schéma d'identification discuté dans le texte. La zone grisée correspond aux dates de récession identifiées à l'aide de l'algorithme de Bry–Boschan appliqué au logarithme du PIB par tête en niveau.

sensiblement tandis que le chômage présente une hausse persistante. Ces effets sont liés à une hausse persistante du taux d'intérêt réel. Le calcul des mesures d'inefficience sur les marchés du travail et des biens suggère toutefois que les marges sur les salaires, qui augmentent fortement après un choc de désinflation, ont joué un rôle non négligeable dans la propagation de ces chocs. Ces résulats apparaissent relativement robustes à d'autres définitions et spécifications des variables, ainsi qu'à d'autres représentations et identifications dans le modèle VAR Structurel. Les réponses éstimées de la zone, si elles sont prises plus ou moins pour argent comptant, peuvent représenter une nouvelle base de faits empiriques pour les modèles DSGE afin notamment d'expliciter plus clairement les mécanismes principaux (rigidités nominales et/ou réelles, degré de substitution des facteurs de production, conduite de la politique monétaire) responsables de ces effets observés.

Annexe A : Construction des données

Les données brutes utilisées dans l'analyse et leur provenance sont détaillées dans le tableau A.1.

Tableau A.1. Données brutes

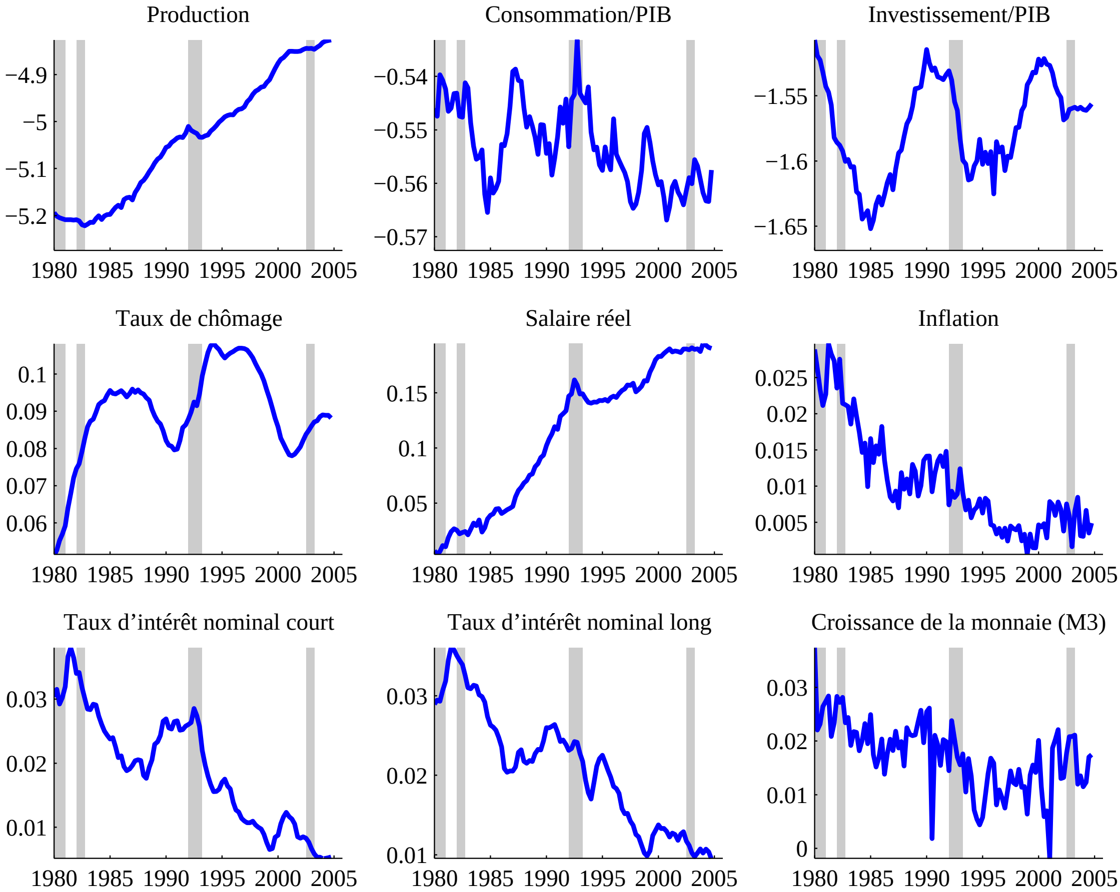
Variable	Symbole	Mnémonique	Source
PIB	Y_t	YER	Base AWM, FAGAN *et alii.* [2005]
Consommation	C_t	PCR	Base AWM, FAGAN *et alii.* [2005]
Investissement	X_t	ITR	Base AWM, FAGAN *et alii.* [2005]
Chômage	u_t	URX	Base AWM, FAGAN *et alii.* [2005]
Salaire	W_t	WRN	Base AWM, FAGAN *et alii.* [2005]
Taux d'intérêt nominal court	i_t^s	STN	Base AWM, FAGAN *et alii.* [2005]
Taux d'intérêt nominal long	i_t^l	LTN	Base AWM, FAGAN *et alii.* [2005]
Déflateur du PIB	P_t	YED	Base AWM, FAGAN *et alii.* [2005]
M3	M_t	—	Bulletin de la BCE
Population	N_t	—	Perspectives économiques, OCDE
Prix du pétrole	—	—	Base Fred II
Taux de change Euro-$	—	EEN	Base AWM, FAGAN *et alii.* [2005]
PIB américain	—		Base FREDII
Déflateur du PIB américain	—-		Base FREDII

Ces données brutes sont ensuite utilisées pour construire les variables utilisées dans l'analyse, selon la procédure décrite ci-dessous

$$\begin{aligned}
y_t &= \log(Y_t/N_t) \\
c_t &= \log(C_t/N_t) \\
x_t &= \log(X_t/N_t) \\
w_t &= \log(W_t) \\
\pi_t &= \Delta\log(P_t) \\
\gamma_t &= \Delta\log(M_t)
\end{aligned}$$

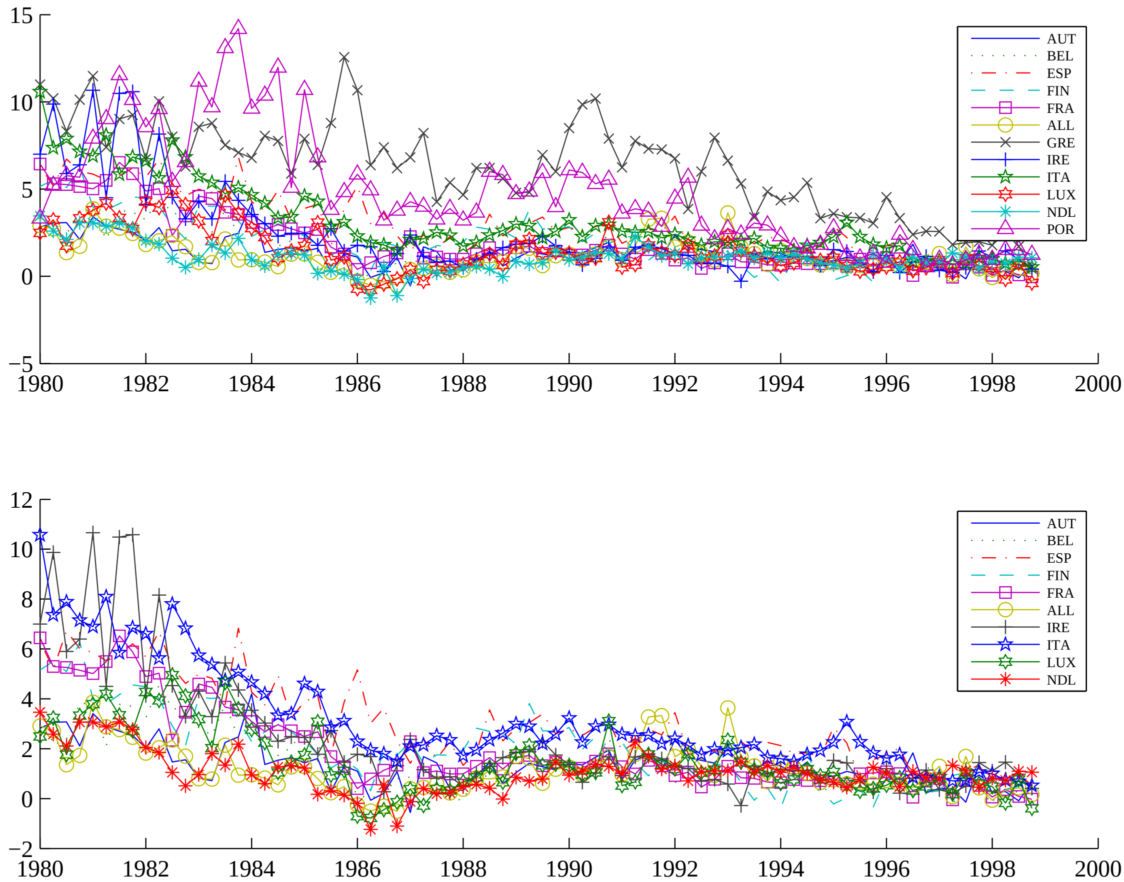
Les trajectoires historiques de y_t, $c_t - y_t$, $x_t - y_t$, u_t, w_t, π_t, i_t^s, i_t^l et γ_t sont reportées sur le graphique A.1.

Figure A.1 – Données utilisées dans l'analyse



Notes : la zone grisée correspond aux dates de récession identifiées à l'aide de l'algorithme de Bry–Boschan appliqué au logarithme du PIB par tête en niveau.

FIGURE A.2 – Trajectoires d'inflation des pays membres



Notes : l'inflation est mesurée comme le taux de croissance de l'indice des prix à la consommation désaisonnalisé. Le panneau inférieur exclut la Grèce et le Portugal.

Références Bibliographiques

ANDERSEN P. et WASCHER W. (1999) - «Sacrifice Ratios and the Conduct of Monetary Policy in Conditions of Low Inflation», *Document de travail n°82*, Banque des Règlements Internationaux.

ANDRES J. et HERNANDO I. (1999) - «Does Inflation Harm Economic Growth ? : Evidence for the OECD Countries», in *The Costs and Benefits of Achieving Price Stability*, M. Feldstein (ed.). NBER and Michigan University Press. Boston.

ANDRES J., HERNANDO I. et LÓPEZ-SALIDO J. (2002) - «The Long-Run Effect of Permanent Disinflations», *Document de travail n°9825*, Banque d'Espagne.

BALL L. (1994) - «What Determines the Sacrifice Ratio ?», in G. Mankiw (Ed.) *Monetary policy*, University of Chicago Press, 155–88.

BANQUE CENTRALE EUROPÉENNE (2004) - *The Monetary Policy of the ECB*, Banque Centrale Européenne.

BLANCHARD O.J. (1997) - «The Medium Run», *Brookings Papers on Economic Activity*, **2**, 89–158.

BLANCHARD O.J. (2003) - «Monetary Policy and Unemployement», remarques à la conférence *Monetary Policy and the Labor Market*, en l'honneur de James Tobin.

BLANCHARD O.J. et SUMMERS L. (1986) - «Hysteresis and the European Unemployment Problem», *NBER Macroeconomics Annual*, **1**, 15–78.

BLANCHARD O.J. et MUET P.A. (1993) - «Competitiveness through Disinflation : An Assessment of French Macroeconomic Policy since 1983», *Economic Policy*, **16**, 12–56.

BLANCHARD O.J. et QUAH D. (1989) - «The Dynamic Effects of Aggregate Demand and Supply Disturbances», *American Economic Review*, **79**, 655–673.

BOMFIM A.N. et RUDEBUSCH G.D. (2000) - «Opportunistic and Deliberate Disinflation under Imperfect Credibility», *Journal of Money, Credit, and Banking*, **32**, 701–21.

BUITER W. et MILLER M. (1985) - «Costs and Benefits of an Anti-Inflationary Policy : Questions and Issues», in V. Argy et J. Neville (Eds.) *Inflation and Unemployment :Theory, Experience and Policy Making*, George Allen & Unwin, Londres.

BULLARD J. et KEATING J. (1995) - «The Long-Run Relationship between Inflation and Output in Postwar Economies», *Journal of Monetary Economics*, **36**, 477–496.

CHANG Y., DOH T. et SCHORFHEIDE, F. (2007) - «Non-Stationary Hours in a DSGE Model», *Journal of Money, Credit, and Banking*, **39**, 1357–1373.

CHRISTIANO L.J., EICHENBAUM M. et EVANS C. (1999) - «Monetary Policy Shocks : What Have We Learned and to What End ?», in M. Woodford et J. Taylor (Eds), *Handbook of Macroeconomics*, North-Holland, chapitre 3.

CHRISTIANO L.J., FITZGERALD T.J. (2003) - «The Band Pass Filter», *International Economic Review*, **44**, 435-465.

CLING, P.P et MEUNIER F. (1986) - «La Désinflation en France : Le Point de Vue de l'Econometre», *Revue Economique*, 1093–1126.

COENEN G. et VEGA J.L. (2001) - «The Demand for M3 in the Euro-Area», *Journal of Applied Econometrics*, **16**, 727–748.

COGLEY T. et SARGENT T. (2007) - «Inflation-Gap Persistence in the U.S.», mimeo.

DE WALQUE G., SMETS, F. et WOUTERS R. (2006) - «Firm-Specific Production Factors in a DSGE Model with Taylor Price Setting», *International Journal of Central Banking*, **2**, 107–154.

DOLADO J., LÓPEZ-SALIDO J. et VEGA J. (2000) - «Unemployement and Inflation Persistence in Spain : Are there Phillips Trade-Offs ?», *Spanish Economic Review*, **2**, 267–291.

DUPAIGNE M., FÈVE P. et MATHERON J. (2007) - «Avoiding Pitfalls in Using Structural VARs to Estimate Economic Models», *Review of Economic Dynamics*, **10**, 238–255.

FAGAN, G., HENRY J. et MESTRE R. (2005) - «An Area-Wide Model (AWM) for the Euro-Area», *Economic Modelling*, **22**, 39–59.

FITOUSSI J.P et PASSET O. (2000) - «Réformes Structurelles et Politiques Macroéconomiques : Les Enseignements des Modèles de Pays», dans *Réduction du Chômage : Les Réussites en Europe*, Rapport n°23 du CAE, la Documentation Française.

FRIEDMAN M. (1968) - «Inflation : Causes and Consequences», in *Dollars and Deficits : Living With America's Economic Problems*, Prentice-Hall.

Galí J. (1999) - «Technology, Employment, and the Business Cycle : Do Technology Shocks Explain Aggregate Fluctuations ?», *American Economic Review*, **89**, 249–271.

Galí J., Gertler M. et López–Salido D. (2003) - «The Euro Area Inefficiency Gap», *Document de travail n°0302*, Banque d'Espagne.

Galí J., Gertler M. et López–Salido D. (2007) - «Markups, Gaps and the Welfare Costs of Business Fluctuations», *Review of Economics and Statistics*, **89**, 44–59.

Ireland P. (2007) - «Changes in the Federal Reserve's Inflation Target : Causes and Consequences», *Journal of Money, Credit and Banking*, **39**, 1851–1882.

Kormendi R. et Meguire P. (1985) - «Macroeconomic Determinants of Growth : Cross Country Evidence», *Journal of Monetary Economics*, **16**, 141–163.

McCandles G.T. et Weber W.E. (1995) - «Some Monetary Facts», *Federal Reserve Bank of Minneapolis Quarterly Review*, **19**, 2–11.

Orphanides A. et Wilcox D.W. (2002) - «The Opportunistic Approach to Disinflation», *International Finance*, **5**, 47–71.

Romer C. et Romer D. (1989) - «Does Monetary Policy Matter ? A New Test in the Spirit of Friedman and Schwartz», *NBER Macroeconomics Annual*, **4**, 121–170.

Romer C. et Romer D. (1994) - «Monetary Policy Matters», *Journal of Monetary Economics*, **34**, 75–88.

Sims C.A. (1998) - «Comment on Glenn Rudebusch's "Do Measures of Monetary Policy in a VAR Make Sense ?"», *International Economic Review*, **39**, 933-941.

Smets F. et Wouters R. (2003) - «An Estimated Dynamic Stochastic General Equilibrium Model of the Euro–Area», *Journal of the European Economic Association*, **1**, 1123–1175

Smets F. et Wouters R. (2005) - «Comparing Shocks and Frictions in US and Euro–Area Business Cycles : a Bayesian DSGE Approach», *Journal of Applied Econometrics*, **20**, 161–183.

Staiger D., Stock J. et Watson M. (2002) - «Prices, Wages, and the US NAIRU in the 1990s», in Alan Krueger and Robert Solow (Eds.), *The Roaring Nineties*, Russel Sage Foundation, New York, 2002, 3-60 .

STOCK J. et WATSON M. (2007) - «Why has U.S. Inflation Become Harder to Forecast ?», *Journal of Money, Banking and Credit*, **39**, 3–34.

VLAAR P. (2004) - «Shocking the Eurozone», *European Economic Review*, **48**, 109–131.

WATSON M. (1994) - «Business-Cycle Durations and Postwar Stabilization of the U.S. Economy», *American Economic Review*, **84**, 24-46.

Decreasing aversion and prudence under ambiguity aversion

Frédéric Cherbonnier
Toulouse School of Economics

Christian Gollier*
Toulouse School of Economics (LERNA and IDEI)

June 30, 2010

Abstract

We examine the conditions under which a risk-averse and ambiguity-averse agent exhibits some standard preference properties, as decreasing aversion and prudence. We explore two decision criteria (maxmin and smooth ambiguity aversion), and two decision problems (acceptance of a lottery and the portfolio choice problem).

Keywords: Decreasing aversion, prudence, portfolio choice, smooth ambiguity aversion, maxmin.

1 Introduction

How does one's attitude towards risk evolve when one becomes wealthier? One of the most ubiquitious assumption in uncertainty economics is that wealthier people are less risk-averse. Various definitions of the concept of decreasing aversion exist in the literature. For example, an agent is said to have decreasing aversion if any risk that is undesirable at some specific wealth level is also undesirable at all smaller wealth levels. Another definition of decreasing aversion is that in the one-risk-free-one-risky-asset portfolio choice problem, the demand for the risky asset is an increasing function of the initially sure wealth of the agent. In the classical expected utility model, these two definitions of decreasing aversion are equivalent, and the necessary and sufficient condition is expressed by the decreasing nature of absolute risk aversion. This universally accepted property of individual risk preferences plays a crucial role in many applications of the expected utility theory, as illustrated in Gollier (2001).

*This research was supported by the Chair of the Financière de la Cité at TSE, and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 230589. **Correspondence**: Toulouse School of Economics, Manufacture des Tabacs, Aile J.-J. Laffont, MF007, 21 allée de Brienne, 31000 Toulouse, France. Email: gollier@cict.fr

In this paper, we explore the concept of decreasing aversion in the context of ambiguity and ambiguity aversion. In most cases, the probability distribution of the risk is not perfectly known. Since Ellsberg (1961), we know that economic agents do not behave accordingly to the subjective expected utility model. Under ambiguous distribution, they do not use a subjectively chosen probability distribution to compute the expected utility of the set of possible acts to determine their optimal strategy. They are ambiguity-averse in the sense that they always prefer a lottery that yields a given payoff with probability p over another lottery that yields the same payoff with an unknown probability whose expectation is p. Gilboa and Schmeidler (1989) were the first to propose a decision criteria that satisfies this behavioural trait, and that generalizes the expected utility model. In short, agents are assumed to have multiple priors whose formation is a characteristic of the preferences of the agent. The agent's ex ante welfare associated to an act is the smallest expected utility generated by this act over the different possible priors. This multi-prior model has recently been complemented by a smooth version of ambiguity aversion by Klibanoff, Marinacci and Mukerji (2005). In this version, the agent's welfare under uncertainty is measured by the certainty equivalent of the different prior-dependent expected utility levels. This certainty equivalent is computed by using a function ϕ that is increasing and concave, and whose degree of concavity is an index of ambiguity aversion. For these two decision criterion, we determine the conditions under which wealthier people are less averse to risk, under the two standard definitions for this concept.

2 Acceptance of risk

In this section, we characterize the conditions under which wealthier people have a larger set of desirable loteries. In other words, when becoming poorer, it is never possible that a previously undesirable lottery becomes desirable. In that case, we say that the agent exhibits Decreasing Aversion (DA). The benchmark model is expected utility. Under this model, decreasing aversion is characterized by the following condition. For any initial wealth z, and for any random variable $\widetilde{x}$ so that the support of $z+\widetilde{x}$ is in the domain of the utility function u, we have that

$$Eu(z+\widetilde{x}) \leq u(z) \implies Eu(z'+\widetilde{x}) \leq u(z') \quad \forall z' \leq z. \tag{1}$$

This means that if an agent with utility u dislikes lottery $\widetilde{x}$ at wealth level z, she must also dislike it at all wealth levels smaller than z. We know since Pratt (1964) that this is true if and only if the utility function u exhibits decreasing absolute risk aversion (DARA). We formalize this by using the following definition.

Definition 1 *We say that a function $f : \mathbb{R} \to \mathbb{R}$ satisfies (weak) Decreasing Concavity (DC) if $-f''/f'$ is non-increasing.*

It is easy to check that f DC means that there exists a concave function g such that $-f' = g \circ f$. In the expected utility model, decreasing aversion holds if and only if u exhibits DC. Notice that condition (1) is equivalent to the single-crossing from below of function v defined as $v(z) = Eu(z + \widetilde{x})$ with respect to function u. Thus, condition (1) is equivalent to

$$Eu(z + \widetilde{x}) = u(z) \implies Eu'(z + \widetilde{x}) \geq u'(z). \tag{2}$$

Suppose alternatively that there is some ambiguity about the true distribution of the payoff of the lottery. For the sake of simplicity, suppose that the payoff of the lottery has n possible distributions, corresponding to random variables $\widetilde{x}_1, ..., \widetilde{x}_n$. We hereafter examine two decision models under uncertainty: maxmin and smooth ambiguity aversion. Under the maxmin model, decreasing aversion requires for any z, and for any set of random variables $(\widetilde{x}_1, ..., \widetilde{x}_n)$ so that the support of $z + \widetilde{x}_\theta$ is in the domain of the utility function u for all $\theta = 1, ..., n$, we have that

$$\min_\theta Eu(z + \widetilde{x}_\theta) \leq u(z) \implies \min_\theta Eu(z' + \widetilde{x}_\theta) \leq u(z') \quad \forall z' \leq z. \tag{3}$$

Proposition 2 *The maxmin criterion implies decreasing aversion if and only if u exhibits decreasing concavity.*

Proof: The necessity comes from the fact that DC is necessary when the $\widetilde{x}_\theta$ are identically distributed. For sufficiency, consider a specific z and let θ_0 denote the argument of $\min_\theta Eu(z + \widetilde{x}_\theta)$. Suppose that

$$Eu(z + \widetilde{x}_{\theta_0}) = \min_\theta Eu(z + \widetilde{x}_\theta) \leq u(z).$$

Because u is DC, this implies that for all $z' \leq z$, $Eu(z' + \widetilde{x}_{\theta_0}) \leq u(z')$. It implies that

$$\min_\theta Eu(z' + \widetilde{x}_\theta) \leq u(z').$$

This concludes the proof of sufficiency. Such proof can easily be extended when facing an infinite number of priors[1]. ■

We can conclude from this result that the multi-prior maxmin model has the property of decreasing aversion if and if the corresponding expected utility model has it also, which is the case if $-u''/u'$ is non-increasing.

We now examine the model of smooth ambiguity aversion, as introduced by Klibanoff, Marinacci and Mukerji (KMM, 2005). This KMM model is based on the same ingredients than the maxmin model. A new ingredient is the vector $(q_1, ..., q_n)$ of non-negative scalars that sum up to unity. Parameter q_θ can be interpreted as the (subjective) probability that $\widetilde{x}_\theta$ describes the true distribution of the lottery. Another ingredient is a new real-valued function ϕ that is

[1] Indeed, if the left condition in (3) is satisfied, then for any $\epsilon > 0$, there exist $\bar{\theta}$ such that $Eu(z + \widetilde{x}_{\bar{\theta}}) \leq u(z + \epsilon)$. Thus, by considering the prior $\widetilde{x}'_{\bar{\theta}} = \widetilde{x}_{\bar{\theta}} - \epsilon$ instead, we can apply the decreasing aversion property and get $Eu(z' + \widetilde{x}_{\bar{\theta}}) = Eu(z' + \epsilon + \widetilde{x}'_{\bar{\theta}}) \leq u(z' + \epsilon)$ for all $z' \leq z$. Thus the right condition in (3) holds at the limit

increasing and concave, so that the welfare if the lottery is accepted is measured by the certainty equivalent of the conditional expected utility $Eu(z+\widetilde{x}_\theta)$:

$$W(z)=\phi^{-1}\left(\sum_{\theta=1}^{n} q_\theta \phi\left(Eu(z+\widetilde{x}_\theta)\right)\right)$$

Because ϕ is assumed to be concave, the existence of ambiguity reduces welfare since, by Jensen's inequality,

$$W(z)\leq \phi^{-1}\left(\phi\left(\sum_{\theta=1}^{n} q_\theta Eu(z+\widetilde{x}_\theta)\right)\right)=Eu(z+\widetilde{x}),$$

where $\widetilde{x}$ is distributed as $(q_1,\widetilde{x}_1;...;q_n,\widetilde{x}_n)$. If the lottery is rejected, the agent enjoys welfare $\phi^{-1}(\phi(u(z)))=u(z)$. Decreasing aversion as defined in this section requires in this case that W single-crosses u from below. This requires that the following property be satisfied:

$$\phi^{-1}\left(\sum_{\theta=1}^{n} q_\theta \phi(Eu(z+\widetilde{x}_\theta))\right)=u(z) \implies \frac{\sum_{\theta=1}^{n} q_\theta \phi'(Eu(z+\widetilde{x}_\theta))Eu'(z+\widetilde{x}_\theta)}{\phi'(W(z))}\geq u'(z). \tag{4}$$

One can easily extract two necessary conditions for decreasing aversion in this framework. The first is that u must be DC, since one possibility is that all $\widetilde{x}_\theta$ be identically distributed. Indeed, in that case, condition (4) is equivalent to (2). The second condition can be derived in the special case in which all $\widetilde{x}_\theta$ are degenerated. Suppose that $\widetilde{x}_\theta$ takes value y_θ almost surely, for all $\theta=1,...,n$. Let random variable $\widetilde{y}$ be distributed as $(y_1,q_1;...;y_n,q_n)$. Then, condition (4) can be rewritten as

$$E\phi\circ u(z+\widetilde{y})=\phi\circ u(z) \implies \quad E(\phi\circ u)'(z+\widetilde{y})\geq(\phi\circ u)'(z).$$

As recalled above, this condition holds for all z and all random variable $\widetilde{y}$ if and only if $\phi\circ u$ exhibits DC. This observation is just a restatement of the fact that the KMM model simplifies to the expected utility model with utility function $\phi\circ u$ when the multiple priors are all degenerated.

This means that both conditions u DC and $\phi\circ u$ DC are necessary for decreasing aversion in the KMM smooth ambiguity model. We hereafter show that this joint condition is also sufficient.

Proposition 3 *Consider the KMM smooth ambiguity aversion model characterized by functions (u,ϕ). It exhibits decreasing aversion if and only if u and $\phi\circ u$ both exhibit decreasing concavity.*

Proof of sufficiency: Let y_θ be the certainty equivalent of $\widetilde{x}_\theta$ under function u, i.e., $u(z+y_\theta)=Eu(z+\widetilde{x}_\theta)$. Let random variable $\widetilde{y}$ be distributed as

$(y_1, q_1; ...; y_n, q_n)$. Suppose that

$$W(z) = \phi^{-1}\left(\sum_{\theta=1}^{n} q_\theta \phi(Eu(z+\widetilde{x}_\theta))\right) = u(z),$$

or equivalently, that $E\phi \circ u(z+\widetilde{y}) = \phi \circ u(z)$. Because u is DC, $Eu(z+\widetilde{x}_\theta) = u(z+y_\theta)$ implies that $Eu'(z+\widetilde{x}_\theta) \geq u'(z+y_\theta)$. It implies in turn that

$$\sum_{\theta=1}^{n} q_\theta \phi'(Eu(z+\widetilde{x}_\theta))Eu'(z+\widetilde{x}_\theta) \geq \sum_{\theta=1}^{n} q_\theta \phi'(u(z+y_\theta))u'(z+y_\theta) = E(\phi \circ u)'(z+\widetilde{y}).$$

We assume that $\phi \circ u$ is DC. It implies that $E(\phi \circ u)'(z+\widetilde{y}) \geq (\phi \circ u)'(z)$. We conclude from the previous equation that

$$\sum_{\theta=1}^{n} q_\theta \phi'(Eu(z+\widetilde{x}_\theta))Eu'(z+\widetilde{x}_\theta) \geq \phi'(u(z))u'(z) = \phi'(W(z))u'(z).$$

It implies that the right condition in (4) is satisfied. This concludes the proof of the sufficiency of u and $\phi \circ u$ being DC. ■

Notice that this necessary and sufficient condition does not require that ϕ exhibits DC. Only the DC of u and $\phi \circ u$ is required. A sufficient condition is that u and ϕ be DC. Indeed, we have that

$$-\frac{(\phi \circ u)''(z)}{(\phi \circ u)'(z)} = -\frac{\phi''(u(z))}{\phi'(u(z))}u'(z) - \frac{u''(z)}{u'(z)}. \tag{5}$$

Because u is increasing and u' is decreasing, the right-hand side of this equality is decreasing in z if both $-\phi''/\phi'$ and $-u''/u'$ are decreasing. This yields the following corollary.

Corollary 4 *Consider the KMM smooth ambiguity aversion model characterized by functions (u, ϕ). It exhibits decreasing aversion if u and ϕ both exhibit decreasing concavity.*

The fact that the DC of ϕ is not necessary can be illustrated by the following counterexample. Suppose that $u(z) = z^{1/2}$. Suppose also that ϕ is such that $\phi'(u) = \exp(-ku^2)$, so that $-\phi''(u)/\phi'(u) = 2ku$. Thus, this ϕ function, which is increasing and concave in the relevant domain, exhibits increasing concavity. Still, $\phi \circ u$ exhibits DC, since using (5), we have that

$$-\frac{(\phi \circ u)''(z)}{(\phi \circ u)'(z)} = \frac{2kz^{1/2}}{2z^{1/2}} + \frac{1}{2z} = k + \frac{1}{2z},$$

which is decreasing. Thus, from Proposition 3, the smooth ambiguity-averse agent (u, ϕ) is decreasingly averse in spite of the fact that ϕ is not DC.

3 Portfolio allocation

In this section, we examine an alternative decision model. The decision maker with initial wealth z can invest in two assets. There is a risk free asset whose return is normalized to zero, and a risky asset whose return is expressed by random variable $\widetilde{x}$. The agent must choose the size α of her investment in the risky asset. In the expected utility model, this decision problem can be written as

$$\max_{\alpha} Eu(z+\alpha\widetilde{x}). \tag{6}$$

We normalize the unit of the risky asset in such a way that it is optimal to invest excatly one monetary unit in the risky asset, so that the first order condition of the above program yields

$$E\tilde{x}u^{'}(z+\tilde{x})=0.$$

In this model, it is easy to check that an increase in the initial wealth raises the dollar investment in the risky asset if and only if u is decreasingly concave, as first shown by Arrow (1963). Indeed, this requires to show that the cross derivative of the objective function in (6) with respect to α and z evaluated at $\alpha^*=1$ is positive. Because u DC means that $-u'$ is a concave function g of u, it implies that

$$E\widetilde{x}u''(z+\widetilde{x})=-E\widetilde{x}g'(u(z+\widetilde{x}))u'(z+\widetilde{x})\geq -g'(u(z))E\widetilde{x}u'(z+\widetilde{x})=0.$$

The inequality above comes from the observation that for all x, $-xg'(u(z+x))\geq -xg'(u(z))$. This demonstrates that the DC property in the expected utility model is necessary and sufficient for two testable results. Namely, the DC of u means that when wealth increases, the set of acceptable lotteries inflates, and the demand for the risky asset increases. In the following, we will consider whether this property is still satisfied when facing an ambiguous distribution.

We suppose now that the distribution of the return of the risky asset is ambiguous. This ambiguity is characterized by n possible random variables $(\widetilde{x}_1,...,\widetilde{x}_n)$. Under the maxmin criterion, the decision program can be rewritten as

$$\max_{\alpha}\min_{\theta} Eu(z+\alpha\widetilde{x}_\theta). \tag{7}$$

The next result provides us with general conditions on the set of priors under which an increase of wealth raises the investment in the risky asset. We first need to define a relation between risks. We say that $\tilde{x}$ dominates $\tilde{y}$ in the sense of the "Risk Dominance" (RD) iff there exist a risk $\tilde{\epsilon}$ non correlated to $\tilde{x}$ such that $\tilde{y}=\tilde{x}+\tilde{\epsilon}$ (this relation is not necessarily an order).

Proposition 5 *Suppose that both u and ϕ are DC. An increase in wealth reduces the demand for the risky asset if one of the following two conditions is satisfied*

- *the set of priors $(\tilde{x}_1, ..., \tilde{x}_n)$ can be ranked according to both risk dominance and central dominance, that is, if $\tilde{x}_\theta \preceq_{RD} \tilde{x}_{\theta+1}$ and $\tilde{x}_\theta \preceq_{CD} \tilde{x}_{\theta+1}$ for all $\theta = 1, ..., n-1$;*
- *the set of priors $(\tilde{x}_1, ..., \tilde{x}_n)$ can be ranked according to SSD.*

Let's consider the optimal investment α^* that solves the maximization problem 7. If the minimum $\min_\theta Eu(z + \alpha\widetilde{x}_\theta)$ is reached at only one prior $\bar{\theta}$, then locally around z, the decision program can be rewritten as $\max_\alpha Eu(z + \alpha\widetilde{x}_{\bar{\theta}})$ and the decreasing aversion property is a consequence of u DC. Hence, in order to check the decreasing aversion property, we only need to consider cases where there exist $\bar{\theta}$ and $\bar{\theta}^{'}$ such that the optimum of the decision problem is reached at

$$Eu(z + \alpha^*\widetilde{x}_{\bar{\theta}}) = Eu(z + \alpha^*\widetilde{x}_{\bar{\theta}'}) \tag{8}$$

which implies that $Eu^{'}(z + \alpha^*\widetilde{x}_{\bar{\theta}})$ and $Eu^{'}(z + \alpha^*\widetilde{x}_{\bar{\theta}'})$ can not be both either strictly positive or strictly negative. When the set of priors can be ranked according to SSD, such equality 8 is ruled out which proves the second part of the proposition. More generally, a sufficient condition for the decreasing aversion property is that $\alpha^{'*}(z) \geq 0$ when the equality 8 holds. Since

$$\alpha^{'*}(z) = -\frac{Eu^{'}(z + \alpha^*\widetilde{x}_{\bar{\theta}}) - Eu^{'}(z + \alpha^*\widetilde{x}_{\bar{\theta}'})}{E\widetilde{x}_{\bar{\theta}}u'(z + \alpha^*\widetilde{x}_{\bar{\theta}}) - E\widetilde{x}_{\bar{\theta}'}u'(z + \alpha^*\widetilde{x}_{\bar{\theta}'})}$$

we see that, when $\tilde{x}_{\bar{\theta}} \preceq_{\text{CD}} \tilde{x}_{\bar{\theta}'}$, the denominator is negative. If $\tilde{x}_{\bar{\theta}} \preceq_{\text{RD}} \tilde{x}_{\bar{\theta}'}$, then $\tilde{x}_{\bar{\theta}}$ is, up to a constant, an increase in risk with respect to $\tilde{x}_{\bar{\theta}'}$, i.e. there exist a non correlated $\tilde{\epsilon}$ (with non zero mean) such that $\tilde{x}_{\bar{\theta}} = \tilde{x}_{\bar{\theta}'} + \tilde{\epsilon}$, and consequently $Eu^{'}(z + \alpha^*\widetilde{x}_{\bar{\theta}}) = Ev^{'}(z + \alpha^*\epsilon)$ where v is the indirect utility function $v(z) = Eu(z + \alpha^*\widetilde{x}_{\bar{\theta}'})$. We know that u DC implies v DC, thus $Ev^{'}(z + \alpha^*\epsilon) \geq v^{'}(z)$ and the numerator is thus positive[2]. ■

Let's now consider the dynamic decreasing aversion problem in the KMM ambiguity framework. The decision problem becomes

$$\max_\alpha E\phi(Eu(z + \alpha\widetilde{x}_\theta)) \tag{9}$$

and the first order condition is therefore

$$E\phi'(Eu(z + \widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u'(z + \widetilde{x}_{\widetilde{\theta}}) = 0, \tag{10}$$

The decreasing aversion property is then satisfied when the derivative of the optimal allocation α^* is positive. Since the denominator in the relation

$$\alpha^{*'} = -\frac{E\phi^{''}(Eu(z + \widetilde{x}_{\widetilde{\theta}}))Eu'(z + \widetilde{x}_{\widetilde{\theta}})E\widetilde{x}_{\widetilde{\theta}}u'(z + \widetilde{x}_{\widetilde{\theta}}) + E\phi'(Eu(z + \widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u^{''}(z + \widetilde{x}_{\widetilde{\theta}})}{E\phi^{''}(Eu(z + \widetilde{x}_{\widetilde{\theta}}))(E\widetilde{x}_{\widetilde{\theta}}u'(z + \widetilde{x}_{\widetilde{\theta}}))^2 + E\phi'(Eu(z + \widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}^2_{\widetilde{\theta}}u''(z + \widetilde{x}_{\widetilde{\theta}})}$$

[2] we assumed here that the denominator is strictly negative. If it were null then, when changing slightly z, the optimum of the decision problem would still be the optimum of the one-prior decision problem $\max_\alpha Eu(z + \alpha\widetilde{x}_{\bar{\theta}})$ for either $\bar{\theta}$ or $\bar{\theta}^{'}$

is positive, the decreasing aversion property is satisfied when 10 implies

$$E\phi^{''}(Eu(z+\widetilde{x}_{\widetilde{\theta}}))Eu'(z+\widetilde{x}_{\widetilde{\theta}})E\widetilde{x}_{\widetilde{\theta}}u'(z+\widetilde{x}_{\widetilde{\theta}})+E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u^{''}(z+\widetilde{x}_{\widetilde{\theta}}) \geq 0 \tag{11}$$

The next result provides us with general conditions on the set of priors under which an increase of wealth raises the investment in the risky asset. Those conditions are similar to the conditions obtained by Gollier (2010) in order to ensure that an increase in ambiguity decreases the investment in the risky asset.

Proposition 6 *Suppose that both u and ϕ are DC. An increase in wealth reduces the demand for the risky asset if the set of priors $(\tilde{x}_1, ..., \tilde{x}_n)$ can be ranked according to both second stochastic dominance (SSD) and central dominance, that is, if $\tilde{x}_\theta \preceq_{SSD} \tilde{x}_{\theta+1}$ and $\tilde{x}_\theta \preceq_{CD} \tilde{x}_{\theta+1}$ for all $\theta = 1, ..., n-1$;*

We assume in the following that both u and ϕ are DC, and that the relation 10 is true. Let's first show that the right term of 11 is positive. Since u is DC, the function $f = u^{'}u^{-1}$ is convex. Thus

$$E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u^{''}(z+\widetilde{x}_{\widetilde{\theta}}) = E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u^{'}(z+\widetilde{x}_{\widetilde{\theta}})f^{'}(u(z+\widetilde{x}_{\widetilde{\theta}})) \geq 0$$

since, when $\widetilde{x}_{\widetilde{\theta}} > 0$, we have $f^{'}(u(z+\widetilde{x}_{\widetilde{\theta}})) > f^{'}(u(z))$. We now consider whether the left term of equation 11 is positive. This condition can be rewritten as

$$E\left[-\frac{\phi^{''}(Eu(z+\widetilde{x}_{\widetilde{\theta}}))}{\phi^{'}(Eu(z+\widetilde{x}_{\widetilde{\theta}}))}Eu'(z+\widetilde{x}_{\widetilde{\theta}})\phi^{'}(Eu(z+\widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u'(z+\widetilde{x}_{\widetilde{\theta}})\right] \leq 0$$

The right terms are similar to the equation 10, and thus takes both positive and negative values. If the set of priors $(\tilde{x}_1, ..., \tilde{x}_n)$ can be ranked according to both second stochastic dominance and central dominance, then the positive values are taken for the highly ranked priors $\widetilde{\theta}$, which have also higher value of $Eu(z+\widetilde{x}_{\widetilde{\theta}})$ and $-Eu^{'}(z+\widetilde{x}_{\widetilde{\theta}})$ according to the second stochastic ranking (since u is DC implies that $-u^{'}$ is concave). ■

We show thereafter that the decreasing concavity property is not enough to ensure that the optimal investment $\alpha^*(z)$ is decreasing. We will also provide an example in which an increase in ambiguity aversion induces more investment in the risky asset, as in Gollier (2009). Those counterexamples will be built by considering any couple of priors $\tilde{x}_1$ and $\tilde{x}_2$ such as the following couple of property is satisfied : $E\tilde{x}_1u^{'}(z+\tilde{x}_1) = E\tilde{x}_2u^{'}(z+\tilde{x}_2) = 0$ and $Eu(z+\tilde{x}_1) < Eu(z+\tilde{x}_2)$. It means that an investor would invest 1 share of $\tilde{x}_i$ for $i = 1, 2$ and that the investor strictly prefers $\tilde{x}_2$ to $\tilde{x}_1$. Let's denote by (P) this condition.

Proposition 7 *An investor is assumed to have utility and ambiguity function both exhibiting decreasing concavity. For any distribution $\tilde{x}_1$ and $\tilde{x}_2$ satisfying condition (P), there exist λ_1 and $\lambda_2 > 0$ such that, if we denote by $\tilde{x}$ the ambiguous distribution defined by the two priors of equal probability $\lambda_1\tilde{x}_1$ and $\lambda_2\tilde{x}_2$, one of the two following condition hold :*

- *The investor would invest more in the ambiguous asset that if he were neutral to ambiguity;*
- *When assuming besides that u is CARA, the optimal investment in the ambiguous asset is not always increasing with the wealth.*

Let's choose $\lambda_1 < 1$ and $\lambda_2 > 1$ such that

$$E\lambda_1\tilde{x}_1 u'(z+\lambda_1\tilde{x}_1) = -E\lambda_2\tilde{x}_2 u^{'}(z+\lambda_2\tilde{x}_2) > 0 \tag{12}$$

and such that $Eu(z+\lambda_1\tilde{x}_1) < Eu(z+\lambda_2\tilde{x}_2)$. The latter relation is true if the λ_i are close enough to one. The former relation means that the investor would invest 1 share of $\tilde{x}$ if he was neutral to ambiguity. This relation can easily be obtained by an appropriate choice of λ_i since the left term of this relation is a function of λ_1 that is strictly decreasing at $\lambda_1 = 1$. We then have

$$\phi'(Eu(z+\lambda_1\widetilde{x}_1))E\lambda_1\widetilde{x}_1 u'(z+\lambda_1\widetilde{x}_1) + \phi'(Eu(z+\lambda_2\widetilde{x}_2))E\lambda_2\widetilde{x}_2 u'(z+\lambda_2\widetilde{x}_2) \geq 0$$

since $\phi'(Eu(z+\widetilde{x}_1)) > \phi'(Eu(z+\widetilde{x}_2)) > 0$. We have thus built a example where an increase in ambiguity aversion induce a higher investment in the risky asset. Similarly, let's now pick $\lambda_1 < 1$ and $\lambda_2 > 1$ such that

$$\phi'(Eu(z+\lambda_1\widetilde{x}_1))E\lambda_1\widetilde{x}_1 u'(z+\lambda_1\widetilde{x}_1) = -\phi'(Eu(z+\lambda_2\widetilde{x}_2))E\lambda_2\widetilde{x}_2 u'(z+\lambda_2\widetilde{x}_2) > 0$$

and such that $Eu(z+\lambda_1\tilde{x}_1) < Eu(z+\lambda_2\tilde{x}_2)$. Thus, the investor facing ambiguity would invest 1 share of $\tilde{x}$. To simplify the notation thereafter, let's replace $\tilde{x}_i$ by $\lambda_i\tilde{x}_i$ (i.e. $\lambda_i = 1$). If u is CARA, i.e. $u(z) = a(1-e^{-Az})$ for $a, A \geq 0$, then $u^{''}(z) = -Au^{'}(z)$, and the derivative of the $E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))E\widetilde{x}_{\widetilde{\theta}}u'(z+\widetilde{x}_{\widetilde{\theta}})$ with respect to the wealth z is equal to

$$\phi^{''}(Eu(z+\widetilde{x}_1))E\widetilde{x}_1 u'(z+\widetilde{x}_1)Eu'(z+\widetilde{x}_1) + \phi^{''}(Eu(z+\widetilde{x}_2))E\widetilde{x}_2 u'(z+\widetilde{x}_2)Eu'(z+\widetilde{x}_2)$$

because

$$\frac{\partial}{\partial z}E\widetilde{x}_1 u'(z+\widetilde{x}_1) = -AE\widetilde{x}_1 u'(z+\widetilde{x}_1).$$

Since $E\widetilde{x}_1 u'(z+\widetilde{x}_1) > 0$, this term is strictly negative if and only if

$$-\frac{\phi^{''}(Eu(z+\widetilde{x}_1))}{\phi'(Eu(z+\widetilde{x}_1))}Eu'(z+\widetilde{x}_1) > -\frac{\phi^{''}(Eu(z+\widetilde{x}_2))}{\phi'(Eu(z+\widetilde{x}_2))}Eu'(z+\widetilde{x}_2)$$

Since $u'(z) = aA - Au(z)$ and since, by assumption, ϕ verifies the decreasing concavity property, the terms of this inequality are the product of two decreasing functions. The inequality is then a consequence of $Eu(z+\widetilde{x}_1) < Eu(z+\widetilde{x}_2)$. ■

4 Prudence (P)

The notion of prudence is another notion, which is linked to precautionary saving. An agent is prudent if any zero-mean future risk on income induces the consumer to save more:

$$E\widetilde{x}_{\widetilde{\theta}} = 0 \implies E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))Eu'(z+\widetilde{x}_{\widetilde{\theta}}) = \phi'(u(z))u'(z).$$

We say that function f satisfies the Convex Marginal (CM) property is f' is convex. In the expected utility framework, it is well known that the consumer is prudent if and only if is CM. By Jensen inequality, it is obvious from looking at degenerate cases that u and $\phi \circ u$ being CM are necessary for prudence in the KMM framework. We hereafter show that the combination of these two conditions is also sufficient.

Let ξ_θ denote the expectation of $\widetilde{x}_\theta$. Because u is concave, we have that $Eu(z+\widetilde{x}_\theta) \leq u(z+\xi_\theta)$. It implies that

$$\phi'(Eu(z+\widetilde{x}_\theta)) \geq \phi'(u(z+\xi_\theta)).$$

Similarly, because u' is convex, we have that $Eu'(z+\widetilde{x}_\theta) \geq u'(z+\xi_\theta)$. These two inequalities imply that

$$E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))Eu'(z+\widetilde{x}_{\widetilde{\theta}}) \geq E\phi'(u(z+\xi_{\widetilde{\theta}}))u'(z+\xi_{\widetilde{\theta}}) = Ev'(z+\xi_{\widetilde{\theta}}).$$

If we assume that $v = \phi \circ u$ exhibits CM , we also have that $Ev'(z+\xi_{\widetilde{\theta}}) \geq v'(z)$ since $E\xi_{\widetilde{\theta}} = 0$. It implies that

$$E\phi'(Eu(z+\widetilde{x}_{\widetilde{\theta}}))Eu'(z+\widetilde{x}_{\widetilde{\theta}}) \geq v'(z) = \phi'(u(z))u'(z).$$

This concludes the proof of the following proposition.

Proposition 8 *(u,ϕ) exhibits prudence if and only if u and $v = \phi \circ u$ are both CM, i.e., u' and v' are convex.*

The following lemma shows that u and ϕ being imprudent is a sufficient condition for (u,ϕ) being imprudent. This is not a necessary condition. Indeed, if we consider the exponential utility function $u = 1/2 - e^{-z}$ and the ambiguity function $\phi = -(1-2z)^{3/2}$ defined for $z \leq 1/2$, then (u,ϕ) is CM while ϕ isn't CM.

Lemma 9 *u CM and ϕ CM implies $v = \phi \circ u$ CM.*

This comes from the observation that $v''' = \phi'''u'^3 + 3\phi''u''u' + \phi'u'''$, which is positive because ϕ' and u' are positive, decreasing and convex. ■

References

Arrow, K.J., (1963), Liquidity preference, Lecture VI in "*Lecture Notes for Economics 285, The Economics of Uncertainty*", pp 33-53, undated, Stanford University.

Dow, J., and S. Werlang, (1992), Uncertainty aversion, risk aversion and the optimal choice of portfolio, *Econometrica,* 60, 197-204.

Ellsberg, D., (1961), Risk, ambiguity, and the Savage axioms, *Quarterly Journal of Economics*, 75, 643-69.

Gilboa, I. and D. Schmeidler, (1989), Maximin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18, 141, 153.

Gollier, C., (2001), *The economics of risk and time*, MIT Press, Cambridge.

Kimball, M.S., (1990), Precautionary savings in the small and in the large, *Econometrica*, 58, 53-73.

Klibanoff, P., M. Marinacci and S. Mukerji, (2005), A smooth model of decision making under ambiguity, *Econometrica* 73(6), 1849-1892.

Rothschild, M. and J. Stiglitz, (1970), Increasing Risk: I. A Definition, *Journal of Economic Theory,* 2, 225-243.

Rothschild, M. and Stiglitz, J.E., (1971), Increasing Risk II: Its Economic Consequences, *Journal of Economic Theory*, 3, pp. 225-243.

Corporate Control and Multiple Large Shareholders*

June 30, 2010

Amrita Dhillon
University of Warwick

Silvia Rossetto
Toulouse School of Economics

*We are grateful to Denis Gromb, Parikshit Ghosh, Alexander Gümbel, Klaus Ritzberger, Arunava Sen, Oren Sussman, Lucy White and Lei Zhang for very useful comments. This research benefited from the support of the research grant "Financière de la Cité". Address for correspondence: Silvia Rossetto, Toulouse School of Economics, 21 Allée de Brienne, 31000 Toulouse, France. E-mail: Silvia.Rossetto@univ-tlse1.fr

Corporate Control and Multiple Large Shareholders

Abstract

Recent evidence on financial structures within firms suggests that concentrated ownership in the form of several large shareholders, rather than dispersed ownership, is the norm. This is puzzling given that often the stakes are too big for optimal diversification and too small to guarantee control. This paper attempts to provide an explanation. We consider a setting where multiple shareholders have endogenous conflicts of interest depending on the size of their stake. Such conflicts arise because larger shareholders tend to be less well diversified and would therefore prefer the firm to pursue more conservative investment policies, while dispersed shareholders prefer high risk, high return policies. A second blockholder (or more) can mitigate the conflict by shifting the voting outcome more towards the dispersed shareholders' preferred investment policy and this raises the share price. Having a larger stake increases the ability to change decisions: however, it also changes the preferences of the shareholders, reducing the incentive to buy larger shares. The main contribution of the paper is to show the conditions under which blockholder equilibria exist when there are *endogenous* conflicts of interest. The model shows how different ownership structures affect firm value and IPO's underpricing.

1 Introduction

TO BE ADDED REFERENCE TO POUND 88 TO JUSTIFY THE LAMBDA ALSO CITE THE OECD REPORT PAGE 39 AND 40 ABOUT THE DIFFICULTY TO VOTE AND DATA ON VOTE PARTICIPATION!

Finance theory has long recognized the important role of the size of shareholders in corporate governance. In order to overcome the free rider problems of dispersed ownership (Grossman and Hart, 1980) a large shareholder can act as a monitor and discipline the manager (Stiglitz (1985), Shleifer and Vishny (1986), Holmstrom and Tirole (1993), Admati, Pfleiderer, and Zechner (1994), Burkart, Gromb, and Panunzi (1997)). From a positive perspective however, optimal diversification implies that no investor should want to invest too high a proportion of his wealth in a single firm. Indeed, what is observed empirically is neither dispersed ownership in the firm nor one large shareholder: the evidence is consistent with *intermediate* sizes of stakes. Indeed, there are investors who hold about 5% to 10% of the firm's shares with commensurate voting power.[1]

Of course, holding 5% to 10% of the firm's shares does not give control over the firm's decisions! In this sense, it remains a puzzle why investors hold these "intermediate" stakes – too small to control the firm, yet not small enough to be well diversified. Our paper investigates this apparent paradox, shows the mechanism by which control is exercised and provides conditions (or firm characteristics) under which we might observe such ownership patterns.

Formally, suppose there is an initial owner or founder of a firm who needs to raise capital to finance a project. The initial owner has a large stake in the firm because of an inability to commit to monitoring without having large enough stakes in the firm. He raises capital through issuing shares and/or debt. There is a set of other investors who are ex-ante identical in their preferences, whose role is to buy shares in the firm after the price is announced by the initial owner. Both the owner as well as other investors are risk averse, but their preferences on risk/return depend only on the size of their stakes in the firm: the larger the stakes the lower the risk/return they prefer.

Once the ownership structure is established, shareholders vote on the riskiness of the investment projects that the firm subsequently undertakes. At the time of buying shares therefore, investors face a trade off between holding a diversified portfolio and having little influence on the firm's

[1] E.g. in eight out of nine largest stock markets in the European Union, the median size of the second largest voting block in large publicly listed companies exceeds five percent (data from the European Corporate Governance Network). The voting power of the second blockholder varies depending on the firm characteristics, but usually a second blockholder is almost always present in every firm across country, sector and type. La Porta, Lopez-De-Silanes, and Shleifer (1999) find that 25% of the firms in various countries have at least two blockholders while Laeven and Levine (2008) find that 34% (12%) of listed Western European firms have more than one (two) large owners where large owners are considered shareholders with more than a 10% stake. Even US firms which are often cited as examples of dispersed ownership have blockholders: 90% of all S&P500 firms have shareholders with more than 5% participation and more generally in US firms, 40% of the equity is owned by blockholders (Holderness, 2009).

decisions, or buying more shares, maybe holding a suboptimal portfolio, but influencing the firm's decisions. This trade off matters because the initial owner prefers less risky and lower return projects because of his large initial share while outside investors would ideally want the highest risk/return, creating an endogenous conflict of interest.[2]

Some investors then buy bigger blocks, so as to guarantee (via higher voting power) that the risk/return profile of the firm is higher. Paradoxically, of course, when they do buy a larger fraction of shares, their preferences move closer to those of the initial large shareholder! This is the main innovation of our paper: *since the conflicts of interest are endogenous, it is not trivial to show that having a larger size will be beneficial to outside investors*, since the large size itself reduces the conflict of interest between the initial owner and large outside investors.

Our main result is that for intermediate values of the initial owner's monitoring costs, *any* subgame perfect equilibrium features multiple blockholders. When monitoring costs are very high the initial owner is in control while when they are very low there is a dispersed ownership structure. In our model, monitoring costs proxy for the frictions that are responsible for the large stake of the founder of the firm. Hence what we show is that when conflicts of interest are endogenous to the size of the stakes, blockholder equilibria exist, but only partially mitigate the conflicts of interest.

The broader message of our paper is that intermediate sized blockholders emerge as a response to the way the firm aggregates preferences of shareholders and contributes to the ongoing discussion on the objective of the firm (see for example Tirole (2001) and Becht, Bolton, and Röell (2003)). Changing the voting rules, or the exact nature of the conflict of interest may offer different results on the size and number of blockholders. But it does not change the main message of the paper, which is essentially a positive one.

In the finance literature (a seminal contribution being Shleifer and Vishny (1986)), the only reason to hold a large block is to have incentives for monitoring the manager. In contrast, we isolate another reason for an initial blockholder to hold large blocks – control of the firm. In contrast to much of the literature (see Section (2)) on the existence of multiple blockholder ownership structures, the initial owner in our model only sets the price of shares: he cannot choose the ownership structure *directly*: this is important given the existence of resale markets in shares. Addressing both the issue of monitoring and control our model shows that an initial owner might prefer to retain more shares than the monitoring incentive requires in reaction to an anticipated

[2]The effects on the risk choices when a controlling shareholder is less diversified can be seen in the choices of the Swedish bank Skandinaviska Enskilda Banken (SEB). The controlling shareholder, the Wallenberg family, has a big part of its wealth invested in the bank. The bank's approach is to be prudent as "sometimes life can turn sour". For this reason it faced the financial crises with a lot of cash which helped it to perform better than its' peers and in general than the stock market (Economist, 2009). For a more systematic study see Faccio, Marchica, and Mura (2009)

blockholder ownership structure.

We get a number of comparative statics results from our model. In our model the fraction of liquidity shareholders who vote can be considered a proxy for minority protection. Depending on country's regulation or the firm's bylaws, investors are able and/or willing to vote. This might be due to a minimum stake required to vote or because of the costly information acquisition. Our model suggests that multiple blockholder ownership structures are more likely to be seen in economic systems where the small shareholders' voting participation is higher, i.e. where investors have more power in firm's decision.

We also predict a relationship between size of the firm and the ownership structure: while the size of firms which have a single large block is usually smaller, there is less of a correlation between ownership structure and size when there is more than one block. This is confirmed by the findings of Laeven and Levine (2008) on size and ownership structure. In small firms we see one main blockholder; while in large corporations complex ownership structures are more likely to be observed.

On the question of how ownership structure affects other variables, our model predicts that firms with multiple blockholders should be characterized by higher risk/return (of course assuming that the salient conflict of interests between shareholders is on the risk profile of the firm). Carlin and Mayer (2000; 2003) and Teodora (2009) confirm this prediction. Firms with more dispersed ownership tend to invest in higher risk projects, like R&D and skill intensive activities. Laeven and Levine (2008) find that firms with several blockholders have a higher Tobin's Q than firms with only one big shareholder. Value increasing role of multiple blockholders have been found in general have been found in many studies (Lehmann and Weigand (2000), Volpin (2002), (Faccio, Lang, and Young, 2001), Maury and Pajuste (2005), Gutierrez and Tribo (2004)).

Intriguingly, our model also predicts that ownership structure affects the IPO's underpricing. In particular to guarantee the subscription of the blockholders who are undiversified, the initial owners sets a low price. This allows the diversified investors to extract some rent.[3] This is confirmed by some empirical studies(Brennan and Franks (1997), Fernando, Krishnamurthy, and Spindt (2004), Goergen and Renneboog (2002) and Nagata and Rhee (2009)).

Finally, the paper contributes to the literature on voting. Typically, the models on voting do not endogenize the individual preferences, the price of votes and hence the voting power of an agent (Dhillon, 2005). When applying voting theories to corporate governance issues, the firm value (and hence share prices) and shareholders decision are closely related. Furthermore, an investor

[3]This result is similar to Stoughton and Zechner (1998) and DeMarzo and Urosevic (2006) who show that IPO underpricing can serve to ensure the participation of large investors who can monitor and hence be value enhancing. However, in their papers control considerations are absent and the role of multiple large investors is not analyzed.

can decide how many shares to buy and their voting decision changes depending on the block he chooses to buy. The price, being set by the initial owner, becomes an endogenous variable that affects and is affected by the existence of a second blockholder and the voting outcome.

The paper is organized as follows. Section 2 discusses the related literature, Section 3 outlines the model. In section 4 we describe the equilibrium concept. Section 5 solves the model and find the possible equilibria. In section 6 we derive the empirical implications of the model. Finally, section 7 concludes. All the proofs are in the Appendix.

2 Related Literature

There is a considerable body of related theoretical literature that focuses on the role of the largest blockholder as a tool to discipline the manager or to take value enhancing actions. (See Holmstrom and Tirole (1993), Admati, Pfleiderer, and Zechner (1994) Burkart, Gromb, and Panunzi (1997), Pagano and Röell (1998), Bolton and Von Thadden (1998), Maug (1998)). Noe (2002) and Edmans and Manso (2008) focus on the role of blockholders as a tool to discipline the manager through the threat of exit which would confer information to outside investors.

A second strand of this literature focuses on the sharing of private benefits of control: Zwiebel (1995) suggests that multiple blockholders emerge as the optimal response to a situation where there are divisible private benefits of control, Bennedsen and Wolfenzon (2000) show that the initial owner in a privately held firm has incentives to choose an ownership structure that commits him to more efficient decisions. Gomes and Novaes (2001) focus on a firm's decision when blockholders have veto power. Bloch and Hege (2001) argue that the existence of two blockholders reduces the appropriation of private benefits of control.

All these approaches essentially view the emergence of multiple blockholders either (i) as a direct choice of ownership structure by the initial owner whereby he can commit himself to higher firm value, (ii) focus on exogenous private benefits of control, and (iii) allow only equity as the method of financing.[4]

In contrast, our paper lets the initial owner choose the ownership structure only *indirectly* through pricing of shares when raising capital: this is equivalent to allowing the initial owner to choose the ownership structure directly and then allowing a resale market for shares. Hence, conditional on having chosen to raise capital at least partly through shares' issue as the method of financing the firm, the owner does not *choose* the ownership structure: it is not a commitment

[4]Demichelis and Ritzberger (2008) relax some of these assumptions. In their paper the conflicts of interest among investors arise because they are not only investors but also consumers. Hence, when they hold a small fraction of shares they prefer to make consumer favorable choices while blockholders are interested in maximizing firm value.

device chosen by the owner as in much of the literature cited above. Second, we explicitly allow strategic interactions between blockholders. This is an important consideration as free-riding between blockholders can be very important when we move from single to multiple blockholders. Thirdly, we endogenize the degree of private benefits as the risk/return choice. This allows us both to highlight how private benefits can differ depending on share participation and to avoid assumptions over the divisibility of private benefits of control. Finally our results are robust to allowing the initial owner to use debt instead of equity as a way of raising funds.

Our paper is also related to Admati, Pfleiderer, and Zechner (1994) and DeMarzo and Urosevic (2006) who analyze the trade off between risk sharing and monitoring. Our paper is not about this trade off, but rather on the aggregation problems inherent when there are conflicts of interest. These conflicts happen to be on risk sharing objectives of the firm.

3 The Model

The initial sole owner of a firm needs a minimum amount of capital, K, to finance a project. He has an endowment of 1 and chooses to invest a fraction w_E of his wealth in the project. w_E is unconstrained as the initial owner can borrow at the risk free rate. The remaining amount $K - w_E$ needs to be raised by issuing equity. We allow the initial owner to raise more capital than needed for the project, i.e. $I \geq K$ and use it to diversify risk, where I is the total amount of capital raised, including the owner's contribution w_E. In this case the extra capital is invested in the risk free asset, which is the only other asset in the economy.[5] Initially we assume that $I = K$; later in section 5.2 we show that this assumption is without loss of generality.

In exchange for capital $K - w_E$, the initial owner offers an aggregate fraction $1 - \alpha_E$ of future profits to the new investors, keeping α_E for himself. Moreover we assume 'one share-one vote' so that a higher α_E implies more control rights for the initial owner. Observe that separating the capital invested and the fraction of shares retained allows the initial owner to take two different decisions, the expected return on the investment and the control rights, so our conclusions would be suitably altered. When $\alpha_E = 0$, the initial owner does not hold any shares in the firm, i.e. he exits from the firm.

The set of (potential) outside investors is denoted M and the number of such investors is also (to save on notation) denoted M. We assume that the set M is partitioned into two types of investors – those who are *active* shareholders, denoted M_A, and those who are *passive* shareholders, such that $\frac{M_A}{M} = \lambda$. Active shareholders are assumed to vote anticipating that their vote is going to have

[5]We use the term 'diversify' loosely to capture the notion that an agent may benefit from reducing his exposure to the project risk.

an impact on the decisions of the firm.[6] Passive shareholders act competitively and take the firm's decisions as given, ignoring their own potential influence. Thus they do not vote.

Investors have initial wealth of 1 and they decide how to divide their wealth between the firm's shares and a risk free asset.[7] Each investor j chooses the fraction of his wealth, w_j, to invest in the firm. There are no financial constraints so both the initial owner and the investors can borrow money at the risk free rate or go short in the firm shares, i.e. w_j is not bounded. For the initial owner this implies that the decision is not only on the ownership structure but also the composition of his portfolio between debt and equity. The higher the debt, the higher is the risk exposure but the higher the control he has. The gross return of the risk free asset is normalized to 1.

Investors and initial owner have identical preferences represented by the following utility function:

$$u_j = -\frac{1}{\gamma}e^{-\gamma Y_j} \tag{1}$$

where $j = \{i, E\}$, i refers to the generic outside investor, E refers to the initial owner, γ is the parameter of risk aversion and $Y_j = f(w_j)$ is the final wealth when a fraction w_j of the wealth is invested in the project.

The timing of the game is as follows: in period 0 the initial owner decides w_E and the fraction of the shares to retain, α_E. This is equivalent to announcing the fraction of shares retained together with the share price, which is given by the capital raised over the cash flow rights tendered, $\frac{K-w_E}{1-\alpha_E}$.

In period 1 investors decide the fraction of shares of the firm, α_i to buy, having observed the share price. A passive investor chooses α_i which optimizes her portfolio taking the voting decision as given, i.e., disregarding the effect her own votes can have on the voting outcome. An active investor, on the other hand, anticipates the effect that her vote has on the voting outcome. Hence her demand for shares will internalize the potential effect of ownership structure on the operating decisions of the firm in period 2.

We assume that there are sufficiently many investors in the market so that there is never a problem of excess supply of shares. If there is under-subscription, then the project cannot go ahead. If there is oversubscription, we assume that this is a stable situation only when no investor who gets shares is willing to sell them at a price lower than the maximum price that an excluded

[6] Institutional investors such as hedge funds can be intepreted as active investors. It is well known that the value creation effect of hedge funds is highly significant both in the short and in the long run. For more evidence on the emergence of active investors see Smith (1996), Brav, Jiang, Partnoy, and Thomas (2008) and Becht, Franks, Mayer, and Rossi (2009).

[7] Since the motivation of this paper is to show that blockholders are sub-optimally diversified, we are interested in the *fraction* of their wealth that is invested in the firm. Introducing heterogeneity in wealth levels would not change the results unless attitudes to risk depend on wealth. In this case the identity of blockholders would be easier to predict: more wealthy investors who are relatively less risk averse are more likely to be blockholders.

investor is willing to pay.[8]

In period 2 shareholders – the initial owner and/or the investors who bought the shares – have to take a decision by voting. This decision is about the risk profile of the firm. For example, we may think of this as a decision about projects' type. The projects differ in their risk- return profile.[9] The project cash flows are affected by a variable X as follows: they are normally distributed with mean $\bar{R}X + f(m)$ and and standard deviation σX, where $f(m)$ is the extra expected cash flow from monitoring (which is chosen in period 3) and $\bar{R} > 0$ and $\sigma > 0$ are generic parameters. Shareholders thus choose a project profile $X \in [0, \bar{X}]$: the higher is X the higher is the risk and return to the firm. Hence, conflicts of interest between shareholders can be captured in a simple way through the uni-dimensional linear efficiency frontier of possible projects. We assume that $\bar{X} > \frac{\bar{R}}{\gamma\sigma^2}$. If $\bar{X}$ is too small, then we may be artificially removing any potential conflict of interest and the main driving force of our model.

At this stage, the ownership structure is common knowledge. The decision X is taken through majority voting. We show later that once the ownership structure is fixed, preferences on X are single peaked (Lemma 1) so that the Condorcet winner[10] in the space $[0, \bar{X}]$ is the median shareholders' choice. Voting is costless. However since only a subset of investors are assumed to vote the Condorcet winner must be chosen from among the ideal points X_j of the voting subset only.

In period 3 the initial owner decides whether to monitor or not. His choice variable is $m \in \{0, 1\}$. In particular we assume, as in Maug (1998), that if the initial owner monitors ($m = 1$), the expected firm value increases by $f(1) = K$ at a monetary cost of $c(1) = \bar{m}K$, with $\bar{m} \leq 1$.[11] If the initial owner does not monitor, $f(0) = 0$ and $c(0) = 0$. We assume that the cost of monitoring is not divisible among shareholders.[12]

Finally at period 4 the payoffs are realized. The time structure of the game is sketched in Fig.1.

Hence, in our model there is both an issue of incentives and of control. Ideally, the initial owner and the investors would like to have as few shares as possible and choose maximum return, even though it comes with high risk. However, the incentive problems associated with monitoring imply that the initial owner faces a trade off between increasing the value of the firm through monitoring

[8] We do not explicitly model the secondary market in shares but we capture some of the spirit of the secondary market by imposing this particular refinement of the Nash equilibrium concept.

[9] It could be thought of as a choice on managers who come with different reputations for risk taking.

[10] The Condorcet winner is the project that wins against every other project in pair-wise majority voting.

[11] If $\bar{m} > 1$ the initial owner would never choose to monitor.

[12] This assumption strengthens rather than weakens our results, as the main driving force as to why blockholders emerge in our model is not the sharing of monitoring costs, but rather the conflict of interest between the initial owner and other potential investors. If in addition, investors can share in monitoring there may be an additional reason for blockholders to emerge.

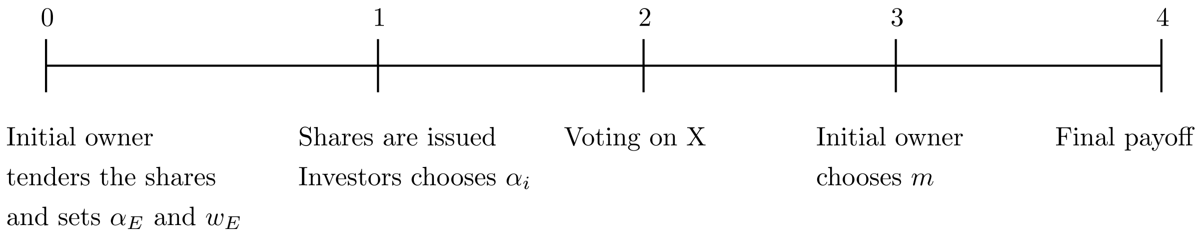


Figure 1: The Time Structure

and diversification. In the event that the initial owner resolves this trade off by holding a less diversified portfolio, his choice of X conditional on his shareholdings is likely to be biased towards a lower X. The first best choice for the investors, on the other hand, is to hold few shares, diversify optimally, and have a high return (high X). Hence there is a conflict of interest. Notice that outside investors are identical and hence have the same preferences: the only *ex-ante* conflict on risk- return is between the initial owner on the one hand and other investors on the other. *Ex-post* conflicts depend on equilibrium ownership structure: it is in this sense that conflicts of interest are endogenous in our model. The choice of how many shares to hold changes the preferences of shareholders and therefore also the conflict of interest.

As a result, active investors too face a trade-off: holding a sufficiently big block means higher control on the voting decisions, but means that diversification is less than optimal. When holding a block an investor incurs a cost that is increasing in the price and in the risk exposure.

Assume first that X is fixed. We first derive the utility function of the initial owner and outside investors. The certainty equivalent representation of the utility function (1) is:

$$U_j = E\left[Y(w_j)\right] - \frac{\gamma}{2} w_j^2 V\left[Y(w_j)\right] \tag{2}$$

where $E\left[Y(w_j)\right]$ and $V\left[Y(w_j)\right]$ are respectively the expected value and the variance of the final wealth.

Investors get a share of the firm value that is proportional to their contribution w_i, i.e. $\alpha_i = \frac{w_i}{K-w_E}(1-\alpha_E)$ (assuming full subscription). Hence the investors' certainty equivalent is:

$$\frac{w_i}{K-w_E}(1-\alpha_E)\left(\bar{R}X + f(m)\right) - \frac{\gamma}{2}\sigma^2 X^2 \left(\frac{w_i}{K-w_E}(1-\alpha_E)\right)^2 + (1-w_i)\cdot 1 \tag{3}$$

The last part of this expression represents the share of wealth invested in the risk free asset, i.e., the opportunity cost of investing in the firm. The first part represents the expected wealth from investing in the firm and the second part represents the dis-utility from the risk of investing in the firm.

Equation (3) can be re-written as:

$$U_i = \alpha_i \left(\bar{R}X + f(m)\right) - \frac{K - w_E}{1 - \alpha_E}\alpha_i - \frac{\gamma}{2}\sigma^2 X^2 \alpha_i^2 + 1 \tag{4}$$

The first part is the expected wealth from investing in the firm, the second part is the price paid and the third is the dis-utility from investing in a risky asset. Investor i will maximize (4) by choice of α_i, given $\alpha_E, K - w_E$ and the beliefs on $X, f(m)$.

Similarly, the initial owner's exponential utility function can be re-written in terms of certainty equivalent as:

$$U_E = \alpha_E(\bar{R}X + f(m)) - \frac{\gamma}{2}\sigma^2 X^2 \alpha_E^2 + 1 - w_E - c(m) \tag{5}$$

The initial owner chooses α_E, w_E in period 1 and m in period 3 to maximize (5) subject to the constraint that he needs to raise the capital, i.e. $K - w_E \leq \sum_i w_i$, or equivalently $1 - \alpha_E \leq \sum_{i=1}^N \alpha_i$ where $\sum_i \alpha_i$ is the sum of the total shares demanded. In addition he needs to satisfy $\alpha_E + \sum_i \alpha_i = 1$. Note that we allow $w_E < 0$ which means that the initial owner can sell shares in return for cash from the other shareholders in order to invest in the risk free asset. $w_i < 0$ on the other hand means that investors are going short in shares, although we show later that this never occurs in equilibrium. Finally when $w_j > 1$ with $j \in \{E, i\}$ it means that the investors borrow money (at the risk free rate) in order to invest in the firm.

We can now derive the ideal point $X_j(\alpha_j) \in [0, \bar{X}]$ for any investor j. We can then determine the payoff functions of players given the ownership structure defined as the vector of shares owned by investors: $\vec{\alpha} = (\alpha_E, \alpha_1, ..., \alpha_k)$ where k is the number of active investors who hold shares in the firm.

Lemma 1 *The preferred choice of X given α_j, for any shareholder $j \in \{i, E\}$, denoted X_j is uniquely defined by:*

$$X_j = \min\left[\frac{\bar{R}}{\gamma\sigma^2\alpha_j}, \bar{X}\right] \tag{6}$$

The choice of X depends only on the investor's shareholdings α_j and not on the decision to monitor. Observe that $\frac{\bar{R}}{\gamma\sigma^2\alpha_j}$ is a one-to one function of α_j. Hence, we can define $\bar{\alpha} \equiv \frac{\bar{R}}{\gamma\sigma^2\bar{X}}$ as the fraction of shares $\bar{\alpha}$ such that $X_j(\bar{\alpha}) = \bar{X}$.

It follows from Lemma 1 above that once $\vec{\alpha}$ is fixed, preferences of investors and the initial owner on X are single peaked. Hence the Condorcet Winner on the set $[0, \bar{X}]$ is well-defined and is given by the preferences of the median shareholder. Denote $X_{med}(\vec{\alpha})$ as the median X when the ownership structure is $\vec{\alpha}$. To save on notation, we suppress the argument $\vec{\alpha}$. For convenience we will denote the median shareholdings as α_{med}.[13]

Notice from equations (4)-(6) that α_E determines both the price paid, $\frac{K-w_E}{1-\alpha_E}$, and (potentially) X_{med} through the share ownership structure. Hence the indirect utility function for active investors depends on α_E, as well as the anticipated m and $\vec{\alpha}$ given α_E and $K-w_E$. Passive investors' indirect utility depends also on α_E, but here X is taken as given.

Pure strategies of the owner are 3-tuples $(w_E, \alpha_E, m(\alpha_E, X_{med}))$ together with a function from α_E to a voting decision over X. Pure strategies of investors are functions from $(\alpha_E, K - w_E)$ to a shareholding α_i and a voting decision over X.[14] This describes an extensive form game, Γ where the set of players are the initial owner and other active investors, the pure strategies and payoffs are as above.

4 Equilibrium definition

We define shareholders as *liquidity shareholders*, or *blockholders* depending on whether they hold an optimally diversified portfolio or not. Liquidity shareholders' shares are denoted by $\alpha_l(X_j)$,which represents the optimal portfolio when the voting outcome is assumed to be X_j. A shareholder is a *blockholder* if in equilibrium he holds a suboptimal portfolio: $\alpha_i > \alpha_l(X_j)$ where α_i is the shareholding of blockholder i.

Our notion of equilibrium is subgame perfect equilibrium of the game described in Fig. 1. Note that because in many potential equilibria more than one investor needs to buy shares for the initial owner to find it worthwhile to start the firm, there is always a No Trade equilibrium. In this equilibrium, no investor buys any shares anticipating that no other investor will buy shares. Below we provide a definition for equilibria with positive trade.

Equilibrium is a monitoring level, $m \in \{0, 1\}$, a fraction α_E^* of shares held by the initial owner, a fraction of wealth invested w_E^*, a decision X_{med} and an allocation of shares among investors, $\vec{\alpha}^*$ such that: (i) α_E^* and w_E^* maximize the utility of the initial owner given the anticipated demand, the anticipated monitoring level, m, and the anticipated ownership structure $\vec{\alpha}(K - w_E, \alpha_E)$. (ii)Each active investor chooses α_i to maximize her utility given $K - w_E^*, \alpha_E^*$, the anticipated m and the anticipated α_{-i}. (iii) Each passive investor chooses α_i to maximize her utility given $K - w_E^*, \alpha_E^*$ and the anticipated m and X_{med}. (iv) In equilibrium there must be full subscription. There can be

[13]Consider the frequency distribution of shares of *initial owner and active investors only* on the set X. The median X is the unique X_j such that exactly half the shares are on either side of it. Since it is common knowledge that passive investors never vote, α_{med} is defined only on the basis of shares of initial owner and active investors.

[14]Since there is pairwise voting, voting is assumed to be sincere and we rule out strategic agenda setting issues.

excess demand in equilibrium as long as no investor who owns shares is willing to sell them at a price lower than the maximum willingness to pay of the excluded investors. (v) The monitoring decision must be optimal for the initial owner given his stake and the vote outcome. (vi) Expectations are rational.

We will distinguish between monitoring and no monitoring equilibria as follows: (No) monitoring equilibria are those where the initial owner is assumed to choose (not) to monitor in the last stage, and anticipating that, he chooses the optimal α_E (and w_E) in period 0.

We are interested in equilibria which are characterized by a conflict of interest among shareholders: such equilibria are of two types: (1) the initial owner is the sole blockholder and is in control of the voting outcome, i.e. $X = X_E$. We call this the *Initial Owner Equilibrium.* (2) the initial owner and a subset of the active investors are blockholders. When there are $n+1$ blockholders, including the initial owner we call it an *n-Blockholder equilibrium.*[15] In case of blockholder equilibria we focus only on *symmetric* equilibria, where each blockholder (other than the initial owner) holds exactly the same shares. However we allow asymmetry among active investors on the decision to be a blockholder. When outside investors are in control holding a diversified portfolio we will call it a *Liquidity Shareholder Equilibrium.* When there is no conflict among any shareholders we call it a *No Conflicts Equilibrium.*

As we see later, the model admits multiple blockholder equilibria. Among these we restrict attention to those equilibria where the fraction of active investors among liquidity investors is fixed and known at $\lambda = \frac{M_A}{M}$. λ therefore captures the anticipated fraction of liquidity shareholders who take part in the voting decisions of the firm. In general we expect λ to be "small" reflecting the proportion of active investors in the market for shares of the firm (this is not needed for the model however).

5 Equilibria

We solve the game by backward induction. The last stage is the monitoring decision. The initial owner can commit to monitor only when he owns enough shares (as e.g. in Shleifer and Vishny (1986)). Observe that the decision to monitor is independent of the voting outcome:

Lemma 2 *The initial owner monitors iff $\alpha_E \geq \bar{m}$.*

The second last stage is the voting game. This is trivial given the vote shares. Each voter votes for his ideal point given his shares (see Lemma 6) and the outcome is the Condorcet winner.

[15] For expositional simplicity, we refer only to investors (and not the initial owner) as blockholders although, strictly speaking, the initial owner is the "first" blockholder.

Finally we come to what we call the *ownership subgame*. In this subgame, investors buy their shares given (α_E, w_E), anticipating the effects of their share ownership on the voting outcome and the monitoring outcome. We have seen that the monitoring outcome is independent of ownership structure except through α_E. We can therefore partition the subgames at this stage into those where $\alpha_E \geq \bar{m}$ and those where it is less. There is a continuum of such subgames. We now define the *Equilibrium Ownership Structure* (EOS) as the Nash equilibrium of the subgame for each pair $(\alpha_E, w_E) \in S \equiv [0,1] \times (-\infty, +\infty)$.

Let U_1^{nBH} denote the value function of a representative blockholder who we name blockholder 1 and who owns α_1 fraction of shares, in an ownership structure which admits n blockholders. Recall by the definition of a blockholder equilibrium that $X_{med} = X_1$. $U_{l,1}$ then denotes the value function of a liquidity shareholder in such an ownership structure.

Definition 1 *An Equilibrium Ownership Structure (EOS) corresponding to a pair* (α_E, w_E) *is an equilibrium of the subgame beginning at the information set* (α_E, w_E). *In particular the following must be satisfied in equilibrium: (1)* $U_1^{nBH} \geq 1$*(if there are any blockholders in equilibrium) and* $U_{l,j} \geq 1$ *(if there are any liquidity shareholders in equilibrium). We call this the Participation Constraint, (2) No active investor wants to unilaterally increase or decrease his shares. We call this the Incentive Constraint. (3) Passive investors maximize their utility conditional on the anticipated* $X = X_{med}$. *(4) No investor who receives shares is willing to sell them at a price lower than the maximum price that any excluded investor is willing to pay.*

The equilibrium concept is standard– that of Nash between active investors, and passive investors act to maximize their utility given their beliefs on X_j, which must be the right beliefs so $X_j = X_{med}$. Refinement (4) is imposed because in case of oversubscription, we would like a rationing rule that allocated shares in a "stable" way. In other words, we want to ensure that no excluded investor can do better by deviating unilaterally. This will turn out to affect the equilibrium price. Note that in our model, the initial owner does not directly determine the ownership structure in equilibrium, he can only influence it through the choice of share price. We believe that this structure is much more plausible: given that there is a secondary market for shares, in practice, the initial owner does not really determine ownership structure directly.

The ownership structure can be of four types based on who is the median shareholder (as discussed above in Section 4): (A) The *Initial Owner EOS*, where $X_{med} = X_E < \bar{X}$ and all investors are liquidity shareholders . (B) *Liquidity Shareholder EOS* where active liquidity investors (who hold a perfectly diversified portfolio) are in control of the firm, $X_{med} = \bar{X}$. (C) *No Conflicts EOS* where $X_{med} = X_E = \bar{X}$, hence there are no conflicts of interest between the initial owner and outside investors and all investors are liquidity shareholders. (D) An *n-Blockholder EOS* , where $X_{med} = X_1$ and n active investors hold a non-perfectly diversified portfolio. More formally:

Definition 2 *A symmetric n-Blockholder ownership structure is one where there are $n > 0$ active investors (blockholders) with shares α_1 each and $N_A \geq 0$ active investors (liquidity investors) with shares $\alpha_l(X_1)$ such that $\alpha_1 > \alpha_l(X_1) > 0$, and $X_{med} = X_1$.*

Given the investors' beliefs on $X_{med} = X_j$ we can determine the demand for shares by passive investors:

Lemma 3 *Let X_j be the belief on the voting outcome and $K - w_E$ the capital demanded. Then liquidity shareholders demand:*

$$\alpha_l(X_j) = \frac{X_j \bar{R} + f(m) - \frac{K - w_E}{1 - \alpha_E}}{\gamma X_j^2 \sigma^2} \tag{7}$$

Lemma 3 shows that the fraction of shares chosen by the passive investors depends on their beliefs on the voting outcome X_j. When they believe the voting outcome is going to be low risk and return, i.e. low X_j, the optimal portfolio they choose has a higher fraction of shares and vice versa. Observe that the value function of liquidity investors, $U_{l,j}$ is always bigger than 1 since there is no restriction on $\alpha_l(X_j)$: if the utility from holding the firms' shares is less than the risk free asset, then $\alpha_l(X_j)$ could be negative, i.e. investors can go short on the shares. We show later that the constraint on full subscription by the initial owner implies that in equilibrium $\alpha_l(X_j) > 0$.

The participation constraint of liquidity investors, $U_{l,j} \geq 1$, also implies that the maximum price they are willing to pay, $\frac{K - w_E}{1 - \alpha_E}$, is higher, the higher is the anticipated risk/return profile of the firm, i.e. the higher is X_j.

Lemma 4 *Liquidity shareholders are better off than blockholders regardless of the monitoring decision.*

This is true by definition: Liquidity investors hold the fraction of shares that provides optimal diversification given the anticipated voting outcome, X_j. Any other shareholding gives lower utility.

The next lemma shows that if the share price is sufficiently high,then the optimal liquidity shareholding is always smaller than $\alpha_{med} = \alpha_j$.[16]

Lemma 5 *Assume $\frac{K - w_E}{1 - \alpha_E} > f(m)$, then $\alpha_l(X_j) \leq \alpha_j$.*

[16] If, to the contrary, the share price is too low, so that $f(m) \geq \frac{K - w_E}{1 - \alpha_E}$, investors will demand all the shares tendered. We show later that this cannot be an equilibrium as the initial owner will always maximize the share price.

We now identify how many shares an investor buys when he decides to hold a block. Recall that λ denotes the proportion of active to total investors in the market.

Lemma 6 *Assume full subscription of shares. If an n-blockholder EOS exists with $\alpha_E > 0$ and $\frac{K-w_E}{1-\alpha_E} > f(m)$, each blockholder holds*

$$\alpha_1 = \frac{\alpha_E(1+\lambda)-\lambda}{(1-\lambda)\,n} \tag{8}$$

This lemma shows that blockholders buy just enough shares to be pivotal, kowing α_E and anticipating the choice of the other $n-1$ blockholders as well as the choice of active liquidity investors. They do not want to hold more shares because that would expose them to higher risk without gaining anything in terms of control. If they hold any less, they do not affect the voting outcome. From Lemmas (5) and (6), it follows that blockholders hold more shares than liquidity shareholders and less than the initial owner, $\alpha_l(X_j) < \alpha_j < \alpha_E$. Hence, X_1 will lie between the preferred choice of the liquidity shareholders and the initial owner. This is the sense in which blockholders mitigate the conflicts of interest between the initial owner and the liquidity shareholders.

We now define the *first best* choice for investor i as the optimal α_i assuming that in the second stage agent i acts as a dictator in the choice of X.

Lemma 7 *Assume $\frac{K-w_E}{1-\alpha_E} > f(m)$, the first best choice of the investors is $X = \bar{X}$ and $\alpha_i = \alpha_l(\bar{X})$.*

This lemma helps to clarify the conflict of interest between investors and the initial owner. Outside investors prefer to diversify maximally, i.e. buy a small amount of shares, $\alpha_l(\bar{X})$, and choose the project with maximum risk and returns, $\bar{X}$. Absent the monitoring constraint, there is no conflict of interest, the only incentive to is maximize monopoly rents while minimizing risk, and this is achieved by selling the firm (as we show later in Proposition 1). However the monitoring constraint forces him to have a different preference ex-post (after shares are allocated) than outside investors. Given the constraint on his shareholdings ($\alpha_E \geq \bar{m}$), he prefers lower return/risk projects while outside investors prefer higher risk/return.

Finally, we solve the first stage maximization for the initial owner given what he anticipates will happen for each (α_E, w_E). This depends on the EOS that is anticipated for each pair (α_E, w_E). These will be used in the derivations of the subgame perfect equilibria: the method of proof is to solve for the Initial Owner's maximization problem, assuming a common belief about the EOS following every information set (α_E, w_E). Since the analysis of the EOS is quite technical, it is done in the Appendix (Lemmas (11)- (13) and Corollaries (3)-(5)). Before moving to the subgame perfect equilibria we need the definitions of some symbols we are going to use for the rest of the paper. Define $\epsilon_j > 0$ as a very small number such that $\epsilon_j = \frac{\gamma X_j^2 \sigma^2}{\eta_j}$ where η_j is the fraction of shares corresponding to one share, when $X_{med} = X_j$. Define:

$$
\begin{aligned}
\underline{w}_E^E(\alpha_E) &\equiv K - \left(\frac{\bar{R}^2}{\gamma\sigma^2\alpha_E} + f(m)\right)(1-\alpha_E) + \epsilon_E &\quad (9)\\
\underline{w}_E^n(\alpha_E) &\equiv K - \left(\bar{R}X_1 + f(m) - \frac{\gamma}{2}X_1^2\sigma^2\alpha_1\right)(1-\alpha_E) &\quad (10)\\
\underline{w}_E^{LS}(\alpha_E) &\equiv K - \left(\bar{R}\bar{X} + f(m)\right)(1-\alpha_E) + \epsilon_{LS} &\quad (11)
\end{aligned}
$$

Suppose α_E is fixed. Then $\underline{w}_E^E(\alpha_E)$ is the minimum w_E that the initial owner needs to invest in the firm in order to guarantee that liquidity shareholders' participation constraint is satisfied when $X_j < \bar{X}$. However at this level $w_E = \underline{w}_E^E$ the demand for shares by liquidity shareholders is zero. We need $\epsilon_E > 0$ to have a strictly positive demand by the liquidity shareholders. $\underline{w}_E^n(\alpha_E)$ on the other hand, is the analogous expression if n blockholders participate. $\underline{w}_E^{LS}(\alpha_E)$ is the analogous expression to $\underline{w}_E^E$ when $X_j = \bar{X}$: it is the minimum w_E that the initial owner needs to invest in the firm to satisfy the participation constraint of liquidity shareholders (and for them to have a strictly positive demand) when $X_j = \bar{X}$.

The first such equilibrium we analyze is the no-monitoring equilibrium: Suppose the initial owner decides not to monitor in the last stage (i.e. $\alpha_E < \bar{m}$). We show that then he always chooses $\alpha_E = 0$ and the unique ownership structure that emerges is a Liquidity Shareholder Ownership Structure: the initial owner either does not invest in the firm or sells it letting the liquidity shareholders be in control of a non-monitored firm. Our interest in the no monitoring payoff stems from the fact that in any monitoring equilibrium, the owner would have to get a higher payoff from monitoring in equilibrium than non-monitoring.

Proposition 1 *Suppose the initial owner chooses not to monitor in period 3, in period 0 he sells the firm when the firm has a positive NPV, otherwise he does not raise the capital and invests in the risk free asset (in both cases, $\alpha_E = 0 < \bar{m}$). The initial owner's value function is given by:*

$$V_E^{NM} = \max(\bar{R}\bar{X} - K + 1, 1) \quad (12)$$

If the initial owner sells the firm, he sets $w_E = \underline{w}_E^{LS}$ and the unique EOS is a Liquidity Shareholder ownership structure, $X = \bar{X}$.

This proposition illustrates the trade-offs faced by the initial owner. Since he acts as a monopolist in the pricing of shares, when he has no constraint on his shareholdings, he can extract the full value of the firm without incurring any risk, by simply selling the firm. However, investors are willing to buy shares only if the expected NPV is positive ($\alpha_E = 0$, $w_E < 0$).[17] Otherwise he prefers not to raise capital ($w_E = \alpha_E = 0$) .

Proposition 1 shows that the choice of $\alpha_E > 0$, $m = 0$ by the initial owner is dominated by the choice of either selling the firm or not raising capital. Hence, in what follows it is sufficient to show that the participation constraint and the non selling constraints are satisfied, to ensure that the initial owner prefers $\alpha_E \geq \bar{m}$, $m = 1$ to any non-monitoring equilibrium.

5.1 Monitoring Equilibria

The existence of blockholder equilibria depend on three necessary conditions (1) a conflict of interest between investors that is generated endogenously by the fact that if they have different shares in the firm, they have different preferences on the risk/return profile of the firm; (2) they are able to influence the voting decision if they strategically buy more shares than a liquidity shareholder would; (3) in equilibrium the initial owner's shareholding is large enough that active liquidity investors in the firm cannot jointly ensure that $X_{med} = \bar{X}$, without holding more than the liquidity shares. If any of these three requirements is not met, then we do not have a blockholder equilibrium.

Section 5.1.4 shows our main result, that when the monitoring costs are intermediate (so the conflict of interest is not too big and not too small and all the conditions (1)-(3) are satisfied) blockholders emerge endogenously and help to mitigate the conflict of interests between the initial owner and the liquidity shareholders. For completeness and ease of exposition however we first analyze the other equilibria.

Hence, Section 5.1.1 analyzes the case where requirement (1) is not satisfied that is where monitoring costs are so low ($\bar{m} \leq \bar{\alpha}$) that the final choice is $X_{med} = X_E = \bar{X}$. Since this is the first best point for outside investors (see Lemma 7), this implies that there is no conflict of interest: hence no blockholders in the ownership structure. Section 5.1.2, on the other hand considers the case where requirement (2) is not satisfied, i.e. where the blockholders cannot influence the voting decision because in equilibrium the initial owner holds a too large fraction of shares. Section 5.1.3 discusses when requirement (3) does not apply. In such a case there is a Liquidity Shareholder equilibrium where liquidity shareholders are in control because the initial owner holds such few shares that the active liquidity shareholders can exert control on the vote outcome without holding a block.

In all of the proofs, we do not allow the initial owner to raise more capital than the minimum needed for the firm i.e. K. So in Section (5.2) we show that this assumption is without loss of

[17] We may interpret $w_E < 0$ as rent for the initial owner for the entrepreneurial idea.

generality.

Before moving to the equilibria we show that investors are willing to receive less shares than what they proportionally contribute. Conditional on monitoring, investing in the firm increases the utility of the investors because it widens the possible portfolios they can choose among. Since the initial owner is a monopolist he can push share prices up to the point where the the participation constraint of investors is satisfied with equality.

Put another way, the initial owner contributes to capital proportionally less than what he receives in cash flow rights: $\alpha_E > \frac{w_E}{K}$ when $m = 1$. This is shown in the next Lemma:

Lemma 8 *Assume that $m = 1$. In any subgame perfect equilibrium, the price per share $\frac{K - w_E}{1 - \alpha_E} > K$.*

In all monitoring equilibria, the initial owner sets the price per share as high as possible to avoid dilution of his shareholdings. If the initial owner monitors, the expected firm value is above the return on the risk-free asset. Hence, the minimum possible price that guarantees the participation of the investors is above the price of the risk-free asset.

5.1.1 No Conflicts Equilibrium

Proposition 2 *Suppose*

$$\bar{m} \in \left[0, \min\left[\bar{\alpha}, \bar{m}_{NC}^{RC}(K), \bar{m}_{NC}^{S}(K)\right]\right] \tag{13}$$

then there exists a No Conflicts (NC) equilibrium where the initial owner monitors, $m = 1$, $\alpha_E = \bar{m}$, $X_{med} = X_E = \bar{X}$ and $w_E = \underline{w}_E^{LS}(K)$.
(The exact values of $\bar{m}_{NC}^{RC}$ and $\bar{m}_{NC}^{S}$ as functions of the parameters are given in the proof in the appendix.)

In this equilibrium, $\bar{m}$ is so low ($\bar{m} \leq \bar{\alpha}$) that the Initial Owner's first best is the same as other investors ($X_E = \bar{X}$), so that there is no conflict of interest. The initial owner retains just enough shares to have the incentive to monitor, i.e. $\alpha_E = \bar{m}$. He does not want to retain more as this would imply a higher risk exposure without any gain in terms of monitoring: moreover he can extract all rents from liquidity shareholders using his position as a monopolist when $\alpha_E = \bar{m}$. Hence, there are no conflicts of interest, issues of control are absent and there is no incentive for any investor to hold a block.

Even when he holds a non-perfectly diversified portfolio because of the monitoring costs, the initial owner would still vote for the maximum risk/return because the "friction" in the model is low. However observe that the No Conflicts equilibrium does not imply that the initial owner has the same shareholdings as the liquidity investors. His shareholdings are determined by the monitoring costs, $\bar{m}$, and depending on the monitoring costs, the initial owner can hold more or less than the liquidity shareholders. In the extreme case when the monitoring costs are 0, he would like to sell the firm. In this way he extracts all the rent from the investment without incurring any risk. The liquidity shareholders instead would always like to hold some shares so as to take advantage of the return premium from holding some risk.

Liquidity shareholders are willing to buy the shares only if the returns are high enough to compensate for the risk, $w_E \geq \underline{w}_E^{LS}$. If the price were higher investors are better off just investing all their wealth in the risk-free asset. At the same time the initial owner sets the highest possible price in order to reduce dilution and risk exposure. Hence $w_E = \underline{w}_E^{LS}$.

The monitoring requirement sets a maximum threshold on the shares that can be distributed to the liquidity shareholders. Given this threshold there is a maximum fraction of wealth that liquidity shareholders are willing to invest. The rest of the capital (if needed) must be pledged by the initial owner ($w_E = \underline{w}_E^{LS}$). The higher the amount of capital the initial owner needs for the project, K, the higher the wealth he needs to pledge.

Note that given that there are no financial constraints, the initial owner could finance the project completely on his own through borrowing money. However, he prefers to rely on outside equity rather than issuing debt in order to limit his risk exposure.

If the value created is very high the initial owner does not need to invest *any* money; indeed, he can be compensated by the investors for the monitoring exerted and the entrepreneurial idea ($\underline{w}_E^{LS} < 0$). These characteristics of the capital invested by the initial owner are common to all equilibria we find.

Figure 2 offers a graphic representation of the No Conflicts Equilibrium. The equilibrium exists when the area in the graph that satisfies both the no selling and raising capital constraints is positive. Since $\bar{m}_{NC}^{RC}, \bar{m}_{NC}^{S}, \bar{\alpha} > 0$, we can conclude that this is the case.

Alternatively the initial owner could choose not to raise the capital or to sell the firm. To be a viable project for the initial owner, the value remaining to the initial owner after compensating the investors, must be high enough to compensate him for the money invested and the risk (i.e. $\bar{m} \leq \bar{m}_{NC}^{RC}$).

At the same time to be willing to remain a shareholder of the firm, rather than sell it outright, the value created by monitoring ($f(m) = K$) has to be high enough. The extra value due to monitoring compensates the initial owner for the direct cost of monitoring as well as the indirect costs related to holding a sub-optimal portfolio. If the extra utility created by monitoring can be achieved by a dispersed ownership structure without monitoring then he would prefer to sell the firm (if $\bar{m} \geq \bar{m}_{NC}^{S}$ and $\bar{R}\bar{X} - K \geq 0$).

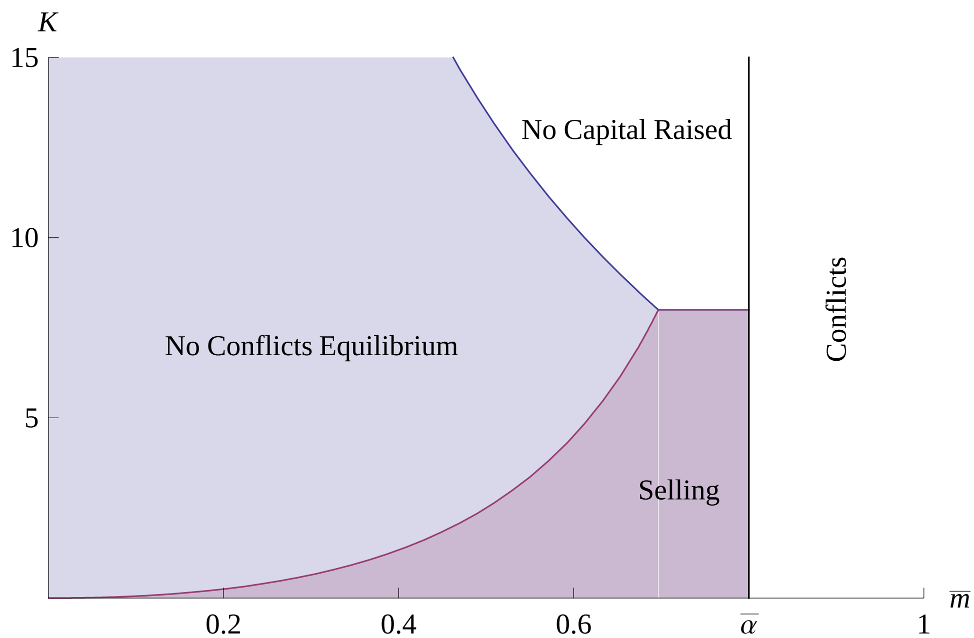


Figure 2: No Conflicts Equilibrium ($\bar{R} = 0.8$, $\gamma = 10$, $\sigma = 0.1$, $\bar{X} = 10$).

5.1.2 Initial Owner Equilibrium

The second equilibrium we consider is the one where requirement (2) is partly relaxed, i.e. outside investors *cannot* or *do not want* to influence the voting decision and there will be no blockholders apart from the initial owner. We obtain the Initial Owner equilibrium in two cases. First, when the initial owner has more than 50% of the shares and hence no outside investor can influence the decision: in this case the Initial Owner equilibrium is unique. The second case occurs when the initial owner holds less than 50% but α_E is high enough such that 1 blockholder would not have a unilateral incentive to deviate from holding liquidity shares.

Proposition 3 *Suppose that*

$$\bar{m} \in \left(\max\left[\bar{\alpha}, \min\left[\frac{1}{2}, \max\left[\hat{\alpha}(1), \tilde{\alpha}(1)\right]\right]\right], \min\left[\bar{m}_E^{RC}(K), \bar{m}_E^S(K), 1\right]\right] \tag{14}$$

then there exists an Initial Owner equilibrium where the initial owner is the only blockholder, $m = 1$*,* $\alpha_E = \bar{m}$*,* $X_{med} = X_E = \frac{\bar{R}}{\gamma\sigma^2\bar{m}}$ *and* $w_E = \underline{w}_E^E(K)$*.*

When $\bar{m} \in \left(\max\left[\bar{\alpha}, \frac{1}{2}\right], \min\left[\bar{m}_E^{RC}, \bar{m}_E^S, 1\right]\right]$ *the Initial Owner equilibrium is unique.*

(The exact values of $\bar{m}_E^{RC}$ *and* $\bar{m}_E^S$ *as functions of the parameters are given in the proof in the appendix.)*

If there is a wide range of projects such that there is a conflict of interests between investors and initial owner (i.e. $\bar{X}$ is sufficiently high) and the monitoring technology is sufficiently productive,

$\bar{m}$ small relative to $f(m)$, such an equilibrium always exists, i.e interval (14) is not empty.

Consider first the case, when the monitoring costs are very high ($\bar{m} \geq \frac{1}{2}$), such that the initial owner is willing to monitor only if he holds more than 50% of the shares. This implies that he is highly exposed to firm risk and because he is in control of the vote outcome he chooses a low risk/return project. Hence there is a conflict of interest between the initial owner and outside investors on the choice of X. No other blockholders emerge because under these conditions control is concentrated in the hands of the initial owner in all circumstances and there is no reason to hold a suboptimally diversified portfolio. So this equilibrium exists if $\bar{m} \geq \frac{1}{2}$ and participation constraints for liquidity shareholders are satisfied ($w_E \geq \underline{w}_E^E$) and as long as it is worthwhile to monitor rather than not (i.e. $\bar{m} \leq \min[\bar{m}_E^{RC}, \bar{m}_E^S]$).

The condition $\bar{m} > \hat{\alpha}(1)$ ensures that then an investor is not willing to unilaterally hold more shares in order to influence the voting decision, given $w_E = \underline{w}_E^E$. If the condition $\bar{m} > \tilde{\alpha}(1)$ is not satisfied then an active investor would be willing to pay a higher price than the equilibrium price given by $(\bar{m}, \underline{w}_E^E)$. Hence a passive liquidity investor, who never votes, would be willing to sell the shares to an active investor who would still hold a perfectly diversified portfolio but could become pivotal simply by going to vote, therefore destroying the Initial Owner equilibrium (if there are sufficiently many such passive investors). Finally, in order to guarantee conflicts of interests between investors and initial owner, we set $\bar{m} > \bar{\alpha}$.

Analogously to the No Conflicts equilibrium, liquidity shareholders are willing to buy shares if the price is low enough to compensate them for the risk they bear. This implies that the amount of capital that they contribute (the demand for shares) depends on their belief on the voting outcome and hence in the risk profile of the firm. This amount of capital can be greater or smaller than the capital needed. The initial owner anticipates this given $X = X_E$ and chooses w_E to maximize his utility. Therefore he may choose $\underline{w}_E^E > 0$, or $\underline{w}_E^E < 0$, and cash in the rest as compensation for the entrepreneurial idea.

Monitoring not only implies a direct cost of $\bar{m}$, but it also has an indirect cost for the initial owner in terms of holding a sub-optimally diversified portfolio. Hence the initial owner is willing to monitor and hold a large fraction of shares only if he is able to increase the firm value sufficiently to compensate both for the monitoring cost and the risk of holding a suboptimal portfolio. When the marginal returns on firm value from monitoring are not very high the initial owner prefers not to be involved in the project. In this case as Proposition 1 stated, there are two possible outcomes. If the extra value created by monitoring is not very high and the firm is very valuable even without monitoring (high $\bar{X}$), i.e. $\bar{m} > \bar{m}_E^S$, the initial owner sells the firm to the investors. In this case the firm will show dispersed ownership (a Liquidity Shareholder equilibrium). Instead, when the value

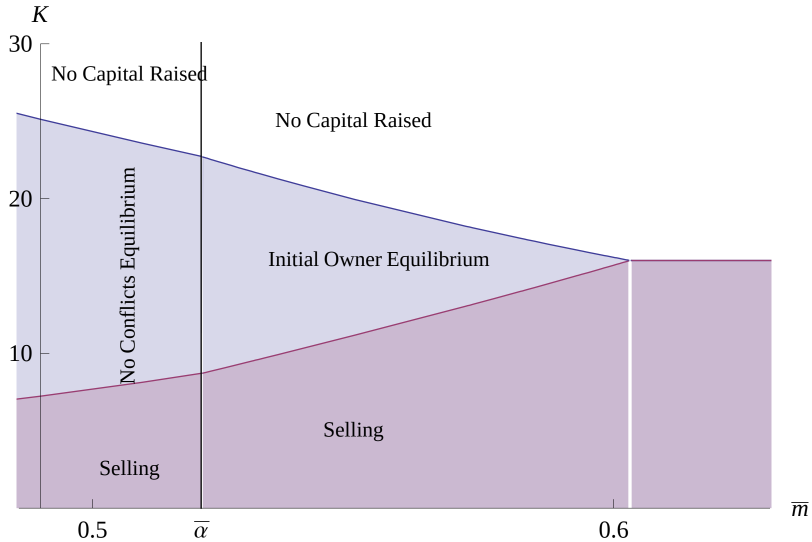


Figure 3: Initial Owner Equilibrium ($\bar{R} = 2$, $\gamma = 12$, $\sigma = 0.2$, $\bar{X} = 10$, $\lambda = 0.05$).

created by the project is not high enough (with or without monitoring) to compensate the initial owner for the capital invested $\bar{m} > \bar{m}_E^{RC}$, the initial owner does not raise the capital and invests all his wealth in the risk free asset. The possible outcomes are shown in Figure 3.

A less obvious result is that an Initial Owner equilibrium also exists if the monitoring costs are *smaller* than $\frac{1}{2}$, although in this case it is not unique. There may be blockholder equilibria as well, but we cannot rule out an Initial Owner equilibrium. To ensure existence of an Initial Owner equilibrium when $\bar{m} < \frac{1}{2}$, we need conditions such that no single investor has a unilateral incentive to deviate to become a blockholder.[18] This case is discussed in Section (5.1.4)–see Corollary (1).

5.1.3 Liquidity Shareholders equilibrium

We now consider equilibria where the liquidity shareholders are in control and there are no blockholders. Intuitively this happens when there are sufficiently many active liquidity shareholders and $\bar{m}$ is not too high. Recall that N_A is the number of active investors in the equilibrium that satisfy $\lambda(1 - \alpha_E) = N_A \bar{\alpha}_l$.

Proposition 4 *Suppose*

$$\bar{m} \in \left(\bar{\alpha}, \min\left[\frac{1}{2}, \tilde{\alpha}(n), \bar{m}_1^{NC}(K), \bar{m}_2^{NC}(K)\right]\right] \tag{15}$$

then there exists a Liquidity Shareholders equilibrium with $N_A + n$ active investors where $m = 1$, $\alpha_E = \bar{m}$, $X_{med} = \bar{X}$ and $w_E = \underline{w}_E^{LS})(K)$.

[18] However it is possible to have blockholder equilibria if $n \geq 2$.

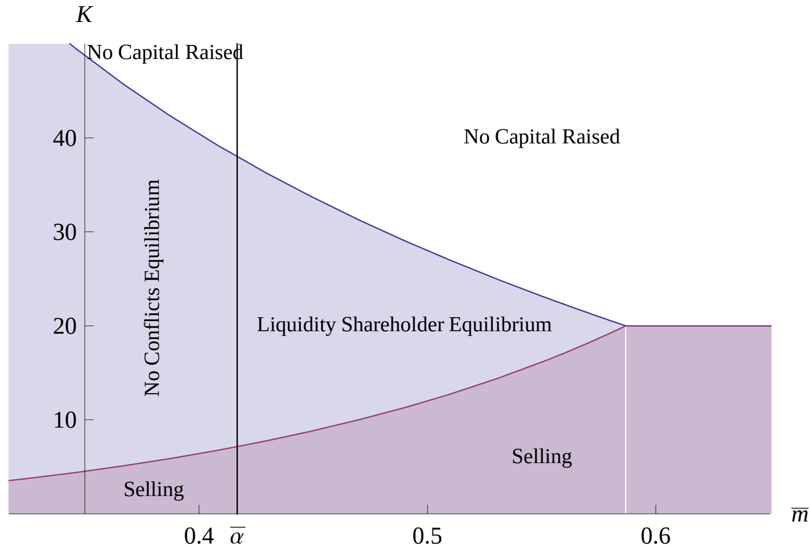


Figure 4: Liquidity Shareholder Equilibrium ($\bar{R} = 2$, $\gamma = 12$, $\sigma = 0.2$, $\bar{X} = 10$, $\lambda = 0.05$).

(The exact values of $m_1^{NC}(K), \bar{m}_2^{NC}(K)$ as functions of the parameters are given in the proof in the appendix.)

A Liquidity Shareholder equilibrium exists when the monitoring costs are relatively low, but not low enough that there are no conflicts (i.e. $\bar{m} > \bar{\alpha}$). In this equilibrium, the initial owner finds it optimal to choose $\alpha_E = \bar{m}$, sufficiently small so that there are enough active liquidity shareholders to be pivotal. In particular we can have two cases: one where the fraction λ of the liquidity shareholders is sufficient to change the vote outcome (i.e. $n = 0$) and the other where there are n active investors in addition to the fraction λ (i.e. N_A) active liquidity investors who vote and ensure that the outcome is $\bar{X}$. If monitoring costs are higher ($\bar{m} > \tilde{\alpha}(n)$), then there is an n blockholder EOS and if $\bar{m} > \frac{1}{2}$ then there is an Initial Owner EOS. Finally, as in the case of the No Conflicts and the Initial Owner equilibrium, when the monitoring costs are higher $\bar{m}_1^{NC}$ than the initial owner prefers to raise no capital and if monitoring costs are higher than $\bar{m}_2^{NC}$, he prefers to sell the firm.

As in the other equilibria, the participation constraint of liquidity shareholders is satisfied (when $X = \bar{X}$), at $w_E = \underline{w}_E^{LS}(K)$.

Notice, that the initial owner could always choose to retain a strictly higher fraction of shares ($\alpha_E > \bar{m}$) to induce a blockholder EOS or even to hold α_E bigger than half, to induce an Initial Owner EOS. In either of these cases, the vote outcome is closer to his own preferred point. However the higher control comes at the expense of both a lower price paid by the investors, more dilution and a less diversified portfolio for the initial owner.

5.1.4 Blockholder equilibria

This section shows our main result: the existence of blockholder equilibria for intermediate monitoring costs. In the following, we discuss two different cases of blockholder equilibria. For one range of parameters, it is possible to have both Initial Owner and n-Blockholder equilibria (Proposition 5). However, we have a stronger result: for some parameter values of monitoring costs, there exist *only* blockholder equilibria. So, even though there are multiple equilibria in terms of the number of blockholders, we know that there will be no non-blockholder equilibria (Corollary 2).

Proposition 5 *Suppose:*

$$\bar{m} \in \left(\max\left[\bar{\alpha}, \tilde{\alpha}(n), \bar{m}_1^E(n,\lambda), \bar{m}_{1,n}^{RC}(n,K), \bar{m}_{1,n}^S(n,K)\right], \min\left[\frac{1}{2}, \hat{\alpha}(n), \bar{m}_2^E(n,\lambda), \bar{m}_{2,n}^{RC}(n,K), \bar{m}_{2,n}^S(n,K)\right]\right] \tag{16}$$

then there exists an n Blockholder equilibrium ($n \in [1, M_A]$) where the initial owner monitors, $m = 1$, $\alpha_E = \bar{m}$, $X_{med} = X_1$ and $w_E = \underline{w}_E^n$.
($\bar{m}_{1,n}^{RC}(n), \bar{m}_{2,n}^{RC}(n), \bar{m}_{1,n}^S(n), \bar{m}_{2,n}^S(n), \bar{m}_1^E(n), \bar{m}_2^E(n)$ are given in the proof in the appendix.)

Proposition 5 demonstrates our main result: i.e there exist blockholder equilibria where n investors prefer to hold a large block of shares (and a sub-optimally diversified portfolio) in order to shift the decision to a higher level of risk/return. This mitigates the conflicts of interests between the initial owner and the investors ($\bar{X} > X_1 > X_E$). There can be multiple EOS, involving different n depending on the beliefs on the EOS.

Now we explain where we get the conditions (16) from. The initial owner could choose to set an $\alpha_E > \frac{1}{2}$ to guarantee himself control and end up in a Initial Owner EOS. Setting $\bar{m} \in [\bar{m}_{1,n}^E, \bar{m}_{2,n}^E]$ guarantees that the initial owner prefers to be in a n-Blockholder equilibrium rather than in an Initial Owner one (he gets a higher price though he gives up control). If the initial owner were to choose a different α_E he may end with a different EOS. Thus, e.g. if $\alpha_E \geq \frac{1}{2} > \bar{m}$ then a dispersed ownership structure is the unique EOS (assuming appropriate prices to guarantee participation). Any α_E between $\bar{m}$ and $\frac{1}{2}$ does not give him control and only dilutes the shares (since we can assume that the EOS following such a choice has at least n blockholders). Hence it is never chosen.

Following from Propositions (2) and (4), if $\bar{m} < \bar{\alpha}$ there are no conflicts and no blockholders; if $\bar{m} \leq \tilde{\alpha}(n)$ then there would be an incentive for passive shareholders to sell shares to active liquidity investors, possibly leading to a dispersed ownership structure with $X_{med} = \bar{X}$ (this condition affects the equilibrium price).

To guarantee that the initial owner prefers to monitor rather than not with this EOS we have the condition $\bar{m} \in [max\left(\bar{m}_{1,n}^{RC}(n), \bar{m}_{1,n}^S(n)\right), min\left(\bar{m}_{2,n}^{RC}(n), \bar{m}_{2,n}^S(n)\right)]$ (See Fig. 5). If the monitoring costs are too high the residual value for the initial owner after compensating the blockholder for the extra risk they bear due to the undiversified portfolio, is not high enough to compensate for holding an undiversified portfolio and the monitoring costs. When the monitoring costs are too low, there are blockholder equilibria with small blocks or liquidity shareholder equilibria where there are enough active investors to determine X. In such a case the risk/return outcome is very high compared to the initial owner's preferred point. Then the risk the initial owner bears is so high and the gain from monitoring is so low that he prefers to sell the firm or not to raise capital.

What if the initial owner were to choose a different w_E? The condition $\alpha_E < \hat{\alpha}(n)$ guarantees that whenever the participation constraint of liquidity shareholders is satisfied, so is that of blockholders. Hence the initial owner cannot prevent the entry of blockholders through setting a lower price – he only loses rent. Hence he sets the highest $w_E = \underline{w}_E^n$ that guarantees participation by investors. Increasing the price, (below $\underline{w}_E^n$) drives out participation (see Lemma 10) and so the initial owner would not be able to raise capital. Finally, $\bar{m}$ must be be smaller than $\hat{\alpha}(n)$ so that no blockholder in a Nash equilibrium with $n-1$ other blockholders has a unilateral incentive to deviate: recall that in a blockholder equilibrium, each blockholder is pivotal. Therefore if a single blockholder reduces his shareholdings, the voting outcome in the next period would be X_E. The incentive compatibility condition that the utility of an investor is higher when he is a blockholder with decision X_1 than a liquidity shareholder with decision $X_E < X_1$, holds if $\alpha_E \leq \hat{\alpha}(n)$.

The reader may find it puzzling that we start with ex-ante identical outside (active) investors, yet only some of them decide to become blockholders.[19] This is because there are multiple Nash equilibria for every n, and the identity of the blockholders could be different in each of these equilibria and the rest of the investors (liquidity shareholders) free ride on these. Like in a (discrete) public goods provision problem, the blockholders contribute to the public good provision (i.e. moving the decision on the project closer to the most preferred point of all outside investors) because given the other shareholders contributions, it is a Nash equilibrium for them to contribute as long as the value of the public good to them is sufficiently high.[20] This translates into the condition that α_E is sufficiently low: as α_E decreases, the incentives to hold larger blocks increases. This is because, in the first place, as α_E decreases, fewer shares are needed in order to gain control over X and hence the cost of holding a block is lower. Second, because of the convexity of X_j

[19]It may be the case that some large institutional investors are *less* risk averse relative to other investors. This destroys the ex-ante symmetry between outside investors, but this actually helps in our model, in the sense that there are some investors who are believed to be natural blockholders giving us a nice focal point equilibrium.

[20]This is also the case in the Nash demand game or the Divide the Dollar game : each of two players announces the share he demands of an amount of money that may be split between them. If the demands can be satisfied, they are. Otherwise neither player receives any money. This game has multiple pure strategy Nash equilibria: any demands that add upto the feasible amount are Nash equilibria.

with respect to α_j (equation (6)), it follows that the smaller α_E is, the larger is the shift in X (for the same $\alpha_E - \alpha_1$), i.e. $X_E - X_1$ is greater. This implies a higher increase in the expected return of becoming a blockholder. Hence there exists a threshold, $\hat{\alpha}(n)$, such that when $\alpha_E \leq \hat{\alpha}(n)$ and $w_E = \underline{w}_E^n$, the utility of being a blockholder is higher than being a liquidity shareholder with the initial owner in control.

Unlike the usual public goods contribution game, however, when blockholders buy a larger block of shares, their preferences over X are closer to those of the initial owner. This is why the presence of blockholders mitigates, but does not remove, the conflict of interest between the initial owner and the outside investors.

Lemma (4) implies that in the blockholder equilibrium, liquidity shareholders free ride on blockholders as they hold the optimal portfolio. The price which satisfies the participation constraints of the blockholders is lower than the maximum price that liquidity shareholders are willing to pay. Therefore the existence of blockholders allows the liquidity shareholders to extract some of the rent, that in all other equilibria goes entirely to the initial owner. In all the other equilibria the initial owner sets the price low enough to satisfy the participation constraint of the liquidity shareholders with equality, and hence he extracts all the rent. As the initial owner is not able to extract all the rent, when switching from an Initial Owner to an n Blockholder equilibrium there is a jump in the level of utility of the initial owner. This discontinuity can be seen in Fig. 5 at $\hat{\alpha}$ which represents the switching point between the Initial Owner to 1 Blockholder equilibrium as the monitoring costs decrease. Because the initial owner's utility decreases at $\hat{\alpha}$, the possibility to sell or not raise money becomes more attractive.

The next corollary demonstrates the existence of an Initial Owner equilibrium where the initial owner retains full control ($\alpha_E = \frac{1}{2} + \eta_E$) in order to avoid an n-Blockholder equilibrium.

Corollary 1 *Suppose*

$$\bar{m} \in \left(\max\left[\bar{\alpha}, \tilde{\alpha}(n), \right], \min\left[\bar{m}_E^{RC}(K), \bar{m}_E^{S}(K)\right]\right) \tag{17}$$

and either $\bar{m} \leq \bar{m}_{1,n}^E(n,\lambda)$*, or* $\bar{m} > \bar{m}_{2,n}^E(n,\lambda)$ *then an Initial Owner Equilibrium exists and is unique where* $m = 1$*,* $\alpha_E = \frac{1}{2} + \eta_E$*,* $X_{med} = X_E$ *and* $w_E = \underline{w}_E^E$*.*

The initial owner is willing to hold $\alpha_E > \frac{1}{2} > \bar{m}$ when the monitoring costs are either very high, $\bar{m} > \bar{m}_{2,n}^E$ or very low $\bar{m} < \bar{m}_{1,n}^E$. The intuition behind this result is that the initial owner faces a trade off between two types of costs: the cost of holding a very risky asset (this cost increases the bigger the gap between α_E and α_1) if he does not control the vote outcome, and the cost from being

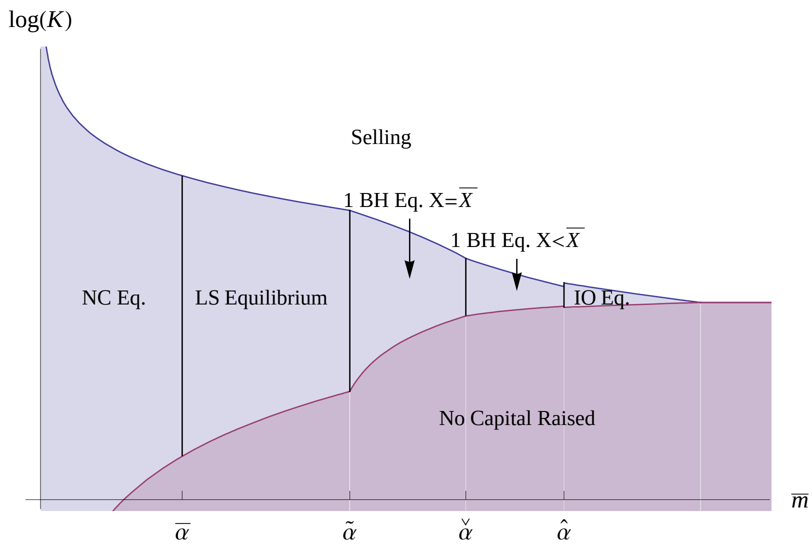


Figure 5: 1 Blockholder equilibrium ($\bar{R} = 1$, $\gamma = 12$, $\sigma = 0.2$, $\bar{X} = 50$, $\lambda = 0.1$). Note that in the graph there is not the boundary $\underline{\alpha}$ and $\bar{m}_2^{1E}$ as they are not binding and far out from the plot range considered.

relatively undiversified (this cost increases with α_E). When the monitoring costs are very high, $\bar{m} > \bar{m}^E_{2,n}$, the cost of increasing α_E a little bit and reducing diversification is very small relative to the gain from controlling the vote outcome and choosing his preferred risk/return combination. Hence he chooses to retain more shares than the monitoring incentive requires and retains control. When the monitoring costs are very low, on the other hand, blockholders face a very low cost in terms of diversification to be in control as now the vote of the liquidity shareholders has relatively more weight. Hence, the initial owner faces a large cost arising from the conflict of interest on the risk/return vote outcome. For this reason he prefers to increase α_E to a point where he can control the vote outcome, $\alpha_E > \frac{1}{2}$. The other conditions have the same rationale as Proposition 5 so we do not repeat them here.

The result is that the initial owner holds more shares than what the monitoring incentives require. An initial owner wants to hold a block in order to retain control and be sure to obtain his preferred choice on the project chosen. Hence the model offers another rationale for why one investor would choose to hold the biggest block in a firm quite apart from the monitoring incentive proposed so far in the literature. (Shleifer and Vishny, 1986)

A possible objection to our main result is that there are multiple equilibria for this configuration of monitoring costs. In particular it can be that for the same monitoring costs there exist both an n Blockholder and an Initial Owner equilibrium. Indeed this is true for n Blockholder equilibria where $n > 1$. This result depends on the multiplicity of EOS for a given (α_E, w_E) combination. However, Corollary 2 shows that for some values of the monitoring costs only n Blockholder equilibria are possible. The intuition behind this result is that when monitoring costs are low enough then the cost of losing diversification is low enough for a single blockholder to deviate and demand a block, thus ruling out an Initial Owner equilibrium. This is because he needs a smaller block to both win the vote (using the vote of the active liquidity investors) and to have a large shift in X. If monitoring costs are too low of course, there may be no conflict between the initial owner and outside investors and hence no desire to hold blocks.

Corollary 2 *Suppose:*

$$\bar{m} \in \left(\max[\bar{\alpha}, \tilde{\alpha}(1), \bar{m}_1^E(1\lambda), \bar{m}_1^{RC}(1)(K), \bar{m}_1^S(1,K)], \min\left[\frac{1}{2}, \hat{\alpha}(1), \bar{m}_2^E(1,\lambda), \bar{m}_2^{RC}(1,K), \bar{m}_2^S(1,K)\right]\right] \tag{18}$$

then there exist n Blockholder equilibria with $n \in [1, M_A]$ where the initial owner is monitoring, $m = 1$, $\alpha_E = \bar{m}$, $X_{med} = X_1 < \bar{X}$, $w_E = \underline{w}_E^1$. No other types of equilibria with positive trade exist for these parameter values.

This corollary is a special case of Proposition 5, so the conditions are the same except that n (blockholders) is replaced with 1 blockholder.

Among the possible blockholder equilibria we might want to consider the one which is most preferred by the initial owner, since he can always "induce" this EOS by appropriate choice of (α_E, w_E).

Proposition 6 *The optimal number of blockholders for the initial owner is:*

$$n^* = \left[\frac{(1+\bar{m})(\bar{m} - \lambda + \bar{m}\lambda)}{2\bar{m}^2(1-\lambda)}\right] > 0 \tag{19}$$

The optimal number of blockholders for the initial owner arises from the trade-off from having few blockholders and a vote outcome closer to his optimal outcome or having many blockholders and being able to sell the shares at a high price, i.e. low w_E. The preferred number of blockholders is a function of the monitoring costs (and hence, in equilibrium, of the amount of shares he retains in a monitoring equilibrium) and the proportion of active liquidity investors.

When the monitoring costs are high a further increase of the monitoring costs would induce the initial owner to prefer less blockholders. In such a case blockholders are already holding a highly undiversified portfolio and the initial owner needs to set a very low price in order to induce them to buy the shares. Decreasing the number of blockholders further does not reduce the share price by much but he gains in terms of risk/return outcome that is it decreases the costs of holding a suboptimal portfolio.

On the other hand, when the monitoring costs are very low an increase in the monitoring costs induces the initial owner to prefer more blockholders. In this case the discrepancy between the preferences of the blockholders and the initial owner is not that high so that increasing the number of blockholders allows the initial owner to raise the price at which the shares are tendered. Hence the initial owner would prefer to have more blockholders.

Finally the effect of the proportion of active investors among the liquidity shareholders, λ, has a negative effect on the optimal number of blockholders preferred by the initial owner. When more liquidity investors vote it becomes cheaper for blockholders to hold sufficiently large blocks so that they are jointly pivotal in the voting. This is very costly for the initial owner in terms of risk exposure. Hence when the λ is higher the initial owner would prefer to set a lower price, but to have the vote outcome closer to his preferred point. Hence the higher the participation of the liquidity shareholders to the vote the lower the number of blockholders the initial owner would like to have.

5.2 Optimal amount of capital raised

In this section we relax the assumption that the initial owner raises just enough capital needed to implement the project, K. We allow the initial owner to invest more capital $I \geq K$ and use the difference, $I - K$ to buy the risk free asset. This would offer him the possibility to achieve the preferred degree of diversification through the firm's investment in the risk free asset: hence the conflict of interest between the initial owner and outside investors may disappear. We show in the proposition below that our results are robust to relaxing this assumption.

Proposition 7 *The initial owner always strictly prefers to raise the minimum amount of capital, i.e.* $I = K$.

Observe that the initial owner acts as a monopolist when setting the share price. Hence, if he increases I, this lowers the price per share and it lowers the risk of the project for the same X. The decrease in price decreases the initial owner's utility in such a way that it more than offsets the increase in utility due to a lower risk of the project. The compensation demanded by the investors in terms of lower price is higher than the gain that the initial owner obtains. Hence the initial owner prefers to hold a suboptimally diversified portfolio rather than issue the shares at a lower price.

6 Comparative Statics and Empirical Implications

In this section we discuss some of the important predictions of our model and discuss how they relate to the empirical literature on ownership structure.

6.1 What does Ownership Structure depend on?

Our first prediction is that the *number* and *size* of the blockholders will depend on the degree of co-ordination required for voting decisions captured by the size of the first blockholder (hence e.g. on the monitoring technology or the degree of complementarity between the firms production and the input of the initial owner). We expect firms to have blockholders only in cases where they hold less than 50% of the stock. No blockholders should be observed in firms that keep the voting control when they release the rest.

Secondly, we find that the degree of concentration in the ownership structure is decreasing in the monitoring costs, $\bar{m}$, or in the size of the first block. For very small monitoring costs (or small size of the initial block) there is dispersed ownership, for intermediate levels of monitoring costs there are multiple large blocks and for high monitoring costs there is a single large block. For example, high monitoring costs are more common in firms dedicated to innovation or R&D where moral hazard issues are much more pervasive. In such firms we should see an initial owner who has control.[21]

Third, we can capture the effect of riskier sectors through $\bar{X}$. High $\bar{X}$ corresponds to industries where risk can be potentially very high. Then our model predicts that blockholder ownership structures emerge in more risky industries. Thus, in more mature industries the projects among which the shareholders can choose is limited to the low risk ones, that is in our models, i.e. low $\bar{X}$. Ceteris paribus, in more mature sectors having a choice of low risk/low return projects, dispersed ownership structures with more than one blockholder are less likely. Anecdotal evidence suggests that in more mature sectors it is more common to see families in control of firms. In very innovative industries, on the other hand, we should see blockholders which ease the conflicts of interests between investors and initial investor and induce more risky decisions. These blockholders are usually represented by institutional investors, e.g. venture capitalists, who professionally look for firms with a high risk/return profile.

There is also an inverse relationship between dispersion in the ownership structure and the size clustering of the projects, K. When K is small, the initial owner may prefer not to raise the capital

[21] Although in our model the monitoring technology can be considered more broadly as a friction which determines the first block, this does not exclude these empirical predictions applying in firms where the largest blockholder exists because of monitoring.

as the monitoring costs are too high to compensate for the risk incurred. On the other hand, when the project size K is very big the extra value added by the monitoring is such a small fraction of the total cost that it becomes more attractive to sell the idea and have a totally dispersed ownership structure.Hence, concentrated ownership should invest in similar size projects, while more dispersed ownership should be common in any type of project size. Even more interesting it is the feature of the ownership structure for intermediate levels of monitoring costs. As Fig. 5 shows dispersed ownership is common in all firms whatever is the siwe (given low monitoring costs). Concentrated ownership instead arises only for projects of specific intermediate size.

Fourth, we analyze the effects of λ on the ownership structure. The effect of higher λ in our model is ambiguous because we do not impose any capital constraints on investors: on the one hand it helps to reduce the costs of participation for blockholders implying, ceteris paribus, a more diversified structure. On the other hand, it becomes more costly for the initial owner to give up control since the vote outcome will be further from his first best. The balance between these two forces determines the equilibrium outcome. Voting participation can be influenced by country regulation or firm bylaws. For example until a few years ago in Germany only shareholders with more than 5% participation could vote. Alternatively, shareholders' vote participation can be related to information disclosure and hence the ability to make informed decisions. Hence, we may interpret higher λ as equivalent to an increasing minority shareholder protection. In financial markets with high minority shareholder protection and where investors are not capital constrained we should expect to see either an initial owner with a big block and many small investors or situations where there are few blocks – the initial owner and some others who need to hold relatively few shares to be able to gain control over the firm's decision.

Finally notice that because the utility of the initial owner is decreasing in λ, the threshold of monitoring costs below which he is willing to invest in the project is lower. Hence minority shareholder protection has a positive and a negative effect on investments. On one side minority shareholder protection leads to higher value projects indirectly through its effect on the ownership structure. However it also reduces the willingness of the initial owner to invest money or to exert monitoring. Hence minority shareholder protection is detrimental to induce the investment, but on the other side if an investment is taken minority shareholder protection is a tool to undertake more value enhancing projects.

6.2 The implications of the Ownership Structure

Let us now look at the predictions we make on how ownership structure influences firm choices.

In our model, the risk/return decision depends not only on whether there is concentration of ownership but rather on the number and size of blockholders: firm value decreases with the size of the first block. This is because for low values of monitoring costs the initial owner prefers a high return as well so both agree on the choice of projects. Ceteris paribus, the smaller the size of the median block the higher is the predicted firm value. These results are consistent with the empirical literature which shows that the effect of the first blockholder on the value of the firm is usually negative, (Barclay and Holderness (1989) and Kirchmaier and Grant (2005) and Lehmann and Weigand (2000)). This is in keeping with our results on the Initial Owner equilibrium. Some of the papers distinguish between the role of the first and subsequent blockholders: The empirical consensus seems to be (again in keeping with our results) that a second blockholder increases firm value. This phenomenon is present across countries and across publicly listed or private firms (Volpin (2002), Maury and Pajuste (2005), Faccio, Lang, and Young (2001), Gutierrez and Tribo (2004), Isakov and Weisskopf (2009) and Laeven and Levine (2008)). Roosenboom and Schramade (2006) studying French IPOs find that when the owner is powerful, the firm is less valued; when the initial owner shares control with other blockholders the value increases. Helwege, Pirinsky, and Stulz (2007) find that firms where insider ownership gets reduced over time or firms with widely held ownership are more highly valued.

Our paper also offers an extra prediction to differentiate our theoretical results from other theories explaining why blockholders emerge: in our model the risk profile of a firm also depends on how widely dispersed is the ownership. There are few empirical papers looking at this relationship. Carlin and Mayer (2000; 2003) and Teodora (2009) find that multiple blockholders are present in high risk firms, while a single blockholder is common in low risk firms.

Our paper offers an explanation for the underpricing observed in IPOs. Empirically, it has been found that IPOs are usually associated with a first day positive return (i.e. the underpricing).[22] In the literature there is no agreement on the reasons why this phenomenon arises (see Jenkinson and Ljungqvist (2001) and Welch and Ritter (2002) for reviews). Brennan and Franks (1997) among others (Boulton, Smart, and Zutter (forthcoming), Nagata and Rhee (2009) and Yeh and Shu (2004)) found that ownership structure contributes to the degree of underpricing. In particular they argue that underpricing can be more severe when the initial owner wants to avoid blocks. However, they note that this is not a stable outcome and over time blocks are formed anyway. The findings of Brennan and Franks (1997) are in line with our predictions. If the initial owner could decide the share allocation and retain control, he would be willing to do so even though this implies a lower price. However, if share trade is allowed this outcome cannot be stable. The findings of Brennan and Franks (1997) contrast instead with the predictions of Stoughton and

[22]The average underpricing is between 15% and 17% though it varies a lot over time

Zechner (1998) and DeMarzo and Urosevic (2006) where the higher the ownership concentration the higher the underpricing. When the initial owner has no discretion on share allocation and trade is allowed, blockholder equilibria cannot be avoided and underpricing occurs in the sense that liquidity shareholders would be willing to pay more than the equilibrium price to buy shares. Our theory predicts therefore that underpricing occurs when the size of the initial block (initial owner) in the shares is not too large (in particular less than the relevant voting threshold) and not too small. In such a case the predictions of our model are similar to those of Stoughton and Zechner (1998) and DeMarzo and Urosevic (2006): underpricing occurs when blockholders are present and it is higher the higher the degree of concentration.

7 Conclusions

This paper attempts to reconcile two well documented empirical regularities in the corporate governance literature: the presence of multiple large shareholders, some of them without a controlling interest on the one hand and increased firm value when such large shareholders are present on the other hand (e.g. Carlin and Mayer (2000; 2003) and Laeven and Levine (2008) among others).

Our model relies on an endogenous conflict of interest over the choice of risk between the initial owner of the firm and outside investors. This problem is similar to a public good contribution game: if there are multiple blockholders which are intermediate in size between the entrepreneur and small shareholders, then the decision on the project is more favorable to all investors, but it comes at a cost to the large shareholders as they hold suboptimal portfolios. Hence blockholders provide a public good to other investors. We show that at very low levels and at very high levels of monitoring costs, we get equilibria where either the initial owner has full control when monitoring costs are high or liquidity shareholders having full control when they are low. For intermediate values we get equilibria with multiple blockholders. The corresponding choice of projects goes from low risk/return (the entrepreneur's first best) when monitoring costs are low, to intermediate levels of risk/return in the blockholder equilibria (depending on how many blockholders there are), to very high risk/return projects when the monitoring costs are very high.

The main contribution of our paper is to generate a multiple blockholder equilibrium with a very simple and standard model of corporate control. Unlike many of the papers that explain this phenomenon, we show that blockholders arise in the absence of any coordination between investors and without a direct choice of share allocation on the part of the initial owner – indeed the initial owner only chooses the ownership structure indirectly through the pricing of shares. This simple model is able to explain a variety of stylized facts about the links between ownership size and firm characteristics.

Although our paper assumes that all outside investors are identical, the paper could be easily extended to the case of heterogenous agents. Indeed this would help to reduce the problem of multiple equilibria. In this case less risk averse investors would be the natural blockholders and we would expect that if this occurs, the risk/return choice is even higher. This is in line with empirical papers which find that the presence of institutions as blockholders enhances firm performance (Ben Dor (2003), Hartzell, Kallberg, and Liu (2008), Barber (2007) and Chen, Harford, and Li (forthcoming)).

References

Admati, A. R., P. Pfleiderer, and J. Zechner (1994): "Large Shareholder Activism, Risk Sharing, and Financial Market Equilibrium," Journal of Political Economy, 102(6), 1097–1130.

Barber, B. M. (2007): "Monitoring the Monitor: Evaluating CalPERS' Activism," Journal of Investing, 16(4), 66–80.

Barclay, M. J., and C. G. Holderness (1989): "Private Benefit from Control of Public Corporations," Journal of Financial Economics, 25(2), 371–395.

Becht, M., P. Bolton, and A. Röell (2003): "Corporate Governance and Control," in Handbook of the Economics of Finance, ed. by G. Constantinides, M. Harris, and R. M. Stulz. North Holland.

Becht, M., J. Franks, C. Mayer, and S. Rossi (2009): "Returns to Shareholder Activism: Evidence from a Clinical Study of the Hermes UK Focus Fund," Rev. Financ. Stud., 22(8), 3093–3129.

Ben Dor, A. (2003): "The Performance of Initial Public Offerings and the Cross Section of Institutional Ownership," Unpublished manuscript.

Bennedsen, M., and D. Wolfenzon (2000): "The balance of power in closely held corporations," Journal of Financial Economics, 58(1–2), 113–139.

Bloch, F., and U. Hege (2001): "Multiple Shareholders and Control Contest," Unpublished manuscript.

Bolton, P., and E. L. Von Thadden (1998): "Blocks, Liquidity, and Corporate Control," The Journal of Finance, 53(1), 1–25.

Boulton, T. J., S. B. Smart, and C. J. Zutter (forthcoming): "IPO Underpricing and the Market for Corporate Control," Journal of International Business Studies.

Brav, A., W. Jiang, F. Partnoy, and R. S. Thomas (2008): "Hedge Fund Activism, Corporate Governance, and Firm Performance," The Journal of Finance, 63(4), 1729–1775.

Brennan, M., and J. Franks (1997): "Underpricing, ownership and control in initial public offerings of equity securities in the UK," Journal of Financial Economics, 45(3), 391–413.

BURKART, M., D. GROMB, AND F. PANUNZI (1997): "Large Shareholders, Monitoring, and the Value of the Firm," Quarterly Journal of Economics, 112(3), 693–728.

CARLIN, W., AND C. MAYER (2000): "How Do Financial Systems Affect Economic Performance?," in Corporate Governance: Theoretical and Empirical Perspectives, ed. by X. Vives. Cambridge University Press, Cambridge, UK.

——— (2003): "Finance, investment, and growth," Journal of Financial Economics, 69(1), 191–226.

CHEN, X., J. HARFORD, AND K. LI (forthcoming): "Monitoring: Which Institutions Matter?," Journal of Financial Economics, Forthcoming.

DEMARZO, P. M., AND B. UROSEVIC (2006): "Ownership Dynamics and Asset Pricing with a Large Shareholder," Journal of Political Economy, 114(4), 774–815.

DEMICHELIS, S., AND K. RITZBERGER (2008): "Corporate Control and Stock Markets," Unpublished manuscript.

DHILLON, A. (2005): "Poliltical Parties and Coalition Formation," in Group Formation in Economics: Networks, Clubs and Coalitions, ed. by G. Demange, and M. Wooders. Cambridge University Press, Cambridge, UK.

ECONOMIST, T. (2009): "Business: The ties that bind; The Wallenbergs," 390(8615), 68.

EDMANS, A., AND G. MANSO (2008): "Governance Through Exit and Voice: A Theory of Multiple Blockholders," Discussion Paper 225/2006, ECGI.

FACCIO, M., L. H. P. LANG, AND L. YOUNG (2001): "Who Disciplines Management in Poorly Performing Companies?," American Economic Review, 91(1), 54–78.

FACCIO, M., M.-T. MARCHICA, AND R. MURA (2009): "Large Shareholder Diversification and Corporate Risk-Taking," Unpublished manuscript.

FERNANDO, C. S., S. KRISHNAMURTHY, AND P. A. SPINDT (2004): "Are Share Price Levels Informative? Evidence from the Ownership, Pricing, Turnover, and Performance of IPO Firms," Journal of Financial Markets, 7(4), 377–403.

GOERGEN, M., AND L. RENNEBOOG (2002): "Prediction of Ownership and Control Concentration in German and UK Initial Public Offerings," in Convergence and Diversity of Corporate Governance Regimes and Capital Markets, ed. by J. McCahery, P. Moerland, T. Raaijmakers, and L. Renneboog, pp. 251–267. Oxford University Press, Oxford, UK.

GOMES, A., AND W. NOVAES (2001): "Sharing of Control as a Corporate Governance Mechanism," Discussion Paper PIER Working paper 01-029, University of Pennsylvania Law School.

GROSSMAN, S. J., AND O. D. HART (1980): "Takeover Bids, the Free-Rider Problem: and the Theory of Corporation," Bell Journal of Economics, 11(1), 42–64.

GUTIERREZ, M., AND J. A. TRIBO (2004): "Private Benefits Extraction In Closely-Held Corporations: The Case For Multiple Large Shareholders," Discussion Paper 53/2004, ECGI.

Hartzell, J. C., J. G. Kallberg, and C. H. Liu (2008): "The Role of Corporate Governance in Initial Public Offerings: Evidence from Real Estate Investment Trusts," Journal of Law and Economics, 51(3), 539562.

Helwege, J., C. A. Pirinsky, and R. M. Stulz (2007): "Why do Firms Become Widely Held? An Analysis of the Dynamics of Corporate Ownership," The Journal of Finance, 62(3), 995–1028.

Holderness, C. G. (2009): "The Myth of Diffuse Ownership in the United States," Review of Financial Studies, 22(4), 1377 – 1408.

Holmstrom, B., and J. Tirole (1993): "Market Liquidity and Performance Monitoring," Journal of Political Economy, 101(4), 678–709.

Isakov, D., and J.-P. Weisskopf (2009): "Family Ownership, Multiple Blockholders and Firm Performance," Unpublished manuscript.

Jenkinson, T., and A. Ljungqvist (2001): Going Public: The Theory and Evidence on How Companies Raise Equity Finance. Oxford University Press, Oxford.

Kirchmaier, T., and J. Grant (2005): "Corporate Ownership Structure and Performance in Europe," Discussion Paper 3.

La Porta, R., F. Lopez-De-Silanes, and A. Shleifer (1999): "Corporate Ownership Around the World," The Journal of Finance, 54(2), 471–517.

Laeven, L., and R. Levine (2008): "Complex Ownership Structures and Corporate Valuations," Review of Financial Studies, 21(2), 579–604.

Lehmann, E., and J. Weigand (2000): "Does the Governed Corporation Perform Better? Governance Structures and Corporate Performance in Germany," European Finance Review, 4(2), 157–195.

Maug, E. (1998): "Large Shareholders as Monitors: Is there a tradeoff between liquidity and control?," The Journal of Finance, 53(1), 65–98.

Maury, B., and A. Pajuste (2005): "Multiple Large Shareholders and Firm Value," Journal of Banking and Finance, 29(7), 1813–1834.

Nagata, K., and S. Rhee (2009): "Ownership, Control and Underpricing in Initial Public Offerings: Evidence from Japan," Unpublished manuscript.

Noe, T. H. (2002): "Investor Activism and Financial Market Structure," Review of Financial Studies, 15(1), 289–318.

Pagano, M., and A. Röell (1998): "The Choice of Stock Ownership Structure: Agency Costs, Monitoring, and the Decision to Go Public," Quarterly Journal of Economics, 113(1), 187–226.

Roosenboom, P., and W. Schramade (2006): "The price of power: Valuing the controlling position of owner-managers in French IPO firms," Journal of Corporate Finance, 12(2), 270–295.

Shleifer, A., and R. W. Vishny (1986): "Large Shareholders and Corporate Control," Journal of Political Economy, 94(3), 461–488.

Smith, M. P. (1996): “Shareholder Activism by Institutional Investors: Evidence for CalPERS,” Journal of Finance, 51(1), 227–52.

Stiglitz, J. E. (1985): “Credit Markets and the Control of Capital,” Journal of Money, Credit and Banking, 12(2), 133–152.

Stoughton, N., and J. Zechner (1998): “IPO-Mechanisms, Monitoring and Ownership Structure,” Journal of Financial Economics, 49(1), 45–77.

Teodora, P. (2009): “Corporate Risk-Taking and Ownership Structure,” Unpublished manuscript.

Tirole, J. (2001): “Corporate Governance,” Econometrica, 69(1), 1–35.

Volpin, P. (2002): “Governance with poor investor protection: evidence from top executive turnover in Italy,” Journal of Financial Economics, 64(1), 61–90.

Welch, I., and J. Ritter (2002): “A Review of IPO Activity, Pricing, and Allocations,” The Journal of Finance, 57(4), 1795–1828.

Yeh, Y. H., and P.-G. Shu (2004): “Corporate Governance and Valuation of IPOs,” Unpublished manuscript.

Zwiebel, J. (1995): “Block Investment and Partial Benefits of Corporate Control,” Review of Economic Studies, 62(2), 161–185.

A Appendices

A.1 The Model

A.1.1 Lemma 1:

Proof. The proof is obvious. We just maximize the objective functions (over X) of the outside investors, equation (4), and of the initial owner, equation (5), given the fraction of shares held, α_j. Concavity ensures uniqueness of the solution. ■

A.2 The equilibria.

A.2.1 Lemma 2:

Proof. At date 3, the ownership structure, $\vec{\alpha}$, and thus X_{med} are already fixed. Given the initial owner's objective function (5), he monitors iff the utility from monitoring is greater than from not monitoring, that is when:

$$\alpha_E\left(X_{med}\bar{R}+K\right)-\frac{\gamma}{2}\alpha_E^2X_{med}^2\sigma^2-\bar{m}K\geq\alpha_EX_{med}\bar{R}-\frac{\gamma}{2}\alpha_E^2X_{med}^2\sigma^2 \tag{20}$$

Rearranging, we get the condition $\alpha_E\geq\bar{m}$. ■

A.2.2 Lemma 3

Proof. Investor l chooses $\alpha_l(X_j)$ to maximize equation (4) where $X_j=X_{med}$. The first order condition implies equation (7). The second order condition is satisfied as long as $X_j>0$, so this is a maximum. ■

A.2.3 Lemma 4

Proof. As the liquidity shareholding $\alpha_l(X_j)$ maximizes the utility of an investor given a vote outcome, any shareholding $\alpha_1 \neq \alpha_l(X_j)$ gives less utility. ■

A.2.4 Lemma 5

Proof. Consider first the case where $\alpha_j \geq \bar{\alpha}$. Hence by Lemma 1 $X_j = \frac{\bar{R}}{\gamma\sigma^2\alpha_j}$ and by Lemma 3 $\alpha_l(X_j) = \alpha_j + \frac{f(m)-\frac{K-w_E}{1-\alpha_E}}{\frac{\bar{R}^2}{\gamma\sigma^2\alpha_j^2}}$. By assumption, $f(m) - \frac{K-w_E}{1-\alpha_E} < 0$ and thus $\alpha_l(X_j) < \alpha_j$.

Now let $\alpha_j < \bar{\alpha}$. By definition, j is the median shareholder, hence $X_{med} = X_j = \bar{X}$. As $f(m) - \frac{K-w_E}{1-\alpha_E} < 0$ a liquidity investor always chooses $\alpha_l(\bar{X}) < \bar{\alpha}$ and hence it is sufficient to show that no active investors hold $\alpha_j < \alpha_l(\bar{X})$. Assume to the contrary investors j holds $\alpha_j < \alpha_l(\bar{X})$. (Active) Investor j can improve his utility by choosing $\alpha_l(\bar{X})$ and voting for $X = \bar{X}$ without changing X_{med}. Contradiction to the equilibrium definition where shareholders are maximizing their utility when X is fixed. This proves that in equilibrium $\alpha_j \geq \alpha_l(X_j)$. ■

A.2.5 Lemma 6:

Proof. Before moving to the proof of the Lemma we have an intermediary step given by Lemma 9 which describes the necessary conditions for an n Blockholder EOS to exist.

Denote the size of the shareholdings of other *active* investors excluding investor 1 as α_{-1}. So $\alpha_{-1} = (n-1)\alpha_1 + N_A\alpha_l(X_1)$.

Lemma 9 *Suppose $\alpha_E > 0$ and $f(m) - \frac{K-w_E}{1-\alpha_E} < 0$. If a n-Blockholder EOS arises then (1) $\alpha_{-1} > 0$; (2) $\alpha_1 + \alpha_{-1} \geq \alpha_E > \alpha_1$; (3) $\alpha_E \geq N_A\alpha_l(X_1)$. In any such equilibrium $X_1 > X_E$ and $N_A\alpha_l(X_1) + n\alpha_1 \geq \alpha_E$.*

Proof.

We now prove the lemma first computing the utility of the median shareholder (a) when $X_{med} = X_1 < \bar{X}$ and (b) when $X_{med} = X_1 = \bar{X}$ and then showing the lemma by contradiction.

(a) Since $X_{med} = X_1 < \bar{X}$, in Definition 2, shareholder 1 is the median shareholder and $\alpha_1 > \bar{\alpha}$. Using equation (4), the utility of the median shareholder is:

$$U_1^{nBH} = 1 + \alpha_1\left(X_1\bar{R} + f(m) - \frac{K - w_E}{1-\alpha_E}\right) - \frac{\gamma}{2}\alpha_1^2 X_1^2\sigma^2$$

By Lemma 1 and the fact that $X_{med} = X_1 < \bar{X}$, this is equivalent to

$$\alpha_1(f(m) - \frac{K - w_E}{1-\alpha_E}) + 1 + \frac{\bar{R}^2}{2\gamma\sigma^2} \tag{21}$$

By assumption, $f(m) - \frac{K-w_E}{1-\alpha_E} < 0$, hence U_1^{nBH} is decreasing in α_1.

(b) Suppose instead that $X_{med} = X_1 = \bar{X}$. By definition $\alpha_l(\bar{X}) < \alpha_1$ in an n-Blockholder EOS and $\alpha_l(\bar{X}) = \text{argmax}\ U_i(\alpha_i)|_{X=\bar{X}}$. Hence, U_1^{nBH} is decreasing in α_1.

Now we are ready to prove the Lemma by contradiction:

(1) Suppose to the contrary that $\alpha_{-1} = 0$ in an n-Blockholder EOS. In order to have $X_{med} = X_1$ we must have $\alpha_1 \geq \alpha_E$, since investor 1 is the only outside investor who votes. From part (a), we know that U_1^{nBH} above is decreasing in α_1 and hence he is better off setting $\alpha_1 = \alpha_E$, otherwise the initial owner becomes the median shareholder. But in this case, $X_{med} = X_1 = X_E$ and *de facto* investor 1 does not affect the vote outcome. Hence he

prefers to be a liquidity shareholder and his optimal shareholding is given by $\alpha_l(X_E)$. Contradiction to the fact that this is an EOS (since investor 1 wants to deviate unilaterally). Hence, $\alpha_{-1} > 0$ in any n-Blockholder EOS.

(2) $\alpha_E > \alpha_1$ follows from the proof of part (1) above. We need to prove that $\alpha_1 + \alpha_{-1} \geq \alpha_E$. Suppose to the contrary that $\alpha_1 + \alpha_{-1} < \alpha_E$. Then $\alpha_E = \alpha_{med}$ and so $X_{med} = X_E$, so this is not an n-Blockholder EOS by Definition 2. Contradiction.

(3) $\alpha_E \geq N_A\alpha_l(X_1)$. Given (2) above, this holds iff $\alpha_1 = 0$ which contradicts Definition 2. This implies that $N_A\alpha_l(X_1) + n\alpha_1 > \alpha_E$ (since $\alpha_1 > 0$ and $\alpha_E \geq N_A\alpha_l(X_1)$). ■

Lemma 9 shows that U_1^{nBH} is decreasing in $\alpha_1 (\geq \alpha_l(\bar{X}))$. So, $n\alpha_1 = \alpha_E - N_A\alpha_l(X_1)$. Recall that the total shares of the firm (assuming full subscription) must add up to 1 and that we consider equilibria where the proportion of active liquidity investors among the liquidity shareholders is given by λ. Hence N_A must satisfy $\lambda(1-\alpha_E-n\alpha_1) = N_A\alpha_l(X_1)$. Hence $n\alpha_1 + \lambda(1 - n\alpha_1 - \alpha_E) = \alpha_E$. Solving for α_1 we get equation (8). ■

A.2.6 Lemma 7:

Proof. By Lemma 1, given the shareholding α_i of investor i the first best X for the investor is given $X_i = \min\left[\frac{\bar{R}}{\gamma\sigma^2\alpha_i}, \bar{X}\right]$. By the proof of Lemma 9, we know that the investors' utility function is decreasing in α_i when $\alpha_i \geq \alpha_l(\bar{X})$. Hence, the first best choice of $\alpha_i = \alpha_l(\bar{X})$. ■

A.2.7 Proposition 1:

Proof. We solve the problem using backward induction. The initial owner's objective function is different depending on the anticipated ownership structure.

The initial owner maximizes the following objective function:

$$\max_{\alpha_E, w_E} U\left(m=0\right) = \bar{R}X_{med}(\alpha_E)\alpha_E - \frac{\gamma}{2}X_{med}(\alpha_E)^2\sigma^2\alpha_E^2 + 1 - w_E \tag{22}$$

Given the objective function, regardless of the ownership structure, he chooses the lowest w_E, that is $w_E = w_E^j(\alpha_E)$ where $w_E^j(\alpha_E) = \{\underline{w}_E^E, \underline{w}_E^n, \underline{w}_E^{LS}\}$.

Substituting for w_E^j in the objective function, it can be checked that $U\left(m=0\right)$ is decreasing in α_E for all w_E^j, for $X_{med} \leq \bar{X}$. Therefore $\alpha_E = 0$ is the optimal choice of the initial owner for *any* ownership structure.

When $\alpha_E = 0$, and there is at least one active investor ($\lambda > 0$), all active investors vote for $\bar{X}$ and hence $X_{med} = \bar{X}$. The participation constraint of outside investors is satisfied if $w_E \geq \underline{w}_E^{LS}$. This is the first best for outside investors hence there is no incentive to become a blockholder. Hence, the Liquidity Shareholder EOS is the unique EOS. Also it is trivial to see that no investor is willing to sell his shares at a price lower than the maximum that an excluded investor is willing to pay.

Furthermore the initial owner sets $w_E = \underline{w}_E^{LS}$ as the lowest w_E^j. Hence, the initial owner's utility is given by:

$$\bar{R}\bar{X} - K + 1 \tag{23}$$

Finally we check if the participation constraint of the initial owner is satisfied: i.e if he invests in the riskfree asset his utility is 1. Hence when the project has a positive NPV he sells the firm, otherwise he does not raise capital. The initial owner's value function is then given by equation (12). ■

A.3 Monitoring Equilibria

A.3.1 Lemma 8:

Proof. Before moving to the proof of the Lemma we show which are the various EOS for different pair of $(\alpha_E, w_E) \in S \equiv [0,1] \times (-\infty, \infty)$.

First, observe that for certain combinations of (α_E, w_E) there always exists a no trade equilibrium where no investors participate in the share issue. Define $\underline{w}_E^{NT} = \begin{cases} \underline{w}_E^E & \text{if } \alpha_E > \max\left[\frac{1}{2}, \bar{\alpha}\right] \\ \underline{w}_E^1 & \text{if } \bar{\alpha} \leq \alpha_E \leq \frac{1}{2} \\ \underline{w}_E^{LS} & \text{if } \alpha_E \leq \bar{\alpha} \end{cases}$.

This is the minimum amount the initial owner needs to invest in order to guarantee that at least one investor is willing to buy shares. The next lemma shows the conditions under which the No Trade equilibrium exists.

Lemma 10 *There always exists a No Trade EOS if* $w_E < \underline{w}_E^{NT}$.

Proof. Suppose to the contrary that there exists a No Trade EOS and $w_E \geq \underline{w}_E^{NT}$. Then a single shareholder can buy $1 - \alpha_E$ shares and ensure full subscription since $w_E \geq \underline{w}_E^{NT}$, contradiction. ■

Define:

$$\underline{\underline{w}}_E(\alpha_E) \equiv K - f(m)(1 - \alpha_E) \tag{24}$$

$$\hat{\alpha}(n) \equiv \max[\hat{\alpha}_1(n), \hat{\alpha}_2(n)] \tag{25}$$

$$\hat{\alpha}_1(n) \equiv \frac{2\lambda}{2(1+\lambda) - n(1-\lambda)} \tag{26}$$

$$\hat{\alpha}_2(n) \equiv \frac{2\bar{\alpha}(1-\lambda)n + \lambda - \sqrt{\lambda^2 + 4\bar{\alpha}(1-\lambda)n\left(\bar{\alpha}(1-\lambda)n - 1 - 2\bar{\alpha}(1+\lambda)\right)}}{2\left(1+\lambda\right)} \tag{27}$$

$$\breve{\alpha}(n) \equiv \frac{n(1-\lambda)\bar{\alpha} + \lambda}{1+\lambda} \tag{28}$$

$$\tilde{\alpha}(n) \equiv \frac{\lambda}{1+\lambda} - \frac{n}{\eta_j} \tag{29}$$

Suppose α_E is fixed, then $\underline{\underline{w}}_E(\alpha_E)$ is the minimum w_E that guarantees that $f(m) > \frac{K - w_E}{1 - \alpha_E}$. Note that when $m = 1$ this condition implies that the price of the risky asset is lower than that of the risk free one. Later we show this is always the case in equilibrium. $\hat{\alpha}(n)$ is the value of α_E such that if $\alpha_E \leq \hat{\alpha}(n)$ then $\underline{w}_E^E \leq \underline{w}_E^n$. In particular if $\alpha_E \leq \hat{\alpha}_1(n)$, then $\underline{w}_E^E \leq \underline{w}_E^n$ where $X_1 < \bar{X}$, and if $\alpha_E \leq \hat{\alpha}_2(n)$, then $\underline{w}_E^E \leq \underline{w}_E^n$ where $X_1 = \bar{X}$. $\breve{\alpha}(n)$ is the value of α_E such that when $\alpha_E \leq \breve{\alpha}(n)$, $\alpha_1 = \bar{\alpha}$. Finally $\tilde{\alpha}(n)$ is the value of α_E such that when $\alpha_E \leq \tilde{\alpha}(n)$ n extra active liquidity shareholders can become pivotal in the voting decision and change it from X_E to $\bar{X}$.

Lemma 11 *There exists an Initial Owner EOS, with* $X_{med} = X_E < \bar{X}$ *for any pair* (α_E, w_E), *satisfying the following conditions:*

$$\alpha_E \in \left(\max\left(\bar{\alpha}, \min\left[\frac{1}{2}, \max\left[\hat{\alpha}(1), \tilde{\alpha}(1)\right]\right]\right), 1\right] \tag{30}$$

$$w_E \in \left[\underline{w}_E^E(\alpha_E), \min\left[\underline{w}_E^1(\alpha_E), \underline{\underline{w}}_E(\alpha_E)\right]\right) \tag{31}$$

Proof. We use Definition 1: An Initial Owner EOS exists for any combination of (α_E, w_E) iff (a) the participation constraint of investors is satisfied; (b) the incentive constraint of active investors is satisfied; (c) $X_{med} = X_E < \bar{X}$ in the EOS and (d) no shareholder is willing to sell his participation at a price lower than the price at which the excluded investors are willing to buy.

(a) The participation constraint for liquidity shareholders is satisfied iff $U_{l,E} \geq 1$. A sufficient condition for $U_{l,E} \geq 1$ is that $\alpha_l(X_E) \geq 0$, i.e. iff $w_E \geq \underline{w}_E^E$. This is the first part of condition (31).

(b) We check that no liquidity investor has an incentive to switch to becoming a blockholder. Notice that $\underline{w}_E^E < \underline{w}_1^E$ whenever $\alpha_E > \hat{\alpha}(1)$. This implies that, given the conditions of the lemma, if a liquidity investor switches to becoming a blockholder, his utility is negative while if he stays as a liquidity investor his utility is strictly positive since $w_E \geq \underline{w}_E^E$. Hence condition (31) implies that $U_{l,E} \geq U_1^{1BH}$ (assuming that $\frac{K-w_E}{1-\alpha_E} > f(m)$ – this is the case since $w_E \leq \underline{\underline{w}}_E$).

(c) $\alpha_E > \bar{\alpha}$ ensures that $X_E < \bar{X}$.

(d) Recall that there are a fraction λ of active investors among the liquidity shareholders. For the EOS we also need to guarantee that none of the investors who hold shares want to sell them given the maximum price that excluded investors are willing to pay. Any active investor who is excluded can do better by buying liquidity shares from a *passive* investor (who does not vote) at a price slightly higher than the initial owner's price, if by voting he is able to become pivotal and change the outcome. This occurs when

$$\lambda(1 - \alpha_E - \bar{\alpha}_l) + \alpha_l(X_E) > \alpha_E$$

which corresponds to $\alpha_E \leq \tilde{\alpha}(1)$. In such a case the participation constraint of the investors is also satisfied since it was satisfied for the investors who originally held the shares.

To rule this out we set $\alpha_E > \max[\hat{\alpha}(1), \tilde{\alpha}(1)]$. However these conditions become irrelevant when the initial owner has the majority of the shares since then, no active investors can change the vote outcome. This implies that $\alpha_E > \min\left[\frac{1}{2}, \max[\hat{\alpha}(1), \tilde{\alpha}(1)]\right]$.

Putting together the constraints on α_E from point (c) and (d), the lower bound on α_E is given by (30).

Observe that the interval in condition (31) is non empty whenever $m = 1$ i.e. $f(m) = K$. This is because then $\underline{w}_E^E < \underline{\underline{w}}_E$, and we already showed that $\underline{w}_E^E < \underline{w}_1^E$ whenever $\alpha_E < \hat{\alpha}(1)$. ■

Lemma 12 *There exists an n-Blockholder EOS, with $X_{med} = X_1 < \bar{X}$, for any pair (α_E, w_E), satisfying the following conditions:*

$$\alpha_E \in \left(\max\left[\breve{\alpha}(n)\right], \min\left[\hat{\alpha}_1(n), \frac{1}{2}\right]\right] \tag{32}$$
$$w_E \in [\underline{w}_E^n(\alpha_E), \underline{\underline{w}}_E(\alpha_E)) \tag{33}$$

where $1 \leq n \leq M_A$.[23]

Proof. An n-Blockholder EOS with $X_1 < \bar{X}$ exists iff the conditions of Definition 1 are satisfied: (a) $U_1^{nBH} \geq 1$, $U_{l,1} \geq 1$; (b) No investor wants to unilaterally increase or decrease his shares; (c) No investor wants to sell his shares at a price lower than the maximum willingness to pay of excluded investors.

Observe that $X_1 < \bar{X}$ iff $\alpha_1 > \bar{\alpha}$ (Lemma 1). From Corollary 6, this condition is equivalent to $\alpha_E > \breve{\alpha}(n)$, the lower boundary of condition (32) .

(a) The participation constraint of blockholders is satisfied iff

$$U_1^{nBH} = \left(\bar{R}X_1 + f(m)\right)\alpha_1 - \frac{K - w_E}{1 - \alpha_E}\alpha_1 - \frac{\gamma}{2}X_1^2\sigma^2\alpha_1^2 + 1 \geq 1 \tag{34}$$

[23]Observe that $\hat{\alpha}(n) > \frac{\lambda}{1+\lambda} > 0$ for all $n \geq 1$. Moreover when $n \geq 2$, $\hat{\alpha}(n) \geq \frac{1}{2}$. So $\frac{1}{2} = \min\left(\hat{\alpha}(n), \frac{1}{2}\right)$ for $n \geq 2$. Finally the interval of α_E is always positive when $n \geq \frac{2((1+\lambda)\bar{\alpha}-\lambda)}{(1-\lambda)\bar{\alpha}}$ and in such a case $\max\left[\bar{\alpha}, \tilde{\alpha}(n)\right] = \bar{\alpha}$.

Rearranging, this is equivalent to $w_E \geq \underline{w}_E^n(\alpha_E)$, the lower boundary of condition (33). By Lemma 4, liquidity shareholders participation constraint is always satisfied whenever the blockholders' is, hence $U_{l,1} \geq 1$.

(b) Claims 1 and 2 show that given conditions (32) and (33), no blockholder wants to change his shares unilaterally from α_1. Claim 3 shows that no liquidity investor wants to increase or decrease unilaterally his shareholding.

Claim 1: Suppose conditions (32) and (33) are satisfied. No blockholder wants to decrease his shares unilaterally from α_1.

Proof. Observe that $w_E \leq \underline{\underline{w}}_E$ by condition (33). By Lemma 6, a blockholder chooses α_1 as given in equation (8). If any blockholder reduces his shares, given α_{-i}, X_{med} shifts to $X_E < X_1 < \bar{X}$ as $\alpha_E > \alpha_1$. In such a case the highest possible utility a blockholder investor can achieve, is given by being a liquidity shareholder. Hence, it is sufficient to consider the deviation to liquidity shareholding only. Hence, the incentive compatibility constraint is $U_1^{nBH} \geq U_{l,E}$, where $U_{l,E}$ denotes the utility of a liquidity shareholder when $X_{med} = X_E$.

Note that, the utility of the blockholders increases monotonically in w_E, while the liquidity shareholders' utility is convex quadratic in w_E with the minimum equal to 1 at $\underline{w}_E^E$. However, because the firms' shares are issued only if the initial owner raises the capital, the liquidity shareholders' utility is convex quadratic in w_E for $w_E \geq \underline{w}_E^E$ and it is 1 for $w_E < \underline{w}_E^E$. Hence $U_1^{nBH} \geq U_{l,E}$ iff $\underline{w}_E^E > \underline{w}_E^n$ and $w_E \leq \bar{w}_E$ where $\bar{w}_E$ the biggest solutions of $U_1^{nBH} = U_{l,E}$. Note also that when $\underline{w}_E^E > \underline{w}_E^n$, $\bar{w}_E$ is real and is smaller than $\underline{\underline{w}}_E$ (this can be seen substituting $\underline{\underline{w}}_E$ in $U_1^{nBH} - U_{l,E}$ and verifying that this is always positive). Hence the only relevant condition to ensure the incentive compatibility is $\underline{w}_E^E > \underline{w}_E^n$. This condition is always satisfied when $2\alpha_1 \leq \alpha_E$ and substituting for α_1 from expression (8), this becomes:

$$\alpha_E \leq \hat{\alpha}_1(n) \tag{35}$$

This is the first upper bound in condition (32) .

Finally because an investor wants to hold a block only if he is pivotal in the vote outcome, we need to impose that $\alpha_E \leq \frac{1}{2}$. Suppose $\alpha_E > \frac{1}{2}$ then $X_{med} = X_E$ always, so there is no Blockholder EOS. This gives the second upper bound in condition (32). ■

Claim 2: Suppose conditions (32) and (33) are satisfied. No blockholder wants to increase his shares from α_1.

Proof. Since $w_E \leq \underline{\underline{w}}_E$ by condition (33) his utility U_1^{nBH} is decreasing in α_1 so the blockholder holds just enough to be pivotal. ■

Claim 3: Liquidity Shareholders cannot gain from unilateral deviation.

Proof. No (active) liquidity shareholder has an incentive to hold shares bigger than α_1 in order to change $X_{med} < X_1$, as this would reduce their utility (which is increasing in X and decreasing in shareholdings when $\alpha_i > \alpha_l(X_j)$).

If the investors choose any $\alpha_i < \alpha_1$ they do not change the outcome, hence $\alpha_l(X_1)$ maximizes their utility.

(c) No shareholder, either liquidity or blockholder, is willing to sell his shares for a price lower than what he paid to the initial owner. Excluded investors are willing to buy the shares at a higher price from the liquidity shareholders, but not from the blockholders as this would shift the vote outcome to X_E. Hence the EOS does not change. ■ ■

Corollary 3 *There exists an n-Blockholder EOS, with $X_{med} = X_1 = \bar{X}$, for any pair (α_E, w_E), satisfying the following conditions:*

$$\alpha_E \in \left(\max\left[\bar{\alpha}, \tilde{\alpha}(n)\right], \min\left[\check{\alpha}(n), \hat{\alpha}_2(n), \frac{1}{2}\right]\right] \tag{36}$$

$$w_E \in [\underline{w}_E^n(\alpha_E), \underline{\underline{w}}_E(\alpha_E)) \tag{37}$$

where $1 \leq n \leq M_A$.

Proof. It follows directly from Lemma 12 and taking into account that when $\tilde{\alpha}(n) < \alpha_E < \check{\alpha}_E(n)$, $\alpha_l(\bar{X}) < \alpha_1 < \bar{\alpha}$. ■

Both Lemma 12 and Corollary 3 do not guarantee uniqueness: it is possible that there is an Initial Owner EOS for the same parameters. However when $n = 1$, we can show that there is no Initial Owner EOS for these parameters: one investor is willing to deviate unilaterally from being a liquidity shareholder and hold a block. Hence in this case the Initial Owner EOS cannot be a Nash equilibrium.

We present this case of a one blockholder equilibrium as Corollary.

Corollary 4 *There exists a* 1*-Blockholder EOS, with* $X_{med} = X_1$*, for any pair* (α_E, w_E)*, satisfying the following conditions:*

$$\alpha_E \in (\max[\bar{\alpha}, \tilde{\alpha}(n)], \hat{\alpha}_1(1)] \tag{38}$$

$$w_E(\alpha_E) \in [\underline{w}_E^1(\alpha_E), \underline{\underline{w}}_E(\alpha_E)) \tag{39}$$

Moreover, there does not exist an Initial Owner EOS for any w_E *when* α_E *is in interval (38).*

Proof. The proof of the first part follows directly from Lemma 12 and Corollary 3, setting $n = 1$ and observing that $\min\left[\hat{\alpha}(1), \frac{1}{2}\right] = \hat{\alpha}(1)$. For the second part, suppose to the contrary that there is an EOS with $X_{med} = X_E$ when α_E is in the interval (38). We know from the proof of Lemma 12, that any active liquidity shareholder has an incentive to switch to becoming a single blockholder when $\alpha_E \leq \hat{\alpha}_1(1)$. Contradiction to the Definition 1 of EOS. ■

Lemma 13 *Let* $n \in [1, M_A - N_A]$*. There exists a Liquidity Shareholder EOS, with* $X_{med} = \bar{X}$ *with* $N_A + n$ *active investors for any pair* (α_E, w_E)*, if:*

$$\alpha_E \in \left(\bar{\alpha}, \min\left(\tilde{\alpha}(n), \frac{1}{2}\right)\right] \tag{40}$$

$$w_E(\alpha_E) \geq \underline{w}_E^{LS}(\alpha_E) \tag{41}$$

Proof. We divide the proof in two parts. We first consider the case $n \geq 1$ and then $n = 0$.

(A) $n \geq 1$

Let $U_{l,\bar{\alpha}}$ denote the value function of a liquidity shareholder when $X = \bar{X}$, and he holds the optimal shareholdings. A liquidity shareholder EOS exists iff the following conditions are satisfied: (1) $U_{l,\bar{\alpha}} \geq 1$; (2) No active investor wants to unilaterally increase or decrease his shares; (3) Passive investors maximize their utility conditional on $X_{med} = \bar{X}$. Moreover all active investors are liquidity shareholders. (4) No investor is willing to sell his shares at any price lower than the maximum that excluded investors are willing to pay.

1. By the proof of Lemma 3, $U_{l,\bar{\alpha}} \geq 1$ iff $w_E \geq \underline{w}_E^{LS}(\alpha_E)$.
2. No active liquidity investor wants to increase or decrease his shareholdings since this is the most preferred point (see Lemma 7), as long as $\alpha_E \leq \tilde{\alpha}_E(n)$, since this condition guarantees that $X_{med} = \bar{X}$.
3. Passive investors hold $\alpha_l(\bar{X})$ which maximizes their utility.
4. No investor is willing to sell his shares to any excluded investors as the maximum price at which excluded investors are willing to buy the shares is the minimum price at which the liquidity shareholders are willing to sell.

(B) $n = 0$

In such a case the interval (40) becomes $\alpha_E \in \left(\bar{\alpha}, \frac{\lambda}{1+\lambda}\right]$. In such a case there are sufficiently many active liquidity shareholders that even with liquidity shares, $\alpha_l(\bar{X})$ they can get their most preferred point and $X_{med} = \bar{X}$. The participation constraint of the liquidity shareholders is satisfied when $w_E \geq \underline{w}_E^{LS}(\alpha_E)$. No active liquidity shareholders find it profitable to hold a suboptimal portfolio to switch the decision to $X < \bar{X}$, as this is their first best. Note that no (passive) investors are willing to sell their shares to the excluded investors. The excluded investors cannot improve the vote outcome and hence the maximum price they are willing to buy is equal to the minimum price the shareholders are willing to sell. ■

Corollary 5 *There exists a No Conflicts EOS, with $X_{med} = X_E = \bar{X}$ iff $\alpha_E \in (0, \bar{\alpha}]$, $w_E \geq \underline{w}_E^{LS}(\alpha_E)$.*

Proof. When $\alpha_E \in (0, \bar{\alpha}]$ there are no conflicts of interests between outside investors and the initial owner, $X_{med} = X_E = \bar{X}$. Active investors hold $\alpha_l(\bar{X})$, and are at their most preferred X. Hence they have no incentive to deviate. The participation constraint of the liquidity shareholders is satisfied when $w_E \geq \underline{w}_E^{LS}$. To prove necessity note that if $\alpha_E > \bar{\alpha}$, $X_E < \bar{X}$ hence there are conflicts between investors. If $w_E < \underline{w}_E^{LS}$, no investors would buy the shares hence it is not an EOS. Finally note no investor is willing to sell shares at a price lower than the maximum that an excluded investor will pay. ■

We are now ready to show the Lemma. First note that in any EOS the utility of the initial owner is decreasing in w_E. Further, from the Lemmas above, all the possible EOS are characterized by outside investors who are either liquidity shareholders or blockholders.

Consider the case where no blockholders exist. Suppose to the contrary, that there is an equilibrium with $\frac{K-w_E}{1-\alpha_E} < K$, $\alpha_l(X_j) > 0$ and $U_{l,j} > 1$. By assumption there are sufficiently many investors in the market, so there always exist passive shareholders who have a strictly positive demand for shares for any $0 < X_j \leq \bar{X}$. Hence, the initial owner can increase his utility by decreasing w_E, for any α_E and still ensure that there is a (smaller) positive demand by passive investors, ensuring full subscription. As the demand of the liquidity shareholders is given by equation (7), the initial owner will do this until $\frac{K-w_E}{1-\alpha_E} > K$. Contradiction.

Consider the case where blockholders and liquidity shareholders exist in equilibrium. In such a case either blockholders or liquidity shareholders will have the most binding constraint. We already showed above that when the liquidity investors participation is more binding then $\frac{K-w_E}{1-\alpha_E} > K$. So it is sufficient to show that this is true when the binding constraint is that of blockholders.

When the blockholders's participation constraint is more binding given $m = 1$ the value function for blockholders given $X_{med} = X_j$ is given by :

$$U_1^{nBH} = \bar{R}X_j\alpha_1 + \left(K - \frac{K - w_E}{1 - \alpha_E}\right)\alpha_1 - \frac{\gamma}{2}X_j^2\sigma^2\alpha_1^2 + 1 \tag{42}$$

By the same logic as for the first part, suppose that there is an equilibrium with $\frac{K-w_E}{1-\alpha_E} < K$. Because $\alpha_1 \leq \max[\alpha_j, \bar{\alpha}]$ when $\frac{K-w_E}{1-\alpha_E} < K$, then the participation constraint of the blockholders is satisfied with strict inequality, i.e. $U_1^{nBH} > 1$. The initial owner can decrease w_E and still satisfy the constraints and ensure full subscription. Contradiction to the fact that it is an equilibrium. ■

A.3.2 Proposition 2:

Proof. Before we prove the next proposition, we need few lemmas which provide expressions for the value function of the initial owner under the alternative ownership structures that could be obtained in a subgame perfect equilibrium.

Lemma 14 *Suppose the conditions for the No Conflicts EOS (Corollary 5) are satisfied and the equilibrium of the game is the No Conflicts equilibrium, then the initial owner sets* $\alpha_E = \bar{m}$*,* $X_{med} = X_E = \bar{X}$*,* $w_E = \underline{w}_E^{LS}$ *and the value function of the Initial Owner is given by:*

$$V_E^{NC} = \bar{R}\bar{X} + 1 - \frac{\gamma}{2}\bar{X}^2\bar{m}^2\sigma^2 - \bar{m}K - \epsilon \tag{43}$$

Proof. By Lemma 2, in any monitoring equilibrium, $\alpha_E \geq \bar{m}$. By Corollary 5, the No Conflicts EOS with $X_{med} = X_E = \bar{X}$ exists if $\alpha_E \in (0, \bar{\alpha}]$ and $w_E \geq \underline{w}_E^{LS}(\alpha_E)$. Therefore the maximization problem of the Initial Owner in the No Conflicts equilibrium is:

$$\max_{\alpha_E, w_E} U_E = (\bar{R}\bar{X} + K)\alpha_E - \frac{\gamma}{2}\bar{X}^2\alpha_E^2\sigma^2 + 1 - w_E - \bar{m}K \tag{44}$$

$$\text{s.t} \quad w_E \geq \underline{w}_E^{LS}(\alpha_E) \tag{45}$$

$$\alpha_E \in [\bar{m}, \bar{\alpha}] \tag{46}$$

The initial owner's utility is decreasing in the wealth invested, w_E. Hence he chooses w_E such that it satisfies the participation constraint of the liquidity investors, (45), at equality. Hence $w_E = \underline{w}_E^{LS}$ where $\underline{w}_E^{LS}$ is given by equation (11). Inserting it in the initial owner's objective function we obtain:

$$\bar{R}\bar{X} + 1 - \frac{\gamma}{2}\bar{X}^2\alpha_E^2\sigma^2 - \bar{m}K - \epsilon \tag{47}$$

This expression is decreasing in α_E. Hence the initial owner will retain just enough shares to satisfy the monitoring constraint with equality: $\alpha_E = \bar{m}$. Inserting $\alpha_E = \bar{m}$ in the initial owners utility function, we have expression (43). ■

Let $\underline{b} = \left(\max\left[\bar{\alpha}, \min\left[\frac{1}{2}, \max\left[\hat{\alpha}(1), \tilde{\alpha}\right]\right]\right]\right) + \eta_E$. Remember that η_E is the fraction correspondent to one share when the vote outcome is X_E.

Lemma 15 *Suppose the conditions of the Initial Owner EOS are satisfied (Lemma 11), and the equilibrium of the game is an Initial Owner equilibrium, then* $X_{med} = X_E < \bar{X}$*,* $w_E = \underline{w}_E^E$*,* $\alpha_E = \max[\bar{m}, \bar{b}]$*, and the value function of the Initial Owner is given by:*

$$V_E^E = \bar{R}X_E + 1 - \frac{\gamma}{2}X_E^2\max[\bar{m}, \bar{b}]^2\sigma^2 - \bar{m}K - \epsilon = \frac{\bar{R}^2}{\gamma\sigma^2}\left(\frac{1}{\max[\bar{m}, \bar{b}]} - \frac{1}{2}\right) + 1 - \bar{m}K - \epsilon \tag{48}$$

Proof. The proof follows the same steps as for the proof of Lemma 14 and it applies Lemmas 2 and 11. Detailed proof is available upon request. ■

Let $\underline{c} = \max[\bar{\alpha}, \tilde{\alpha}_E(n)] + \bar{\eta}$ where $\bar{\eta}$ is the fraction correspondent to one share when the vote outcome is $\bar{X}$.

Lemma 16 *Suppose the conditions of the n-Blockholder EOS are satisfied (Lemma 12 and Corollary 3) and the equilibrium of the game is an n Blockholder equilibrium with* $X_{med} = X_1$*, the initial owner sets* $\alpha_E = \max[\underline{c}, \bar{m}]$*,* $w_E = \underline{w}_E^n$ *and the value function of the Initial Owner is given by:*

$$V_E^n = \bar{R}X_1 + 1 - \bar{m}K - \frac{\gamma}{2}X_1^2\sigma^2(\max[\underline{c}, \bar{m}]^2 + \alpha_1 - \alpha_1\max[\underline{c}, \bar{m}]) \tag{49}$$

Proof. The proof follows the same steps as for the proof of Lemma 14 and it applies Lemmas 2 and 12. Detailed proof is available upon request. ■

Lemma 17 *Suppose the conditions for the Liquidity Shareholder EOS are satisfied and there exists a Liquidity Shareholder equilibrium with monitoring (Lemma 13). Then,* $\alpha_E = \max(\bar{\alpha}+\bar{\eta},\bar{m})$*, there are at least* $N_A + n$ *active investors,* $w_E = \underline{w}_E^{LS}$ *and the value function of the Initial Owner is given by:*

$$V_E^{LS} \equiv \bar{R}\bar{X} + 1 - \bar{m}K - \frac{\gamma}{2}\bar{X}^2\sigma^2 \max[\bar{m},\bar{\alpha}+\bar{\eta}]^2 - \epsilon \tag{50}$$

Proof. The proof follows the same steps as for the proof of Lemma 14 and it applies Lemmas 2 and 13. Detailed proof is available upon request. ■

Finally define $\bar{m}_{NC}^{RC}$ and $\bar{m}_{NC}^{S}$ the values such that when $\bar{m}$ is smaller the initial owner prefers monitoring rather than selling or not raising capital. Hence:

$$\bar{m}_{NC}^{RC} \equiv \frac{\sqrt{K\left(K+2\bar{X}^2\gamma\sigma^2\right)} - K}{\bar{X}^2\gamma\sigma^2} \tag{51}$$

$$\bar{m}_{NC}^{S} \equiv \frac{\sqrt{2\bar{R}\gamma\sigma^2\bar{X}^3 + K^2} - K}{\bar{X}^2\gamma\sigma^2} \tag{52}$$

We are now ready to prove the proposition. We solve the game by backward induction. The initial owner chooses α_E and w_E, anticipating the ownership structure. His problem can be broken into the following: (1) $\alpha_E \in [\bar{m},\bar{\alpha}]$, (2) $\alpha_E \in (\bar{\alpha},1]$, (3) $\alpha_E \in [0,\bar{m}]$.

We first describe the beliefs on the EOS and the corresponding value functions in each interval.

Case (1). By Corollary 5 there exists a No Conflicts EOS. Hence in this interval the beliefs of all players on the EOS are the No Conflicts EOS when $w_E \geq \underline{w}_E^{LS}$. As $w_E \leq \underline{w}_E^{LS} < \underline{w}_E^{NT}$ for any possible X there exists a No Trade EOS by Lemma 10, and we assume that the belief is on the No Trade EOS. The initial owner's value function is given by equation (43) when $w_E \geq \underline{w}_E^{LS}$ and by the no trade value function, $V_E^{NT} = 1$ in case $w_E < \underline{w}_E^{LS}$.

Case (2). By Lemmas 11, 12 and 13 the possible EOS in this interval are the Initial Owner, the n Blockholder or the Liquidity Shareholder ones if $w_E \geq \underline{w}_E^{j}$ where $j = \{IO, n, LS\}$. Again if $w_E < \underline{w}_E^{j}$ the No Trade EOS exists and $V_E^{NT} = 1$. If an Initial Owner EOS exists the initial owner sets $\alpha_E = \max[\bar{m},\underline{b}] = \underline{b}$ as $\bar{m} \leq \bar{\alpha}$ and his value function, V_E^{E} is given by equation (48). If an n Blockholder EOS exists, Lemma 16 shows that the initial owner's utility is decreasing in α_E. Observe that for an n Blockholder equilibrium to exist $\tilde{\alpha}(n) > \bar{\alpha}$ otherwise the initial owner would choose $\alpha_E = \bar{\alpha}$. Hence $\underline{c} = \tilde{\alpha}(n)$ and the value function V_E^{n} is given by (49). If a Liquidity Shareholders EOS arises, Lemma 17 shows that $\alpha_E = \max[\bar{\alpha}+\bar{\eta},\bar{m}] = \bar{\alpha}+\bar{\eta}$. Hence the value function V_E^{LS} is given by equation (50).

Case (3). In this interval the Liquidity Shareholders EOS exists as we showed in Proposition 1, hence we assume that the belief is that if $w_E \geq \underline{w}_E^{LS}$ the Liquidity Shareholders EOS emerges. Lemma 2 shows that the initial owner chooses not to monitor in this interval, and by Proposition 1 $\alpha_E = 0$ and the initial owner's value function is $V_E^{NM} = \max\left[1, \bar{R}\bar{X} - K + 1\right]$. If $w_E < \underline{w}_E^{LS}$ then the belief is on the No Trade EOS (Lemma 10) and the corresponding value function is $V_E^{NT} = 1$.

The initial owner will choose α_E to maximize his value function across the intervals (1)–(3) above.

First consider Case (2), ignoring V_E^{NT} for the moment: It is easy to see from (48) that $V_E^{E}|_{\alpha_E=\underline{b}} < V_E^{E}|_{\alpha_E=\bar{m}} = V_E^{NC}$. Hence the initial owner is better off in a No Conflicts equilibrium than in an Initial Owner equilibrium. If an n-Blockholder EOS exists, it is easy to see from equation (49) that $V_E^{n}|_{\alpha_1<\bar{m}} < V_E^{n}|_{\alpha_1=\bar{m}}$. $V_E^{n}|_{\alpha_1=\bar{m}} < V_E^{NC}$ iff

$$\frac{1}{\alpha_1}\left(1+\alpha_E - \frac{\alpha_E^2}{\alpha_1}\right) < \frac{2}{\bar{\alpha}} - \frac{\bar{m}^2}{\bar{\alpha}^2}$$

which is always true as $\bar{m} \leq \bar{\alpha} \leq \alpha_E$. Hence the initial owner is better off in a No Conflicts equilibrium than in an n-Blockholder equilibrium. If a Liquidity Shareholder EOS exists then by Lemma 17 $V_E^{LS}|_{\alpha_E > \bar{\alpha}} < V_E^E|_{\alpha_E = \bar{m}} = V_E^{NC}$. Hence the initial owner is better off in a No Conflicts equilibrium than in an Liquidity Shareholders equilibrium.

Now consider Case (3): The initial owner's value function is $V_E^{NM} = \max\left(1, \bar{R}\bar{X} - K + 1\right)$. Hence he prefers to monitor iff $V_E^{NC} \geq V_E^{NM}$, i.e. iff $V_E^{NC} \geq \bar{R}\bar{X} - K + 1$ and $V_E^{NC} \geq 1$.

This first condition is satisfied when $\bar{m} \in [c, \bar{m}_{NC}^S]$, where $c < 0$. Hence $V_E^{NC} \geq \bar{R}\bar{X} - K + 1$, whenever $\bar{m} < \bar{m}_{NC}^S$.[24] When the above condition is not satisfied the initial owner sells out the firm. The second condition, $V_E^{NC} \geq 1$, is satisfied when $\bar{m} \in [d, \bar{m}_{NC}^{RC}]$, where $d < 0$. Hence, if $\bar{m} > \bar{m}_{NC}^{RC}$ then the initial owner does not raise capital.

Finally, consider the No Trade equilibrium. Clearly this gives the same value to the initial owner as not raising capital, so under the conditions of the proposition, the No Conflicts equilibrium is preferred by the initial owner. ■

A.3.3 Proposition 3:

Proof. The proof follows the same steps as the proof of Proposition 2. Let

$$\bar{m}_E^{RC} \equiv \frac{1}{2}\left(1 - \frac{\bar{R}\bar{X}}{K}\right) - \frac{\bar{R}^2}{4K\gamma\sigma^2} + \frac{\sqrt{16K\bar{R}^2\gamma\sigma^2 + \left(\bar{R}^2 + 2\gamma\sigma^2\left(\bar{R}\bar{X} - K\right)\right)^2}}{4K\gamma\sigma^2} \tag{53}$$

$$\bar{m}_E^S \equiv -\frac{\bar{R}^2}{4K\gamma\sigma^2} + \frac{\bar{R}\sqrt{\bar{R}^2 + 16K\gamma\sigma^2}}{4K\gamma\sigma^2} \tag{54}$$

We break up the maximization problem of the initial owner into the following cases: (1) $\alpha_E \in [\max[\underline{b} + \eta_E, \bar{m}], 1]$, (2) $\alpha_E \in [0, \bar{m}]$. We first describe the beliefs on the EOS and the corresponding value functions in each interval.

Case (1). By Lemma 2 the initial owner monitors. By Lemma 11 an Initial Owner EOS exists in this interval of α_E as long as w_E satisfies condition (31) and we assume that the anticipated EOS is the Initial Owner EOS for $w_E \geq \underline{w}_E^E$. Lemma 15 implies then that: $\alpha_E = \bar{m}$, $w_E = \underline{w}_E^E$ and the initial owner's value function, V_E^E, is given by equation (48). Otherwise when $w_E < \underline{w}_E^E$ the No Trade EOS (with value function V_E^{NT}) is anticipated.

Case (2). This case is the same as in Proposition 2, Case (3). The initial owner's value function is $V_E^{NM} = \max[\bar{R}\bar{X} - K + 1, 1]$.

Maximizing across intervals of Cases (1) and (2) the initial owner will choose $\alpha_E = \bar{m}$ as long as $V_E^E \geq \max(V_E^{NM}, V_E^{NT}) = V_E^{NM}$. This occurs when $\bar{m} \leq \min\left[\bar{m}_E^{RC}, \bar{m}_E^S, 1\right]$.

Note that when $\bar{m} > \left[\frac{1}{2}, \bar{\alpha}\right]$ this is the unique equilibrium (under the conditions of the proposition), induced by the uniqueness of the Initial Owner EOS. ■

A.3.4 Proposition 4:

Proof. Following the same steps as in the proof of Proposition 2, we break up the maximization problem into the following intervals of α_E : (1) $\alpha_E \in \left[\bar{m}, \min\left(\tilde{\alpha}_E(n), \frac{1}{2}\right)\right)$; 2) $\alpha_E \in \left[\tilde{\alpha}_E(n), \min\left[\hat{\alpha}(n), \frac{1}{2}\right], 1\right]$; (3) $\alpha_E > \min\left[\hat{\alpha}(n), \frac{1}{2}\right]$; (4) $\alpha_E \in [0, \bar{m})$. As before, we first describe the beliefs on the EOS in each interval and the corresponding value functions.

Case (1). By Lemma 2, $m = 1$. Then all investors anticipate monitoring in the last stage. We will assume the following beliefs about the EOS in date 1: if $w_E \geq \underline{w}_E^{LS}$ then the anticipated EOS is the Liquidity Shareholder EOS which exists by Lemma 13. If $w_E < \underline{w}_E^{LS}$ the EOS is the No Trade EOS with value function V_E^{NT} (Lemma 10). By

[24] Note that as ϵ is a very small number we just consider a strict inequality.

Lemma 17, if a Liquidity Shareholder equilibrium exists the initial owner's value function, V_E^{LS}, is given by equation (50).

Case (2). If this interval is non-empty, and $w_E \geq \underline{w}_E^n$ there exists an n-Blockholder EOS by Lemma 16 and Corollary 3. The proof of Lemma 16 shows that the initial owner's minimizes α_E and the initial owner's utility function is continuous between these two intervals of α_E and is given by expression (49). Hence he prefers to minimizes α_E, i.e. $\alpha_E = \tilde{\alpha}(n)$. The value function is therefore given by V_E^n, expression (49) with $\underline{c} = \tilde{\alpha}(n) + \bar{\eta}$. If $w_E < \underline{w}_E^n$ then the belief on the EOS is the No Trade EOS, with value function V_E^{NT}.[25]

Case (3). In this case the unique EOS is the Initial Owner EOS for $w_E \geq \underline{w}_E^E$. Using the proof of Lemma 15 we know that the initial owner minimizes α_E, i.e. $\alpha_E = \underline{d} \equiv \min[\hat{\alpha}(n), \frac{1}{2}]$ and the value function is given by V_E^E. If $w_E < \underline{w}_E^E$ the belief on the EOS is the No Trade EOS, with value function V_E^{NT}.

Case (4). This interval is the same as in Proposition 2, Case 3. By Proposition 1, the unique equilibrium is the no monitoring equilibrium and the value function is given by V_E^{NM} (expression (12)).

We now show that the initial owner chooses $\alpha_E = \bar{m}$, i.e. a Liquidity Shareholder EOS. This is true whenever $V_E^{LS} \geq \max(V_E^E, V_E^{NM}, V_E^n, V_E^{NT})$. Because the initial owner's value function is decreasing in α_E, $V_E^E|_{\alpha_E = \underline{d}} < V_E^E|_{\alpha_E = \bar{\alpha}} = V_E^{NC}$. Hence the liquidity shareholder ownership structure of Case (1) is preferred over the initial owner one. Second, as in the proof of Proposition 2, $V_E^n|_{\alpha_E = \tilde{\alpha}(n)} < V_E^n|_{\alpha_E = \bar{m}} < V_E^{NC}$. Hence the liquidity shareholder ownership structure of Case (1) is preferred over the n Blockholder ownership structure of Case (2). Moreover $V_E^{NM} \geq V_E^{NT}$. Hence we only need to check that $V_E^{LS} \geq V_E^{NM}$ and this is true iff $\bar{m} \leq \min\left(\bar{m}_{NC}^{RC}, \bar{m}_{NC}^{S}\right)$. ■

A.3.5 Proposition 5:

Proof. Define as $\bar{m}_{1,n}^{RC}(n)$ and $\bar{m}_{2,n}^{RC}(n)$ the first two biggest solutions of the equation $V_E^n = 1$ and $\bar{m}_{1,n}^{S}(n)$ and $\bar{m}_{2,n}^{S}(n)$ the two biggest solutions of the equation $V_E^n = \bar{R}\bar{X} + 1 - K$. Let:

$$\bar{m}_1^E(n) \equiv \frac{\lambda(1+\lambda) + n(1-\lambda)\left(1 - \sqrt{8(3n-1)\lambda^2 + (4-24n)\lambda} + 1\right)}{11n^2(1-\lambda)^2 + (n(1-\lambda) - (1+\lambda))^2} \tag{55}$$

$$\bar{m}_2^E(n) \equiv \frac{\lambda(1+\lambda) + n(1-\lambda)\left(1 + \sqrt{8(3n-1)\lambda^2 + (4-24n)\lambda} + 1\right)}{11n^2(1-\lambda)^2 + (n(1-\lambda) - (1+\lambda))^2} \tag{56}$$

$\bar{m}_1^E(n)$ and $\bar{m}_2^E(n)$ are the two monitoring costs for which the initial owner is indifferent between being in a n Blockholder equilibrium or in an Initial Owner equilibrium holding the majority of the shares, i.e. $V_E^n = V_E^E\left(\alpha_E = \frac{1}{2}\right)$. For monitoring cost values within this interval $[\bar{m}_1^E(n), \bar{m}_2^E(n)]$ the initial owner prefers to be in an n Blockholder equilibrium.

Following the steps of the proof of Proposition 2 above, we break up the maximization problem of the initial owner into the following intervals of α_E : (1) $\alpha_E \in \left[\bar{m}, \frac{1}{2}\right]$; (2) $\alpha_E \in (\frac{1}{2}, 1]$; (3) $\alpha_E \in [0, \bar{m}]$. We first describe the beliefs on the EOS and the corresponding value functions in each of these intervals.

Case (1). By Lemma 2, $m = 1$. Then all investors anticipate monitoring in the last stage. We assume the following beliefs about the EOS at date 1: if $w_E \geq \underline{w}_E^n$ then the anticipated EOS is the n Blockholder equilibrium which exists by Lemma 12. By Lemma 16 in such a case the initial owner's value function, V_E^n is given by equation (49). If $w_E < \underline{w}_E^n$ then the EOS is the No Trade EOS with corresponding value function V_E^{NT}.

Case (2). By Lemma 2, $m = 1$. In this interval whenever $w_E \geq \underline{w}_E^E$ there exists an Initial Owner EOS. By Lemma

[25] In this interval there can be also an Initial Owner EOS if $n > 1$ and $w_E \geq w_E^E$. In such a case the proof that shows that the initial owner prefers the Liquidity Shareholder EOS follow the same steps as Case (3).

15, he minimizes α_E, i.e. $\alpha_E = \frac{1}{2}$ and his value function becomes:

$$V_E^E = \frac{3}{2}\frac{\bar{R}^2}{\gamma\sigma^2} + 1 - \bar{m}K - \epsilon \tag{57}$$

Case (3). This is the same as Proposition 2, Case 3 and generates a value of V_E^{NM}.

Now we show that the conditions under which the initial owner chooses $\alpha_E = \bar{m}$, i.e. Case (1).

We first check that $V_E^n \geq V_E^E\left(\alpha_E = \frac{1}{2}\right)$. This occurs when:

$$\frac{1}{\alpha_1}\left(1 - \frac{\bar{m}^2}{\alpha_1} + \bar{m}\right) \geq 3$$

This condition is satisfied iff:

$$\frac{1 + \bar{m} - \sqrt{1 + 2\bar{m} - 11\bar{m}^2}}{6} \leq \alpha_1 \leq \frac{1 + \bar{m} + \sqrt{1 + 2\bar{m} - 11\bar{m}^2}}{6} \tag{58}$$

Substituting α_1 we obtain that $V_n^E \geq V_E^E$ iff $\bar{m} \in [\bar{m}_{1,n}^E, \bar{m}_{2,n}^E]$. Note also that in order to guarantee real values of condition (58) $\bar{m}$ has to be below 40%. This means that in order to guarantee an n-Blockholder equilibrium $\bar{m} < 40\%$.

Second we check that $V_n^E \geq V_{NM}^E$:

(i) $V_n^E \geq 1$ iff:

$$n(1-\lambda)\lambda\bar{R}^2 + \bar{m}\left(\bar{m}n^2(1-\lambda)^2\bar{R}^2 - n(1-\lambda)(\lambda\bar{m} + \bar{m} + 1)\bar{R}^2 + 2K\gamma(\lambda\bar{m} + \bar{m} - \lambda)^2\sigma^2\right) < 0 \tag{59}$$

The left hand side is a third degree inequality which goes from $-\infty$ to ∞, and it is positive at $\bar{m} = \frac{\lambda}{1+\lambda}$. Note also that when $\bar{m} = \frac{1}{2}$, the left hand side can be either positive or negative. Hence of the 3 potential roots for which the left hand side is equal to 0, we are interested for the two biggest ones which are defined as $\bar{m}_{1,n}^{RC}$ and $\bar{m}_{2,n}^{RC}$ and the negative values are between these two values, that is $\bar{m}_{1,n}^{RC} < \bar{m} < \bar{m}_{2,n}^{RC}$.

(ii) $V_n^E \geq \bar{R}\bar{X} - K + 1$ iff:

$$\begin{aligned}&2\bar{\alpha}K\gamma(1+\lambda)^2\sigma^2\bar{m}^3 + \left(\bar{R}^2\left(2(1+\lambda)^2 - \bar{\alpha}n(1-\lambda)(\lambda+1)\right) - n(1-\lambda) - 2\bar{\alpha}K\gamma\left(3\lambda^2 + 4\lambda + 1\right)\sigma^2\right)\bar{m}^2 \\ &+ \left(\bar{\alpha}\left(2K\gamma\lambda(3\lambda+2)\sigma^2 - n\bar{R}^2(1-\lambda)\right) - 4\bar{R}^2\lambda(1+\lambda)\right)\bar{m} + \lambda\left(2\lambda\bar{R}^2 + \bar{\alpha}\left(n\bar{R}^2(1-\lambda) - 2K\gamma\lambda\sigma^2\right)\right) < 0\end{aligned} \tag{60}$$

The left hand side has the same features of the left hand side of condition (59). Hence this condition is satisfied when $\bar{m}_{1,n}^S < \bar{m} < \bar{m}_{2,n}^S$ where $\bar{m}_1^S$ and $\bar{m}_2^S$ are the biggest solutions of the left hand side set equal to zero. ∎

A.3.6 Corollary 1:

Proof. It follows directly from Propositions 3 and 5. ∎

A.3.7 Corollary 2:

Proof. This is special case of Proposition 5. Note that in such a case no Initial Owner equilibrium can arise as one active investor is always willing to unilaterally deviate and hold a block and become pivotal. ∎

A.3.8 Proposition 6:

Proof. The initial owner's value function has a maximum for $n = n^*$. As $\bar{m} > \frac{\lambda}{1+\lambda}$, $n > 0$. ∎

A.3.9 Proposition 7

Proof. When the initial owner can raise an amount of capital $I \geq K$ and invest the remaining amount in the risk free asset his objective function becomes:

$$\alpha_E(X\bar{R} + K + I - K) - \frac{\gamma}{2}\alpha_E^2 X^2 \sigma^2 - w_E - \bar{m}$$

The objective function of the investors is instead:

$$(1 - \alpha_i \frac{I - w_E}{\alpha_I}) + \alpha_i \left[X\bar{R} + K + I - K\right] - \frac{\gamma}{2}\alpha_i^2 X^2 \sigma^2$$

Repeating the same steps of Propositions 2, 3, 4, 5, we obtain the optimal w_E. Inserting it in the initial owner objective function, we obtain that the initial owner's objective function is decreasing in I. ■

Good Cop, Bad Cop: Complementarities between Debt and Equity in Disciplining Management*

Alexander Guembel
Toulouse School of Economics (IAE, IDEI)

Lucy White
Harvard Business School and CEPR

11 May, 2010

*We thank Matthias Dewatripont, James Dow, Denis Gromb, Thomas Mariotti, Colin Mayer, Anjan Thakor, Jean Tirole and three anonymous referees for helpful comments. We would also like to thank conference and seminar participants at the AFA 2003, INSEAD, Goethe University in Frankfurt, SOAS, the Stockholm School of Economics, the CEPR workshop on The New Economy and the Oxford Finance Summer Symposium. We are responsible for any errors and omissions in the paper.

Correspondence address: Alexander Guembel, Toulouse School of Economics, Manufacture des Tabacs, allée de Brienne 21, 31000 Toulouse, France, Tel.: +33 561 12 85 74, e-mail: guembel@cict.fr or lwhite@hbs.edu

Good Cop, Bad Cop: Complementarities between Debt and Equity in Disciplining Management

Abstract

In this paper we demonstrate an inherent conflict that can arise in a firm between inducing *ex ante* efficient monitoring and liquidation decisions by outside claimholders. This tension arises because the choice of liquidation decision when firm prospects are uncertain will influence incentives for monitoring to produce information about firm prospects. We show that when high levels of outside monitoring are desirable in order to induce managerial effort, it can be useful to follow an inefficient liquidation policy, because this will provide greater incentives for the monitor. This result in turn has implications for firm capital structure: the quantity of information generated about firm prospects - and hence firm value - can be improved by *splitting* a firm's cash flow into a 'safe' claim (debt) and a 'risky' claim (equity) rather than selling a single claim, precisely because of the *ex post* conflicts of interest between claimholders that this creates. This generates a partial answer to the puzzle raised by Tirole (2001) as to why firms issue multiple securities when this leads to ex post conflicts of interest.

Keywords: Debt, Equity, Soft Budget Constraint, Monitoring.

JEL Classification: D82, G3

1. Introduction

Most firms issue multiple claims in order to finance their activities. This is puzzling, because multiple claims with differing cash flow and control rights generate costly externalities between security-holders. A geared firm, for example, may suffer from asset substitution whereby equity-holders may wish to increase the riskiness of assets in order to transfer wealth from creditors to themselves (Jensen and Meckling (1976)). Similarly, debt overhang may lead to underinvestment by equity-holders (Myers (1977)). Some theories of multiple securities (e.g., Allen and Gale (1988), Gorton and Pennacchi (1990), Boot and Thakor (1993), Fulghieri and Lukin (2001)) show that investor heterogeneity can render multiple claims optimal. Most obviously, trade-off theory says that tax shields provide a rationale for debt but that some equity must also be issued to avoid excessive bankruptcy costs. But all of these theories leave unanswered the corporate governance implications of multiple claims. In particular, as Tirole (2001) argues, if one accepts that the holders of different claims face conflicts of interest with respect to the firm decisions, one might expect firms to try to achieve the best of both worlds by selling a single homogeneous claim to an intermediary which internalizes these conflicts and then achieves the benefits from diverse claims by selling on multiple claims with different cash flows to various different clienteles.

The objective of this paper is to provide a theory of multiple securities based on the corporate governance *benefits* of conflicting interests between multiple claimants. We consider a set-up where the provider or providers of capital have to take two different actions. Firstly, it will be necessary to decide at an interim date whether to liquidate or continue the firm's operations. Secondly, before this interim date arrives, but after the manager has chosen his effort level, it is possible to perform costly monitoring to provide information about the firm's expected value at the interim date, that is, about whether or not liquidation will enhance or destroy value. Intuitively, since there are two different activities for outsiders to perform, it may be useful to have two different outside claimants with different claims. Yet this is not obvious since an aggregate claimant holding all the returns to the firm would have efficient incentives both to collect and to act upon information, and one might expect the benefits to collecting information to be greatest when the informed party has control over decision-making (see, e.g., Aghion and Tirole, 1997). Thus the existence of multiple activities to be carried out by outsiders does not *per se* generate a need for multiple outside claimants. But we show in this paper that it may do so if it is not *ex ante* optimal for the firm to set monitoring and control decisions to the *ex post* optimal levels, because then selling the firm to a single claimant will not achieve the desired outcome.

We demonstrate that there is an inherent tension between providing strong monitoring and liquidation incentives, so that monitoring incentives can be improved by the credible adoption of an *inefficient* liquidation policy. It follows from this that it can be optimal to split the function of exercising control over the firm and monitoring it between two different providers of capital with differing cash flow rights, which provide the right incentives for each of their roles. In other words, the externality which the claim-holder with control over the liquidation decision exerts on the monitoring claim-holder can be seen as part of the corporate design: it may indeed be inefficient to separate claims *ex post*, but it improves efficiency *ex ante*.

Of course, committing to a value-destroying strategy has a cost as well as an increased monitoring benefit, and the cost is that the organisation will typically choose the value-destroying action too often (whenever the monitor is uninformed). Why would one ever wish to impose an ex post inefficient liquidation policy? In our paper, we consider one particular application where this can be useful: the difficulty of inducing managerial effort. Our model thus contains two or potentially three active parties: the manager, the monitor and the party taking the liquidation/continuation decision. We begin by showing, that, as one would expect, the manager will exert more effort if the decision to continue or liquidate the firm is more informed - so that the more informed is the continuation decision, the lower the expected agency rents that must be paid to the manager. Therefore it is useful for the monitor to acquire more information about the future prospects of the firm. We then come to the key point of our paper: we show that the ex post incentive for the monitor to collect information under the efficient continuation policy may be lower than it would be under the inefficient continuation policy. The intuition for this result is that, if monitoring itself is unverifiable, the monitor can only be rewarded according to the final cash flows - which depend on the continuation decision. If the monitor's acquisition of information does not change the continuation decision, it does not change the cash flows which he receives, and so he has no incentive to collect that information. Now, if the firm can commit to employing a continuation decision that is expected to be value-destroying unless the monitor obtains information, then the monitor's information will have a larger expected impact on cash flows than if the firm were committed to following the value-maximising policy. So there is a trade-off between employing a liquidation policy that maximizes the firm's future value and inducing high monitoring effort.

The existence of this trade-off allows us to generate a theory of firm capital structure. In particular, we show that *if* it is desirable to employ an inefficient liquidation decision to motivate high levels of monitoring, then the firm must issue multiple securities: one debt-like (with limited upside) and one equity-like (with limited downside), owned by *independent* claim-holders. Further, when this split is desirable, there will necessarily be an ex post conflict between the claim-holder who is given incentives to monitor and the claim-holder who holds control over the continuation/liquidation decision. We then go on to show that there is a range of model parameters where it is indeed optimal to divide functions, cash flows, and control rights in this way, hence providing one possible resolution to the puzzling question of why firms issue multiple securities outlined above.

In order to draw out the empirical implications of our theory, we need to distinguish between two possible regimes, depending on whether (uninformed) liquidation at the interim date enhances or destroys value. Consider first a firm that is sufficiently profitable ex ante that in the absence of further information the presumption is that operations should be continued. We call this the 'soft budget constraint' case, because the problem is that investors anticipate continuation and therefore have little reason to produce costly information. To be more specific, after the manager has taken his action, the only reason for investors to collect information is to be able to liquidate the firm if prospects turn out to be sufficiently bad – but if the manager has worked, it is unlikely that prospects are bad. The amount of information which a single principal (holding 100% of the claims on the firm) would find it optimal to collect may therefore be too low to provide much incentive to the manager. Therefore a mechanism is required to allow the principal to commit to collecting more information than is ex post optimal. One such mechanism is to divide cash flows into two parts. First, the firm should issue a "tough" claim: "debt", which has control over the firm at the interim date. With little to gain from the upside, but exposure to the downside, the debt-holder is willing to liquidate the firm unless a monitor generates information showing that the firm is doing better than expected. On the other hand, the firm must issue an option- like "equity" claim for the monitor to hold since the monitor has an incentive to produce costly information only if his upside from continuing the firm is sufficiently high relative to his pay-off from liquidation. An equity-holder would be too soft in his continuation policy if he had control in the absence of information, because then he would not liquidate but rather continue the firm in the absence of information. Indeed, the threat of inefficient liquidation makes the equity-holder's monitoring incentives more high-powered. Intuitively, therefore, it may be optimal to split the cash flow rights precisely in order to generate externalities between claim-holders. Moreover, the equity claim needs to be concentrated in the hands of one agent so as to prevent a free rider problem among equity holders to monitor. Our model therefore makes the novel prediction that levering firms up to give creditors substantial control (highly levered firms such as LBO companies, firms with a high proportion of short term debt and firms in financial distress) can create value, but only if equity ownership becomes more concentrated as well. (This contrasts with Jensen's (1986) free cash flow theory, for example, where levering up per se is valuable because it forces managers to pay out cash rather than wastefully re-investing it.)

Our model also applies to the opposite kind of firm, that is, one for which the expected prospects are poor, so that in the absence of news that the firm is performing better than expected, it is optimal to liquidate the firm. We call this the 'start-up' case, because the parameters are such that these firms' value comes from the option to abandon or liquidate at the interim date, together with high cash flows in the unlikely event that the firm is successful. Monitoring incentives in these firms stem from the desire to identify the good state in which the firm should be continued. However, given that the ex ante probability that the firm in fact turns out to be highly profitable is small, the *expected* payoff from identifying the good state may also be relatively small, and so, correspondingly, are monitoring incentives. Nevertheless, monitoring may still be optimal ex ante, because it mitigates the managerial moral hazard problem. If liquidation is too frequent, the manager has no incentive to exert effort. We show that if control is now allocated to a soft claimant (equity) who will continue the firm in the absence of information, monitoring incentives for a tough (risky debt) claimant increase, because they are now driven by the larger payoff that can be obtained from avoiding value-destroying continuation in the bad state. Our theory may help explain the finding of Guedj (2005) and Guedj and Scharfstein (2005) that drug trials inside large companies are terminated more often than when they are carried out by small bio-tech firms that are allied with larger companies. Since they also find that these earlier terminations are, on the whole, efficient, this raises the question why alliances adopt inefficiently soft termination policies. Our theory provides one possible explanation for this finding: allowing the smaller firm some control over the continuation decision simultaneously provides good effort incentives for the bio-tech's manager and good monitoring incentives for the large pharmaceutical firm, whose contract is structured such that it needs to find hard evidence in order to terminate an unprofitable trial.

The plan of this paper is as follows. Section 2 describes the basic model and assumptions. Section 3 derives optimal incentive contracts for the manager, the monitor and the party charged with making the liquidation decision. In Section 4 we show that the optimal policy may require issuing financial claims that are strictly decreasing in the firm's cash flows. Section 5 considers the case where security payoffs must be non-decreasing in cash flows and shows that a division of labour between a monitoring and a controlling agent may be necessary to induce monitoring under some circumstances. It shows that to achieve this the firm must issue a pair of securities that can be interpreted as debt and equity. Section 6 then shows under which circumstances inducing monitoring is worth the sacrifice in the efficiency in the continuation decision. We introduce the possibility of renegotiation in section 7 and section 8 discusses some related literature on security design. Section 9 discusses and interprets our results, and provides empirical implications. Section 10 concludes. Appendix A shows robustness of our main results with respect to two model extensions: (i) renegotiation between multiple claim-holders over the liquidation decision at the interim date, and (ii) allowing the manager to perform the monitoring function. All proofs are relegated to Appendix B.

2. The Basic Model

This section starts with a description of the model set-up, followed by a discussion of key assumptions. We base our model on a simplified version of the set-up used by Dewatripont and Tirole (1994), but extend their model to allow for endogenous monitoring activity.[1]

There are three dates $t = 0, 1, 2$. At date 0, the owner of a production technology can set up a firm to undertake a project requiring up-front investment I. There are two possible date 2 values R_ω of the firm (project), depending on the realization of a random variable $\omega \in \{l, h\}$, where $R_h > R_l$. There is a continuum of risk-neutral agents in the economy who can potentially buy claims in and provide financing for the firm, as well as undertaking monitoring of the firm's prospects and taking decisions as to its continuation (to be explained below). In addition, the firm must employ a manager to run the project. The manager has a zero reservation utility, is risk-neutral, has no wealth and enjoys limited liability. At $t = 0$, the manager of the firm chooses an unobservable effort $e \in \{\underline{e}, \overline{e}\}$, where $0 \leq \underline{e} < \overline{e} \leq 1$. If the manager chooses high effort $\overline{e}$, he pays a non-monetary cost $k(\overline{e}) = \gamma$, whereas low effort $\underline{e}$ comes at zero cost $k(\underline{e}) = 0$. If effort is high, the probability of the high state $\omega = h$, is given by $\overline{e}$, and if effort is low, the probability of the high state is $\underline{e}$. We denote the improvement in the probability of the high state by $\Delta e \equiv \overline{e} - \underline{e}$. The manager also receives a private benefit from control given by b if the firm continues to operate into the second period $t = 2$. For simplicity we assume that all agents' discount rate is zero.

At the interim date 1, after financing but before outcomes, information about the firm's future prospects may become available. This takes the form of a signal $s \in \{l, h, \varnothing\}$ observed by all investors about the future state of the world ω. With probability $\underline{\theta}$ the signal is perfectly informative about the true state of the world ($s = \omega$). With complementary probability $1 - \underline{\theta}$ the signal is uninformative ($s = \varnothing$). There is a monitoring technology which can increase the probability that information is generated. If an agent uses the monitoring technology, then we will say that that agent acts as 'monitor'; in that case the probability of an informative signal $s = \omega$ becoming available increases to $\overline{\theta} > \underline{\theta}$. Again, with complementary probability $1 - \overline{\theta}$ no signal is produced ($s = \varnothing$). We denote by $\Delta\theta \equiv \overline{\theta} - \underline{\theta}$ the improvement in the probability of producing an informative signal that results from monitoring. Monitoring accrues a cost $f(\overline{\theta}) = c \geq 0$ and $f(\underline{\theta}) = 0$ to the agent undertaking this task. We assume that the monitoring activity itself cannot be publicly observed. Moreover, although the signal realization can be observed by all investors, we assume that it cannot be verified by a court. Therefore a monitoring agent will have to be rewarded as a function of the cash flows and not as a function of the signal he produces.

Also at date $t = 1$, but after the signal has been observed, an agent can choose an action C (continue operations) or S (stop) at date 1. Choosing action S corresponds to terminating the project and liquidating the firm, yielding the liquidation value L, implying that the manager will not receive his private benefit from continuation. Choosing action C instead means the firm continues until date 2, when realized returns will be R_h or R_l according to the state of the world, and the manager will receive b. We assume that the manager's presence is essential for the continuation of the firm, so that the firm cannot

[1] In Dewatripont and Tirole (1994) there is no need for a monitor because all claimants on the firm automatically receive a costless signal about the firm's future value.

be continued without him, for example because running the project requires particular skills that only the current manager has.

We refer to the agent who has the right to choose which action is taken as the "controlling stakeholder". Because we have assumed that signal realization is unverifiable, the notion of control has content (i.e., control is valuable) since it is impossible to do away with the controlling party by writing a contract specifying ex ante which action will be taken after which signal. For this reason, the assumption that the state of the world is observable but unverifiable has been central to the literature on incomplete contracts.[2] In our model, the allocation of control (i.e., the right to choose C or S) will be determined endogenously in order to maximise the value of the firm. Note however that the unverifiability of the signal means that the allocation of control will have to be uncontingent in our model (in contrast, for example, to Dewatripont and Tirole 1994).[3]

There is an interaction between the action choice and the need for a monitor. In particular, one reason to collect costly information at the interim date is that the continuation decision will be more informed. We denote by $A : \{l, h, \varnothing\} \to \{C, S\}$ a mapping from the signal realization onto the liquidation decision, i.e. $A(s)$ specifies a signal-contingent liquidation policy. Since we assume that the signal is non-contractible, any such policy will have to be incentive compatible for the controlling stakeholder at the time of taking the decision.

We assume that $R_h > L > R_l + b$ so that it is efficient for liquidation to occur in the low $\omega = l$ state but not in the high state $\omega = h$.

2.1 *The Optimization Problem*

We assume that the ownership rights for the project initially belong to a risk-neutral investor who maximizes his expected payoff by choosing (i) a wage contract for the manager, (ii) a monitoring contract, (iii) and a control contract, specifying who has the right to take the $t = 1$ continuation decision. We assume that the manager has no wealth, but the monitor and the controlling agent are not wealth constrained. Hence, the initial project owner needs to determine a price at which he sells the monitoring and control contracts. Finally, the initial owner needs to decide whether to allocate the monitoring contract and the control rights to the same or to different agents. We denote this choice by an indicator variable $\lambda = 0$, if the continuation decision and the monitoring are performed by one and the same agent, or $\lambda = 1$ if the two functions are performed by two different agents.

The firm's cash flow is verifiable. Contractual payments can therefore be made con-

[2] Corporate finance papers that use this assumption in models of control allocation include Aghion and Bolton (1992), Dewatripont and Tirole (1994), Berglöf and von Thadden (1994), Rajan and Winton (1995), Harris and Raviv (1995), Bolton and Scharfstein (1996) and von Thadden, Berglöf and Roland (2003). Economic theorists have debated the merit of this assumption (see, among others, Tirole (1999), Hart and Moore (1999) and Maskin and Tirole (1999)). More recently, the assumption has been defended on behavioural grounds (Hart and Moore, 2007) and by showing that the mechanisms that may overcome the incompleteness problem are not robust to small deviations from common knowledge of the publicly observed signal (Aghion, Fudenberg and Holden, 2007).

[3] It is, however, straightforward to extend our model to allow for a "more realistic" contingent allocation of control rights. We do not pursue this extension because our purpose is not to explain contingent control (which has already been done by Dewatripont and Tirole 1994 and others) but rather to highlight the division of monitoring and decision-making functions among different claimholders.

tingent on the realization of a variable $\varphi \in \{l, L, h\}$ indicating whether cash flows were $CF_{\varphi=l} = R_l, CF_{\varphi=L} = L$ or $CF_{\varphi=h} = R_h$. All contracts can thus be described as three dimensional vectors, specifying what the signatory to the contract will receive contingent on each possible cash flow realization. The wage contract for the manager is denoted by $\mathbf{w} = (w_l, w_L, w_h)$, and because of limited liability it must satisfy $\mathbf{w} \geq \mathbf{0}$. The monitoring contract is $\mathbf{m} = (m_l, m_L, m_h)$ and the control contract is $\mathbf{a} = (a_l, a_L, a_h)$. We denote the price at which the monitoring and control contracts are offered by $P_{\mathbf{m}}$ and $P_{\mathbf{a}}$, respectively.

Initially, we make no restrictions on the payments $\mathbf{m}$ and $\mathbf{a}$ and to derive optimal contracts for this benchmark. In Section 5 we will impose the restriction that contractual payments must be non-decreasing in firm cash flows.

The manager then chooses his effort level e^* so as to maximize expected utility:

$$e^* \in \arg\max E\left(w_\varphi\right) - k\left(e\right) \tag{1}$$

The manager's participation constraint can then be written as:

$$E\left(w_\varphi | e^*\right) - k\left(e^*\right) \geq 0. \tag{2}$$

The monitor chooses the level of monitoring so as to maximize his expected utility:

$$\theta^* \in \arg\max E\left(m_\varphi\right) - f\left(\theta\right) \tag{3}$$

Similarly, the controlling party chooses the continuation decision so as to maximize his expected payoff:

$$A\left(s\right) \in \arg\max E\left(a_\varphi | s\right), \qquad s \in \{l, h, \varnothing\}. \tag{4}$$

The participation constraints of the monitoring and controlling stakeholders depend on whether they are the same or different agents. Consider first the case where the monitor and the controlling stakeholder are different agents ($\lambda = 1$). The respective participation constraints can be then written as

$$E\left(m_\varphi | \theta^*\right) - f\left(\theta^*\right) - P_{\mathbf{m}} \geq 0, \tag{5}$$

$$E\left(a_\varphi\right) - P_{\mathbf{a}} \geq 0. \tag{6}$$

If the same party who monitors also takes the continuation decision ($\lambda = 0$), the participation constraint changes, because the stakeholder buys only one claim. In particular, his maximum willingness to pay for the claim will be determined by the more binding of the two constraints (5) and (6), which for $\mathbf{m} = \mathbf{a}$ is clearly the constraint (5). In that case, the initial owner receives only one payment for the single claim he sells, given by (5). The total amount of external finance raised can thus be written as $P_{\mathbf{m}} + \lambda P_{\mathbf{a}}$ and it may (or may not) exceed the financing need I.

The initial owner solves the following problem[4]:

$$\max_{\lambda, \mathbf{w}, \mathbf{m}, \mathbf{a}, P_{\mathbf{m}}, P_{\mathbf{a}}} V = E\left(CF_\varphi - w_\varphi - m_\varphi - \lambda a_\varphi\right) + P_{\mathbf{m}} + \lambda P_{\mathbf{a}} - I \tag{7}$$

$$s.t.$$

$$V \geq 0$$

$$(1) - (6).$$

[4]If $V < 0$ for all possible values of the choice variables, then the initial owner will not invest. In order to simplify notation we do not introduce another choice variable to denote this possibility.

The above formulation imposes no restrictions on the identity of the monitor and controller; that is, the investor can choose himself or an outsider as the monitoring and/or the controlling party. In the appendix we will also analyse the case that the manager himself may act as monitor or the party in control of the continuation decision.

As a solution concept for the game we use pure strategy Perfect Bayesian Nash equilibrium, i.e., all agents maximize their expected payoffs given their belief about the strategy chosen by the other players. In equilibrium these beliefs are correct. Moreover, all agents update their beliefs according to Bayes' rule.

3. Optimal Contracting

We begin this section by stating the intuitive result (proved in the appendix) that the set of optimal liquidation strategies must have the following property.

Lemma 1 *It is optimal to structure claims such that $A(h) = C$ and $A(l) = S$.*

We can thus restrict attention to control contracts that always induce continuation of the firm after the high signal ($A(h) = C$), and always induce liquidation of the firm after a low signal ($A(l) = S$). Where liquidation policies differ is whether they continue or liquidate the firm *in the absence of information.* We will say that the firm's liquidation policy is 'soft' if the firm is continued when no signal is available, $A(\varnothing) = C$, and 'tough' if instead the firm is liquidated in the absence of a signal: $A(\varnothing) = S$. For ease of exposition we only allow for pure strategies as liquidation policies. In section 6 we briefly discuss why allowing for mixed liquidation strategies would not change our qualitative results.

Since we are interested in the role of monitoring in facilitating the resolution of the managerial moral hazard problem, we focus on a setting in which it is optimal to induce high managerial effort, if the project is to be financed at all.[5] If low effort were optimal, monitoring would still be valuable as it would improve the quality of the interim continuation decision. However, there would be no interaction between monitoring, control and managerial effort, which is the focus of this paper. A sufficient condition for low effort to be undesirable is that an external provider of capital could not break even in expectation from funding the firm. We can write the expected net cash flows from operating the project (excluding private benefits and the non-monetary cost of managerial effort) under the soft liquidation policy as

$$NCF^C(e,\theta) = eR_h + (1-e)(1-\theta)R_l + (1-e)\theta L - f(\theta) - I,$$

and under the tough policy as

$$NCF^S(e,\theta) = e\theta R_h + (1-e\theta)L - f(\theta) - I.$$

It is then impossible to (at least) break even under low effort if the following conditions hold: When the soft liquidation policy $A(\varnothing) = C$ is employed, the project has a negative

[5] Note that we do not yet make assumptions to ensure that the firm has positive net cash flows when the manager does exert effort. It is more convenient to impose the corresponding restrictions directly in Section 6 when we derive our result (proposition 4). This is because we do not require that the project has positive value when the manager exerts effort under any liquidation policy (and choice of monitoring level), but only that it has positive value under the ex ante optimal liquidation policy.

net cash flow if

$$NCF^{C}\left(\underline{e},\overline{\theta}\right) < 0, \quad \text{and} \tag{8}$$
$$NCF^{C}\left(\underline{e},\underline{\theta}\right) < 0. \tag{9}$$

Similarly, when the tough liquidation policy $A(\varnothing) = S$ is used, the analogous conditions are:

$$NCF^{S}\left(\underline{e},\overline{\theta}\right) < 0, \quad \text{and} \tag{10}$$
$$NCF^{S}\left(\underline{e},\underline{\theta}\right) < 0. \tag{11}$$

We assume throughout the paper that conditions (8)-(11) hold.

In the following, we will distinguish between parameter values where the expected cash flow generated by continuation in the absence of information is larger than the liquidation value (i.e., continuation is efficient from a purely financial perspective) and those where it is smaller (so that continuation is inefficient from a purely financial perspective). Let $\overline{R} \equiv \overline{e}R_h + (1-\overline{e})\,R_l$ denote the expected cash flow generated from continuation when $s = \varnothing$ and the manager has chosen $\overline{e}$. We will say that the soft liquidation policy enhances firm value if $\overline{R} > L$ and that it destroys value if $\overline{R} < L$, and conversely for the tough liquidation policy.[6]

3.1 *Wage payments*

We now derive the wage contract that maximises expected firm value taking the liquidation policy as given and assuming that the manager is not also the monitor (this latter case is covered in the appendix). Since the manager does not have wealth to pay ex ante for the agency rent that he will extract ex post, the firm's founder will be interested in choosing the wage contract which induces the manager to exert high effort at the lowest possible expected cost given the liquidation policy $A(\varnothing) = C$ or $A(\varnothing) = S$.[7] As we will explain below, the amount that the manager will need to be paid to induce effort depends on the probability with which he anticipates that the informative signal will be available (θ). Thus we write the optimal wage contract as a function of the manager's belief about θ.[8] We denote by w_{φ}^{C} and w_{φ}^{S} the wage payments under the soft and the tough liquidation policy, respectively.

[6] Note that our focus on value at the interim date excludes the manager's private benefits. Interim efficiency would be defined with respect to a comparison between $\overline{R} + b$ and L.

[7] To the extent that the manager has wealth that would enable him to pay for the rents that he will earn ex ante, it is clearly desirable for the initial owner to require him to do so. The presence of managerial wealth would change the initial owner's objective function by encouraging to internalise the effect of the liquidation policy on the manager's private benefits. However, as we will see below, the amount of agency rents that the manager can expect to earn is endogenous and will depend on capital structure, monitoring and control choices induced by the initial owner, so allowing for managerial wealth would significantly complicate our analysis. In general, the presence of mangerial wealth can be expected to slacken the manager's incentive constraint and so reduce the need for inefficient continuation decisions - and hence the need for external control. See Grinstein (2003) for theory and evidence on this point.

[8] Remember that the monitoring activity is not contractible, so wage payments are not actually conditioned on θ. The notation merely indicates that the optimal wage policy depends upon the expected monitoring intensity.

Lemma 2 *The wage contracts that result in the lowest expected payment to the manager whilst rendering high managerial effort incentive compatible are as follows:*

(a) under the soft liquidation policy $(A(\varnothing) = C)$

$$w_h^C(\theta) = \max\{\frac{\gamma}{\Delta e} - \theta b, 0\}, \tag{12}$$
$$w_L^C(\theta) = w_l^C(\theta) = 0,$$

and

(b) under the tough liquidation policy $(A(\varnothing) = S)$

$$w_h^S(\theta) = \max\{\frac{\gamma}{\theta \Delta e} - b, 0\}, \tag{13}$$
$$w_L^S(\theta) = w_l^S(\theta) = 0,$$

where $\theta \in \{\underline{\theta}, \overline{\theta}\}$.

When the private benefits of control b are large ($b \geq \frac{\gamma}{\theta \Delta e}$), implicit incentives are sufficient to motivate the manager to exert high effort $\overline{e}$. That is, the threat to liquidate the firm after receiving negative information $s = l$ results in a loss of private benefits and provides an incentive for the manager to exert effort. This is in line with previous work on the ex ante effort incentives provided by a termination threat.[9] On the other hand, the liquidation policy when the decision maker is uninformed ($s = \varnothing$) does not affect the manager's effort incentive compatibility constraint. This is because the manager's effort level cannot change the liquidation decision when the latter is uninformed. Hence, the liquidation threat improves implicit incentives in our setting to the extent that a signal is obtained, and not when there is no information.

Implicit incentives thus work more effectively when the signal quality is higher. In (12) and (13), a larger θ reduces the requirement for explicit wage payment. Monitoring can therefore help to resolve the agency problem between the investor and the manager. Again, this is a standard result in this type of model (see, for example, v. Thadden, 1995).

Note that in our set-up the liquidation policy has, by construction, no direct impact on managerial incentives. All effects will therefore stem from the liquidation policy's impact on monitoring incentives.[10]

3.2 *Monitoring and control incentives*

Consider now the contract that induces monitoring, again as a function of the monitor's belief about the level of effort e exerted by the manager. As before, we distinguish between the soft and tough liquidation policy cases:

[9] See, for example, Dewatripont and Tirole, 1994, Fluck, 1998, and others. Khanna (1998) studies a herding model where shareholders contracting with a managers can include a penalty component in the contract, which may be similar to termination in its effect on incentives. We thank the referee for drawing our attention to this paper.

[10] This contrasts with Dewatripont and Tirole (1994) who assume an information structure that allows for inefficient termination to improve incentives directly. In their model, a "medium" signal indicates that termination is slightly inefficient, but at the same time the probability that the manager has worked hard is not very high, so it is worthwhile to sacrifice efficiency to improve managerial incentives.

Lemma 3 *Under the soft liquidation policy $A(\varnothing) = C$ the monitor chooses $\overline{\theta}$ if and only if*

$$m_L^C - m_l^C \geq \frac{c}{(1-e)\Delta\theta}. \tag{14}$$

Under the tough liquidation policy $A(\varnothing) = S$ the monitor chooses $\overline{\theta}$ if and only if

$$m_h^S - m_L^S \geq \frac{c}{e\Delta\theta}. \tag{15}$$

The monitor's incentives derive from the payoff received *in the states in which the signal changes the liquidation decision*. Under the soft liquidation policy, the default choice is to continue the firm, so the monitor has no incentive to learn that the firm's prospects are good (R_h), because then the firm will be continued anyway and he will receive m_h whether or not he provided information. The payoff m_h therefore does not enter the monitor's incentive compatibility constraint. The monitor's information will change the decision only when it is negative (suggesting that the firm will earn only R_l from continuation), so any incentive compatible payoff to the monitor must reward the monitor when the firm is liquidated (m_L) compared to when it is continued in the bad state (m_l).

Conversely, under the tough liquidation policy, a positive signal that the outcome will be R_h now improves the liquidation decision by allowing value-enhancing continuation. The monitor's incentive payment m_h therefore needs to reward him for identifying the high state instead of accepting the 'default' payment m_L. As a result, the monitor's payoff in the low state m_l is immaterial for his monitoring decision.

Finally, it is straightforward to set out the payoff requirements a_h, a_L, and a_l to the controlling stakeholder, such that either the soft or the tough policy is incentive compatible. Assume that in case of indifference, the controlling stakeholder chooses the action that maximizes expected future cash flows generated by the firm.[11] Since (according to lemma 1) we always want to continue the firm after the high signal, but liquidate it after the bad signal we require

$$a_h \geq a_L \geq a_l. \tag{16}$$

The soft liquidation policy is incentive compatible if the controlling party prefers continuation in the absence of information:

$$\overline{e}a_h + (1-\overline{e})a_l \geq a_L, \tag{17}$$

whereas the tough liquidation policy is incentive compatible when the above inequality (17) is reversed.

4. Optimal Policy Choice for the initial owner: a benchmark

In this section we solve for the optimal liquidation and monitoring policies in the set-up described above. This will serve as a benchmark for our later analysis, in which we will impose an additional restriction on the monotonicity of claims. First we show that,

[11] We will discuss the role of this assumption when we allow for renegotiation in section 7.

depending on parameter values, it may be optimal to employ either the tough or the soft liquidation policy, and to induce either high or low monitoring. Then we consider how securities must be designed in order to make the optimal policy incentive compatible.

4.1 *Optimal policy choice*

We first write down payoffs to the initial owner as a function of the liquidation policy and the level of monitoring. Since providers of capital are risk-neutral and not wealth-constrained, neither the monitor nor the controlling party can capture any agency rent for their activity. Moreover, providers of capital are in perfect competition, so the initial owner will optimally offer payments so as make the participation constraints (5) and (6) just binding (or just (5) if $\lambda = 0$). It follows that at the optimum $P_{\mathbf{m}} + \lambda P_{\mathbf{a}} = E(m_\varphi + \lambda a_\varphi) - f(\theta)$. Therefore any extra cash raised over and above I has no direct implication for the initial owner (that is, it does not affect the payment that he will receive for the sale of the firm). In principle, however, allowing the owner to raise funds in excess of I might be important if it relaxes some incentive compatibility constraints (though it turns out that it does not do so).

We can write the initial owner's expected payoff depending on whether payments are structured such that the resulting liquidation policy is soft or tough and as a function of the level of monitoring induced. We denote the corresponding payoff by $V^C(\theta)$ and $V^S(\theta)$ respectively; these can be calculated from (7) as

$$V^C(\theta) = NCF^C(\overline{e}, \theta) - \overline{e} w_h^C(\theta), \tag{18}$$

$$V^S(\theta) = NCF^S(\overline{e}, \theta) - \overline{e}\theta w_h^S(\theta). \tag{19}$$

Consider first which liquidation policy is preferred for a given level of monitoring. The following is straightforward to verify:

$$V^C(\underline{\theta}) \geq V^S(\underline{\theta}) \Longleftrightarrow V^C(\overline{\theta}) \geq V^S(\overline{\theta}) \Longleftrightarrow \overline{R} \geq L.$$

Intuitively, the soft liquidation policy maximizes the initial owner's expected payoff if and only if $\overline{R} \geq L$, that is if and only if continuation is expected to be more valuable than stopping in the absence of information.

Now let us consider when the initial owner would prefer to induce monitoring. Suppose $\overline{R} \geq L$ and therefore the soft policy is preferred by the initial owner. The condition $V^C(\overline{\theta}) \geq V^C(\underline{\theta})$ can then be re-written as

$$L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta} - \frac{\overline{e}\left(w^C(\underline{\theta}) - w^C(\overline{\theta})\right)}{(1-\overline{e})\Delta\theta}. \tag{20}$$

The left-hand side of the inequality measures the improvement in expected cash flows that results from a more informed liquidation decision. The first expression on the right-hand side of the inequality measures the cost of monitoring, while the second expression accounts for the reduction in expected wage payments that monitoring generates.

Now suppose $\overline{R} < L$. We can write down a similar condition for high level of monitoring under the tough policy, i.e., the condition $V^S(\overline{\theta}) \geq V^S(\underline{\theta})$ can be re-written as:

$$R_h - w^S(\overline{\theta}) - L \geq \frac{c}{\overline{e}\Delta\theta} - \frac{\underline{\theta}\left(w^S(\underline{\theta}) - w^S(\overline{\theta})\right)}{\Delta\theta}. \tag{21}$$

Again, we have the expected increase in cash flows from monitoring (net of the wage payment in the high state), while the right-hand side accounts for the cost of monitoring and the reduction in expected wage payments.

We thus get four possible regimes depending on whether the soft or the tough liquidation policy is optimal, and on whether monitoring is optimal. The following proposition summarizes this result.

Proposition 1 *The optimal liquidation and monitoring regime is as follows. The soft liquidation policy is optimal if and only if $\overline{R} \geq L$. The high level of monitoring is optimal under the soft liquidation policy if (20) holds and under the tough policy if (21) holds.*

4.2 *Security Design*

We now consider how the optimum can be implemented. In particular, we are interested in whether the initial owner can achieve the optimum by issuing a single type of claim, or whether multiple claims (with potentially conflicting incentives) must be issued. Suppose first that $\overline{R} \geq L$ and (20) holds so that the soft policy with monitoring is optimal. Suppose also, for the sake of argument, that (apart from the manager, who receives wage payments) there is only one claimant on all of the firm's cash flows, so that this claimant takes both the monitoring and the liquidation decisions. This claimant must therefore receive the following (residual) claim: $m_h = R_h - w^C\left(\overline{\theta}\right)$, $m_L = L$ and $m_l = R_l$. It is then clear that the constraint (20) is easier to satisfy than (14). That is to say, there are cases where monitoring increases the initial owner's expected payoff ex ante, but it is not incentive compatible for the monitor to monitor once he has purchased the claim from the owner. This is because the anticipation of monitoring reduces the wage payment necessary to induce high managerial effort. But once the manager has actually chosen the high effort level, it is no longer worthwhile to incur the monitoring cost. To put it differently, monitoring has an ex ante effect (reducing expected wage payments) and an interim effect (better liquidation decision), but the incentives that a monitor has are purely determined by its interim effect.

One way to restore the monitoring incentive compatibility is to increase the monitor's payoffs between states L and l to more than $L - R_l$. This, however, can only be achieved by introducing another claimant (a "budget breaker") whose payoffs are decreasing in the firm's underlying cash flows. Thus the optimum cannot be implemented by issuing a single claim.

Under the tough policy a very similar situation can arise. To see this, suppose that $\overline{R} < L$ and the aggregate claimant receives $m_h = R_h - w^S\left(\overline{\theta}\right)$, $m_L = L$ and $m_l = R_l$. But then there are again parameter values such that (21) holds, but (15) is violated. Similar to the previous case, restoring monitoring incentive compatibility requires setting $m_h - m_L$ above $R_h - w^S\left(\overline{\theta}\right) - L$, which necessitates a budget breaker whose payoffs are decreasing in firm cash flows.

The above reasoning is summarized in the following proposition.

Proposition 2 *When*

(a) $\overline{R} \geq L$ *and* $\frac{c}{(1-\overline{e})\Delta\theta} > L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta} - \frac{\overline{e}\left(w^C(\underline{\theta}) - w^C\left(\overline{\theta}\right)\right)}{(1-\overline{e})\Delta\theta}$, *or*

(b) $\overline{R} < L$ *and* $\frac{c}{\overline{e}\Delta\theta} > R_h - w^S\left(\overline{\theta}\right) - L \geq \frac{c}{\overline{e}\Delta\theta} - \frac{\underline{\theta}\left(w^S(\underline{\theta}) - w^S\left(\overline{\theta}\right)\right)}{\Delta\theta}$

then if a single agent is to perform both monitoring and controlling functions, any optimal financial structure must also feature a claimant (budget breaker) whose payoff is strictly decreasing in the firm's cash flows.

This proposition can be interpreted as explaining why some firms choose to issue securities which are decreasing in firm cash flows. One way to implement the solution outlined above would be to have the budget breaker buy a put option on the firm. (For other explanations of why firms may issue put options, see Gibson, Povel and Singh (2004), Atanasov et al (2004), and Chakraborty and Yilmaz (2009).) Whilst the practice of firms issuing put options on their own stock is not widespread, in 1991 the SEC issued a "no action" letter which allowed companies to write put options on their own stock under certain conditions. In particular, they must have announced an on-going share repurchase programme: the outstanding put options are then considered to be part of this repurchase programme. As pointed out by Atanasov et al (2004), this may make it difficult for firms to use put-options as a means to raise finance. Clearly, if we take our model literally, the initial owner cannot announce a share repurchase programme at the same time that he is issuing shares to the monitoring party. Therefore, in the next section, we investigate an alternative means to solve the principal's problem in which we impose the constraint that the firm issues claims which are monotonic in firm value.[12] We believe that this leads to a more realistic prediction about the type of claims that firms will issue.

5. Monitoring feasibility with non-decreasing claims

We now turn to an analysis of the case where the firm cannot issue claims that are decreasing in underlying cash flows. This assumption is in line with much of the existing literature on optimal security design[13] and is consistent with the observation that financial contracts between a firm and its claimholders in practice usually display the property that payoffs are non-decreasing in underlying firm cash flows. It is outside the scope of this model to explain this feature of contracts; instead we take it as a constraint and investigate its implications for the interaction between monitoring and liquidation incentives. Before doing so, we briefly note that, in addition to the legal constraints on contracting mentioned in section 4.2 above, the literature has suggested a number of reasons why in practice financial securities' payoffs are usually non-decreasing. Firstly, decreasing payoffs may provide some security holders with an incentive to sabotage the firm.[14] Secondly, if it is possible to manipulate performance upwards (for example by hidden borrowing) then payoffs that are decreasing in performance might be undermined

[12] Note that financial market participants can also have decreasing payoffs when they hold a short position in a firm's shares. However, the total size of such positions tends to be a relatively small fraction of firm value. According to Desai et.al. (2002) the median short interest as a fraction of shares outstanding in Nasdaq firms was between 0.08% and 0.16% in the years 1988-1994. Hence, it is doubtful whether allowing in our analysis for short positions or derivatives that may have decreasing payoff functions would relax the monitor's incentive compatibility constraint substantially, once one takes into account the relatively small fraction of such (unhedged) instruments outstanding in practice.

[13] See for example Harris and Raviv, 1989, Innes, 1990, Nachman and Noe, 1994 and Duffie and DeMarzo, 1999. For a recent contribution which eschews the monotonicity assumption, see Hellwig (2009).

[14] Such sabotage need not be direct. For example, in Goldstein and Guembel (2008), speculators can make manipulative trades that drive down share prices and ultimately firm value.

by such manipulation. For instance, in our model, the agent charged with monitoring and controlling in section 4.2 above has clear incentives to lend secretly to the firm to try to conceal the poor performance R_L. Avoiding sabotage or manipulation may be costly, so removing the incentives for such actions from securities' payoff structures may be optimal.

In this section we outline the parameter values for which it is possible to write an incentive compatible monitoring contract under the additional constraint of non-decreasing claims. We say that $\overline{\theta}$ is *feasible*, if the parameter values are such that it is possible to write a contract that renders $\overline{\theta}$ incentive compatible for the monitor. Lemma 3 characterizes incentive compatible payments to a monitor, where the incentive compatibility constraint depends on the liquidation policy. The constraint that financial claims be non-decreasing now imposes the additional restriction that

$$m_h^S - m_L^S \leq R_h - w_h^S\left(\overline{\theta}\right) - L$$

under the tough policy, and

$$m_L^C - m_l^C \leq L - R_l$$

under the soft liquidation policy.[15]

Suppose now that $\overline{R} \geq L$ so that it would be optimal to use the *soft* policy in the benchmark of the previous section. If $L - R_l < \frac{c}{(1-\overline{e})\Delta\theta}$, then it is not possible to induce monitoring under the soft policy when all claims must be non-decreasing. In this case, however, monitoring may be feasible if control is allocated to an agent other than the monitor who has incentives to implement the *tough* liquidation policy. We will explain the reasons behind this in the next subsection.

A similar argument applies to the case where $\overline{R} < L$, and $\frac{c}{\overline{e}\Delta\theta} > R_h - w^S\left(\overline{\theta}\right) - L$. The tough liquidation policy is optimal but under this policy monitoring is not incentive compatible. By contrast, monitoring may be incentive compatible under the soft liquidation policy, but implementing this policy requires control to be given to an agent other than the monitor.

The next proposition shows for which parameter regions monitoring is feasible and when feasibility requires a split of control from monitoring.

Proposition 3 *If*

(i) $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$ *and* $R_h - L \geq \frac{c}{\overline{e}\Delta\theta} + w_h^S(\overline{\theta})$ *then* $\overline{\theta}$ *is feasible regardless of the liquidation policy. Control rights can (but need not) be allocated to the monitor* $(\lambda = 0, 1)$.

(ii) $L - R_l < \frac{c}{(1-\overline{e})\Delta\theta}$ *and* $R_h - L > \frac{c}{\overline{e}\Delta\theta} + w_h^S(\overline{\theta})$ *then* $\overline{\theta}$ *is feasible only if the tough liquidation policy is chosen and control rights are allocated to another stakeholder* $(\lambda = 1)$,

(iiia) $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$ *and* $\frac{c}{\overline{e}\Delta\theta} + w_h^C(\overline{\theta}) \leq R_h - L < \frac{c}{\overline{e}\Delta\theta} + w_h^S(\overline{\theta})$ *then* $\overline{\theta}$ *is feasible only if the soft policy is chosen. Control rights can be allocated to the monitor* $(\lambda = 0, 1)$.

(iiib) $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$ *and* $w_h^C(\overline{\theta}) < R_h - L < \frac{c}{\overline{e}\Delta\theta} + w_h^C(\overline{\theta})$ *then* $\overline{\theta}$ *is feasible only if the soft policy is chosen and control rights are allocated to another stakeholder* $(\lambda = 1)$.

(iv) For all other parameters $\overline{\theta}$ *is not feasible.*

[15] There are, of course, the additional constraints $m_L^S - m_l^S \geq 0$ and $m_h^C - m_L^C \geq 0$. Since the payments m_l^S and m_h^C do not affect incentive compatibility, these constraints can always be satisfied.

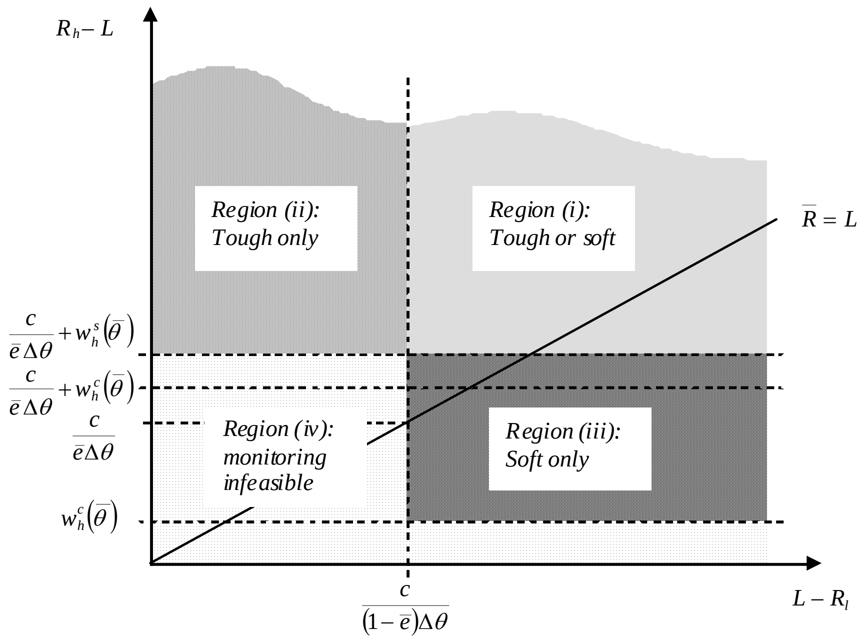


Figure 1: Shows parameter regions for which monitoring is feasible under (i) either liquidation policy, (ii) only the tough policy, (iii) only the soft policy, or (iv) never.

Proposition 3 identifies four regions. It may be feasible to induce monitoring under (i) either liquidation policy, (ii) the tough policy only, (iii) the soft policy only, or (iv) monitoring is never feasible. These regions are depicted in figure 1. Proposition 3 also sets out conditions under which, in each of the regions of figure 1, it will be necessary to split the functions of monitoring and control. These additional partitions are illustrated in figure 2. It can be seen that for much of the illustrated range, it will be necessary to divide functions in this way, and that whether it is necessary to do so depends on the expected cash flows from continuation compared to liquidation (shown by the line $\overline{R} = L$). In the remainder of this section, we provide some intuition for why it is sometimes necessary to split the function of monitoring from control over the liquidation decision, and how this can be done using capital structure. We have two cases to consider: monitoring under the tough liquidation policy and monitoring under the soft liquidation policy.

5.1 *The Tough Liquidation Policy and the Soft Budget Constraint Problem*

Part (ii) of proposition 3 above suggests that when we are to the left of the vertical line $\frac{c}{(1-\bar{e})\Delta\theta}$ in figure 2, a tough liquidation policy is necessary to induce monitoring. But above the diagonal line $\overline{R} = L$, the tough liquidation policy is also value-destroying. The proposition claims that the two functions - liquidation choice and monitoring cannot both be performed by a single investor. The reason for this is the following.

Corollary 1 *Suppose parameters are in region (ii) and the monitor's contract satisfies the incentive compatibility constraint (15). If the monitor also had control rights he would always continue the firm at the interim date following $s = \varnothing$, i.e., he would implement* *the soft liquidation policy. Anticipating this he would not monitor. In consequence, if the firm owner wants to ensure that monitoring is undertaken, he will have to make sure that the tough policy, which is ex post value-destroying, is adopted.*

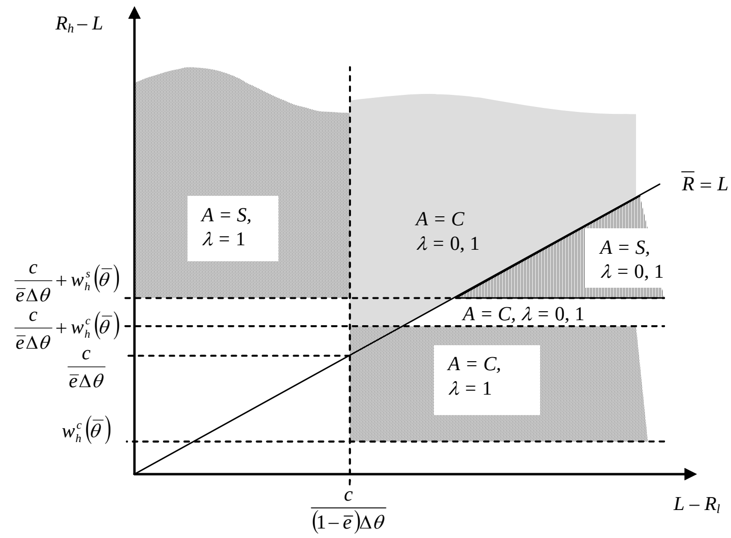


Figure 2: Shows the regions of $R_h - L$ and $L - R_l$ for which monitoring is feasible under the soft and the tough liquidation policies. To the East of $\frac{c}{(1-\overline{e})\Delta\theta}$ monitoring is feasible under the soft, and to the North of $\frac{c}{\overline{e}\Delta\theta} + w_h^s(\overline{\theta})$ under the tough liquidation policy. When monitoring is feasible under both policies, the diagram proposes the policy that maximizes expected future cash flows (soft above the diagonal line $\overline{R} = L$ and tough below it). Moreover, the regions indicate whether a split of control from the monitor is necessary to achieve monitoring ($\lambda = 1$).

To gain intuition for this result, consider for simplicity the case where the original investor remains the only claimholder in the firm. In region (ii) a single claimholder would find it optimal to continue the firm in the absence of information. For reasons that will become obvious in the following, we will refer to this as the 'soft budget constraint' case, and it occurs when

$$\overline{e}(R_h - w_h^c(\overline{\theta})) + (1 - \overline{e})R_l > L. \tag{SBC}$$

The investor's incentives to monitor in this situation are given by (14) with $m_L = L$ and $m_l = R_l$:

$$(1 - \overline{e})(L - R_l) \geq \frac{c}{\Delta\theta}. \tag{22}$$

When all claims on the firm are held by a single party, that party's incentives to monitor come from the fact that with better information, the loss of value $(1 - \overline{e})(L - R_l)$ from mistaken continuation can be avoided more often. This observation can be contrasted with monitoring incentives if the tough policy were followed (e.g., equation (15)). In that case monitoring incentives come instead from ensuring continuation when

$s = h$, and a single claimant's monitoring incentive compatibility constraint would be given by:

$$\overline{e}(R_h - w_h^c(\overline{\theta}) - L) \geq \frac{c}{\Delta\theta}. \tag{23}$$

Comparing the two expressions, we can see that if the "downside" from continuing, $(1-\overline{e})(L-R_l)$ is relatively small compared to the "upside" $\overline{e}(R_h - w_h^c(\overline{\theta}) - L)$, then a single investor has less incentive to monitor under the soft continuation policy than under the tough continuation policy. A single investor would benefit from committing to a larger monitoring effort in order to reduce managerial wages. But unfortunately, under the assumption of a soft budget constraint, he cannot do so when he also has control of the continuation decision since he prefers ex post to follow the soft policy. Intuitively, this is why splitting monitoring from control can be helpful.

The starkness of the dilemma is revealed by rearranging (SBC) to show that it is equivalent to $\overline{e}(R_h - L - w_h^c(\overline{\theta})) > (1-\overline{e})(L-R_l)$, i.e., the LHS of 22 < the LHS of 23. In other words, *the investor would have more incentive to monitor the firm under the tough policy precisely when the soft policy is ex post optimal.* This inherent conflict arises because a project with large upside and limited downside automatically features (a) a strong incentive to continue and (b) very limited incentives to monitor to avoid inefficient continuation. The only way to improve incentives to monitor is delegate control of the liquidation decision such that the project may be inefficiently liquidated in the absence of information, because then monitoring incentives will instead come from the desire to benefit from the relatively large upside of the project.

5.2 *Implementation of the Tough Policy Using Capital Structure*

We next explore how the separation of ownership and control just outlined can be implemented using capital structure.

Lemma 4 *When $L - R_l < \frac{c}{(1-\overline{e})\Delta\theta}$, the regime $\left\{A(\varnothing) = S, \overline{\theta}, \overline{e}\right\}$ can be implemented by splitting the firm's capital structure into debt and equity, where debt has face value D, with $R_l < D \leq \min\left\{\frac{L-(1-\overline{e})R_l}{\overline{e}}, R_h - w_h^S(\overline{\theta}) - \frac{c}{\overline{e}\Delta\theta}\right\}$. Moreover, control rights are assigned to the debt-holder and monitoring is carried out by the equity-holder.*

From proposition 3 we know that the only way in which monitoring can be implemented for $L - R_l < \frac{c}{(1-\overline{e})\Delta\theta}$ is by splitting the capital structure into two claims, with one claim-holder performing the monitoring, and the other having control over the liquidation decision. When the goal is to implement a tough liquidation policy, an optimal capital structure is to give control to the holder of a debt-like claim (the 'bad cop'), which implements a tough liquidation policy even though continuation might be efficient - because this claim-holder does not benefit from the upside of continuation. The claim is debt-like in that it is relatively flat and has control over the liquidation decision. On the other hand, monitoring is carried out by the holder of an equity-like claim (the 'good cop'). This separation of cash flows generates an externality between claim-holders at the interim date whereby the creditor, who has control, shuts down the firm in a state of the world ($s = \varnothing$) in which the equity-holder would have preferred continuation. However, as explained above, this interim inefficient shut-down decision is necessary in order to provide the equity-holder with sufficient incentives ex ante to produce information about the firm's prospects.

Note that our model does not allow us to pin down an optimal security uniquely. We focus on the claims set out in lemma 4 because these claims implement the optimum and can readily be interpreted as debt and equity. However, there are other claims that could implement the above solution, for example, the controlling party could hold a claim that pays 0 when cash flows are R_l and a positive constant when cash flows are either L or R_h (subject to constraints similar to those identified in lemma 4). Note, however that for the parameter range identified in lemma 4, all optimally designed securities will feature the negative externalities between claimholders that we highlight.

5.3 *Soft liquidation and monitoring incentives*

We now consider the region identified by proposition 3, in which it is feasible to induce monitoring only under the *soft* liquidation policy: region (iiib) where $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$ and $w_h^C(\overline{\theta}) < R_h - L < \frac{c}{\overline{e}\Delta\theta} + w_h^C(\overline{\theta})$. Referring back to figure 2 and the proposition, it is necessary to split control from the monitoring function in this region. Notice that, in direct contrast to the soft budget constraint problem outlined in the previous section, in this region parameters are such that the project is not worth continuing in the absence of information, so the soft liquidation policy destroys value and it would be better to follow the tough liquidation policy.

Corollary 2 *Suppose parameters are in region (iiib) and the monitor's contract satisfies the incentive compatibility constraint (14). If the monitor also had control rights he would always stop the firm at the interim date following $s = \varnothing$, i.e., he would implement the tough liquidation policy. Anticipating this he would not monitor. In consequence, if the firm owner wants to ensure that monitoring is undertaken, he will have to make sure that the soft policy, which is ex post value-destroying, is adopted.*

Intuitively, the problem is now the reverse of that outlined in section 5.1. If a single claim-holder were entitled to all the returns on the firm, he would want to liquidate in the absence of information, precisely because the expected upside gain is, by hypothesis, relatively small compared to the downside loss. But this implies that the payoffs available to provide monitoring incentives under the efficient tough liquidation policy are limited, because the benefit of monitoring then is that sometimes one receives the high signal and learns that in fact the project is indeed worth continuing, and by hypothesis this happens rarely (or the expected gains are small). By contrast, if the value-destroying soft policy were used, incentives to monitor come from avoiding the downside which, since continuation destroys value, must be relatively large. So, once again, *it may be necessary to use the soft liquidation policy to induce monitoring precisely when that policy destroys value.* This is what the lower right rectangle in figure 2 illustrates.

If, in this region of the parameter space, all claims on the firm were held by a single investor, he could not commit to a soft continuation policy, and hence would be unable to implement a high monitoring effort. However, monitoring can be achieved by allocating control to a party that is soft in liquidation, i.e., a junior claimant, and getting a senior claimant to monitor. The threat of inefficient continuation now forces the senior claimant to find negative information to persuade the controlling party to liquidate the firm. In this case the claims that ensure monitoring and the appropriate liquidation incentives are given in the following lemma.

Lemma 5 *In region (iiib), $\overline{\theta}$ is feasible only if claims (capital structure) are split into a junior claim with control rights, such as common equity, and a risky senior claim, for example debt or preferred shares, with face value L.*

Compared to the case in section 5.2 where the liquidation policy had to be excessively tough to induce monitoring there is still a gain to splitting firm returns into debt (or another senior claim) and equity (or another junior claim with control rights, such as preferred stock or convertible debt with control rights attached). Again, the purpose is to generate an ex post externality between claims that will both result in an inefficient liquidation policy, and, as a consequence, motivate greater monitoring effort. Relative to the previous case, however, the monitoring and control functions are now switched. The capital structure which now achieves monitoring requires control to be placed in the hands of the residual, junior claim, for example common equity. This claim benefits from the upside of continuation, but is worthless in liquidation and therefore ensures that the liquidation decision is soft. Monitoring is best carried out by a senior claim that stands to lose $L - R_l$ in the bad state, but has limited upside potential. This could be for example risky debt, or preferred shares that have a claim on the firm in liquidation. Interestingly, the equity-holder can still be thought of as the "good cop", because he implements a soft liquidation policy, which is congruent with the manager's interests. The debt-holder on the other hand is a "bad cop" in the sense that he incurs a high effort level trying to uncover negative information about the firm's performance, which is obviously not in the manager's interest.

6. Optimal Choice of Monitoring Level and Liquidation Policy

In the previous section we highlighted two situations where in order to induce monitoring, it was necessary to employ a liquidation policy that destroyed value at the interim date. Monitoring then carries an indirect cost of destroying value, in addition to the direct monitoring cost c. This result naturally prompts the question whether it is actually worthwhile to induce monitoring? In this section, we turn to a complete characterization of when it is optimal to choose a particular liquidation policy and allocation of control in order to achieve monitoring. Note that we have already ruled out by assumption that any regime with low effort is desirable. We can therefore restrict attention to four regimes: two high monitoring regimes $\{A(\varnothing) = S, \overline{\theta}, \overline{e}\}, \{A(\varnothing) = C, \overline{\theta}, \overline{e}\}$, and two regimes without monitoring $\{A(\varnothing) = S, \underline{\theta}, \overline{e}\}, \{A(\varnothing) = C, \underline{\theta}, \overline{e}\}$. The following proposition sets out the parameter values for each regime, using the parameter regions defined in proposition 3.

Proposition 4 *(a) In region (i) the equilibrium level of monitoring is $\overline{\theta}$. In region (iv) the equilibrium level of monitoring is $\underline{\theta}$ (provided the project has positive NPV). In both regions the liquidation policy is the soft policy if and only if*

$$\overline{R} \geq L. \tag{24}$$

(b) In region (ii) ($\overline{\theta}$ is feasible only under the tough policy), the equilibrium level of monitoring is $\overline{\theta}$ and the liquidation policy is tough $\{A(\varnothing) = S, \overline{\theta}, \overline{e}\}$ if and only if

$$\overline{e}\left(w_h^C\left(\underline{\theta}\right) - \overline{\theta} w_h^S\left(\overline{\theta}\right)\right) + (1 - \overline{e})\,\Delta\theta\,(L - R_l) \geq c + \left(1 - \overline{\theta}\right)\left(\overline{R} - L\right), \tag{25}$$

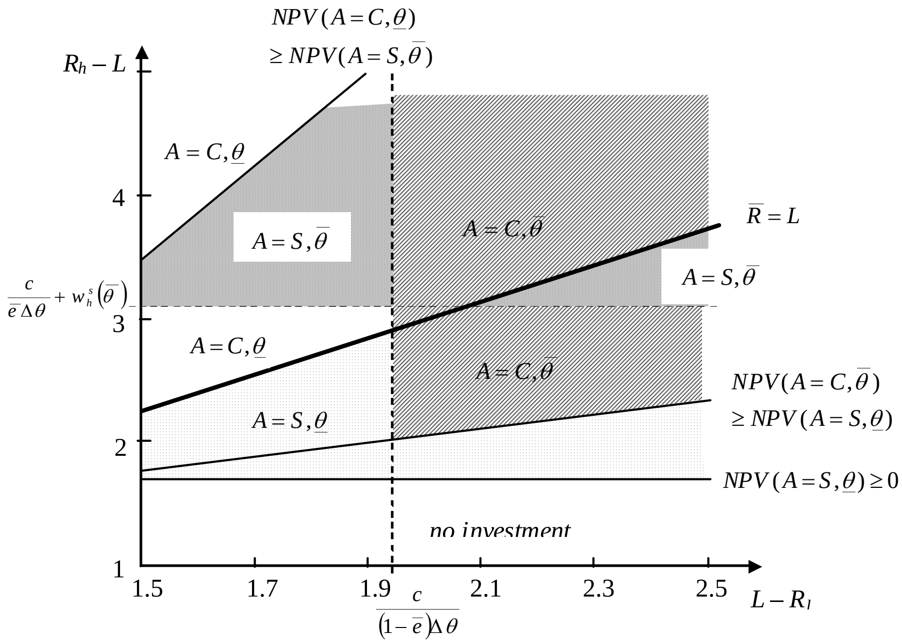


Figure 3: Depicts the optimal liquidation and monitoring policies as a function of $L - R_l$ and $R_h - L$. The model parameters are: $\overline{e} = 0.4, \underline{e} = 0.03, \overline{\theta} = 0.8, \underline{\theta} = 0.5, c = 0.35,$ $\gamma = 0.5, b = 1.5, I = 2.6$ and $L = 2.5$.

and

$$\overline{e}\overline{\theta}\left(R_h - w_h^S\left(\overline{\theta}\right) - L\right) \geq I - L + c. \tag{26}$$

(c) In region (iii) ($\overline{\theta}$ is feasible only under the soft policy), the equilibrium level of monitoring is $\overline{\theta}$ and the liquidation policy is soft $\{A(\varnothing) = C, \overline{\theta}, \overline{e}\}$ if and only if

$$\overline{e}\left(\underline{\theta}w_h^S\left(\underline{\theta}\right) - w_h^C\left(\overline{\theta}\right)\right) + \overline{e}\Delta\theta\left(R_h - L\right) \geq c + \left(1 - \overline{\theta}\right)\left(L - \overline{R}\right), \tag{27}$$

and

$$\overline{e}\left(R_h - w_h^C\left(\overline{\theta}\right) - L\right) - \left(1 - \overline{e}\right)\left(1 - \overline{\theta}\right)\left(L - R_l\right) \geq I - L + c. \tag{28}$$

The different optimal policies in different parts of the parameter space set out in proposition 4 are illustrated in figure 3 for a specific numerical example, which takes figure 2 and adds the constraints from proposition 4.[16] The first part of the proposition makes the obvious point that when monitoring is incentive compatible under either (or neither) liquidation policy, the only determinant for optimality is whether continuation in the absence of information destroys or increases value. Thus in the top right quadrant of figure 3, monitoring is optimal ($\theta = \overline{\theta}$) and the continuation policy is the one that maximises future cash flows: $A = S$ below the line $\overline{R} = L$, and $A = C$ above it.

The interesting regions from our point of view are those where the upside from continuation is large relative to the downside or vice versa, that is, the top left quadrant and the bottom right quadrant. In these regions it may be optimal to follow a value-reducing

[16]We ignore non-binding constraints in Figure 3 so as to make the picture more accessible. See the figure legend for the chosen parameter values.

continuation policy in order to induce monitoring, and hence to split financial claims and monitoring and control. The figure and the proposition indicate that this is indeed the optimal policy as long as making the "wrong" liquidation decision is not too costly, that is, as long as we are not too far from the thick $\overline{R} = L$ line. Thus, for example, it is optimal to follow the tough continuation policy in the dark upper left-hand region of figure 3, even though this policy is inefficient, because it makes monitoring incentive compatible and hence reduces wage payments to the manager. However, eventually, as this liquidation policy becomes more inefficient (i.e., moving towards the top left in the diagram), it is preferable to give up on inducing monitoring and induce managerial effort through greater wage payments, because liquidating the firm in the absence of information is too costly. Note that equation (26) in part (b) of the proposition guarantees that the firm has positive NPV under the tough policy with monitoring.

It can also be the case that the soft liquidation policy is optimal when being tough enhances value. Condition (27) characterizes the set of parameters under which this is the case. The cost and benefits of using this policy in order to improve monitoring incentives are analogous to the previous case. There is again the direct cost of monitoring, and the cost of implementing an interim continuation decision that in expectation destroys value (in region (iii) we have $\overline{R} < L$). In that case liquidating the firm in the absence of information is optimal at the interim date, but continuation is necessary to render monitoring incentive compatible. The benefit of monitoring is, just like before, a reduction in wage payments and an improvement in the quality of the liquidation decision. It is optimal to induce monitoring by employing the soft policy as long as this does not destroy too much value: that is, close to the $\overline{R} = L$ line; but as we get too far below and to the right of this line ($L - R_l$ increases and $R_h - L$ drops), continuing in the uninformed state is more costly than simply paying the manager a higher wage to induce effort. Eventually, as these parameters drop further, financing the firm is no longer a positive NPV project. The lowest horizontal line in figure 3 shows the boundary for positive NPV projects with the tough policy and no monitoring - no investment is optimal below this line. The boundary of the region we are interested in, where implementing the soft policy and monitoring is optimal and requires a split in control, is marked by the upward sloping line above it, which guarantees that the soft policy with monitoring generates more value than the tough policy without monitoring. Similarly, for optimality we also require a positive project NPV under the soft policy, which is captured in (28).

In the above treatment we restricted attention to *pure* liquidation strategies. Allowing for mixed strategies, however, would strengthen our main results. Note that a mixed liquidation strategy could generally improve upon the pure strategy solution. This is because the monitor's incentive compatibility constraint could be satisfied in regions (ii) and (iiib) even though the value-destroying liquidation decision was adopted with probability less than 1. It is then optimal to choose the lowest probability of value-destroying liquidation that maintains monitoring incentive compatibility. The amount of information generated by the monitor would then remain constant, but less value would (in expectation) be destroyed. Note, however, that implementing the mixed liquidation strategy will still require a separation of the monitoring function from control. The controlling stakeholder is only willing to implement a mixed strategy if he is indifferent between liquidating and continuing the firm. But precisely because he is indifferent as to the continuation decision he has no incentive to produce information. To put it differently, the monitor must have a cash flow sensitive claim in order to have an incentive to monitor, but as a result of this sensitivity to liquidation policy he cannot implement a mixed liquidation strategy. Allowing for mixed strategies would enlarge the parameter set for which a separation of monitoring from control is optimal and hence strengthen our results. We do not pursue this extension here as it would considerably complicate the analysis without changing the overall message of the paper: that financial claims must be split in firms where it is difficult, yet desirable, to induce claim-holder monitoring.

7. Renegotiation

In section 5 it was shown that an inefficient liquidation policy is sometimes necessary in order to render the costly monitoring activity incentive compatible. This raises the question of whether our results are robust to allowing for the renegotiation of the continuation decision. It turns out that our main results are robust to allowing for renegotiation between claim-holders, as long as the stakeholder with control rights has some bargaining power at the renegotiation stage. The reasons for this are essentially the same as in Dewatripont and Tirole (1994). To provide some intuition for why this is so, we consider a special case of our model which closely follows their set-up. We discuss renegotiation more generally in the Appendix.

Consider the special case of section 5.1 whereby the interim efficient liquidation policy is soft, but the tough policy is required for monitoring incentive compatibility. (The argument for the converse case of section 5.3 follows similarly.) In this case the potentially relevant parties in renegotiation are a debt-holder with a claim D, an equity-holder (the monitor), (potentially) further non-monitoring equity-holders, and the manager. The simplest case to consider is the one where the monitor is the residual claimant on the entire surplus from continuation: that is, there are no other outside equity-holders and parameters are such that $w_h^s(\theta) = 0$; i.e., the monitoring technology is sufficiently good that the loss of private benefits if the manager does not exert high effort is sufficient to motivate him.[17]

Let us now investigate what happens when the debt and the equity-holder are allowed to renegotiate. Suppose the signal $s = h$ is received at $t = 1$. In that case there is agreement between the monitor and the controlling party to continue the firm. Hence, there is no renegotiation and each claimants' payoffs remain as previously modelled. The same is true when the signal is $s = l$ and all now agree that the firm should be liquidated.

Suppose now that $s = \varnothing$. Since we are in the 'soft budget constraint' case, gains from renegotiation exist. Without renegotiation, the debt-holder would choose S, and so the parties will wish to bargain to reach efficiency. We model the renegotiation outcome as being achieved by Nash bargaining over the surplus from continuation $\overline{R} - L$, where the creditor captures a fraction α of the surplus. To be precise, we assume that the equityholder needs to pay $\alpha\left(\overline{R} - L\right)$ to the creditor in order to achieve continuation and neither claimholders' cash flow rights are altered. In this case the equityholders payoffs

[17] As will become clear below, setting $m_h = R_h - D$ maximises the monitor's incentives to collect information under these circumstances.

(if he monitors) can be written as

$$\overline{e}\left(R_h - D\right) + \left(1-\overline{e}\right)\left(\overline{\theta}\max\left\{L-D,0\right\} + \left(1-\overline{\theta}\right)\max\left\{R_l - D, 0\right\}\right)$$
$$-\left(1-\overline{\theta}\right)\alpha\left(\overline{R}-L\right) - c.$$

This yields the following incentive compatibility constraint for the equity holder.

$$\left(1-\overline{e}\right)\left(\max\left\{L-D,0\right\} - \max\left\{R_l - D,0\right\}\right) + \alpha\left(\overline{R}-L\right) \geq \frac{c}{\Delta\theta}.$$

Under the monotonicity requirement incentives are maximized by setting D just above R_l. This modifies somewhat the indeterminacy of the face value of debt in lemma 4: there the face value of debt did not matter for $D \leq L$, because the worst state $\varphi = l$ never occurred under this policy, and the monitor's incentives were only affected by the difference in payoffs between the $\varphi = h$ and $\varphi = L$ states. Under renegotiation the soft policy will be implemented and therefore the monitor's incentives are maximized if he carries all the risk between the $\varphi = L$ and $\varphi = l$ states, so that it is now optimal to set the face value of debt as close as possible to R_l. Given this, the incentive compatibility constraint can be re-written as

$$L - R_l + \alpha\frac{\left(\overline{R}-L\right)}{\left(1-\overline{e}\right)} \geq \frac{c}{\left(1-\overline{e}\right)\Delta\theta}.$$

This corresponds exactly to the familiar constraint monitoring (IC) constraint under the soft liquidation policy (14), except that incentives are augmented by renegotiation to the extent that the equityholder needs to share the surplus from continuation when he does not produce information.

Allowing for renegotiation strengthens our results in the following sense. If renegotiation is possible, then - provided that the equity-holder's bargaining power is not sufficiently strong to negate the effect - splitting financial claims into a monitoring and a controlling claim is more likely to be optimal than in the absence of renegotiation. The reason for this is simply that renegotiation ensures that the threat of inefficient termination will never actually be carried out. So one of the costs of splitting financial claims is removed, and the initial owner can sell the financial claims for a higher price than in the absence of renegotiation. The benefit of splitting claims - larger monitoring incentives - is retained because renegotiation ensures that when no information is generated, rents are transferred (to the controlling, non-monitoring claimant) rather than destroyed as they are without renegotiation.

Our result suggests that changes in law or security design which improve the bargaining power of creditors in renegotiation may be valuable in generating more information about the firm and thence providing greater incentives to managers. Similarly, when splitting claims the initial investor may wish to consider implementing mechanisms to limit the monitor's ex post bargaining power and enhance that of the debt-holder, and more generally to make renegotiation between the various parties more difficult. Thus issuing multiple creditor claims, so that some creditors may have an incentive to 'hold out' in bankruptcy, may in fact encourage equity-holder monitoring (by increasing the expected fraction of the surplus that creditors extract (Bolton and Scharfstein 1996)), although a formal analysis of this is outside the scope of this paper.[18] Note, by contrast, that the dispersion of equity claims will generally be undesirable in this case, because it may dilute monitoring incentives. We will return to this point when we discuss the empirical implications of our theory in section 9.

Finally, since a country's bankruptcy code significantly affects bargaining power in renegotiation, our theory may throw light on the monitoring incentives provided to claimholders by different financial systems. In particular, under chapter 11 of US bankruptcy law, equity-holders are effectively able to stay in control of a firm even when it is in financial distress; creditors' ability to seize the company's assets and liquidate them is severely limited. There is a large literature on the pros and cons of chapter 11. Some of the main trade-offs identified in this literature to date essentially boil down to the following: while chapter 11 may mitigate ex post inefficiencies from excessive liquidation, it may be detrimental for managerial incentives ex ante, because debt cannot exercise the effective threat of liquidation envisaged by, e.g., Dewatripont and Tirole (1994). Our model suggests another problem: since creditors have little power in renegotiation, equity-holders may have a severely reduced incentive to monitor the firm's prospects, further dulling managerial incentives. Therefore, consistent with what is observed, managerial pay will have to be more performance sensitive (and, by limited liability, higher) in the US than in countries such as the UK with more creditor-friendly bankruptcy codes (e.g., Conyan and Murphy, 2000).

8. Related Literature

The trade-off between ex ante incentives and ex post inefficiency alluded to in the introduction and throughout this paper is well understood in the literature (see for example Crémer, 1995). Our paper is novel in applying this line of argument to the *monitoring incentive of a security holder.* In contrast, the existing literature (e.g., Aghion and Bolton (1992), Berglöf and v. Thadden (1994), Dewatripont and Tirole (1994), Bolton and Scharfstein (1996) and v. Thadden, Berglöf and Roland (2003)) sees the threat of inefficient closure decisions *per se* largely as a way of providing *incentives for managers.* Capital structure is then designed to affect directly the relationship between a provider of capital and the manager. A residual claim (i.e., equity) then arises in these models in order to pick up the slack in claims on cash flows, but without any apparent governance role.

To highlight the differences between our paper and this literature, we set our model up such that the manager's incentives were directly affected *only* by the monitoring effort, and by construction *not* at all by an ex post inefficient shut down policy (recall from section 3 that the continuation policy followed when the monitor is uninformed does not affect managerial incentives). In our model ex post inefficient shutdown serves only to provide stronger incentives for the monitoring claim-holder. It affects the manager only indirectly, by ensuring that the monitor will collect more information about whether or not the firm is performing well. Thus the mechanism which we focus on is different to

[18] This observation relies on the fact that creditors cannot credibly hold out for a larger return when the state is known, for the reasons laid out above. On renegotiation design more generally, see Aghion, Dewatripont and Rey (1994), and on the design of bankruptcy procedures in particular, see Aghion, Hart and Moore (1992). Eraslan (2007) undertakes a detailed analysis of the bargaining game between the various stakeholders in bankruptcy.

that in the existing literature.

This structure allows us to focus on a problem ignored in the above papers: the interaction between incentives for liquidation policy and for firm monitoring to generate the information on which the liquidation decision must be based. This is not to say that we believe that direct incentive effects from inefficient, but informed, termination play no role in practice. On the contrary, we wish to highlight that for this effect to play any role it is necessary that the termination decision be informed, and in that sense our approach is strongly complementary to these papers. The tension between monitoring and liquidation incentives identified in our paper arises naturally once one allows for costly and endogenous monitoring by providers of capital.

A number of theories have considered monitoring incentives by providers of capital (e.g., Winton (1993), Rajan and Winton (1995), Park (2000)). However, in contrast to our work, these papers focus on either equity or debt contracts and explain specific contractual features such as limited liability or debt covenants as a result of monitoring incentives. These theories effectively assume that monitoring is carried out either by equity-holders (Winton (1993)) or by creditors (Rajan and Winton (1995), Koskinen (2000), Park (2000)). An exception is Laux (2001) who endogenises monitoring incentives among different claim-holders, focusing on the implications of monitoring incentives in within-firm hierarchies. By contrast we investigate how securities have to be designed in order to provide the right monitoring incentives and show that monitoring will be carried out either by equity- or by debt-holders, depending on whether the firm's continuation value is high or low compared to its liquidation value.

Moreover, our assumption that a provider of capital cannot commit to monitoring distinguishes our theory from a number of other papers on the topic (e.g., Diamond (1984), Gale and Hellwig (1985), v.Thadden (1995), Repullo and Suarez (1998)). They assume that a monitor can implement a time-inconsistent monitoring policy. In these papers, as here, monitoring is useful ex ante to mitigate a managerial moral hazard problem, but once the manager has taken his equilibrium action, there is little point incurring the cost of monitoring. Our work provides an important foundation for this assumption, since we show how capital structure can be used to solve the time inconsistency problem associated with monitoring. Most importantly, our paper helps us to resolve the puzzle, mentioned in the introduction, as to why multiple claims are issued despite obvious negative externalities between them.

As mentioned in the introduction, a small number of papers share our goal of rationalising the issuance of multiple securities.[19] Allen and Gale (1988), Gorton and Pennacchi (1990), Boot and Thakor (1993), and Fulghieri and Lukin (2001) show that investor heterogeneity or limited arbitrage can render multiple claims optimal. These papers, however, do not address the governance role of financial contracts. Our model does not rely on investor heterogeneity or limited arbitrage, and we show that the debt and equity claims that the firm issues must be held by different investors to create the proper monitoring incentives. In this respect, our paper is most closely related to the recent work

[19] Other papers rationalise multiple claims when the entrepreneur is able to choose both the mean and the variance of returns (Cornelli and Yosha 2003, Biais and Casamatta 1999). But these moral hazard problems on which they focus could equally well be solved by issuing a convertible debt claim rather than debt plus equity (that is, there is no requirement that the debt and equity claims be held by different investors).

by Axelson, Strömberg and Weisbach (2009); but the investors in their paper are passive - their focus is rather on creating appropriate incentives for the manager of a buy-out fund.

There is also a literature on blockholders as active monitors (e.g., Shleifer and Vishny (1986), Burkhart, Gromb and Panunzi (1997)). In these papers, monitoring is interpreted as a general value-enhancing action, but not, as here, directly modelled as information acquisition. This distinction is important, because we show that the incentives to acquire information may run counter to the incentives to act on the information in a way that maximises firm value. These theories do not consider the interaction between equity and debt and do not predict a complementarity between concentrated equity holdings and high gearing ratios (see, however, Mahrt-Smith (2005), discussed in the next section).

9. Empirical Implications

The above analysis shows that an aggregate claim on a firm's cash flows may not always have sufficient incentives to monitor the firm's prospects. This is because interim monitoring incentives derive only from the improvement in the quality of the liquidation decision at an interim date. However monitoring may add further value ex ante, because it mitigates the managerial moral hazard problem. Our theory thus applies to firms where external monitoring of firm performance is thus likely to be useful in improving managerial incentives.

Under these circumstances, we have shown that monitoring incentives may sometimes be improved by splitting firm returns into two (or more) financial claims: one which has the role of monitoring the manager and the other which has the role of taking liquidation or continuation decisions. We saw that the analysis splits into two cases depending on whether parameters are such that it is ex ante desirable to implement an inefficiently tough or soft continuation decision. We explain below that these two parameter constellations are likely to apply to LBO candidates and start-up firms, respectively. We believe that these settings are likely to be ones where our assumption that the monitor's signal about future firm prospects is observable but not verifiable in a court is likely to be appropriate.

9.1 *Tough policy*

Our theory argues that for firms for which continuation is expected to be profitable there can be a soft budget constraint which reduces potential monitors' incentives to collect information. By contrast, when a firm increases its leverage from a low level to levels which would put creditors in control at an interim date in the absence of the arrival of good news, this will result in increased incentives for shareholders to collect information about the firm, and thence improved incentives for the firm's management. If monitors' information is subsequently reflected in firm stock prices through informed trading by the monitor, then the theory predicts that these more highly levered firms' stock prices should be relatively informative about firm performance.[20] This problem applies particularly to firms which have a lot of tangible assets (i.e., $L - R_l$ is small, assets do not lose value dramatically over time).

[20] An extension of the model where monitors do indeed trade on their information in the stock market is available from the authors on request and appeared in an earlier draft of this paper.

9.1.1 A Theory of LBOs

We are not aware of any direct tests of the proposition that a firm's leverage is directly related to the informativeness of its stock price. However, unlike previous work, in highlighting the tension between the monitoring and control functions, our paper does provide a basis for understanding the relationship between firms' capital structure and ownership structure decisions. In particular, a firm that allocates control to a (tough) creditor in order to improve the equity-holders' monitoring incentive needs to ensure that the equity stake is in the hands of agent(s) that do not face excessive free-rider problems in their monitoring decision. Hence, our theory predicts that companies with a high debt burden (and therefore effective creditor control) should have more concentrated equity ownership.

Creditors may exercise significant control over firms, even well before control rights are transferred. Stronger creditor control can be achieved by issuing short term debt at levels above short term cash flows, and sufficiently high that the firm may find it difficult to roll over or refinance the debt unless it can be demonstrated that the firm's prospects are rosy.[21] Coval et.al. (2002) show that in practice it can be difficult for solvent firms to roll over short term debt. Similarly, firms in financial distress are often subject to significant creditor control (see Franks and Mayer (2005)). According to our theory, it is important that such firms have concentrated equity ownership in order to ensure that sufficient information about managerial performance is generated, otherwise (in the absence of other tax issues or agency problems) the high leverage will be counterproductive. Franks, Mayer and Renneboog (2001) provide evidence consistent with our prediction. They study firms in financial distress and find a positive relationship between equity ownership concentration and high gearing ratios. We are not aware of any empirical studies that investigate the relationship between the proportion of a firm's short-term debt and equity ownership concentration.

Interestingly, the combination of high gearing ratios with concentrated equity ownership is also exactly the pattern observed in leveraged buy-outs (LBOs); see Kaplan and Stein (1993). These firms are also very often firms which have substantial tangible assets - in fact, anecdotal evidence suggests that LBOs without significant tangible assets are much less likely to be successful. This is consistent with our finding that levering up is likely to be particularly valuable for firms for which $L - R_l$ is small. Existing theories of capital structure have been silent on why companies do not simply lever up to achieve the managerial incentive (and tax) effects of high gearing ratios. For example, taking on a hefty debt-burden is sufficient to eliminate the moral hazard problem associated with excessive free cash flows identified by Jensen (1986). Instead, leveraging-up is very often accompanied by a drastic change in ownership structure, whereby typically a publicly listed company with dispersed ownership is taken private in a highly-leveraged transaction. The end result is often equity ownership concentrated in the hands of a buy-out specialist. Denis (1994), for example, provides evidence that buyout specialists fulfill an important monitoring function that complements the increased leverage in an LBO. Similarly, Cotter and Peck (2001) argue that buyout specialists have a particular role as active monitors in the firms they acquire. Our model explains why this may be

[21] An alternative would be to issue longer term debt with covenants that are easily triggered in the short run.

so: if leveraging is to achieve greater equity monitoring, this will be ineffective unless equity-holding is relatively concentrated.

Looking at it the other way, why do buy-out specialists typically use so much debt to finance their transactions? A popular hypothesis is simply that debt is somehow cheaper and so using more debt enhances the return to equity - but such a hypothesis must of course explain the apparent violation of Modigliani-Miller. Axelson et. al. (2009) provide an elegant hypothesis relating to the conflict of interest between the limited and the general partners of the buy-out fund. Our model provides a complementary explanation: the incentives of the equity holder to monitor the firm are much enhanced if the firm is highly levered, and this extra monitoring can raise firm value. Our model can thus help us to understand the coincidence of concentrated equity ownership and high leverage in firms with large upside ($R_h - L$) and limited downside ($L - R_l$).

Conversely, we would expect to see below average performance by firms that do have effective creditor control (high gearing) but retain a dispersed equity ownership structure. These firms would not generate active monitors and therefore managers have worse incentives to exert effort, and creditors will liquidate more frequently. Cotter and Peck (2001) show that buyout specialist controlled LBOs have a 15% incidence of subsequent financial distress compared with 80% for those with other outside equity holders. In the majority of the latter cases, the outside investors do not purchase a controlling interest and view their investment as passive. Hence, it appears that the benefits from equity ownership concentration are more significant in the presence of effective creditor control. This finding is consistent with our model.

These predictions distinguish our theory from other theories of debt in which equity occurs only to pick up the slack after the optimal security has been designed (e.g., Dewatripont and Tirole (1994)). In those theories the ownership structure of the remaining equity claim plays no role. An exception is Mahrt-Smith (2005) who shows that debt arises as a complementary security to equity in order to counter-balance potential opportunistic actions by the equity holders. He predicts that equity ownership concentration should be correlated with tightly-held debt, and dispersed equity with publicly-held debt. This leaves open the question why LBOs are financed significantly through public bond issues. Our model can help explain this observation, insofar as dispersed debt ownership renders debt renegotiation more costly, and therefore makes the creditor tougher, complementing strong equity monitoring.

9.2 *Soft policy: a theory of start-up finance*

Note that a project that is best liquidated in the absence of information would also be a negative NPV project by our assumption that $I > L$ unless information sometimes allows the project to be liquidated in the bad state. One could think of this type of project as a negative NPV project that turns positive once one takes into account the option to abandon it in the bad ($\omega = l$) state. The value of the option to abandon is given by $\theta(1-\overline{e})(L-R_l)$, where $\theta \in \{\underline{\theta}, \overline{\theta}\}$. From this it follows that for projects with a very high likelihood of success ($\overline{e}$ is high) the option to abandon adds little value. The types of projects for which the option to abandon is sufficiently valuable to turn around the NPV from negative to positive, must therefore have a relatively small likelihood of success and a sufficiently high payoff R_h in case of success that the NPV is positive overall.

Start-up firms that invest in a new, high risk technology are typically thought of as having this type of payoff. The chances of success are low, but the returns to a successful start-up company are very high. In contrast to our model, these firms do not have large fixed assets which can be salvaged through liquidation, but since they typically rely on staged finance, we can reinterpret our model in a way which is consistent with the characteristics of such firms. In particular, consider that liquidation saves the funds which would otherwise be invested in the next stage of firm finance, which can be substantial. Suppose that at the initial date, a capital provider needs to invest $I - L$ in the first stage that starts off a project. Subsequently, an entrepreneur puts effort e into the project, which affects the final project payoff just like before. After this, a monitor can produce an informative signal about the future payoff, again as in our previous treatment. However, in contrast to the previous set-up, suppose that in order to take the project forward a further investment of magnitude L is required. In that case, if the stage 2 investment is undertaken, total investment is I and an analogous condition for value-enhancing stage 2 investment is $\overline{R} > L$. If the stage 2 investment is not made, total payoffs, net of the stage 1 investment are $I - L$, just as in the previous treatment, where the firm could be liquidated at the interim date, yielding an overall payoff to the providers of capital equal to $I - L$. Using this reinterpretation we can apply the intuitions from the previous model to analyse the finance of start-up companies by re-labelling the liquidation decision as a stage 2 financing decision.

Our theory can thus make predictions about the allocation of control rights, monitoring and the design of securities in the context of start-up firms. Suppose that, in line with the previous discussion, the project has a negative NPV in the absence of the option to abandon the project at stage 2. This implies that the project should be abandoned at stage 2 if no information has been produced about it. If the firm were to implement this investment strategy (stop unless you have positive information), there would be little incentive to acquire information: since the project is unlikely to have turned out well, the expected reward from identifying the high state are limited. On the other hand, if the firm applies an *overinvestment* policy in the form of going ahead with stage 2 financing unless negative information has become available, the party that has committed to providing the funds for stage 2 has a strong incentive to gather information that the state of the world is bad, $\omega = l$, so as not to have to finance stage 2 of the project. This provides the manager with better incentives, raising value overall despite the overinvestment. Moreover, the party that provides stage 2 financing should do so in the form of a senior claim. The residual, junior, claim should rest with the party with control rights over continuation, so that this party has an incentive to invest in the absence of information, thus implementing the soft continuation policy.

How is the division of monitoring and control structure implemented in start-up firms in practice? Start-up firms have two main sources of finance: alliances with larger firms and venture capitalists. Guedj (2005) and Guedj and Scharfstein (2005) provide evidence on decisions to continue drug development in small bio-tech firms allied with large firms; and inside large pharmaceutical firms. They argue that, in contrast to drug trials inside large firms, the decision to continue development in alliances is governed by detailed contracts. They show that drug trials inside large firms are more likely to be terminated than those in alliances, and that these terminations are largely efficient (ex post value-enhancing). Why then, do alliances write such apparently inefficient contracts? Guedj and Scharfstein's findings are consistent with our theory, which suggests that, given their option-like pay-off, bio-tech start-ups are good candidates for implementing the soft liquidation policy even though the tough one may be efficient. To be more specific, within-firm drug trials, although they have efficient ex post continuation, are likely to suffer from poor ex ante incentives for managers precisely because control is not split from monitoring. If the pharmaceutical company claims all the returns on the drug development project; it invests little in monitoring since the project is unlikely to succeed and as a consequence the managers have low-powered incentives as the continuation decision is less correlated with their efforts. In alliances, however, more control rests with the entrepreneur and outside board members of the start-up firm, who have incentives to continue. In order to stop the project, the pharmaceutical company must come up with hard evidence that the project will fail. Thus it has more incentives to monitor the project and the start-up entrepreneur will have higher-powered incentives. As we have shown, under certain parameter restrictions (in particular, for the set of projects with a low probability but very high revenues from success), such an ex post inefficient split of monitoring from control can be ex ante value-enhancing. Nevertheless, since the (efficient) tough policy is used inside firms whereas the (inefficient) soft policy is used between firms, one should see that projects inside firms are terminated more frequently, and that this termination, taken in isolation, is efficient.

Venture capitalists also finance start up firms and govern the continuation decision partly by contract. Note that venture capitalists frequently hold relatively senior claims, such as preferred shares, rather than common equity. This makes them suitable candidates to perform the monitoring function if the soft continuation policy is used, but, according to our model, makes it difficult for them to commit to continuing as frequently as the soft policy requires. We contend that venture capitalists structure their arrangements with firms to help achieve such a commitment. VC financing typically takes the form of an original contract which stipulates the total amount of capital committed, and how the committed capital is to be distributed over the investment stages. But the commitment that is embodied in such a business plan is limited because of the large number of unforeseen contingencies that may arise. Thus in practice the allocation of control in financed firms is very important (Kaplan and Strömberg (2003)).[22] In a majority of cases studied by Kaplan and Strömberg (75%), VCs *do not* have board control (and even less so in early financing rounds), the balance of power being held by independents who have been jointly agreed by the venture capitalists and the entrepreneur. If these independent board members are inclined to follow a soft policy (e.g., their compensation is junior to that of the venture capitalist) then a picture emerges that looks very much like the one described in our model.[23] In particular our model can help explain why VC contracts

[22]Stage financing is often contractually contingent on a variety of performance measures or other targets (for example technological specifications of the product). These allow the venture capitalist some degree of commitment to continuation. We are, however, concerned with the contingencies that cannot be contracted upon ex ante.

[23]An alternative commitment mechanism - which is often used, and is also consistent with our model, is the following. The incumbent VC commits to the provision of continuation finance if a new lead investor with a senior claim can be convinced to provide part of the financing at the interim date. Because of new the claim's seniority over the old claim, the new lead investor would always be willing to participate, unless the incumbent VC generates negative information $s = l$. In practice, the reputation of the incumbent VC is clearly also important in this commitment.

with start-up firms are written so as to provide a certain degree of commitment to future investment: finance is provided unless the venture capitalist can convince board members that this will lead to a very bad outcome. According to our theory this may be necessary in order to provide VCs with the incentive to produce the information that allows them to abandon the project in the bad state. Without this commitment, VCs might be tempted to abandon projects too easily, which undermines their incentive to produce information about their future prospects.

10. Conclusions

In this paper we have investigated the question of why firms may choose to issue different types of securities to different claim-holders, despite the fact that this creates potential conflicts of interest among them. The fundamental idea underlying our analysis is that creating externalities which are costly ex post can nevertheless be beneficial for ex ante incentives. We studied the particular case where a firm faces an important continuation decision at an interim date; and we have shown that it can be beneficial to split the firm's cash flows into payments to one agent who controls the continuation decision and another who attempts to provide information that will be relevant to that decision. We have identified situations where dividing these two functions results in more information gathering than would selling a uniform type of claim to a single claim-holder.

This is not obvious a priori for the following reason. Incentives to produce information generally depend on the sensitivity of returns to profits, and budget-balance dictates that the maximum sensitivity of returns to profits is given by the aggregate claim. (Unless firms can issue securities that are negatively correlated with firm value, which seems to be rare in practice). Our paper shows that, when claims are monotonic in firm value, incentives can be improved by dividing up financial claims in such a way that gives one claim-holder a (credible) incentive to *destroy* value in the absence of information by choosing an inefficient action. Our analysis suggests that issuing two types of securities: one "debt-like" and one "equity-like" is optimal.

Our analysis is likely to be particularly relevant to firms which are relatively opaque, so that the costly collection of information about firm prospects is likely to be particularly valuable. We identify two different sets of circumstances in which splitting claims improves firm value. In the first, "soft budget constraint" case, continuation is expected to be profitable provided that the manager has worked. The issue that arises in such firms is that, given that the manager has worked, there is little reason to spend resources collecting information before deciding to continue - so the manager may be tempted not to work as information that he has performed badly is unlikely to be gathered at the interim date. In such firms it is useful to divide up claims into a (safe) debt claim with control rights and a (risky) equity claim to increase monitoring effort. The creditor's liquidation decision will be tougher than would be interim efficient, but this is precisely the reason why the equity-holder's monitoring incentive is improved: only by generating good news will he be able to prevent liquidation. We believe that this analysis is relevant to mature firms with relatively strong cash flows, which are more likely to suffer the soft budget constraint problem. Levering up such firms may improve incentives for information-gathering and hence managerial incentives, but only if equity-holding is sufficiently concentrated that the equity-holders actually respond to the improved incentives rather than simply free-riding. We argue that this combination - levering up plus concentrated equity-holding - predicted by our theory to be particularly beneficial in firms with relatively high liquidation values and strong cash flows, is in fact characteristic of successful LBOs.

The other set of firms to which our analysis is relevant are those at the other end of the spectrum, where it is considered to be unlikely that continuation has positive expected value, but ex ante investment is nevertheless worthwhile, once one accounts for the value of the option to abandon. This category includes start-up firms. In line with out theory, there is mounting evidence that start ups are often continued inefficiently (Guedj 2005, Guedj and Scharfstein 2005). This is because venture capitalists or alliance partners have made implicit or explicit commitments to fund the firm unless prospects are very poor. One interpretation of this evidence is that such excessive continuation of negative NPV projects provides incentives to managers; our theory shows that one way in which it does this is that it provides strong incentives for senior claimants such as VCs to monitor the firm's prospects in order to improve the correlation between the closure decision and the lack of managerial effort.

The mechanism at work in our model allows us to highlight a fundamental trade-off which is probably of wider importance than the simple application given here. The difficulty with motivating monitors, as opposed to other agents, is that it is often the case that if all other decisions are taken optimally and agents act according to plan, then there is typically little benefit to monitoring, and hence, unless there is a budget breaker willing to subsidise this activity, little reason to monitor. If it is possible, however, to delegate the information contingent action to a third party with incentives to take an inefficient action, then monitoring incentives can be improved. Thus *there is a natural division of labor between those who collect information and those who act on it.* Non-financial examples might include the desirability of separating the function of teaching from that of examining students; the separation of judging criminals (the judiciary) from that of apprehending them (the police); and the separation of lobby groups who collect information from politicians who act upon it.

Appendix A

Renegotiation with other parties at the bargaining table and managerial incentive pay In the analysis of section 7 we made two assumptions to simplify the argument. The first of these was that the manager did not earn an incentive wage and the second, related, assumption, was that only the controlling party and the monitor were involved in the renegotiation. It should be clear that the same kind of argument can be made without these assumptions, but one would need to make assumptions about how the gains from renegotiation are to be divided and which of the various parties which benefit from renegotiation are present at the bargaining table. If the monitor is not the only party to receive a monetary benefit from continuation (for example, there are other outside equity-holders and/or the manager earns an incentive wage), and yet only the monitor is involved in the renegotiation with the controlling party, then efficient renegotiation becomes more difficult. In particular, even though continuation is efficient, the monitor himself may not gain enough surplus from continuation to compensate the controlling party for his loss from continuation, since some of the surplus from continuation will go to the manager (in the form of a bonus wage) and to any other equity-holders. Specifically, the monitor gains $\overline{e}m_h$ from continuation, where m_h is at least $\frac{c}{\overline{e}\Delta\theta}$ and so the surplus available from continuation is $\overline{e}(\frac{c}{\overline{e}\Delta\theta}+\max\{0, D-L\})-(1-\overline{e})(\min\{L, D\}-R_l)$, which, if d is small, can be negative even though the total surplus from continuation $\overline{e}(R_h-L)-(1-\overline{e})(L-R_l)$ is positive. The difficulty is that other parties who gain from continuation (e.g. the manager and any other equity-holders) may not be present at the bargaining table to make concessions. Thus having dispersed equity-holders could, for example, make renegotiation more difficult, and thus provide better incentives for a monitoring blockholder.

The case where the manager earns an incentive wage and is present at the renegotiation table with the monitor and the controller is more complex, because if the manager anticipates making wage concessions in the event of a lack of information, this will affect his incentives ex ante. In fact, for the reasons indicated above, if possible, it will be optimal to design the renegotiation such that the manager is not able to make concessions on his wage, as the fact that he is unable to control the information flow means that making his wage contingent on it will simply serve to blunt his incentives; whereas maximizing the concessions which must come from the monitor improves the latter's incentives.[24] It is straightforward to see that one could set out the ex ante incentive constraints for the monitor and the manager when both anticipate being present at the bargaining table ex post and show again that splitting claims will improve information collection as long as the monitor is forced to make a large enough concession to the controller.

Inside monitoring In this section we consider the case where the manager can also be the monitor. The optimal wage contracts are then given as follows.

Proposition 5 *Suppose the manager is also the monitor and the liquidation policy is $A(\varnothing)=C$. The wage contract that results in the lowest expected payment to the manager and that induces $\overline{e}$ and $\overline{\theta}$ satisfies*

$$w_h^{C,m} = \frac{\gamma}{\Delta e} + \overline{\theta}\frac{c}{(1-\overline{e})\,\Delta\theta}, \tag{29}$$

$$w_L^{C,m} = b + \frac{c}{(1-\overline{e})\,\Delta\theta}, \tag{30}$$

$$w_l^{C,m} = 0. \tag{31}$$

When the manager is the monitor and the liquidation policy is $A(\varnothing)=S$ the wage contract that results in the lowest expected payment to the manager and that induces $\overline{e}$ and $\overline{\theta}$ satisfies

$$w_h^{S,m} = \max\{\frac{\gamma}{\overline{\theta}\Delta e} - b, \frac{c}{\overline{e}\Delta\theta} - b, \frac{\gamma + c}{\overline{\theta}\Delta e + \underline{e}\Delta\theta} - b, 0\}, \tag{32}$$

$$w_L^{S,m} = w_l^S(\theta) = 0.$$

A simple comparison between (29), (30) and (12) shows that the wage contract under the soft liquidation policy is more expensive when the manager is also the monitor compared to when an outsider performs the monitoring function. Firstly, if the manager is

[24]This provides an interesting argument for why managers' bonuses should not be renegotiated when a firm is in financial distress.

the monitor he needs to be given a positive payment when the firm is liquidated ($w_L > 0$), compared to a zero wage when an outsider monitors. This is because monitoring incentives under the soft continuation policy are provided by the difference in payoff when the firm is worth L compared to R_l. Rewarding the manager, however, when the firm is liquidated, reduces his incentive to exert effort ex ante. In order to restore effort incentives, the manager's payoff in state R_h must increase even further. This increases the agency rent to the manager. Further, unlike the other investors, the manager does not have cash to compensate the initial owner for this increase in agency rent ex ante. It is therefore never optimal to make the manager also the monitor when a soft liquidation policy is applied. In other words, when the default is that the firm will continue in the absence of bad news, it is never appropriate to try to design a compensation scheme to induce the manager to collect such bad news. It is always cheaper to have outsiders monitor the firm and suggest that it should be liquidated.

The potential advantage from allocating the monitoring function to the manager becomes apparent under the tough liquidation policy. It derives from the fact that the manager is motivated by his private benefits to produce positive information about the future state of the world. When private benefits are high, the manager may be willing to incur the monitoring cost with little or no additional monetary reward. This makes it cheaper to get the manager to act as monitor than to delegate the task to an outsider. On the other hand, the manager's incentive compatibility constraint is complicated by the fact that he can now deviate simultaneously to low effort *and* to no monitoring. If this incentive compatibility constraint is harder to satisfy than the original incentive compatibility constraint, then the wage payment necessary to induce the manager-monitor to exert effort and to monitor may be so high as to render it more desirable to allocate the monitoring task to an outsider.

Proposition 6 *When* $\frac{\gamma}{\underline{\theta}\Delta e} < \frac{c}{\overline{e}\Delta\theta}$ *it is never optimal to rely on the manager to act as monitor.*

When the effort incentive problem is relatively easy to resolve compared to the monitoring incentive problem under the tough liquidation policy ($\frac{\gamma}{\Delta e}$ is small compared to $\frac{c}{\overline{e}\Delta\theta}$) then it is cheaper to allocate the monitoring function to an outsider rather than to the manager. This is effectively because the manager can extract an agency rent, while the outside monitor cannot. This is because the outside monitor has to pay in order to acquire the stake that then provides him with the incentives to monitor. Hence, the agency rent extracted by the monitor is priced into the contract ex ante. This is not the case for the manager, because his limited wealth renders bidding for employment impossible.

The importance of this proposition lies in the fact that it shows a positive role for *outside* equity,[25] and hence confirms that under some circumstances the firm must issue

[25] The fact that nearly all firms issue outside equity is not well explained by most traditional theories of capital structure. They would predict that for incentive, selection and tax reasons, firms should rather issue debt-like claims to outsiders and with insiders retaining the equity. What is the function of outside equity? If outside equity is present only as a passive provider of funds (the last resort in the pecking order), why is so much attention lavished on firms' leverage ratios (and why is low leverage seen as desirable, whereas the pecking order would suggest that firms should rather issue a lot of debt)? In our theory, when the manager is not the monitor, the balance of outside equity and debt matters because these two claims must work in tandem to make appropriate monitoring and control decisions.

multiple outside claims. Under the tough policy, outside equity collects information in a more cost effective way than inside equity. Under the soft policy, outside equity may also act as monitor if the upside from continuation is sufficiently large (the upper right quadrant of figure 3); otherwise the monitoring function should be allocated to a more senior claimant whilst outside equity is charged with making the continuation decision.

The proposition also provides an interesting slant on the leveraged buyout and venture capital examples discussed above. The highly successful LBOs and MBOs of the 1980s were mostly undertaken on highly profitable firms. Our model suggests that levering up the firm can be helpful in terms of improving managerial incentives under these circumstances; and that under these circumstances it is possible that either the manager or an outside shareholder can act as the monitor, providing positive information to the banks who would otherwise foreclose the debt. It also helps us to understand the anecdotal evidence that MBOs and LBOs on unprofitable firms is less successful. Our theory suggests that in these firms it may be more appropriate to employ a soft policy, with (concentrated) debt holders - not managers or equity holders - acting as monitors. Similarly, notice that it is initially not obvious why investors' stakes in start-up firms are typically so concentrated since the entrepreneurs running these firms have such high-powered incentives that it is not clear that moral hazard or adverse selection is a real issue. Our theory helps us to understand this since we have shown that in firms that are unprofitable apart from the option to abandon, the manager (entrepreneur) cannot credibly act as a monitor. Monitoring will be carried out more effectively by an outside claimant holding a senior claim such as a venture capitalist.

Appendix B

Proof of lemma 1 There are three effects that the liquidation policy has. (i) It affects expected cash flows directly, (ii) it affects wage payments to the manager that are necessary to satisfy his incentive compatibility constraint, (iii) it may affect the feasibility of inducing monitoring.

Regarding (i) it is obvious from the assumption $R_h > L > R_l + b$ that (gross) cash flows are maximized from choosing $A(h) = C$ and $A(l) = S$.

Regarding (ii) we need to consider 2*2*2 liquidation policies, depending on whether $A(h), A(l)$ and $A(\varnothing)$ are set to C or S. For each of the 8 possible policies we can write down the manager's incentive compatibility constraint and then calculate the contract that induces high effort at the lowest cost. Then we can calculate expected wage payments in each case. Denote by $W_{A(\varnothing)}^{A(h),A(l)}$ the expected wage payment for the corresponding case. If $A(\varnothing) = C$ it can be shown (performing the calculations described above) that

$$\begin{aligned} W_C^{C,S} &= \max\left\{0, \frac{\overline{e}}{\Delta e}\gamma - \overline{e}\theta b\right\}, \\ W_C^{C,C} &= \frac{\overline{e}}{\Delta e}\gamma, \\ W_C^{S,C} &= \frac{\overline{e}}{\Delta e}\gamma + b, \\ W_C^{S,S} &= \frac{\overline{e}}{\Delta e}\gamma. \end{aligned}$$

Hence, setting $A(h) = C$ and $A(l) = S$ minimizes expected wage payments.

We proceed in the same way for the case where $A(\varnothing) = S$. If $A(h) = S$ and $A(l) = S$ then it is impossible to satisfy the manager's incentive compatibility constraint: the firm is always liquidated and effort is in that case useless. For the remaining cases we can write expected wage payments as follows

$$\begin{aligned}
W_S^{C,S} &= \max\left\{0, \frac{\overline{e}}{\Delta e}\gamma - \overline{e}\theta b\right\}, \\
W_S^{C,C} &= \frac{\overline{e}}{\Delta e}\gamma, \\
W_S^{S,C} &= \frac{\overline{e}}{\Delta e}\gamma + \theta\overline{e}b + (1-\theta)\left(\frac{\gamma}{\theta\Delta e} + b\right).
\end{aligned}$$

Again, it is obvious that the liquidation policy $A(h) = C$ and $A(l) = S$ minimizes expected wage payments.

Finally, regarding (iii) deviating from the policy $A(h) = C$ and $A(l) = S$ reduces the amount of cash that is available for distribution to the monitor in the state where he has produced information. It follows immediately that deviating from this policy therefore cannot improve monitoring incentives.

Proof of lemma 2 The manager's incentive compatibility constraint (1) under the soft liquidation policy $A(\varnothing) = C$ is given by

$$\begin{aligned}
&\overline{e}(w_h + b) + (1-\overline{e})(\theta w_L + (1-\theta)(w_l + b)) - \gamma \\
&\geq \underline{e}(w_h + b) + (1-\underline{e})(\theta w_L + (1-\theta)(w_l + b)).
\end{aligned}$$

This can be rewritten as

$$\Delta e\,(w_h + \theta b - \theta w_L - (1-\theta)w_l) \geq \gamma.$$

Expected incentive compatible wage payments are minimized by setting $w_L = w_l = 0$. Solving the binding incentive compatibility constraint for w_h yields (12).

Under the tough liquidation policy $A(\varnothing) = S$, the incentive compatibility constraint becomes

$$\begin{aligned}
&\overline{e}\theta(w_h + b) + (1-\overline{e}\theta)w_L - \gamma \\
&\geq \underline{e}\theta(w_h + b) + (1-\underline{e}\theta)w_L.
\end{aligned}$$

Or

$$\Delta e\theta\,(w_h + b - w_L) \geq \gamma.$$

Again, setting $w_L = w_l = 0$ minimizes expected wage payments and yields w_h given in (13).

Proof of lemma 3 Under the soft liquidation policy, the monitor's incentive compatibility constraint (3) can be written as

$$\begin{aligned}
&em_h + (1-e)\left(\overline{\theta}m_L + (1-\overline{\theta})\,m_l\right) - c \\
&\geq em_h + (1-e)\left(\underline{\theta}m_L + (1-\underline{\theta})m_l\right).
\end{aligned}$$

This can be rewritten as

$$(1-e)\left(m_L^C(e)-m_l^C(e)\right)\geq \frac{c}{\Delta\theta}. \tag{33}$$

Under the tough policy (3) is given by

$$\begin{aligned} & e\overline{\theta}m_h+\left(1-e\overline{\theta}\right)m_L-c \\ & \geq e\underline{\theta}m_h+\left(1-e\underline{\theta}\right)m_L \end{aligned}$$

This is the same as

$$e\left(m_h^S(e)-m_L^S(e)\right)\geq \frac{c}{\Delta\theta}. \tag{34}$$

Proof of proposition 2 Consider first region (a). To make monitoring incentive compatible one claimant needs to be given cash flows such that (14) holds. But this implies that, either $m_L > L$ or $m_l < R_l$ (or both). Simple adding up constraints then imply that the budget breaker will have to have a higher payoff when the firm's cash flow is R_l than when it is L.

Consider region (b). In order to satisfy the monitoring incentive compatibility constraint one needs to set, either $m_h > R_h - w^S\left(\overline{\theta}\right)$ or $m_L < L$ (or both). Again, there would have to be a budget breaker whose claim is less valuable when firm cash flows are R_h compared to when they are L.

Proof of proposition 3 Consider payments that satisfy the monitor's incentive compatibility constraints region by region.

Region (i)

Suppose the soft liquidation policy is implemented and payments to the monitor are therefore given by (14). Moreover, monotonicity of payments requires that $L-m_L \geq R_l - m_l$. This together with (14) implies $L-R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$. When the monitor and the party with control rights are different agents, the soft policy can be implement by making payments $a_h \geq 0$, $a_L = a_l = 0$, where $a_h > 0$ only if continuation destroys value. The latter point follows from our assumption that in case of indifference the interim value enhancing policy is implemented. When continuation destroys value, a strictly positive payment needs to be given to the controlling stakeholder to induce continuation. When the monitor also takes the continuation decision, the soft policy can be implemented if

$$\overline{e}\left(m_h-m_L\right)\geq(1-\overline{e})\left(m_L-m_l\right)\left(\geq\frac{c}{\Delta\theta}\right).$$

Setting $m_h - m_L$ to its maximum $R_h - w_h^C(\theta) - L$ implies that $A(\varnothing) = C$ is incentive compatible for the monitor if

$$\overline{e}\left(R_h-w_h^C(\theta)-L\right)\geq\frac{c}{\Delta\theta}.$$

Since $w_h^C(\theta) \leq w_h^S(\theta)$ monitoring can then be incentive compatible under the soft policy if the monitor is also the controlling stakeholder in region (i).

Suppose now that the tough liquidation policy is implemented. Then payments $m_h - m_L$ can be given so as to satisfy (15) if $R_h - w_h^S(\theta) - L \geq \frac{c}{\overline{e}\Delta\theta}$. If the monitor is also in control of the continuation decision he is willing to choose the tough policy if

$$(1-\overline{e})\left(m_L-m_l\right)\geq\overline{e}\left(m_h-m_L\right)\geq\frac{c}{\Delta\theta}.$$

Setting $m_L - m_l$ to its maximum value of $L - R_l$ yields that the monitor's claims can be structured such that he is willing to implement the tough policy if $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$, which is exactly the limit of region (i). As before, if control is given to a different stakeholder than the monitor, it is trivial implement the tough liquidation policy by setting $a_h = \varepsilon$, $a_L \in [\overline{e}\varepsilon, \varepsilon]$ and $a_l = 0$, where $\varepsilon > 0$ is required if the tough policy destroys value.

Note that trivially, when parameters are on the border of region (i), there are no cash flows available in some states for distribution to a controlling party. In that case control cannot be split from the monitor if a value destroying liquidation policy is supposed to obtain.

Region (ii)

If the soft policy is implemented, monitoring can only be made incentive compatible if $L - R_l \geq \frac{c}{(1-\overline{e})\Delta\theta}$. It follows directly that in region (ii) $\overline{\theta}$ cannot be incentive compatible if the soft policy is implemented. Suppose the monitor's payments are structured such that he has an incentive to monitor and has and incentive to implement the tough policy. This requires

$$\frac{c}{\Delta\theta} \leq \overline{e}\left(m_h - m_L\right) \leq \left(1 - \overline{e}\right)\left(m_L - m_l\right).$$

Since the maximum of $m_L - m_l$ is just $L - R_l$, this yields a contradiction with the restriction of region (ii) that $L - R_l < \frac{c}{(1-\overline{e})\Delta\theta}$. Hence, whenever the monitor would choose $\overline{\theta}$ he would also choose to continue the firm, rather than liquidating it.

When the monitor does not have control rights, his incentive compatibility constraint can be satisfied by setting $m_L = m_l = 0$, and $m_h = \frac{c}{\overline{e}\Delta\theta}$. Moreover, the party with control rights is willing to liquidate the firm in the absence of information when $a_h = \varepsilon$, $a_L \in [\overline{e}\varepsilon, \varepsilon]$ and $a_l = 0$. If all the remaining cash goes to a residual claimant, monotonicity of his claim requires

$$R_h - w_h^S\left(\overline{\theta}\right) - \frac{c}{\overline{e}\Delta\theta} - \varepsilon \geq L - \overline{e}\varepsilon \geq R_l. \tag{35}$$

If we set $\varepsilon = 0$ the controlling stakeholder chooses $A = S$ if that action enhances value. In that case (35) is satisfied in region (ii). If $A = S$ destroys value, we need to choose a small $\varepsilon > 0$. From the strict inequalities in region (ii) it follows that there exists a small ε such that (35) holds, and therefore claims can clearly be structured so that monitoring and implementation of the tough policy are incentive compatible for two different parties.

Region (iii)

If the tough policy is implemented, monitoring can only be made incentive compatible if $R_h - L - w_h^S(\theta) \geq \frac{c}{\overline{e}\Delta\theta}$. It follows directly that in region (iii) $\overline{\theta}$ cannot be incentive compatible if the tough policy is implemented. If the monitor takes the liquidation decision, he is willing to monitor and implement the soft policy if

$$\overline{e}\left(m_h - m_L\right) \geq \left(1 - \overline{e}\right)\left(m_L - m_l\right) \geq \frac{c}{\Delta\theta}.$$

This is possible if $\overline{e}\left(R_h - w_h^C(\theta) - L\right) \geq \frac{c}{\Delta\theta}$. Hence, in the region $\frac{c}{\overline{e}\Delta\theta} + w_h^C(\theta) \leq R_h - L \leq \frac{c}{\overline{e}\Delta\theta} + w_h^S(\theta)$ the soft policy can render monitoring incentive compatible and it can be implemented by allocating control to the monitor. If $\overline{e}\left(R_h - w_h^C(\theta) - L\right) < \frac{c}{\Delta\theta}$ monitoring incentive compatibility implies that the monitor would choose the tough liquidation policy if he had control rights. In this region the soft policy can be implemented by setting the controlling stakeholder's payoffs to $a_h = \varepsilon$, $a_L = a_l = 0$. Moreover, in the high state there must be sufficient cash available to pay the manager, and hence we require $R_h - L \geq w_h^C(\theta)$, which needs to hold with strict inequality if $\varepsilon > 0$ is needed.

Region (iv)

It follows directly from the arguments provided for region (ii) and (iii) that in region (iv) monitoring cannot be rendered incentive compatible under either liquidation policy.

Proof of lemma 4 The controlling party (the debt holder) receives payoffs $a_l = R_l$, $a_L = \min\{D, L\}$, and $a_h = D$. In order to ensure that the debt holder chooses $A(\varnothing) = S$, D must be strictly greater than R_l but no larger than $\frac{L-(1-\overline{e})R_l}{\overline{e}}$ (see (17)). The monitor's payoffs are $m_h = R_h - w_h^S(\overline{\theta}) - D$ and $m_L = \max\{0, L - D\}$. If $D \leq L$ we get $m_h - m_L = R_h - w_h^S(\overline{\theta}) - L$ so that in region (ii) monitoring is incentive compatible. If $D > L$ we get $m_h - m_L = R_h - w_h^S(\overline{\theta}) - D$. Monitoring incentive compatibility then requires the additional constraint $D \leq R_h - w_h^S(\overline{\theta}) - \frac{c}{\overline{e}\Delta\theta}$.

Proof of lemma 5 We first need to verify that if the monitor has a senior claim with face value L, his payoff satisfies the incentive compatibility constraints (14). A senior claim with face value L, is worth $m_L = L$ in liquidation but only $m_l = R_l$ when the firm continues in the bad state. The payoff to the monitor when the firm is continued in the high state is $m_h = m_L$. This contract clearly satisfies (14). The equity-holder is the residual claimant and receives $a_h = R_h - w_h^C(\overline{\theta}) - L, a_L = 0, a_l = 0$, which satisfies (17). Note further that it is in fact irrelevant whether the monitor or some senior claimant holds the safe part of the monitor's claim, i.e., the same monitoring and control policy could be implemented by selling a riskless claim with face value R_l to a senior claimant, allocating subordinated debt or preferred shares to the monitor, and leaving control in the hands of the most junior claimant (equity).

Proof of proposition 4 Note that the firm value depends on whether the monitoring and the control functions are carried out by the same or by two different stakeholders, but it does not depend on whether one of these stakeholders is the initial owner, or an outsider. This is because the initial owner has all the bargaining power when offering the monitoring and control contracts. He will therefore optimally set a price $P_{\mathbf{m}}$ for the monitoring contract such that (5) is binding: $P_{\mathbf{m}} = E(m_\varphi|\theta^*) - f(\theta^*)$. Moreover, the same is true for the controlling stakeholder, so that $P_{\mathbf{a}} = E(a_\varphi)$. The initial owner then has to put up the remaining capital $K = I - P_{\mathbf{m}} - P_{\mathbf{a}}$ and receives the value of the firm minus the payments to the monitor, the controlling stakeholder and the manager: $E(CF_\varphi - m_\varphi - a_\varphi - w_\varphi) - K$. The initial owner's objective function (7) can then be written as $E(CF_\varphi - w_\varphi) - I - f(\theta^*)$. This is the same expected value as if the initial investor awarded the monitoring or control contract (but not both) to himself. Hence, in the following we can disregard the identity of the parties who hold the contracts and can restrict attention to whether $\lambda = 0$ or $\lambda = 1$..

Consider the initial owner's objective as a function of the monitoring level under the soft and the tough liquidation policies. If the soft policy applies, his objective function is given by

$$V^C(\theta) \equiv \overline{e}\left(R_h - w_h^C(\theta)\right) + (1-\overline{e})(\theta L + (1-\theta)R_l) - f(\theta) - I. \tag{36}$$

Under the tough policy it is given by

$$V^S(\theta) \equiv \overline{e}\theta(R_h - w_h^S(\theta)) + (1-\overline{e}\theta)L - f(\theta) - I. \tag{37}$$

From (12) and (13) follows that expected wage payments are the same under either policy, if the level of monitoring is the same. Hence, from a direct comparison between (36) and (37), we conclude that the soft policy is preferred over the tough policy at the same level of monitoring, if (24) holds.

Moreover,

$$V^C\left(\overline{\theta}\right) \geq V^C\left(\underline{\theta}\right)$$
$$\Leftrightarrow$$
$$(1-\overline{e})\,\Delta\theta\,(L-R_l) \geq c - \overline{e}\left(w_h^C\left(\underline{\theta}\right) - w_h^C\left(\overline{\theta}\right)\right),$$

which always holds in region (i). Moreover,

$$V^S\left(\overline{\theta}\right) \geq V^S\left(\underline{\theta}\right)$$
$$\Leftrightarrow$$
$$\overline{e}\Delta\theta\,(R_h - L) \geq c - \overline{e}\left(\underline{\theta}w_h^S\left(\underline{\theta}\right) - \overline{\theta}w_h^S\left(\overline{\theta}\right)\right),$$

which always holds in region (i). Hence, the initial owner chooses contracts so that the soft policy will be implemented if and only if $V^C\left(\overline{\theta}\right) \geq V^S\left(\overline{\theta}\right)$.

Consider now region (ii). Since $\overline{\theta}$ is incompatible with the soft policy, the equilibrium regime is now determined by $\max\left\{V^C\left(\underline{\theta}\right), V^S\left(\overline{\theta}\right), V^S\left(\underline{\theta}\right)\right\}$ and the constraint that $V \geq 0$. Clearly, in region (ii) $V^C\left(\underline{\theta}\right) > V^S\left(\underline{\theta}\right)$. Hence, the tough policy with high monitoring is chosen if and only if $V^S\left(\overline{\theta}\right) \geq V^C\left(\underline{\theta}\right)$ and $V^S\left(\overline{\theta}\right) \geq 0$. Using (36) and (37) yields (25) and (26).

Consider region (iii). Now $\overline{\theta}$ is not feasible under the tough policy and the equilibrium is determined by $\max\left\{V^C\left(\overline{\theta}\right), V^C\left(\underline{\theta}\right), V^S\left(\underline{\theta}\right)\right\}$ and $V \geq 0$. When $\overline{R} < L$ then $V^C\left(\underline{\theta}\right) < V^S\left(\underline{\theta}\right)$. Hence, the soft policy is chosen if $V^C\left(\overline{\theta}\right) \geq V^S\left(\underline{\theta}\right)$, which is exactly the constraint (27). The constraint $V^C\left(\overline{\theta}\right) \geq 0$ is then given by (28). If $\overline{R} \geq L$, then $V^C\left(\underline{\theta}\right) \geq V^S\left(\underline{\theta}\right)$. The soft policy and high monitoring is then chosen if $V^C\left(\overline{\theta}\right) \geq V^C\left(\underline{\theta}\right)$, which is the same as

$$(1-\overline{e})\,\Delta\theta\,(L-R_l) \geq c - \overline{e}\left(w_h^C\left(\underline{\theta}\right) - w_h^C\left(\overline{\theta}\right)\right). \tag{38}$$

This always holds in region (iii).

Proof of proposition 5 Consider first the case where the soft liquidation policy is implemented. We then get three incentive compatibility constraints:

$$\begin{aligned}
&\overline{e}\,(w_h+b) + (1-\overline{e})\left(\overline{\theta}w_L + \left(1-\overline{\theta}\right)(w_l+b)\right) - c - \gamma \geq \\
&\quad \underline{e}\,(w_h+b) + (1-\underline{e})\left(\overline{\theta}w_L + \left(1-\overline{\theta}\right)(w_l+b)\right) - c, && (39)\\
&\quad \overline{e}\,(w_h+b) + (1-\overline{e})\left(\underline{\theta}w_L + \left(1-\underline{\theta}\right)(w_l+b)\right) - \gamma, && (40)\\
&\quad \underline{e}\,(w_h+b) + (1-\underline{e})\left(\underline{\theta}w_L + \left(1-\underline{\theta}\right)(w_l+b)\right). && (41)
\end{aligned}$$

Simple calculation shows that satisfying (39) and (40) implies that (41) is also satisfied. The contract which pledges most income to the principal sets $w_l = 0, w_L = b + \frac{c}{(1-\overline{e})\Delta\theta}$, and $w_h = \frac{\gamma}{\Delta e} + \overline{\theta}\frac{c}{(1-\overline{e})\Delta\theta}$.

Consider now the tough liquidation policy. Again three incentive compatibility constraints need to be satisfied for the entrepreneur.

$$
\begin{aligned}
\overline{e}\overline{\theta}\left(w_h+b\right)+\left(1-\overline{e}\overline{\theta}\right)w_L-\gamma-c \geq & \\
& \overline{e}\underline{\theta}\left(w_h+b\right)+\left(1-\overline{e}\underline{\theta}\right)w_L-\gamma, \\
& \underline{e}\overline{\theta}\left(w_h+b\right)+\left(1-\underline{e}\overline{\theta}\right)w_L-c, \\
& \underline{e}\underline{\theta}\left(w_h+b\right)+\left(1-\underline{e}\underline{\theta}\right)w_L.
\end{aligned}
$$

We can thus write the minimum payments to the entrepreneur as $w_l = w_L = 0$, and

$$
w_h^E = \max\left\{0, \frac{\gamma}{\overline{\theta}\Delta e}-b, \frac{c}{\overline{e}\Delta\theta}-b, \frac{\gamma+c}{\overline{\theta}\overline{e}-\underline{\theta}\underline{e}}-b\right\}.
$$

Proof of proposition 6 We only need to consider the case when the tough liquidation policy is implemented. The firm value in this case is given by $V^E = \overline{e}\overline{\theta}\left(R_h - w_h^E\right) + \left(1-\overline{e}\overline{\theta}\right)$. Compare this to the case where an outsider monitors who receives payment m_h in the high state and nothing otherwise. Monitoring incentive compatibility requires $m_h \geq \frac{c}{\overline{e}\Delta\theta}$. The monitor is willing to provide capital up to the amount $I_M = \overline{e}\overline{\theta}m_h - c = c\frac{\underline{\theta}}{\Delta\theta}$. The ex ante firm value is then given by $V^T = \overline{e}\overline{\theta}\left(R_h - m_h - w_h\right) + (1-\overline{e}\overline{\theta})L + I_M$, which can be written as

$$
V^T = \overline{e}\overline{\theta}\left(R_h - w_h^T(\overline{\theta})\right) + (1-\overline{e}\overline{\theta})L - c.
$$

When $w_h^E = \frac{\gamma+c}{\overline{\theta}\overline{e}-\underline{\theta}\underline{e}} - b$, (that is the case when $\gamma\underline{e}\Delta\theta < c\overline{\theta}\Delta e$) then the firm value is higher when monitoring is delegated to an outsider under the following condition: $\overline{e}\overline{\theta}\left(R_h - w_h^T(\overline{\theta})\right) + (1-\overline{e}\overline{\theta})L - c > \overline{e}\overline{\theta}\left(R_h - w_h^E\right) + \left(1-\overline{e}\overline{\theta}\right)L$. This can be rewritten as $\gamma\overline{e}\Delta\theta < c\underline{\theta}\Delta e$.

References

[1] Aghion, P. and P. Bolton, 1992, "An Incomplete Contract Approach to Financial Contracting," *Review of Economic Studies*, 59, 473-494.

[2] Aghion, P., Dewatripont, M. and P. Rey, 1994, "Renegotiation Design with Unverifiable Information" *Econometrica*, 62, 257-282.

[3] Aghion, P., D. Fudenberg, R. Holden, 2007, "Subgame Perfect Implementation with Almost Perfect Information," unpublished manuscript.

[4] Aghion, P., Hart, O. and J. Moore, 1992, "The Economics of Bankruptcy Reform" *Journal of Law, Economics and Organization*, 8, 523-546.

[5] Aghion, P. and J. Tirole, 1997, "Real and Formal Authority in Organizations," *Journal of Political Economy*, 105, 1-29.

[6] Allen, F. and D. Gale, 1988, "Optimal Security Design," *Review of Financial Studies*, 1, 229-263.

[7] Atanasov, V., Gyoshev, S., Szewczyk, S. and G. Tsetsekos, 2004, "Why Financial Intermediaries Buy Put Options from Companies", mimeo, Drexel University.

[8] Axelson, U., Strömberg, P. and Weisbach, M., 2009 "Why are Buyouts Leveraged? The Financial Structure of Private Equity Funds" *Journal of Finance*, 64(4), 1549-1582.

[9] Berglöf, E. and E.-L. von Thadden, 1994, "Short-term, versus Long-term interests: Capital Structure with Multiple Investors" *Quarterly Journal of Economics*, 139, 1055-1084.

[10] Biais, B. and C. Casamatta, 1999, "Optimal Leverage and Aggregate Investment" *Journal of Finance*, 54, 1291-1232.

[11] Bolton, P. and D.S. Scharfstein, 1996, "Optimal Debt Structure and the Number of Creditors," *Journal of Political Economy*, 104, 1-25.

[12] Boot, A. and A. Thakor, 1993, "Security Design," *Journal of Finance*, 48, 1349-1378.

[13] Bull, C., 1987, "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics*, 102, 147-159.

[14] Burkhart, M., D. Gromb, and F. Panunzi, 1997, "Large Shareholders, Monitoring and the Value of the Firm," *Quarterly Journal of Economics*, 112, 693-728.

[15] Chakraborty, A. and Yilmaz, B. 2009 "Adverse Selection and Convertible Bonds" mimeo, Wharton School, University of Pennsylvania.

[16] Conyon, M., and K. Murphy, 2000, "The Prince and the Pauper: CEO Pay in the United States and the United Kingdom", *Economic Journal*, 110, F640-F671.

[17] Cornelli, F. and O. Yosha, 2003, "Stage Financing and the Role of Convertible Securities" *Review of Economic Studies*, 70, 1-32.

[18] Cotter, James, F. and Sarah, W., Peck, 2001, "The Structure of Debt and Active Equity Investors: The Case of the Buyout Specialist," *Journal of Financial Economics*, 59, 101-147.

[19] Coval, J., Greenwood, R. and P. Tufano, 2002, "Williams, 2002" Harvard Business School case no. 9-203-068; see also R. Greenwood, 2002, "Williams, 2002" Harvard Business School teaching note no. 5-204-127.

[20] Crémer, J., 1995, "Arms Length Relationships," *Quarterly Journal of Economics*, 140, 275-295.

[21] DeMarzo, P. and D. Duffie, 1999, "A Liquidity-Based Model of Security Design," *Econometrica*, 67, 65-99.

[22] Denis, D.J., 1994, "Organizational Form and the Consequences of Highly Leveraged Transactions: Kroger's Recapitalization and Safeway's LBO," *Journal of Financial Economics*, 36, 193-224.

[23] Dewatripont, M. and J. Tirole, 1994, "A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence," *Quarterly Journal of Economics*, 139, 1027-1054.

[24] Dewatripont, M. and J. Tirole, 1999, "Advocates," *Journal of Political Economy*, 107, 1-39.

[25] Diamond, D., 1984, "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51, 393-414.

[26] Eraslan, H., 2007, "Corporate Bankruptcy Reorganizations: Estimates from a Bargaining Model" *International Economic Review*, 49, 659-681.

[27] Fluck, Z., 1998, "Optimal Financial Contracting: Debt Versus Outside Equity," *Review of Financial Studies*, 11, 383-418.

[28] Franks, J. Mayer, C. and L. Renneboog, 2001, "Who Disciplines Management in Poorly Performing Companies?" *Journal of Financial Intermediation,* 10, 209-248.

[29] Franks, J. and C. Mayer, 2005, "Ownership Matters," mimeo, University of Oxford.

[30] Fulghieri, P. and D. Lukin, 2001, "Information Production, Dilution Costs and Optimal Security Design," *Journal of Financial Economics*, 61, 3-42.

[31] Gale, D. and M. Hellwig, 1985, "Incentive Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, 52, 647-663.

[32] Gibson, S., Povel, P. and R. Singh, "The Information Content of Put Warrant Issues", mimeo, Carlson School of Management, University of Minnesota.

[33] Goldstein, I. and A. Guembel, 2008, "Manipulation and the Allocational Role of Prices," *Review of Economic Studies*, 75, 133-164.

[34] Gorton, G. and G. Pennacchi, 1990, "Financial Intermediaries and Liquidity Creation," *Journal of Finance*, 45, 49-71.

[35] Grinstein, Y., 2003, "The Disciplinary Role of Debt and Equity Contracts: Theory and Tests", *Journal of Financial Intermediation*, 15, 419-443.

[36] Grossman, S. and O. Hart, 1986, "The Costs and Benefits of Ownership: A Theory of Lateral and Vertical Integration," *Journal of Political Economy*, 94, 691-719.

[37] Guedj, I., 2005 "Ownership versus Contract: How Vertical Integration Affects Investment Decisions in Pharmaceutical R&D", mimeo, M.I.T.

[38] Guedj, I. and D. Scharfstein, 2005, "Organizational Scope and Investment: Evidence from Drug Development Strategies of Biopharmaceutical Firms", mimeo, M.I.T.

[39] Habib, M., and D.B. Johnsen, 2000, "The Private Placement of Debt and Outside Equity as an Information Revelation Mechanism," *Review of Financial Studies*, 13, 1017-1055.

[40] Harris, M. and A. Raviv, 1989, "The Design of Securities," *Journal of Financial Economics*, 24, 255-287.

[41] Harris, M. and A. Raviv, 1995, "The Role of Games in Security Design," *Review of Financial Studies*, 8, 327-367.

[42] Hart, O. and J. Moore, 1990, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119-1158.

[43] Hart, O. and J. Moore, 1999, "Foundations of Incomplete Contracts," *Review of Economic Studies*, 66, 115-138.

[44] Hart, O. and J. Moore, 2007, "Incomplete Contracts and Ownership: Some New Thoughts," forthcoming *American Economic Review.*

[45] Hellwig, M., 2009, "A Reconsideration of the Jensen-Meckling Model of Outside Finance," *Journal of Financial Intermediation*, 18, 495-525.

[46] Innes, R., 1990, "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 52, 45-67.

[47] Jensen, M., 1986, "The Agency Costs of Free Cash Flow: Corporate Finance and Takeovers," *American Economic Review*, 76, 323-329.

[48] Jensen, M. and N. Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Cost and Ownership Structure," *Journal of Financial Economics*, 3, 305-360.

[49] Kaplan, Jerry, 1996, "Start Up: A Silicon Valley Adventure" New York: Penguin Books

[50] Kaplan, Steven, N. and Jeremy C. Stein, 1993, "The Evolution of Buyout Pricing and Financial Structure in the 1980s," *Quarterly Journal of Economics*, 108, 313-357.

[51] Kaplan, Steven, N. and Per Strömberg, 2003, "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts," *Review of Economic Studies*, 70, 281-316.

[52] Khanna, N., 1998, "Optimal Contracting with Moral Hazard and Cascading", *Review of Financial Studies*, 11, 559-596.

[53] Koskinen, Y., 2000, "Debt and Information Acquisition in the Presence of a Large Shareholder," unpublished mimeo, Stockholm School of Economics.

[54] Laux, C., 2001, "Project Specific External Financing and Headquarters Monitoring Incentives," *Journal of Law, Economics and Organization*, 17, 397-412.

[55] MacLeod, B. and J. Malcomson, 1989, "Implicit Contracts, Incentive Compatibility and Involuntary Unemployment," *Econometrica*, 57, 447-480.

[56] Mahrt-Smith, Jan, 2005, "The Interaction of Capital Structure and Ownership Structure," *Journal of Business*, 78, 787-815.

[57] Maskin, E. and J. Tirole, 1999, "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, 66, 83-114.

[58] Nachman, D.C. and T.H. Noe, 1994, "Optimal Design of Securities under Asymmetric Information," *Review of Financial Studies*, 7, 1-44.

[59] Park, C., 2000, "Monitoring and Structure of Debt Contracts," *Journal of Finance*, 55, 2157-2195.

[60] Rajan, R. and A. Winton, 1995, "Covenants and Collateral as Incentives to Monitor," *Journal of Finance*, 50, 1113-1146.

[61] Repullo, R. and J. Suarez 1998, "Monitoring, Liquidation, and Security Design", *Review of Financial Studies*, 11(1), 163-187.

[62] Shleifer, A. and R.W. Vishny, 1986, "Large Shareholders and Corporate Control," *Journal of Political Economy*, 94, 461-488.

[63] Tirole, J., 1999, "Incomplete Contracts: Where Do We Stand?" *Econometrica*, 67, 741-781.

[64] Tirole, J., 2001, "Corporate Governance," *Econometrica*, 69, 1-36.

[65] v. Thadden, E.-L., 1995, "Long-Term Contracts, Short-Term Investment and Monitoring," *Review of Economic Studies*, 62, 557-575.

[66] v. Thadden, E.-L., E. Berglöf and G. Roland, 2003, “Optimal Debt Design and the Role of Bankruptcy,” unpublished manuscript.

[67] Winton, A., 1993, “Limitation of Liability and the Ownership Structure of the Firm,” *Journal of Finance*, 48, 487-512.

Appendix III

The book

Pricing the future:
The economics of discounting
and sustainable development

Christian Gollier[1]

Toulouse School of Economics

June 30, 2010

[1] This project is supported by the Chair of the Financière de la Cité at TSE.

Table of contents

Introduction

Many books, from Adam Smith's *Wealth of Nations*, Arnold Toynbee's *Study of History*, David Landes' *Wealth and Poverty of Nations*, and Angus Maddison's *Phases of Economic Development*, to Gregory Clark's *Farewell to Alms* and Jared Diamond's *Collapse*, have attempted to describe how civilizations rise, flower and then fall. This observed dynamics originates from a myriad of individual and collective costly investments taken in a broad sense, from the accumulation of capital, education, infrastructure, legal systems, or the protection of property rights. This vast literature is retrospective. In this book, I take a prospective standpoint. I analyse the problem of the selection of the investment projects that should be implemented to maximize the intergenerational welfare. The solution of this problem heavily relies on our understanding and beliefs about the dynamics of civilizations.

Future generations in the public debate

Much of life is made of investments. In this book, I examine the economic tools to evaluate actions that entail costs and benefits that are scattered through time. This is useful to optimize the impacts of our current sacrifices for a better future, both at the individual and collective levels.

The publication in 1972 of "The Limits to Growth" by the Club of Rome marked the emergence of a public awareness about collective perils associated to the sustainability of our development. Since then, citizens and politicians were confronted to a never ending list of environmental problems: nuclear wastes, genetically modified organisms, biodiversity, restoration of contaminated lands and brown fields, or exhaustion of natural resources, for example. The increased concentration of greenhouse gases in the atmosphere due to deforestation and the combustion of fossil energy is likely to affect our environment for many centuries. Experts from the Intergovernmental Panel on Climate Change tell us that this will raise the sea level, increase the frequency of extreme climatic events as droughts and cyclones, in addition to an increase of at least 5°C in the average temperature of the earth if the remaining stocks of coal, petrol and natural gas are burned. All those questions raise the crucial challenge of determining what we should and should not do in favour of future generations. This challenge is also crucial in other debates, as the ones on pension reforms, on the level of public debt, or on the investment in public infrastructures, in education, and in research.

The public decision makers are not the only one to face complex choices in the face of long-term environmental risks. Some large firms and altruistic citizens are now mobilized to contribute to a more sustainable development. Financial markets offers specific "socially responsible" funds (SRI), and new institutions have been created to supply extra-financial analyses to measure the companies' performance in the field of sustainable development. In particular, financial markets are often criticized for their short-termism, and SRI funds claim that they will restore a desirable level of long-termism in their assets evaluation rules and portfolio strategy. To say the least, these institutions together with managers of SRI funds face difficulties to converge towards an agreed-upon definition of the concept of short-termism or sustainable development, and towards a methodology to translate these concepts into operational rules for assets pricing. The absence of methodological transparency clearly limits the development of these products.

Today, the judge, the citizen, the politician and the entrepreneur are concerned by the sustainability of our development, but they don't have a strong scientific basis for the valuation of their actions and for decision-making. The objective of this book is to provide a simple framework to organize the debate on *what should we do for the future?*

What do we already do for the future?

For thousands of years since the homo sapiens emerged as the dominating species on earth, they consumed almost everything that they collected or produced over the seasonal cycle. They stayed blocked at their minimum level of subsistence for generations. The absence of the notion of private property or the inadequacy of a legal system to guarantee that what an individual saves belongs to him has been a strong incentive to consume the entire production years after years, for several millennia.

It is however clear that human beings are conscious of their own future. At the individual level, they use to make compromise between their immediate needs and their aspiration for a better future. When they are young, they invest in their human capital. Later on, they save for their retirement. They plan their own future and those of their offspring by having a bequest motive to their capital accumulation. They invest in their health by doing sport, brushing their teeth, eating healthy food. In short, they sacrifice some of their immediate pleasures for the benefit of future pleasures. At the collective level, this generated the enormous accumulation of physical and intellectual capital that the western world experienced over the last three centuries, once the individual property rights on these assets were guaranteed by a strong enough central states. New institutions, like corporations, banks, and financial markets, were created for the governance of these investments. This was a powerful engine for economic growth and prosperity. With a real growth rate of GDP per capita around 2% per year, we consume 50 times more goods and services than 200 years ago.

States and governments have also intervened in this process. They invested in public infrastructures like roads, schools, or hospitals. They heavily invested in public research whose scientific discoveries quickly diffused in the economy. At the collective level, these investments diverted some of the wealth produced in the economy to finance them rather than the immediate consumption of non-durable goods.

In this book, I want to address the difficult question of whether the allocation and the intensity of these sacrifices in favour of the future are socially efficient or not. There is indeed a myriad of ways to improve the future. Investing in the productive capital of the economy is one way, and it itself contains a multitude of options. But one can also makes the future better by limiting the extraction of exhaustible resources, by preserving the environment, by limiting emissions of greenhouse gases, or by improving the educational system for example. It is crucial that we allocate our sacrifices in favour of the future in the way that maximizes its impact on the welfare of future generations. In other words, it is crucial to put a hierarchy in this set of opportunities. This looks like 'mission impossible".

Cost-benefit analysis

Economists have developed a relatively simple and transparent toolbox to solve this challenge. The so-called cost-benefit analysis (CBA) is a set of valuation techniques that is aimed at putting priorities in the set of investment opportunities in such a way to be compatible with social welfare. One key ingredient in this toolbox is the discount rate, which can be interpreted as the minimum rate of return of a safe project to make it socially desirable to implement. This discount rate may be a function of the duration of the project, but it is absolutely crucial that one use the same discount rate to evaluate safe projects with the same duration. This will guarantee that the allocation of our current sacrifices will have a maximum impact on our future. In our decentralised economies, that is in theory organized by the existence of an interest rate to which all economic agents are confronted. In a competitive economy, the interest rate is the rate of return of the marginal capital. The uniqueness of the interest rate, if it exists, implies that only those investment projects with a rate of return greater than the interest rate will be implemented. Indeed, an investor should always compare the return of his investment project to the opportunity cost of capital, i.e., to the return of the alternative strategy which consists in investing in the productive capital in the economy.

The discount rate gives a price to time. With a discount rate of 4%, one kilogram of wheat next year has a value of only 960 grams of wheat today. The decision rule on the comparison of the internal rate of return and the discount rate can be restated equivalently as the one based on the comparison of the present value of the future benefits and the current cost. If the difference, which is called the net present value (NPV), is positive, the investment project is socially desirable.

The level of the discount rate

I want to address more specifically the question of the value of time, i.e., of the level of the discount rate. A high discount rate implies that few investment projects will successfully pass the test of a positive NPV. At the collective level, the outcome will be a low level of investments and savings. Natural resources will be quickly extracted because of the low NPV of the strategy of extracting them later. Emissions of CO_2 will not be abated because of the low present value of the damages that they will generate in the distant future. On the contrary, a reduction of the discount rate enlarges the subset of profitable investment opportunities. This means that a larger share of the wealth of nations will be invested rather than consumed. Thus, the level of the discount rate plays the key role of determining the best compromise between the present and the future.

Let me illustrate this point by considering the case of climate change. Nordhaus (2008) claims that a discount rate of 5% is socially efficient. Using an integrated assessment model, he obtained that the net present value of the future damages generated by one more ton of CO2 emitted today is 8 dollars. This means that none of the big technical projects to curb our emissions, as sequestration, wind, solar, or biofuel technologies, are socially desirable, because they all generates a cost per ton of CO2 saved larger than 8 dollars. Nordhaus concludes that climate change can be solved efficiently only through a massive investment in green research and development. On the other side, Stern (2006) implicitly used a smaller discount rate of 1.4%, thereby favouring immediate actions to fight climate change. By how much? He ended up with a NPV of future damages around 85 dollars. With this value of carbon, it is efficient to immediately implement at least some of the already available green technologies, as windmills for example. This means a massive reallocation of capital in the economy: old technologies – in particular in the energy sector -- should be made obsolete faster; consumers should replace their old cars as soon as possible, and they should refrain to spend money on vacations in order to invest in better isolating their house for example. In any case, this greener economic growth will not be beneficial to the welfare of current generations.

In 2004, a Danish statistician named Bjorn Lomborg, asked a prestigious group of economists, some with a Nobel Prize, to evaluate a set of big international projects favourable to humanity. The so-called "Copenhagen Consensus" (Lomborg (2004)) that came out of this process put top priority to public programs yielding immediate benefits (fighting malaria and AIDS, improving water supply,...), and recommended that environmental projects (climate change) be implemented only after all these other projects will get satisfactory funding. The use of a relatively large discount rate can explain this conclusion, together with the recognition that some of the most basic needs for a decent life are still not satisfied on earth in the early twenty-first century.

The case of the distant future

Suppose that the rate of return r of capital in the economy be constant. The continuously compounded value of 1 dollar invested over t years in the productive capital of the economy is $\exp(rt)$. The exponential nature of compounded interests comes from the fact that the interest obtained in the short run will themselves generate interests in the future. Reversing the argument, this means that the present value of 1 dollar in t years must be equal to $exp(-rt)$. As said before, if the interest rate is *4%*, the present value of 100 dollars next year is approximately *96* dollars. However, the net present value of 100 in 200 years equals an extremely small *4* cents. This means that one should not be ready to invest more than 4 cents today for an investment project that yields 100 dollars in 200 years. This example illustrates the origin of a long dissension between economists and ecologists. Standard ACB tools generate an almost uniform policy recommendation: Don't care about the long-term impacts of one's actions! Only the short-term costs and benefits influence the social desirability of an investment. In other words, CBA – more generally economic theory – drives short-termism in our society.

Economists have recently been working on two questions related to this dissension. First, a discount rate of 4% may be too high. To evaluate this point, one needs to think about the determinants of the discount rate, which is the main objective of this book. Second, it could be socially efficient to use a rate of 4% to discount cash flows occurring in the short run, and only 2% to discount cash flows occurring in the distant future. In other words, there is no a priori reason to use the same discount rate for different time horizons. I also address in this book the question of the term structure of the discount rate.

Recent changes in the discount rate around the world

The debate about the level the discount rate to be used to evaluate public investment projects has been fierce in the 1960s and the 1970s in most developed countries. In the United States, the debate originated in the water resources sector during the 1950s (Krutilla and Eckstein (1958)), but it quickly spread to other public policy debates, most notably energy, transportation, and environmental protection in the large. During the Nixon Administration, the Office of Management and Budget tried to standardize the widely-varying discounting assumptions made by different agencies and issued a directive requiring the use of a *10%* rate (U.S. Office of Management and Budget, OMB (1972)). This rate has since been revised downward first to about *7%* in 1992. It was argued at that occasion that the "*7% is an estimate of the average before-tax rate of return to private capital in the U.S. economy*" (OMB (2003)). In 2003, the OMB also recommanded to examine the case of a discount rate of *3%*, in addition to the *7%* mentioned above. This new rate of *3%* was justified by the "*social rate of time preference. This simply means the rate at which society discounts future consumption flows to their present value. If we take the rate that the average saver uses to discount future consumption as our measure of the social rate of time preference, then the real rate of return on long-term government debt may provide a fair approximation*" (OMB, (2003)). The *3%* then corresponds to the average real rate of return of 10-year Treasury notes between 1973 and 2003.

In the United Kingdom, the HM Treasury (2003) issued general guidance rules to evaluate public policies in the Green Book. It recommands the use of a discount rate of *3.5%*, a rate that is justified by the Ramsey rule that we will examine in chapter 2. Interestingly enough, this discount rate is reduced to *2%* for cash flows accruing more than 125 years into the future, and even to *1%* for more than 200 years. This reduction of the discount rate for the distant future is justified by the high degree of uncertainty surrounding the distant future.

From 1985 to 2005, France used a discount rate of 8% to evaluate public investments, which implied that most investments had a negative net present value. As a consequence, lobbies escaped the pressure of the experts in charge of the evaluation of public policy to get to the political decision without much use of the CBA, or they used to embellish the prospect of future social benefits of their investment projects. In fact, the choice of the 8% was itself in part justified by this intrinsic optimism. In 2004, the French government commissioned Daniel Lebègue, then a high-level civil servant, to produce a report on the discount rate. The outcome was the Lebègue Report (2005), which recommended to use a real discount rate of 4%. Moreover, using the recent development in the scientific literature, it also recommended to reduce the discount rate to only 2% for cash flows occurring in more than 30 years.

International institutions have also addressed the question of the discount rate. For example, the World Bank traditionally uses a discount rate in the range of *10-12%*. It is justified "*as a notional figure for evaluating Bank-financed projects. This notional figure is not necessarily the opportunity cost of capital in borrower countries, but is more properly viewed as a rationing device for World Bank funds*" (Operational Core Services Network Learning and Leadership Center, 1998).

Recent developments in the literature

Weitzman (1998): questionnaire to economists

For most of the XXth century, a single reference existed to drive the economic theory of the discount rate. Ramsey (1928) discovered a simple formula that links the growth of the economy and some psychological traits of consumers to the socially efficient discount rate. This so-called "Ramsey rule", which is quite simple and intuitive, played a crucial role in the shaping of the rules used to evaluate public investments. On the other side, the simple arbitrage argument evoked above suggested to use the observed interest rate on financial markets – which must equal the rate or return of marginal capital in the economy at equilibrium – as the socially efficient discount rate. Combining the two approaches yielded the well-known neoclassical theory of economic growth explored first by Solow (1956).

The modern theory of finance has also investigated much the level of the equilibrium interest rate and the shape of its term structure. Hundreds of articles have been published on this term structure, which using sophisticated mathematical tools relying on simple arbitrage arguments based on exogenous stochastic dynamics of short term risks. Given the limited economic ingredients contained in those theories, I will not devote much space presenting them in this book. The theory of finance contains many puzzles. One of them is the "risk free rate puzzle", which states that the theory predicts an equilibrium interest rate which is much larger than the one that has been observed on markets during the last century (Weil, 1989).

An intense debate emerged in the end of the nineties about whether it is socially efficient to use a discount rate for the distant future that is different from the one used to discount cash flows occurring within the next few years. The root of this literature, which has generated much controversy, is Weitzman (1998). I believe that much of this controversy is now resolved, which in part justifies the writing of this book.

References

HM Treasury, (2003), The Green Book – Appraisal and evaluation in central government, London.

Krutilla, J. V. and O. Eckstein, (1958), *Multiple purpose river development*. Baltimore, MD: Johns Hopkins Press.

Lebègue, D, (2005), Révision du taux d'actualisation des investissements publics, Commissariat Général au Plan,
http://www.plan.gouv.fr/intranet/upload/actualite/Rapport%20Lebegue%20Taux%20actualisation%2024-01-05.pdf

Lomborg, B., (2004), *Global Crises, Global Solutions*, Cambridge University Press.

Nordhaus, W.D., (2008), *A Question of Balance: Weighing the Options on Global Warming Policies,* Yale University Press, New Haven, CT.

Operational Core Services Network Learning and Leadership Center, (1998), *Handbook on Economic Analysis of Investment Operations*, Washington, DC: The World Bank.

Ramsey, F.P., (1928), A mathematical theory of savings, *The Economic Journal*, 38, 543-59.

Solow, R.M., (1956), "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70(1), pp. 65-94.

Stern, N., (2006), *The Economics of Climate Change: The Stern Review,* Cambridge University Press, Cambridge.

US Office of Management and Budget, (1972), Circular N. A-94 (Revised) To the Heads of Executive Department Establishments, Subject: Discount Rates to be Used in Evaluating Time Distributed Costs and Benefits. Washington: Executive Office of the President.

US Office of Management and Budget, (2003), Circular N. A-4 To the Heads of Executive Department Establishments, Subject: Regulatory Analysis. Washington: Executive Office of the President.

Weil, P., (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Weitzman, M. L., (1998), Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate, *Journal of Environmental Economics and Management* 36 (3): 201-208.

PART I

The simple economics of discounting

Three ways to determine the discount rate

Description of the economy

Let us consider a simple economy composed of several identical individuals who live for two periods, “today” and “the future”. These periods are indexed respectively by *0* and *t*. At the beginning of the first period, each agent is endowed with a quantity *w* of the single consumption good. Let us call this good “rice”. Rice can be consumed immediately, or it can be saved to be planted for a crop in the future. This means that rice is also an asset and a capital yielding a benefit for the future. Let us assume that planting *k* units of rice today yields *f(k)* units of grain in the future. We assume that function *f* is increasing and concave, and that *f(0)=0*. The derivative of *f* is the marginal productivity of capital, which is thus assumed to be positive and decreasing.

How should these individuals allocate their initial endowment between immediate consumption and saving/investment for the future? What should they sacrifice of their present to improve their future? In order to answer this question, one should first determine the consumers’ lifetime objective. At this stage, we take the general view that they evaluate their lifetime utility as $U(c_0,c_t)$, where c_0 and c_t are the level of consumption of rice respectively today and in the future. The bivariate utility function *U* is assumed to be increasing in its two arguments. Increasing consumption increases welfare. It is also assumed to be concave. This implies in particular that the marginal utility of rice in periods 0 and *t* is decreasing. The effect on welfare of one more grain of rice is larger when the consumption level is low than when it is high. The concavity of *U* also implies that there is a preference for consumption smoothing over time. If the two consumption plans *(1, 3)* and *(3,1)* are equally preferred, then the consumption plan *(2,2)* is certainly preferred to any of them.

Optimal consumption plan

One can use the standard graphical representation of this problem. In Figure 1, we have drawn the set of feasible consumption plans, whose upper frontier is represented by the lower concave curve. This is the locus of consumption plans *(w-k,f(k))*: when k is saved from the initial endowment w of rice, one can consume $c_0=w-k$ in the first period, and $c_t=f(k)$ in the second period. We also represented the indifference curve defined by equation $U(c_0,c_t)=U_A$ that is tangent to this feasibility frontier. All plans represented by points above this curve yield an intertemporal welfare that is larger than U_A. It clearly appears that the preferred consumption plan in the feasible set is plan A, which yields an intertemporal welfare U_A. There is no feasible consumption plan that generates an intertemporal welfare larger than that.

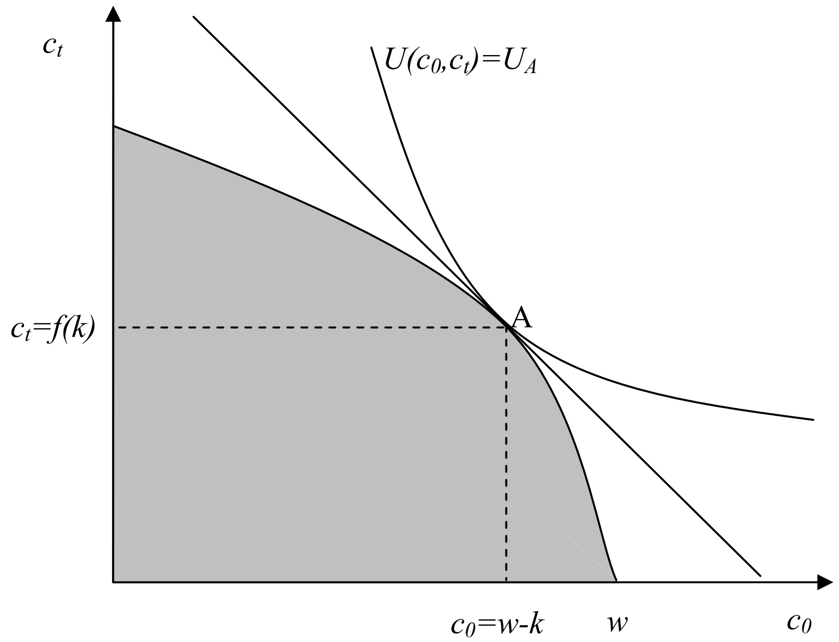


Figure 1: The optimal consumption plan

The optimal consumption plan A is characterized by the tangency of the feasibility frontier and the indifference curve. Technically, it is written as

$$f'(k)=\frac{U_0(c_0,c_t)}{U_t(c_0,c_t)}, \tag{1.1}$$

where U_i is the partial derivative of U with respect to c_i. Condition (1.1) is the first order condition of the problem of maximizing $U(w-k,f(k))$ with respect to k. The left-hand side of equation (1.1) is the marginal productivity of capital, i.e., the increase in future consumption when one more unit of rice is invested in the productive capital of the economy. It measures the (absolute value of the) slope of the feasibility frontier, evaluated at A. The right-hand side of this equality is the marginal rate of substitution between current and future consumption. It tells us by how much future consumption must be increased to compensate the sacrifice of one unit of current consumption. It measures the slope of the indifference curve at A.

Condition (1.1) has a simple economic intuition. It states that the sacrifice of one unit rice today yields an increase *f'(k)* in future consumption of rice that just compensates for the initial sacrifice. If one would have selected another plan on the frontier in the southeast of A, i.e., where *k* is smaller, the same sacrifice today would have yield a future benefit that does more than compensate the initial sacrifice. This is because the smaller *k* implies at the same time a larger marginal productivity of capital and a smaller marginal rate of substitution. This latter observation comes from the fact that in the southeast of A, consumption is very unequal over time, which implies that one is ready to sacrifice more in favour of the future.

It is useful to convert equality (1.1) between the marginal productivity and the marginal rate of substitution into an equality between rates of return. To do this, let us define

$$\rho_k = t^{-1} \ln f'(k) \text{ and } \rho_u = t^{-1} \ln \frac{U_0(c_0, c_t)}{U_t(c_0, c_t)} \tag{1.2}$$

ρ_k characterizes the rate of return of capital, since investing 1 at rate ρ_k during *t* years yields exactly $\exp(\rho_k t) = f'(k)$ in the future. Similarly, if the minimum future benefit to accept to reduce current consumption by 1 is U_0 / U_t, ρ_u characterizes the minimum *rate* of return of an investment to raise intertemporal welfare. We refer to ρ_u as the welfare-preserving rate of return of saving. Optimality condition (1.1) can be restated as requiring that $\rho_k = \rho_u$. The optimal consumption plan is such that the rate of return of capital equals the welfare-preserving rate of return.

The interest rate

Because all individuals are assumed to have the same initial endowment and the same intertemporal preferences, they will all select consumption plan A in autarky. Suppose that a frictionless credit market opens, in which agents can exchange one unit of rice today against a gross return $R = \exp(\rho t)$ expressed in units of rice delivered in the future, where *t* is the number of years between the present and the future. Observe that one can interpret ρ as the risk free interest rate in the economy. Because agents have the possibility to transfer wealth by investing in their own rice technology, a simple arbitrage argument leads to the conclusion that

$$e^{\rho t} = f'(k). \tag{1.3}$$

Suppose alternatively that *R* be larger than the marginal productivity of capital. This would imply that all agents will be willing to reduce their investment in their own rice technology to invest it on the credit market that yields a larger return. This would induce an excess supply of credit on financial markets. This cannot be an equilibrium, and the interest rate will go down.

The existence of a credit market transforms the individual feasibility condition represented by the grey area in the figure by a budget constraint corresponding to the straight line in the same figure. Its slope equals –*R*. By construction, this transformation of the constraint faced by each consumer in the economy does not change their optimal bundle.

We conclude that the competitive equilibrium on financial markets is such that the interest rate equals the rate of return of productive capital in the economy: $\rho = \rho_k$.

The discount rate

Let us know consider the crucial question in this book. Suppose that an entrepreneur, the government or a consumer is contemplating a new collective investment project. This project has an initial cost $-\varepsilon$ per capita, and it will yield a sure benefit εe^{rt} per capita in the future. One recognizes *r* as the internal rate of return of the project. In our framework in which the single consumption good is rice, this investment project could be using a fraction of the initial endowment in rice to manipulate some of the rice's genes, yielding a new production function on that specific investment. More generally, this section is devoted to examining how one should value new investment projects in the economy, like a new transportation infrastructure, investing in education, fighting climate change, or extracting a natural resource.

What is the minimum rate of return of the project under scrutiny that would make it desirable from the collective point of view? The answer to this question is usually referred to as the efficient discount rate. To characterize it, one needs to know whether the initial cost will be financed by a corresponding reduction in current consumption, or by a corresponding reduction in everyone's investment in the rice technology.

Suppose first that the initial cost is financed by a reduction of initial consumption. How does this collective investment modify the people's intertemporal welfare? Because we assume that ε is small, one can use standard differential calculus to get

$$\Delta U = -\varepsilon U_0(c_0, c_t) + \varepsilon e^{rt} U_t(c_0, c_t). \tag{1.4}$$

To get the minimum rate of return that makes the project socially desirable, one should equalize ΔU to zero. This implies that the socially efficient discount rate *r* is such that

$$e^{rt} = \frac{U_1(c_1, c_2)}{U_2(c_1, c_2)} \tag{1.5}$$

This means that the efficient discount rate is equal to the welfare-preserving rate of return: $r = \rho_u$.

Suppose alternatively that the collective investment project is financed by a corresponding reduction in the productive capital in the economy. Trivially, the project is socially desirable only if its internal rate of return is larger than the return of productive capital in the economy. This seemingly innocuous observation is crucial and deep-rooted in most economists' brain: evaluations must also be made by comparisons, and one should take into account of the opportunity cost of funds. This means that the discount rate must equal the rate of return of capital: $r = \rho_k$. This condition garantees that the marginal investment project is socially better than investing in the productive capital in the economy. Requesting that the Net Present Value (NPV) of a project be positive is equivalent to checking that this project does better for the future than all other projects available in the economy, at least at the margin.

Because consumption plans are optimized, we know that $\rho_k = \rho_u$, so that it is indifferent to know whether the initial cost is financed by a reduction in consumption or in other productive investments to characterize the socially efficient discount rate. To sum up, we have that

$$r = \rho = \rho_k = \rho_{u.} \tag{1.6}$$

Notice that we could have gone straight to the point that the efficient discount rate must be equal to the interest rate by observing that any agent can finance the initial cost by borrowing it today on the credit market. This will yield a reimbursement at date *t* equalling $\varepsilon \exp(\rho t)$, where ρ is the interest rate. Obviously, the project is efficient if its benefit at date *t* net of this reimbursement – which is refered to as the Net Future Value (NFV) -- is non negative. The critical internal rate of return is thus defined as yielding a zero NFV:

$$NFV = -\varepsilon e^{\rho t} + \varepsilon e^{rt} = 0. \tag{1.7}$$

This rule is better known as the NPV rule by multiplying the above equality by $\exp(-\rho t)$:

$$NPV = -\varepsilon + \varepsilon e^{rt} e^{-\rho t} = 0, \tag{1.8}$$

which holds if and only if $r = \rho$. This approach is very natural for any specific agent. When performing this analysis, she does not need to know whether this investment will just crowd out other investments in the economy, or whether it will reduce aggregate consumption in the economy.

Summary

In this chapter, we have shown that the socially efficient discount rate can be estimated in there different ways:

- The discount rate r is the interest rate ρ observed on financial markets. This interest rate gives us the crucial information about the intensity of the willingness to transfer wealth into the future in the economy.

- The discount rate r is the rate of return of capital. Indeed, one should invest in a new project only if its rate of return is larger than the alternative strategy to invest in the productive capital.

- The discount rate is the welfare-preserving rate of return of saving. Indeed, one should invest in a new project only if the negative welfare impact of the reduced current consumption to finance the initial cost is more than compensated by the positive welfare impact generated by the future benefit of the project.

We also noticed that these three definitions of the discount rate are fully compatible with each others when consumption plans are optimized and credit markets are frictionless.

The Ramsey rule

Why do we need a model?

The most obvious way to determine the efficient discount rate is to make it equal to the rate of return of risk free capital. This is referred to as the interest rate, which measures the opportunity cost of funds in the economy. This is certainly a good reference when the cash flows to be discounted occur in the next few months or years. Most corporations and public institutions use as their discount rate the rate at which they can borrow on financial markets. Most often this rate contains a risk premium because the projects in which they invest are risky and their cash flows are correlated with the systematic risk in the economy. It is often suggested that corporations use a rate around 15% to evaluate their investment project. Because of the risk premium that this rate contains, this is not what we call the discount rate. The discount rate is the rate at which a *sure* future benefit must be discounted to measure its present value. Thus, one needs to observe the real rate at which a truly risk free agent can borrow. Because of their unlimited ability to tax their citizens' incomes, the safest agents on this planet are governments in the western world, in particular the United States. The probability of default of these agents is small, in particular in the short term, and the short-term inflation rate is almost deterministic. Thus, the observed real rate of return on short-term public debt instruments provides a clever basis to fix the short-term discount rate.

The uncertainties surrounding the inflation and the probability of default of the borrowers in the long term imply that the rate of return on public debt instruments with longer maturities gives us a noisy signal about the equilibrium rate of return of truly safe assets of corresponding maturities. These uncertainties imply that financial markets are tainted with frictions, inefficiencies, and bubbles. This implies that models are useful to construct a scientific basis for the discount rate.

Notice also that there does not exist any public debt instrument with maturities longer than 20 or 30 years. Moreover, as is well-known from the overlapping-generation models of the theory of growth, future generations cannot trade on present credit markets, which make them intrinsically inefficient. Therefore, we don't have any clear benchmark on financial markets to help us determining the rate at which distant cash flows should be discounted. When different generations bear the costs and the benefits of the investment under scrutiny, the utility function U considered in the previous section should be reinterpreted as the social welfare function. In that framework, U characterizes the collective preferences towards the allocation of consumption across time.

In the following, we use the approach based on the welfare-preserving rate of return, which will produce the famous Ramsey rule. This approach can also be interpreted as an attempt to predict what the equilibrium interest rate should be in an economy with perfect financial markets and paternalistic investors. In other words, our aim is to price risk free assets according to a welfare-compatible interpretation of the notion of sustainable development..

Additive time preferences

In the previous chapter, we have examined a simple sure investment project yielding only two cash flows, i.e., a cost today and a benefit at some specific date *t*. We have seen that the minimum rate of return that makes this project socially desirable to implement because it raises intertemporal welfare is

$$r = \frac{1}{t}\ln\frac{U_0(c_0,c_t)}{U_t(c_0,c_t)}. \tag{2.1}$$

This socially efficient discount rate would also be the equilibrium rate of return of a zero-coupon bond with maturity *t*. In this chapter, we calibrate this simple equation. Two ingredients are required here, one on the shape of the intertemporal utility function *U*, and the other on the economic growth from c_0 to $c_1 > c_0$.

One first important simplifying assumption is that *U* is additive with respect to time. Namely, we assume that there exist two functions, *u* and v_t from $\mathbb{R}$ to $\mathbb{R}$ such that

$$U(c_0,c_t) = u(c_0) + v_t(c_t). \tag{2.2}$$

One can interpret equation (2.2) as follows: the agent evaluates his intertemporal welfare by adding his immediate utility $u(c_0)$ generated by consuming c_0 with the anticipated utility $v_t(c_t)$ generated by consuming c_t in the future. This means in particular that the level of initial consumption c_0 has no effect on the utility of consumption at date *t*. There is nothing like a formation of consumption habits or an hysteresis in feelings. This assumption is important because it allows us to isolate the two dates *0* and *t* in the evaluation of the corresponding welfare-preserving discount rate. If there would be some hysteresis, the entire consumption plan between *0* and *t* would have an effect on the marginal value of consumption at date *t*.

Exponential psychological discounting

Since Ramsey (1928), economists assume that agents are impatient, i.e., that they value their future utility less than current utility. Immediate pleasures are favoured more than the same ones, but experienced in the future. This is done by assuming that there is a single function u that links the level of current consumption to the level of current utility, and that the lifetime utility is a discounted flow of current and future utilities. In other words, the additive specification (2.2) is considered in the special case with $v_t(c) = \exp(-\delta t)u(c)$ for all c. More generally, the intertemporal welfare function is assumed to be a weighted sum of the flow of future felicities, the weight associated to any maturity t being $f(t) = \exp(-\delta t)$. Parameter δ is the rate of pure preference for the present, or the rate of impatience. Some economists refer to it as the "discount rate". Indeed, it is a discount rate, since it is used to discount the flow of future *utility*. But it is not the discount rate in the usual sense, the one that is by the experts in charge of the practical evaluation of projects, which is the rate at which one discounts the flow of future *cash flows*. Of course, there must be a link between the rate of impatience δ and the discount rate, that we denote r in this book.

The choice of the exponentially decreasing function $f(t) = \exp(-\delta t)$ for the discount factor multiplying the future utility relies on a simple argument of time consistency. Consider the same investment problem as in the previous chapter, with an initial cost to be incurred at date *0* and with a benefit at date *t*. But rather than examining the value of the project at date *0*, we do it at some date $-\tau<0$ prior to its implementation. Suppose that no new information about the quality of the project and about the environment of the investor is expected between $-\tau$ and *0*. Therefore time consistency requires that what is optimal to do from the point of view of date $-\tau$ is also optimal to do from the point of view of date *0*. Planning is rational. From the initial date $-\tau$, the duration of time before enjoying utility $u(c_0)$ is τ years, so that a discount factor $\exp(-\delta\tau)$ must be attached to that flow of utility in the welfare function. Similarly, the duration of time before enjoying utility $u(c_t)$ is $\tau + t$ years, so that a discount factor $\exp(-\delta(\tau+t))$ must multiply it in the welfare function. We conclude that this intertemporal function at date $-\tau$ can be written as

$$e^{-\delta\tau}u(c_0)+e^{-\delta(\tau+t)}u(c_t)=e^{-\delta\tau}\left(u(c_0)+e^{-\delta t}u(c_t)\right)=e^{-\delta\tau}U(c_0,c_t). \tag{2.3}$$

Thus, the objective function at date $-\tau$ is the product of a constant by the objective function at date *0*. Therefore, any project that raises the welfare $U(c_0,c_t)$ as evaluated at date *0* also raises welfare when evaluated at date $-\tau$, which guarantees time consistency. The exponential nature of the discount factor in the intertemporal welfare function guarantees that the relative "exchange rate" of utility for any pair of dates is insensitive to the passing of time. Other specifications as the hyperbolic one with $f(t)=(1+at)^{-1}$ induce time inconsistent behaviours.

Rate of impatience

There is a simple way to estimate the rate of impatience δ. Suppose that you believe that your income next year will be as it is this year. What is the minimum interest rate that would induce you to save? The answer to this question is what we called the welfare-preserving rate of return, which is defined by equation (2.1). Under the above assumptions with $c_0=c_t$, we obtain that $U_0/U_t=\exp(-\delta t)$, so that $r=\delta$. Thus, the rate of impatience is equal to the minimum interest rate that induces people to save when their income profile is flat.

There is no convergence among experts toward an agreed-upon or unique rate of impatience. Frederick, Loewenstein and O'Donoghue (2002) conducted a meta-analysis of the literature on the estimation of the rate of impatience. Rates differ dramatically across studies, and within studies across individuals. For example, Warner and Pleeter (2001), who examined actual households' decisions between an immediate down-payment and a rent, found that individual discount rates vary between 0% and 70% per year! Thus the calibration of δ is problematic if our objective is positive, i.e., if one wants to explain real behaviours.

As long as consumption at date *0* and *t* concerns a given person, impatience is a psychological trait that economists should take as given. However, many experts in our field have criticized impatience. Arrow (1999) cites various classical authors on this matter. The most well-known citation is from Ramsey (1928) himself: "It is assumed that we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination." Harrod: "Pure time preference [is] a polite expression for rapacity and the conquest of reason by passion." Koopmans: "[I have] an ethical preference for neutrality as between the welfare of different generations." Solow: "In solemn conclave assembled, so to speak, we ought to act as if the social rate of pure time preference were zero." The general view is thus that a small or zero discount rate should be used when the flow of utility over time is related to different generations. The fact that I discount my own felicity next year at *2%* does not mean that I should discount my children's felicity next year at *2%*. In fact, there is no moral reason to value the utility of future generations less than the utility of the current ones. As explained by Broome (1991), good at one time should not be treated differently from good at another, and the impartiality about time is a universal point of view. The normative doctrine is that the rate of time preference is zero. Later on in this book, we will take a *normative* stand to set δ at zero, because present and future utilities will be allocated to different generations of consumers.

Aversion to intertemporal inequality of consumption

We have seen in the previous section that the concavity of the intertemporal welfare function U characterizes a preference for the smoothing of consumption over time. In the additive case that we examine here, this is translated into the concavity of the utility function u. We define the following local measure of the degree of concavity of the utility function u:

$$R(c) = -\frac{cu''(c)}{u'(c)}. \tag{2.4}$$

We hereafter refer to this index as the relative aversion to intertemporal inequality. Let us explain why. Suppose that the consumption plan (c_0, c_t) is unequally distributed over time. Suppose more particularly that future consumption is larger than current consumption: $c_t > c_0$. How much one would be ready to pay today to increase consumption by one in the future? This should not be much, for two reasons: impatience and aversion to consumption fluctuations. In the absence of any of these two effects, one would exchange one for one. Let k be the maximum reduction in current consumption that is compatible with the unit increase in future consumption. It must satisfy the following indifference condition:

$$ku'(c_0) = e^{-\delta t}u'(c_t). \tag{2.5}$$

Assume that $t=1$, and that δt and $c_1 - c_0$ are small. Using a first-degree Taylor approximation of $u'(c_1)$ around c_0 and using the approximation $\exp(-\delta t) \simeq 1 - \delta t$ implies that

$$ku'(c_0) \simeq (1-\delta t)\left(u'(c_0) + \frac{c_1 - c_0}{c_0} c_0 u''(c_0)\right) \quad (2.6)$$

This can in turn be approximated as

$$k \simeq 1 - \delta - \frac{c_1 - c_0}{c_0} R(c_0) \quad (2.7)$$

This equation is helpful to estimate your relative aversion to intertemporal inequality $R(c_0)$. Suppose that your rate of impatience is $\delta=0$, and that you anticipate an increase in future consumption by 10%. In spite of this increase, you are considering the transfer of consumption to the future. What is the maximum reduction k of consumption that you are ready to sacrifice today to increase future consumption by 1 dollar? The answer to this question gives us an estimation of your relative aversion to intertemporal inequality, since by (2.7), $R(c_0)=1-10k$. For example, answering *90* cents to the question yields a relative aversion $R=1$, whereas an answer of *80* cents yields a relative aversion $R=2$.

There is still much consensus on the intensity of relative aversion to intertemporal inequality. Using estimates of demand systems, Stern (1977) found a concentration of estimates of R around *2* with a range of roughly *0-10*. Hall (1988) found an R around *10*, whereas Epstein and Zin (1991) found a value ranging from *1.25* to *5*. Pearce and Ulph (1995) estimate a range from 0.7 to 1.5. Following Stern (1977) and our own introspection, we will hereafter consider $R=2$ as a reasonable value for the relative aversion.

When different generations are concerned by the investment project to be evaluated, the choice of the discount rate entails interpersonal comparisons of utility. In that case, function U is interpreted as a social welfare function, and the concavity of u characterizes the aversion to interpersonal inequality. Is the level of R affected by this shift in analysis? In this literature, it is generally assumed that our normative attitude towards consumption inequalities should not depend upon the nature of the comparisons of consumption levels. Under the common paternalistic view, one should evaluate the impact on social welfare of an intertemporal inequality of consumption exactly as if it would be an interpersonal inequality. We claim that the two problems are equivalent by nature. If one is ready to pay up to *80* cents to increase own consumption by one dollar next year in spite of an anticipated *10%* increase in consumption, one should also be ready to give up *80* cents in order to offer one dollar to another person that is *10%* wealthier than us. Thus, we maintain that $R=2$ is a sensible level of relative aversion to intertemporal inequality even in the intergenerational context.

The power utility

Economists and econometricians often limit their analysis by using a specific utility function in their model. They usually favour exponential, quadratic, logarithmic or power utility functions. In this book as in the modern theory of finance, we will often consider the special case of the power utility function:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}. \tag{2.8}$$

Parameter γ is positive and different from *1*. When $\gamma=1$, we take $u(c) = \ln(c)$, since it can be verified that the limit of (2.8) when γ tends to *1* is the logarithmic utility function. Because $u'(c) = c^{-\gamma}$, these utility functions are increasing and concave. Moreover, the index *R* of relative aversion to intertemporal inequality is constant, and is equal to γ.

As usual, this is not an innocuous assumption. The constancy of the relative aversion means in particular that the answer *k* to the above question depends only upon the growth rate of consumption, and not upon the initial level of consumption. This assumption may be questioned, in particular given the fact that there must be some minimum level of subsistence. In the same vein, the utility function (2.8) implies that the marginal utility tends to infinity when consumption tends to zero. It implies that one would be ready to sacrifice almost 100% of one's current wealth in order to increase by one dollar wealth in a state of nature in the future where consumption vanishes totally. We don't think that this is realistic, which implies that we will be quite cautious in the use of the classical power model entailing Armageddon scenarios.

The Ramsey rule

We can combine the above ingredients to cook the dish served in this chapter. Rewriting equation (2.1), the efficient discount rate must be equal to

$$r = \frac{1}{t}\ln\frac{u'(c_0)}{e^{-\delta t}u'(c_t)} = \delta - \frac{1}{t}\ln\frac{u'(c_t)}{u'(c_0)}. \tag{2.9}$$

A Taylor expansion of $u'(c_t)$ around c_0 yields

$$r \simeq \delta + \frac{c_t - c_0}{tc_0} R(c_0). \tag{2.10}$$

Equations (2.9) and (2.10) show that the socially efficient discount rate has two components. It is the sum of the rate of impatience and of a wealth effect, which is positive when people expect a positive growth of their consumption. This wealth effect is approximately equal to the product of the yearly growth rate of consumption and of the relative aversion to intertemporal inequality. This approximation is exact in the special case of the power utility function. Indeed, plugging $c_t = c_0 \exp(gt)$ and $u'(c) = c^{-\gamma}$ in equation (2.9) yields

$$r = \delta + \gamma g, \tag{2.11}$$

where g is the yearly growth rate of consumption between dates *0* and *t*. This is the well-known Ramsey rule, which links the efficient discount rate to two "taste" parameters (the rate of impatience δ and the relative aversion γ) and the growth rate of the economy.

When people expect that the economy will grow fast in the future, their aversion to intertemporal inequality makes them reluctant to sacrifice further the present in favour of the already better future. They will be willing to do so only if the rate of return of their investment is large enough to compensate the induced increase in intertemporal inequality and their pure preference for the present. This behaviour can be observed on financial markets. When households have better expectations about their future income, they reduce their savings, which implies in turn an increase in the equilibrium interest rate *r*. On the contrary, the expectation of a recession induces them to save more, which implies a reduction in the equilibrium interest rate. In short, the interest rate varies pro-cyclically.

What does come out from this approach?

Several experts have used the Ramsey rule (2.11) to make recommendations on the choice of the discount rate to evaluate public policies, in particular towards climate change. The easiest proposal to memorize is by Weitzman (2007), who recommend the use of a trio of twos:

$$\delta=2\%, \quad g=2\% \text{ and } \gamma=2. \tag{2.12}$$

We share the view of Weitzman that "these numbers at least pass the laugh test". They yield a discount rate of *6%*. Nordhaus (2008) uses *5%* by using a smaller rate of impatience $\delta=1\%$.

Stern (2007) has often been criticized for using a much smaller discount rate around *r=1.4%*. In fact, because the impacts of global warming cannot be considered as marginal, the standard evaluation technique based on the net present value cannot be used. This is why Stern (2007) did not actually use any discount rate. Rather, he measured the monetary equivalent of the impact of climate change on the intertemporal welfare function. However, he considered the following trio of parameter values:

$$\delta=0.1\%,\ g=1.3\% \text{ and } \gamma=1. \tag{2.13}$$

The choice of the rate of time preference at *0.1%* comes from the moral stand of time impartiality – each to count for one, and none for more than one --, and from the possibility of extinction for which the probability of occurrence is fixed at *0.1%* per year. Observe also that Stern assumes a logarithmic utility function, whose relative risk aversion($\gamma=1$) is a lower bound of what is considered as realistic values for *R*. Trio (2.13) yields a discount rate *r=1.3%*, which is considered as a radical position by a vast majority of economists. It drives the conclusion of the Stern Review urging governments around the world to act immediately and strongly to reduce emissions of greenhouse gases.

Following the publication of the Green Book (2003), the UK recommends a discount rate of 3.5% based on the following calibration of the Ramsey rule:

$$\delta=1.5\%,\ g=2.0\% \text{ and } \gamma=1. \tag{2.14}$$

In France, the « Rapport Lebègue » (2005) has been endorsed by the French government in order to use a 4% discount rate for all cash flows with a maturity less than 30 years. This recommendation is based on the following calibration of the Ramsey rule:

$$\delta=0\%,\ g=2\% \text{ and } \gamma=2. \tag{2.15}$$

For time horizons longer than 30 years, a forward discount rate of 2% is used[2], on the basis of arguments that will be made explicit in Part 2 of this book.

Conclusion

The Ramsey rule (2.11) gives us the efficient discount rate based on the estimation of the welfare-preserving rate of return of saving. It relies on three parameters: the rate of impatience, the relative aversion to intertemporal inequality, and the growth rate of the economy. We have justified that intertemporal preferences, when they concern different people, should guarantee the impartiality of time. Thus, the collective rate of impatience

[2] Thus, the discount factor to be used for a maturity *t* larger than 30 is $e^{-(0.04*30+0.02(t-30))}$.

should be zero. We have also advocated a relative aversion to intertemporal inequality around $R=2$. Under these assumptions, the socially efficient discount rate should be around twice the growth rate of consumption per capita. Because the mean growth rate of consumption per capita has been approximately *2%* per year in the western world over the last two centuries, the extrapolation of this fact would justify using a real discount rate of *4%*. However, the calibration of the growth rate g in the Ramsey rule is problematic, because of the large uncertainty surrounding the evolution of our economies in the years, decades and centuries to come. In the next chapter, we explain how to overcome this difficulty.

References

Arrow, K. J. (1999), Discounting and Intergenerational Equity, in Portney and Weyant (eds), *Resources for the Future*.

Broome, J., (1992), *Counting the Cost of Global Warming*, White Horse Press, Cambridge.

Epstein, L.G., and S. Zin, (1991), Substitution, Risk aversion and the temporal behavior of consumption and asset returns: An empirical framework, *Journal of Political Economy*, 99, 263-286.

Frederick, S., G. Loewenstein and T. O'Donoghue, (2002), Time discounting and time preference: A critical review, *Journal of Economic Literature*, 40, 351-401.

Hall, R.E., (1988), Intertemporal substitution of consumption, *Journal of Political Economy*, 96, 221-273.

HM Treasury, (2003), The Green Book – Appraisal and evaluation in central government, London.

Nordhaus, W.D., (2008), *A Question of Balance: Weighing the Options on Global Warming Policies,* Yale University Press, New Haven, CT.

Pearce D and Ulph D (1995), *A Social Discount Rate For The United Kingdom*, CSERGE Working Paper No 95-01 School of Environmental Studies University of East Anglia Norwich

Ramsey, F.P., (1928), A mathematical theory of savings, *The Economic Journal*, 38, 543-59.

Rapport Lebègue, (2005), Révision du taux d'actualisation des investissements publics, Commissariat Général au Plan, Paris.
http://www.plan.gouv.fr/intranet/upload/actualite/Rapport%20Lebegue%20Taux%20actualisation%2024-01-05.pdf

Stern, N., (1977), The marginal valuation of income, in M. Artis and A. Nobay (eds*), Studies in Modern Economic Analysis*, Blackwell: Oxford.

Warner, J.T., and S. Pleeter, (2001), The personal discount rate: Evidence from military downsizing programs, *American Economic Review*, 95:4, 547-580.

Weitzman, M.L., (2007), The Stern review on the economics of climate change, *Journal of Economic Literature,* 45 (3), 703-724.

Extending the Ramsey rule to risk

A decision criterion under risk

Uncertainty is a day-to-day fact. We don't really know today how tomorrow will look like, and, for many of us, the more distant future may be extremely uncertain. This makes the dynamic optimization problem of our lifetime welfare complex. In particular, determining the optimal level of savings requires estimating the future utility gain of this transfer of wealth in a context in which little is known about future incomes. This problem is at the core of the question of what should be done for the future.

When the growth rate of consumption is unknown, the intensity of the wealth effect described in the previous chapter cannot estimated, and the Ramsey rule (2.11) is unable to produce a coherent solution to the choice of the discount rate. Estimating the growth rate for the coming year is already a difficult task. No doubt, any estimation of growth for the next century/millennium is subject to potentially enormous errors. The history of the western world before the industrial revolution is full of important economic slumps, as the one due to the invasion of the Roman Empire, or the one due to the Black Death in the mid XIVth century. The recent debate on the notion of a sustainable growth is an illustration of the degree of uncertainty we face to think about the future of Society. Some will argue that the effects of the improvements in information technology have yet to be realized, and the world faces a period of more rapid growth. On the contrary, those who emphasize the effects of natural resource scarcity will see lower growth rates in the future. Some even suggest a negative growth of the GDP per head in the future, due to the deterioration of the environment, population growth and decreasing returns to scale. They claim that the wealth effect goes the other direction, so that everything should be made to improve the future. This uncertainty at least casts some doubt on the relevance of the wealth effect to justify the use of a large discount rate.

In order to address the question of the role of uncertainty on the selection of the discount rate, we need to characterize its impact on welfare. We hereafter follow the classical approach relying on the Bernoulli-von Neumann-Morgenstern expected utility theory. More specifically, we assume that when the consumption level c_t at date *t* is uncertain, the ex ante welfare at that date is measured by the expected utility of this uncertain consumption. Thus, seen from date *0*, the social welfare in the economy is written as

$$V = u(c_0) + e^{-\delta t} Eu(c_t), \tag{3.1}$$

where the expectation operator E is related to the probability distribution of the random variable c_t. The expected utility criterion relies on a very intuitive "independence axiom". Consider three different acts, A, B and C. A could be to go to the movie, B could be to go to the restaurant, and C is to stay home. Under this axiom, if one prefers A with certainty rather than B with certainty, one also prefer the lottery which yields A with probability *p* to the lottery which yields B with the same probability, the alternative being to get C with probability *1-p* in the two cases. In spite of its intuitive appeal, the Allais' paradox tells us that some agents may violate this axiom in certain circumstances. Because our aim is mostly normative, we believe that public policies should rely on the independence axiom.

Risk aversion

An agent is risk-averse if he always prefers the expected payoff of a lottery to the lottery itself. In the expected utility model, it is well-known that the concavity of the von Neumann-Morgenstern utility function characterizes the aversion to risk of the decision maker. Indeed, by Jensen's inequality, the concavity of *u* implies that $Eu(c_t)$ is smaller than $u(Ec_t)$. Because marginal utility is decreasing, a mean-preserving reduction in risk increases expected utility. For example, if future consumption is *80* or *120* with equal probabilities, decreasing marginal utility implies that increasing consumption by *20* in the bad state increases utility more than the reduction of utility when reducing consumption by *20* in the good state. Therefore, this risk elimination is welfare-improving ex ante.

Let $z = Ec_t$ and $\varepsilon_t = (c_t - z)/z$ denote respectively the expected consumption and the relative risk in date *t*. Let also π denote the risk premium, which is defined as the maximum price that one is ready to pay for the elimination of ε_t, expressed as a fraction of expected consumption:

$$u(z(1-\pi)) = Eu(z(1+\varepsilon_t)). \tag{3.2}$$

The level of π measures the degree of risk aversion, with $\pi = 0$ corresponding to risk neutrality. The well-known Arrow-Pratt approximation allows us to link π to the variance σ_t^2 of ε_t and to the index of concavity of *u*, which is $R(c) = -cu''(c)/u'(c)$:

$$\pi \simeq 0.5\sigma_t^2 R(z) \tag{3.3}$$

It is obtained through Taylor approximations of the two sides of equation (3.2) around z. The relative risk premium is approximately equal to half the product of the variance of the relative risk and of the index of relative risk aversion *R*.

Equation (3.3) gives us a new opportunity to estimate the degree of concavity of *u*. Suppose that one's consumption is subject to a fifty-fifty chance of an increase or a reduction by *10%*. What fraction of consumption one is ready to pay to eliminate this risk? Since σ_t^2 equals *1%* in this case, the answer to this question should approximately be equal to *0.5%* of *R*. For example, when relative risk aversion equals *2*, this fifty-fifty chance of a gain or a loss of *10%* of consumption is equivalent to a sure loss of $\pi \simeq 1\%$. This test is a reassurance that *R=2* is a reasonable level of concavity of the utility function.

What is the quality of the Arrow-Pratt approximation (3.3)? Of course, its quality is reduced when the size of risk ε_t is increased. There is however one special case in which approximation (3.3) is exact, whatever the size of the risk is. Because this special case is almost universally used in the theory of finance and latter on in this book, it is good to write it as a formal Lemma.

Lemma: *Suppose that x is normally distributed with finite mean μ and variance σ^2. Consider any scalar $A \in \mathbb{R}$. Then, we have that*

$$Ee^{-Ax} = e^{-A(\mu - 0.5A\sigma^2)}. \tag{3.4}$$

In other words, the Arrow-Pratt approximation (3.3) is exact when the risk is normally distributed and the utility function is exponential.

We prove this lemma in the appendix of this chapter.

It is noteworthy that in this additive model, the concavity of *u* plays two different roles: the aversion to intertemporal inequality, and the aversion to risk. This property of the so-called

Discounted Expected Utility model has often been criticized in the literature because the attitudes towards risk and time are often considered to have different natures. This is clearly a handicap to explain how people behave in the face of risk and time. However, from a normative point of view, the use of decreasing marginal utility to explain the two types of aversion is quite appealing. It makes sense to link the resistance to transfer wealth to a wealthier future and to a wealthier state of nature to the property that marginal utility is decreasing.

Prudence and precautionary saving

What is the impact of uncertainty on our willingness to improve the future? Before examining this question at a global level, it is useful to go back to the level of the individual. The most obvious action that we do in favour of our own future is to save. So, it is useful to explore the effect of the uncertainty affecting our own future incomes on our saving behaviour. This will be helpful to tell us how we should collectively behave in the face of our uncertain collective destiny. After all, any collective risk will percolate into risks that must be borne by individuals. The intuition suggests that the uncertainty surrounding the future should raise our willingness to save. This is the concept of precautionary saving introduced by Keynes, which has been revisited since then by Leland (1968), Drèze and Modigliani (1972) and Kimball (1990), among else.

Consider an individual who has a flow of income y_0 at date *0*, and y_t at date *t*. His optimal saving *s* solves the following maximization program:

$$\max_s V(s) = u(y_0 - s) + e^{-\delta t} Eu(y_t + e^{rt} s), \tag{3.5}$$

where r is the interest rate. Under the concavity of *u*, the objective function *V* is concave in *s*, and the following first-order condition is necessary and sufficient:

$$V'(s) = -u'(y_0 - s) + e^{(r-\delta)t} Eu'(y_t + e^{rt} s) = 0 \tag{3.6}$$

We compare two cases. In the safe case, y_t equals some constant *z* with certainty. Without loss of generality, suppose that the optimal saving is zero in that case. In the uncertainty case, $y_t = z(1+\varepsilon_t)$, where ε_t is a zero-mean relative risk on future income. Compare to the certainty case, this future risk raises the optimal saving if and only if it raises *V'(0)*. This is the case if and only if

$$Eu'(z(1+\varepsilon_t)) \geq u'(z). \tag{3.7}$$

Because risk $z\varepsilon_t$ has a zero mean, this is the case if and only if *u'* is convex. Marginal utility must be decreasing at a decreasing rate. Following the terminology introduced by Kimball (1990), we say that an agent is prudent if his marginal utility is convex. Prudence is the necessary and sufficient condition to guarantee that individuals want to save more when the future becomes more uncertain.

Let us define the precautionary premium ψ as the sure relative reduction in future income that has the same effect on saving than the future risk on income:

$$u'(z(1-\psi)) = Eu'(z(1+\varepsilon_t)). \tag{3.8}$$

We say that $z(1-\psi)$ is the precautionary equivalent of $z(1+\varepsilon_t)$. Comparing equations (3.8) and (3.2), observe that the precautionary premium ψ of *u* is the risk premium of *–u'*, which is increasing and concave under prudence. By analogy, we can rewrite equation (3.3) as

$$\psi \simeq 0.5\sigma_t^2 P(z), \tag{3.9}$$

where $P(z) = -zu'''(z)/u''(z)$ is the index of relative prudence (Kimball (1990)). Thus, adding a zero-mean relative risk to future consumption has an effect on current saving that is approximately equal to half the product of the variance of this risk and of the index of relative prudence.

There has not been much attempt to estimate the degree of prudence. Usually, researchers considered a family of utility functions with a single parameter. In that case, the choice of this parameter is determined by the assumed degree of risk aversion of the decision maker, and the degree of relative prudence is derived from it. For example, consider the case of the power utility function, with $u'(c) = c^{-\gamma}$, which implies that $u''(c) = -\gamma c^{-\gamma-1} < 0$ and $u'''(c) = \gamma(\gamma+1)c^{-\gamma-2} > 0$. It yields $R(c) = \gamma$ and $P(c) = \gamma + 1$. For power functions, relative prudence equals relative risk aversion plus one. If we take *R=2*, we obtain *P=3*. Facing a fifty-fifty chance of gaining or loosing *10%* of future income has an effect on current saving that is approximately equivalent to the effect of a sure reduction of future income by *1.5%*.

Is the convexity of marginal utility a natural assumption to make? Observe first that we already assumed that marginal utility is positive and decreasing. It implies that it must be convex at least locally for large consumption levels. Observe also, but this is not a very convincing argument, that all classical utility functions used in economics exhibit a convex marginal utility. This is the case for the exponential, the power and the logarithmic ones. The quadratic utility function has a linear marginal utility.

Two positive arguments are in favour of prudence. The first is of course that we observe that people increase their saving when their future becomes more uncertain. See for example the econometric analysis by Guiso, Jappelli and Terlizzese (1996). Second, people are downside risk-averse, which is another term for prudence. Downside risk aversion can be better understood by following the definition proposed by Eeckhoudt and Schlesinger (2006). Suppose that your future consumption is either a low z_l or a high z_h, with equal probabilities. Suppose that you are forced to bear a zero mean risk in one of these two states. Do you prefer to allocate it to the low-consumption state or in the high-consumption one? If you answer that it is better to put it in the high-consumption state, you must be downside risk-averse, i.e. prudent. Indeed, it means that

$$\frac{1}{2}Eu(z_h+\varepsilon)+\frac{1}{2}u(z_l)\geq\frac{1}{2}u(z_h)+\frac{1}{2}Eu(z_l+\varepsilon), \tag{3.10}$$

or equivalently, that

$$Eu(z_h+\varepsilon)-Eu(z_l+\varepsilon)\geq u(z_h)-u(z_l). \tag{3.11}$$

Rewriting this inequality as

$$\int_{z_l}^{z_h}\left[Eu'(z+\varepsilon)-u'(z)\right]dz\geq 0, \tag{3.12}$$

we see that the preference for putting risk in the higher income state requires that marginal utility be convex.

The extended Ramsey rule as an approximation

The uncertainty surrounding the growth of consumption affects the welfare-preserving rate of return of saving. Let us consider a marginal investment that has a unit cost today and that yields a sure benefit $\exp(rt)$ at date *t*. It preserves the intertemporal welfare *V* defined by (3.1) if and only if

$$-u'(c_0)+e^{-\delta t}e^{rt}Eu'(c_t)=0. \tag{3.13}$$

This can be rewritten as

$$r=\delta-\frac{1}{t}\ln\frac{Eu'(c_t)}{u'(c_0)}. \tag{3.14}$$

Now, remember that the existence of the relative risk $\varepsilon_t=(c_t-Ec_t)/Ec_t$ on future consumption has an effect on expected marginal utility that is equivalent to a sure relative reduction of consumption by the precautionary premium . Technically, using (3.8), the above equation can be rewritten as

$$r=\delta-\frac{1}{t}\ln\frac{u'((1-\psi)Ec_t)}{u'(c_0)}. \tag{3.15}$$

We are thus back to the certainty case that we examined in the previous chapter. We can for example rewrite approximation (2.10) as follows:

$$r\simeq\delta+\frac{(1-\psi)Ec_t-c_0}{tc_0}R(c_0). \tag{3.16}$$

This is reminiscent of the Ramsey rule with an impatience effect and the wealth effect, but the latter is reduced by risk. This reduction ψ can be approximated by using equation (3.9). Alternatively, we can use a second-degree Taylor approximation of $u'(c_t)$ around c_0 in (3.14) to get

$$r\simeq\delta+t^{-1}E\left(\frac{c_t-c_0}{c_0}\right)R(c_0)-\frac{1}{2}t^{-1}Var\left(\frac{c_t-c_0}{c_0}\right)R(c_0)P(c_0). \tag{3.17}$$

This is the extended Ramsey rule. As in the standard Ramsey rule (2.10), we have an impatience effect and a wealth effect. The third term in the right-hand side of the above equation is what we call the precautionary effect. It tends to reduce the discount rate. Its intensity is proportional to the product of relative prudence by relative risk aversion, and to the annualized variance of the growth rate of consumption between *0* and *t*.

This confirms the intuition that the uncertainty affecting the future tends to raise our willingness to invest for that future. This is done by reducing the discount rate.

The extended Ramsey rule in the lognormal case

The extended Ramsey rule described by (3.17) can be obtained as an exact solution in an important special case. Let us consider a one year horizon (*t=1*). Suppose that

$$c_1 = c_0 e^x, \tag{3.18}$$

where x is the continuously compounded growth rate of consumption, i.e. the increase in the logarithm of consumption. Let us assume that x is normally distributed with mean μ and variance σ^2. Notice that, using equation *(3.4)* with *A=-1*, it implies that the expected total growth rate of consumption between dates *0* and *1* is $g = \ln(Ec_1 / c_0) = \mu + 0.5\sigma^2$.

Suppose also that the representative agent in the economy has a power utility function, with $u'(c) = c^{-\gamma}$. This implies that

$$\frac{Eu'(c_1)}{u'(c_0)} = \frac{Ec_0^{-\gamma} e^{-\gamma x}}{c_0^{-\gamma}} = Ee^{-\gamma x}. \tag{3.19}$$

Now, we can use the lemma summarized by equation *(3.4)* to rewrite the right-hand side of the above equation as $\exp(-\gamma(\mu - 0.5\gamma\sigma^2))$. Plugging this in the pricing formula (3.14) yields

$$r = \delta + \gamma\mu - 0.5\gamma^2\sigma^2. \tag{3.20}$$

We prefer to rewrite this formula by using the expected growth rate *g*:

$$r = \delta + \gamma g - 0.5\gamma(\gamma + 1)\sigma^2. \tag{3.21}$$

This exact extended Ramsey rule combines the three components of the efficient discount rate: impatience, the wealth effect, and the precautionary effect. The wealth effect is positive and is the product of the expected growth rate of consumption by the relative aversion to intertemporal inequality. The precautionary effect is negative, and is equal to half the product of three factors: relative aversion γ, relative prudence $\gamma + 1$, and the variance of the growth rate of consumption.

Calibration of the extended Ramsey rule

In the previous chapter in which we discarded risk, we justified the use of $\delta = 0$, $\gamma = 2$ and *g=2%*. This justifies using a discount rate of *4%* per year. How much smaller than *4%* should the discount rate be to take account of the future risk? In order to answer this question for a one-year horizon, one needs to estimate the volatility of the yearly growth rate of consumption.

Kocherlakota (1996), using United States annual data over the period 1889-1978, estimated the standard deviation σ of the growth of consumption per capita to 3.6% per year. Assuming normality and an expected growth rate of *2%*, this means that there is a *95%* probability that the actual growth rate of consumption next year will be between *-5%* and *+9%*. Using $\sigma^2 = (0.036)^2$ and $\gamma = 2$ yields a precautionary term in the extended Ramsey rule (3.21) equalling *-0.4%*. The precautionary effect reduces the efficient rate at which one should discount cash flows occurring next year from *4%* to *3.6%*.

μ	σ	δ	γ
2%	*3.6%*	*0%*	*2*

Table : Benchmark calibration of the extended Ramsey rule

Conclusion

It is commonly accepted that individuals are ready to make more effort for their future when this future becomes more uncertain. Keynes was the first to mention this idea by pointing out the precautionary motive to saving. What is desirable at the individual level is also desirable at the collective one. A Society which wants to reinforce the incentive to invest for the future should select a smaller discount rate to perform the socioeconomic evaluation of the set of all possible investment projects.

Calibrating the uncertainty affecting the short-term macroeconomic growth on U.S. data over the last century justifies reducing the short-term discount rate by *0.4%*. In short, taking into account of short-term risk, the efficient short-term discount rate should be reduced from *4%* to *3.5%*. This can be considered as a marginal reduction. In the next few chapters, we re-examine this question by considering long-term risk and discount rates.

APPENDIX

Lemma: *Suppose that x is normally distributed with finite mean* μ *and variance* σ^2. *Consider any scalar* $A \in \mathbb{R}$. *Then, we have that*

$$Ee^{-Ax} = e^{-A(\mu - 0.5A\sigma^2)}. \tag{3.22}$$

Proof : Suppose that $u(c) = -\exp(-Ac)$. If *c* is normally distributed with mean μ and variance σ^2, we have that

$$Eu(c) = \frac{-1}{\sigma\sqrt{2\pi}} \int \exp\left(-Ac\right) \exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right) dc.$$

Rearranging the integrant, we obtain

$$Eu(c) = -\exp\left(-A\left(\mu - \frac{A\sigma^2}{2}\right)\right) \frac{1}{\sigma\sqrt{2\pi}} \int \exp\left(-\frac{(c-\left(\mu - A\sigma^2\right))^2}{2\sigma^2}\right) dc.$$

Because the integral of a density function equals *1*, this implies that

$$Eu(c) = -\exp\left(-A\left(\mu - \frac{A\sigma^2}{2}\right)\right) = u\left(\mu - \frac{A\sigma^2}{2}\right).$$

This concludes the proof of the lemma. □

References

Drèze, J.H., and F. Modigliani, (1972), Consumption decisions under uncertainty, *Journal of Economic Theory* 5, 308-335.

Eeckhoudt, L., and H. Schlesinger, (2006), Putting risk in its proper place, *American Economic Review,* 96:1, 280-289.

Guiso, L., T. Jappelli and D. Terlizzese, (1996), Income risk, borrowing constraints, and portfolio choice, *American Economic Review,* 86, 158-172.

Kimball, M.S., (1990), Precautionary saving in the small and in the large, *Econometrica* 58 (1990), 53-73.

Leland, H.E., (1968), Saving and uncertainty: The precautionary demand for saving, *Quarterly Journal of Economics*, 465-473.

PART II

The term structure of discount rates

Random walk and mean-reversion

The term structure of the discount rate

We have concluded from the first part of this book that there is some solid scientific basis to recommend to use a rate of *3.5%* to discount cash flows occurring in the next few years. Does it implies that one should use the same rate to discount all cash flows, independent of to time of occurrence? The theoretical answer to this question is no, at least at this degree of generality. For the sake of a simple notation, we have earlier referred to *r* as "the" discount rate. But in fact, this *r* should be indexed by the maturity of the cost or benefit that one wants to discount. This can be seen for example in the general pricing formula (3.14) that we now rewrite as

$$r_t = \delta - \frac{1}{t}\ln\frac{Eu'(c_t)}{u'(c_0)}. \tag{4.1}$$

Because the right-hand side of this equality depends in general upon *t*, so does the left-hand side. In fact, the pricing formula (4.1) gives us the entire term structure of the discount rates.

Before going into the details, it is good to get an intuition of the determinants of this term structure. As we have seen before, the discount rate is determined by two competing effects: the wealth effect and the precautionary effect. When comparing two different horizons, *t* and *t'>t*, the intensity of each of these two effects may differ between these two dates, and may therefore affect the shape of the term structure.

The simple case arises when the growth rate is a constant *g*, and will remain so in the future. Assuming a constant relative aversion γ, the pricing formula (4.1) implies that $r_t = \delta + \gamma g$, so that the term structure is completely flat. Consumption increases exponentially with time, which implies that the discount factor $\exp(-r_t t)$ must decrease exponentially. This requires that the discount rate r_t be constant.

The case of diminishing expectations

Suppose that for example that there is no risk in the future but that the growth rate of the economy, which has been x_{-1} last year, will decrease at a constant rate to $\mu < x_{-1}$. More specifically, suppose that there exists a constant $\phi \in [0,1]$ such that

$$\begin{cases} c_{t+1} = c_t e^{x_t} \\ x_t = \phi x_{t-1} + (1-\phi)\mu. \end{cases} \tag{4.2}$$

This deterministic dynamics of consumption illustrates the simple concept of diminishing expectations. The idea is either that we have been particularly lucky in the recent past with a high rate of growth, but the future will go back to normal, i.e. to the historical growth rate μ. Or, we may believe that the current level of growth is unsustainable, and that the economy will have to adapt to a lower, sustainable, growth rate μ. Whatever the interpretation is, we obtain that

$$\ln c_t - \ln c_0 = \mu t + (x_0 - \mu)\frac{1-\phi^t}{1-\phi}. \tag{4.3}$$

In this certainty case with diminishing expectations, and assuming a power utility function, we can rewrite pricing formula (4.1) as

$$r_t = \delta + \gamma \frac{\ln c_t - \ln c_0}{t} = \delta + \gamma \left[\mu + (x_0 - \mu)\frac{1-\phi^t}{t(1-\phi)} \right], \tag{4.4}$$

in which $x_0 = \phi x_{-1} + (1-\phi)\mu$. The first equality in (4.4) tells us that the wealth effect is proportional to the annualized growth of log consumption. This yields in particular the following discount rates in the short and long terms:

$$\begin{cases} r_1 = \delta + \gamma x_0 \\ r_\infty = \delta + \gamma \mu \end{cases} \tag{4.5}$$

In between, the efficient discount rate decreases smoothly at a constant rate. When expectations are diminishing, the term structure is downward sloping. This is because the wealth effect is strong for the short term, but it goes down for longer time horizons.

Remember that the socially efficient discount rate is also the equilibrium interest rate that one would observe on frictionless capital markets. The above analysis thus tells us that the shape of the so-called yield curve, which is the term structure of interest rate, is a crucial source of information about what economic agents believe about the economic dynamics in the future. A downward yield curve provides the information that people believe that the economy will experience a downturn in the future. On the contrary, an upward sloping yield curve is typical of an accelerating economy.

The same ideas apply for longer time horizons. If one believes than the growth rate that our economies experienced during the last two centuries is just unsustainable, this should be taken into account in our evaluation of investment projects. The term structure of the discount rates should be decreasing. This will favour investment projects that have large positive benefits in the distant future in comparison to projects with more immediate projects. In short, the decreasing term structure of discount rates will be favourable to sustainable development.

If the current growth rate of the economy is *2%*, but its sustainable growth rate is believed to be only *0.5%*, then the above pricing formula with $\delta = 2\%$ and $\gamma = 2$ yields discount rates of *4%* and *1%* respectively for the short and long terms.

Our point of view on this approach is that economic growth is indeed subject to business cycles that must be account when shaping the term structure of discount rates for the short and medium terms. This means in particular that the short discount rate should be revised periodically to take into account of the expectations about the short and medium macroeconomic dynamics. However, we are not on the side of catastrophists, who believe that our economic growth is unsustainable. We believe that the chance that our Society will experience a growth even larger than what we experienced since the beginning of the industrial revolution is as large as the chance of experiencing of growth rate smaller than that. This does not mean that we should not focus a special attention to the distant future, quite to the contrary, as we show in the next few chapters.

Decreasing term structure and time consistency

It is often suggested in the literature that economic agents are time inconsistent if the term structure of the discount rate is decreasing. What is crucial for time consistency is the constancy of the rate of impatience δ, which is one of the cornerstone of the classic analysis presented in this book. We have seen above that this assumption is compatible with the possibility that the monetary discount rate be decreasing. Other illustrations of this fact will be presented later on in this book. Let us re-examine this question under the simple framework of diminishing expectations as modelled by the deterministic dynamic process (4.2).

An agent is time consistent if the plan that is optimal at time *t* remains optimal for all future date *t'>t*. To illustrate, consider an investment that costs one monetary unit at date *T* and that generates a single benefit *k* at time $T+\tau$. Evaluating this project from date *0*, investing is optimal if and only if its net present value is positive, i.e., if

$$-e^{-r_T T} + ke^{-r_{T+\tau}(T+\tau)} \geq 0. \tag{4.6}$$

This is equivalent to

$$-1 + ke^{r_T T - r_{T+\tau}(T+\tau)} \geq 0. \tag{4.7}$$

Assuming that the agent faces the exogenous consumption process (4.2), the term structure r_t given by (4.4) should be used at date *0* to discount these cash flows in equation (4.7). Suppose that this condition is satisfied, so that, seen from today, it is optimal to implement the project at date *T*.

Consider now the decision problem at date *T*, when comes the time to invest into the project. To solve this problem, we need to determine the discount rate that should be used at date *T* to discount the cash flow *k* occurring τ periods later. Let $r_{T\to T+\tau}$ denote this discount rate. Thus, seen from date *T*, it is optimal to invest in the project if and only if

$$-1 + ke^{-r_{T\to T+\tau}\tau} \geq 0. \tag{4.8}$$

The problem of time consistency is about whether conditions (4.7) and (4.8) are equivalent, independent of *k*. Obviously, this requires that $-r_{T\to T+\tau}\tau = r_T T - r_{T+\tau}(T+\tau)$. At date *T*, the level of x_T equals

$$x_T = \mu + \phi^T(x_0 - \mu). \tag{4.9}$$

Duplicating the analysis presented in the previous section to the context of date τ implies that

$$r_{T\to T+\tau}\tau = \delta\tau + \gamma\left[\mu\tau + (x_T - \mu)\frac{1-\phi^\tau}{(1-\phi)}\right] = \delta\tau + \gamma\left[\mu\tau + (x_0 - \mu)\frac{\phi^T\left(1-\phi^\tau\right)}{(1-\phi)}\right]. \tag{4.10}$$

It is easy to check that this is equal to $r_T T - r_{T+\tau}(T+\tau)$, which implies that the decision criterion to be used at date *T* is not different from the one to be used at date *0*. The decision process is thus perfectly time consistent, although the term structure of discount rates is not flat.

Random walk

We will hereafter be neutral about the expected growth of the economy. More specifically, we will assume in general that the expected growth rate in the distant future is not different from the short term one. This neutralizes the role of the wealth effect on the term structure. What remains is the term structure of the precautionary effect.

In that vein, the simplest assumption that can be made is a random walk. This means that the growth rate observed in the past does not provide any information about the growth rate that we will experience in the future. More specifically, suppose that the growth rate of the economy follows an independent and identically distributed (iid) process over time:

$$\begin{cases} c_{t+1} = c_t e^{x_t} \\ x_0, x_1, \ldots\ iid. \end{cases} \tag{4.11}$$

This implies that the pricing formula (4.1) can be rewritten as

$$r_t = \delta - \frac{1}{t} \ln \frac{Eu'\left(c_0 \prod_{\tau=0}^{t-1} e^{x_\tau}\right)}{u'(c_0)}. \tag{4.12}$$

To keep things simple at this stage, consider the case of a power utility function with relative aversion γ. The above equation can then be rewritten as

$$r_t = \delta - \frac{1}{t} \sum_{\tau=0}^{t-1} \ln\left(Ee^{-\gamma x_\tau}\right). \tag{4.13}$$

Because the process is iid, this can be rewritten as

$$r_t = \delta - \ln\left(Ee^{-\gamma x_1}\right). \tag{4.14}$$

Thus, in the case of power utility functions and an iid process for the growth rate of the economy, the term structure of the efficient discount rate is completely flat. In the special case of a normal distribution for *x*, the extended Ramsey rule (3.21) gives us the level of this constant discount rate. Up to our knowledge, Hansen and Singleton (1983) were the first to obtain this result.

This case, which is the discrete version of a Brownian motion for the growth of the economy, serves as a benchmark for the analysis of the term structure of discount rates. It is thus important to understand its nature. When the growth rate of the economy follows a random walk with a constant positive trend, the wealth effect goes up exponentially with the time horizon. If *g=2%*, one will be in expectation *2%* wealthier next year, and *5000%* wealthier in

200 years. This exponentially increasing wealth effect justifies taking an exponentially decreasing discount factor, i.e., a constant discount rate. Similarly, the random walk in the growth rate entails an exponentially increasing uncertainty on future consumption. This is equivalent to a linearly increasing variance for c_t. Indeed, we have that

$$Var(\ln c_t - \ln c_0) = Var\left(\sum_{\tau=0}^{t-1} x_\tau\right) = t\sigma^2. \tag{4.15}$$

The exponentially increasing precautionary effect that this implies should impact the discount factor exponentially. In other words, it should affect the discount rate uniformly with respect to the time horizon. Combining these two elements implies that the term structure of discount rates is flat.

A simple extension: Mean-reverting growth process

Following Bansal and Yaron (2004), one can easily combine the two growth processes that we have considered in this chapter in the following way:

$$\begin{cases} c_{t+1} = c_t e^{x_t} \\ x_t = \mu + y_t + \varepsilon_{xt}, \\ y_{t+1} = \phi y_t + \varepsilon_{yt,} \end{cases} \tag{4.16}$$

where ε_{xt} and ε_{yt} are independent and serially independent with mean zero and variance σ_x^2 and σ_y^2, respectively. The state variable y_t exhibits some persistence. Parameter ϕ, which is between *0* and *1*, represents the persistence in the expected growth rate process. When ε is uniformly zero, we are back to the story of deterministic "diminishing expectations". When ϕ is zero, we are back to a pure random walk. This autoregressive model of degree 1 – an AR(1) – illustrates the notion of mean-reversion. Suppose that the expected growth rate equals its historical level μ ($y_0 = 0$), and that a positive shock ε_{y0} affects the expected growth rate between dates *0* and *1*, so that y_1 is larger than *0*. Contrary to the random walk, this shock will have some persistence. For example, the expected growth rate between dates *t* and *t+1* will be $Ex_t \left| \varepsilon_{y0} \right. = \mu + \phi^{t-1}\varepsilon_{y0}$. However, in the long run, the expected growth rate will revert to the mean. But at each date, a new shock may affect the growth rate of the economy.

We can determine the efficient term structure in this case by characterizing the distribution of c_t. By forward induction of (4.16), we obtain that

$$\ln c_t - \ln c_0 = \mu t + y_0 \frac{1-\phi^t}{1-\phi} + \sum_{\tau=0}^{t-1} \frac{1-\phi^{t-\tau-1}}{1-\phi} \varepsilon_{y\tau} + \sum_{\tau=0}^{t-1} \varepsilon_{x\tau}. \tag{4.17}$$

Because the ε are normally distributed, so is $\ln c_t - \ln c_0$. Its mean is the sum of the first two terms in the right-hand side of the above equality. Its annualized variance equals

$$t^{-1} Var(\ln c_t) = \frac{\sigma_y^2}{(1-\phi)^2} \left[1 + 2\frac{\phi^t - 1}{t(1-\phi)} + \frac{\phi^{2t} - 1}{t(\phi^2 - 1)} \right] + \sigma_x^2. \tag{4.18}$$

Observe that the annualized variance of log consumption tends to $(\sigma_y^2/(1-\phi)^2) + \sigma_x^2$, which is larger than the short run uncertainty measured by $Var\ x_1 = \sigma_x^2$. The long-run risk is increasing in the degree of persistence of shocks on the exspected growth rate of concumption. This is because of the positive serial correlation in growth rates. More generally, the analysis of the right-hand side of (4.18) shows that the annualized variance of future log consumption goes up smoothly from σ_x^2 to $(\sigma_y^2/(1-\phi)^2) + \sigma_x^2$ when *t* goes from *1* to infinity.

Suppose that *u* is a power function with relative aversion γ. The pricing formula (4.1) can thus be rewritten as

$$r_t = \delta - \frac{1}{t} \ln E\left[e^{-\gamma(\ln c_t - \ln c_0)} \right]. \tag{4.19}$$

Because of the normality of $\ln c_t - \ln c_0$, we can use property *(3.4)* to obtain that

$$r_t = \delta + \gamma t^{-1} E\left[\ln c_t - \ln c_0\right] - 0.5\gamma^2 t^{-1} Var(\ln c_t). \tag{4.20}$$

Using the properties of the mean and variance of log consumption, we finally characterize the term structure of the discount rate as follows :

$$r_t = \delta + \gamma \left[\mu + y_0 \frac{1-\phi^t}{t(1-\phi)} \right] - 0.5\gamma^2 \left[\frac{\sigma_y^2}{(1-\phi)^2} \left(1 + 2\frac{\phi^t - 1}{t(1-\phi)} + \frac{\phi^{2t}-1}{t(\phi^2-1)} \right) + \sigma_x^2 \right]. \tag{4.21}$$

This equation can be rewritten as follows:

$$r_t = \delta + \gamma\mu - 0.5\gamma^2 \left[\sigma_x^2 + \frac{\sigma_y^2}{(1-\phi)^2} \right] + \left[y_0 \frac{1-\phi^t}{t(1-\phi)} - 0.5\gamma^2 \frac{\sigma_y^2}{(1-\phi)^2} \left(2\frac{\phi^t - 1}{t(1-\phi)} + \frac{\phi^{2t}-1}{t(\phi^2-1)} \right) \right]. \tag{4.22}$$

Observe that the last bracketed term of this equation is the only one that depends upon *t* and that it vanishes when *t* tends to infinity. It is the transitory term which shapes the term structure. The first three terms determine the long term discount rate. The long term wealth effect is still measured by $\gamma\mu$. The long term precautionary effect is magnified by mean-reversion. It yields

$$r_\infty = \delta + \gamma\mu - 0.5\gamma^2 \left[\sigma_x^2 + \frac{\sigma_y^2}{(1-\phi)^2} \right] \tag{4.23}$$

We conclude here that persistent shocks on the growth rate of the economy reduces the rate at which very distant cash-flows must be discounted. This is because of the increased long term risk that the positive correlation of growth rate generates. This effect is increasing in the degree of persistency ϕ of shocks on growth rates.

The term structure is given by the last term in (4.22). The term in y_0 corresponds to the "diminishing expectations" story that we explained earlier in this chapter. It yields a decreasing shape to the term structure if the economy is currently experiencing a growth rate above its historical mean. When this effect is switched off by assuming that $y_0 = 0$, the short term discount rate equals

$$r_1 = \delta + \gamma\mu - 0.5\gamma^2\sigma_x^2, \tag{4.24}$$

The second term under brackets in (4.21) tells us how the discount rate goes down from r_1 to $r_\infty < r_1$. The annualized variance of log consumption is increasing with the time horizon when there is persistence. This tends to make the term structure decreasing.

Let $r_1(t)$ denote the rate that should be used at date *t* to discount cash flows occurring at date *t+1*. This is the short-term interest rate. Notice that the short-term interest rate in this model also follows an AR(1) process since, using the pricing formula (4.20) for *t=1* yields

$$\begin{cases} r_1(t) = \delta + \gamma\mu - 0.5\gamma^2\sigma_x^2 + \gamma y_t \\ y_{t+1} = \phi y_t + \varepsilon_{yt}. \end{cases} \tag{4.25}$$

Vasicek (1977) was interested in determining the shape of the yield curve by using the standard arbitrage method in finance under the assumption of an AR(1) for the short term interest rate. He got equilibrium interest rates for different maturities that are equivalent to our formula (4.21). The degree of persistence ϕ is the same for the economic growth and for the short term interest rate. This is interesting because the degree of persistence of the latter has been well documented in the literature on the term structure of interest rate.

Bansal and Yaron (2004) consider the following calibration of the model in order to fit annual growth data for the United States over the period 1929-1998. Using the month as the unit period, they obtained $\mu = 0.0015$, $\sigma_x = 0.0078$, $\sigma_y = 0.00034$, and $\phi = 0.979$. Using this ϕ yields a half-life time of 32 months. This implies that this model is useful to justify differences in discount rates for maturities expressed in years, but not really for maturities expressed in decades or centuries. In other words, the Vasisek model and mean-reversion in the growth rate has been useful to explain the term structure of interest rates for maturities that are treated by financial markets, i.e. 2 or 3 decades.

In the following figure, we describe how the term structure of interest/discount rates evolves along the business cycle. In addition to the above Bansal-Yaron's parameter values, we assume that the rate of impatience is $\delta = 0$, relative aversion is $\gamma = 2$. Three term structures are represented in this figure. When the recent growth rate is exactly at its historical mean ($y_0 = 0$, which corresponds to an annual growth rate of *1.8%*), the yield curve is decreasing. This slope describes the precautionary effect of the increasing annualized variance of future log consumption. During the recession illustrated by a low growth rate ($y_0 = -0.1\% / month$, which corresponds to an annual growth rate of *0.6%*), the yield curve is upwards sloping. This mostly expresses the accelerating wealth effect generated by the increasing expectations due to mean-reversion. On the contrary, when the economy is booming with $y_0 = 0.1\% / month$ (corresponding to an annual growth rate of *3%*), the yield curve is decreasing because of the diminishing expectation. The long term interest rate is not affected by the business cycle because the long term growth rate in this model is deterministic, and because the long-term uncertainty remains constant along it.

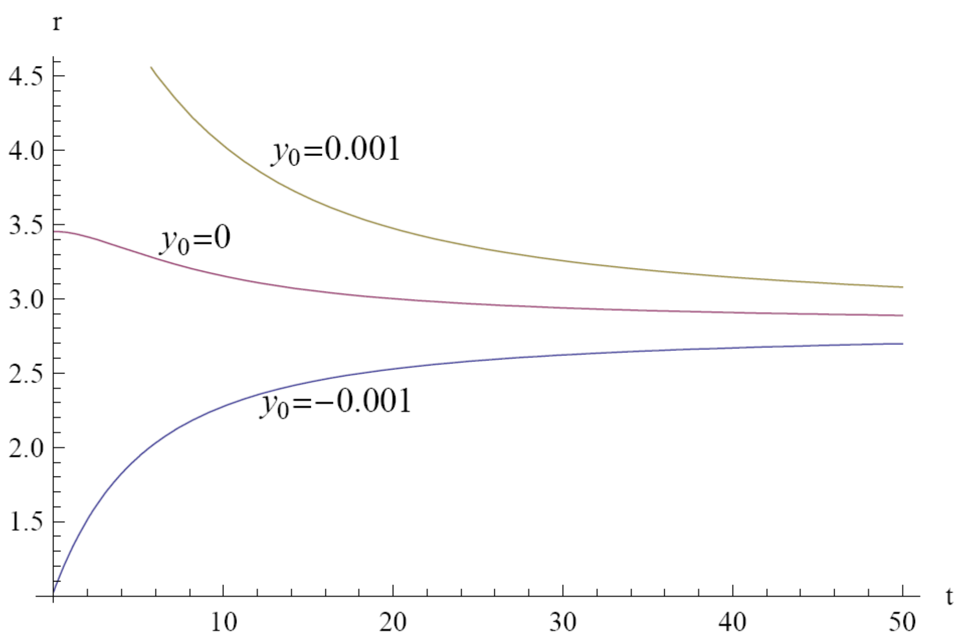


Figure : The efficient discount rate (in %) as a function of the maturity *t* (in years). Using the month as the unit period, the parameter values are $\delta = 0$, $\mu = 0.0015$, $\sigma_x = 0.0078$, $\sigma_y = 0.00034$, $\phi = 0.979$ and $\gamma = 2$.

Conclusion

The shape of the term structure of discount rates is determined by the way the wealth effect and the precautionary effects evolve with the time horizon. When the growth rate of consumption follows a random walk, the expected consumption increases exponentially. The corresponding wealth effect justifies that the discount factor decreases exponentially, i.e. that the discount rate be constant. Similarly, the variance of *log* consumption increases linearly, which yields an exponentially increasing effect on the discount factor. This justifies a constant precautionary effect on the discount rate. This yields a crucial result for the theory of efficient discount rates: *When the growth rate of the economy follows a random walk and when relative aversion is constant, the discount rate should be independent of the maturity of the project to be evaluated.*

A simple extension of the random walk for the growth rate of the economy is when the growth rate follows an autoregressive process of degree 1. Mean-reversion has two consequences on the above result. First, the term structure will be sensitive to the business cycle. When the economy is booming, the short term interest rate is large because of the persistent wealth effect. But because the economy is expected to revert to a smaller growth rate in the future, so the wealth effect becomes relatively less powerful for the distant future, which yields a downward shaped term structure. The opposite effect arises in a downturn. The second effect of mean-reversion is to introduce some positive serial correlation of the growth rate. Compared to the case of a random walk, that tends to magnify the long term risk of the economy. This reinforces the precautionary effect associated to long term. Therefore it tends to make the term structure downward sloping on average.

References

Bansal, R., and A. Yaron, (2004), Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance* 59, 1481–1509.

Hansen, L. and K. Singleton, (1983), Stochastic consumption, risk aversion and the temporal behavior of assets returns, *Journal of Political Economy*, 91, 249-268.

Vasicek, 0., (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-188.

Markov switches and extreme events

The economic history of the world has one obvious feature: For thousands of years, the consumption per capita remained close to the minimum level of subsistence. Under the Malthus' Law, any technical progress led to an increase in the population rather than to an improvement in welfare. For example, Clark (2007) estimates that the daily wage in Babylon (1880-1600 B.C.) was around 15 pounds of wheat. In the golden age of Pericles in Athens, it was around 26 pounds. In England around 1780, it was only 13 pounds.

This doom ended up in the western world around 1800 with the industrial revolution that moved the trend of growth rate of consumption per capita from *0%* to *2%*. It is not the place here to explain the origin of this radical transformation of the economic dynamics. Let us rather focus on the consequences of this observation for our aim in this book. When we think about climate change or nuclear wastes for example, and more generally about sustainable development, we think about time horizons in the range of several centuries. In order to forge our attitude towards the generations who will live in that distant time horizons, we need to form beliefs about their level of prosperity. It is rather myopic to use historical data limited to more or less one century to build our beliefs about the growth of the economy over the next several centuries. We believe that our economies undergo radical transformations at different frequencies, some of them being very low. One such radical transformation was called the "industrial revolution" and had a long lasting effect on economic growth. Who knows whether a reversion to the pre-industrial age, at least in terms of the absence of growth, is possible in some distant future? Other less persistent – but more frequent – events observed in the past were related to wars or great economic depressions. It is important to put the potentiality of these events into the picture.

The role of extreme events on the level of discount rates

The easiest way to examine the effect of extreme events on discount rates is to assume a random walk without assuming the normality of the growth rate per year. The random walk implies that the term structure is flat. Suppose that the increase in log consumption follows an iid process characterized by a random variable x. With a small probability p, there is a catastrophe that reduces the log of consumption by a large λ. This is an extreme event. Otherwise, this is business as usual, with an increase in log consumption that is drawn from random variable x_{bau}. In short we assume that

$$\ln x_{t+1} - \ln x_t \sim (p, \ln(1-\lambda); 1-p, x_{bau}) \tag{5.1}$$

Under the assumption of constant relative aversion, the efficient discount rate equals

$$r_1 = \delta - \ln\left[p(1-\lambda)^{-\gamma} + (1-p)Ee^{-\gamma x_{bau}} \right] \tag{5.2}$$

Assuming that x_{bau} is normally distributed with mean μ_{bau} and variance σ_{bau}^2 allows us to rewrite this equation as follows:

$$r_1 = \delta - \ln\left[p(1-\lambda)^{-\gamma} + (1-p)e^{-\gamma\mu_{bau} + 0.5\gamma^2\sigma_{bau}^2} \right]. \tag{5.3}$$

Of course, the existence of this possible catastrophe reduces the intensity of the wealth effect, and it raises the intensity of the precautionary effect, thereby reducing the efficient discount rate.

Barro (2006) collected data on extreme macroeconomic events across different countries during the last century. His analysis of these events "suggests a disaster probability of *1.5-2%* per year with a distribution of declines in per capita GDP ranging between *15%* and *64%*." In the figure below, we retained a probability of *2%*, and we examined the level of the (flat) discount rate for different intensities of declines of GDP. For the remainder, we retained the standard values for the trend and volatility in the BAU scenario, and for the preference parameters.

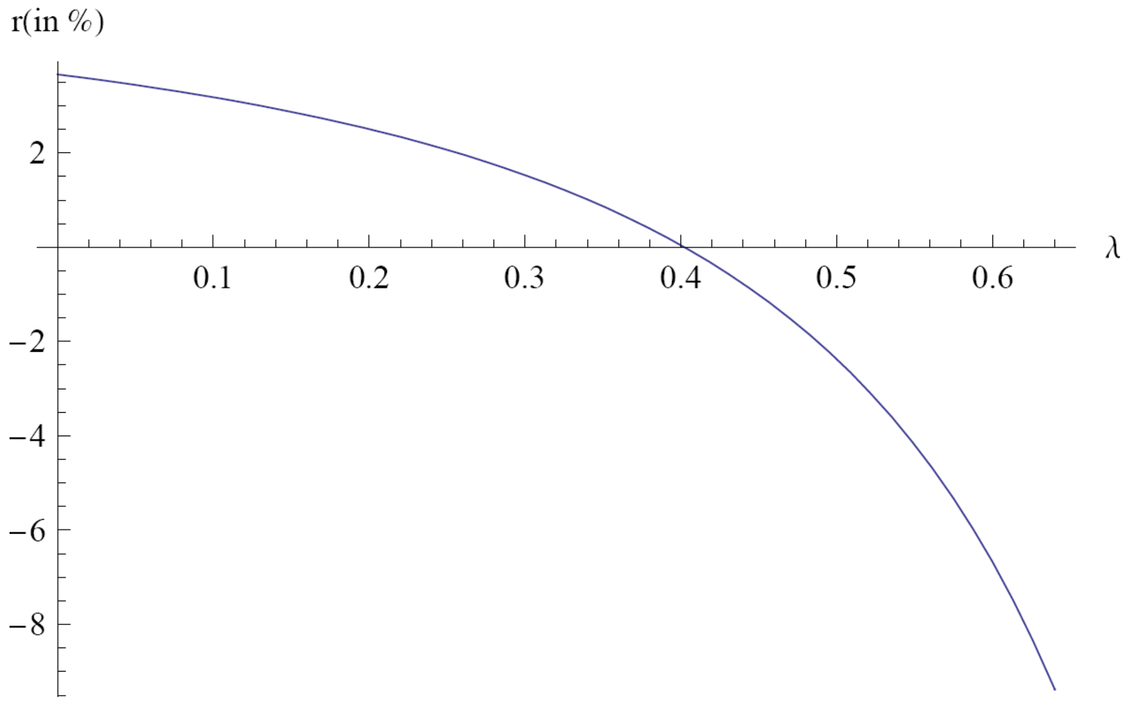


Figure: The efficient discount rate for different size λ of the catastrophe.

Parameter values: $\delta = 0$, $\mu_{bau} = 2\%$, $\sigma_{bau} = 3.6\%$, $\gamma = 2$, $p = 2\%$.

For low intensities of the loss, we get the discount rate of *3.6%* obtained before. However, when the size of the loss exceeds *40%*, the efficient discount rate becomes negative, and sinks quite quickly in deep negative waters. When λ tends to *100%*, the efficient discount rate tends to *-100%*. In spite of the small probability of the catastrophe, one is ready to sacrifice the current welfare at any prices to escape the risk to get to zero consumption. This is due to the fact that the marginal utility tends to infinity when consumption tends to zero, which is a specific property of power utility functions. As coined by Weitzman (2007), people "*dread the thickened-left-tail heightened probability of a negative-growth disaster that they find scary, disruptive, and without precedent*".

Two-state Markov process

In the previous section, we assumed that the economic growth rate follows a random walk. Catastrophes have a permanent effect on the level of consumption, but not on the growth rate of consumption. In this section, we are considering an alternative stochastic process in which the growth rate of consumption is subject to persistent shocks. In the long run, small but persistent shocks on the growth rate have dramatic consequences on the level of consumption. China, which was by far the wealthiest nation in the end of the XVth century experienced a permanent reduction of its growth rate until the early 1990th. This implied that it became one of the poorest nations in the world in the late 50th, facing for example a dramatic famine during the Great Leap Forward, killing more than 30 millions people. Over the last 20 years or so, China's growth rate switched to around 10% per year.

To model this type of dynamic process, a two-state Markov chain for the trend of the economic growth is considered in this section. There are two regimes, *s*=*g* and *s*=*b*, yielding different expected changes in log consumption μ^g and μ^b, with $\mu^g > \mu^b$. In each period, there is a constant regime-dependent probability π^s less than ½ that the state will reverse. We can thus describe this stochastic process as follows:

$$\begin{cases} c_{t+1} = c_t e^{x_t} \\ x_t = \mu^{s_t} + \varepsilon_t \\ P\left[s_{t+1} = b \middle| s_t = g\right] = \pi^g; \quad P\left[s_{t+1} = g \middle| s_t = b\right] = \pi^b \end{cases} \tag{5.4}$$

where ε_t is iid normal with mean zero and variance σ^2.

Suppose that relative risk aversion is a constant γ, and let us denote –*g*=*b* and –*b*=*g*. We have that

$$\frac{E\left[u'(c_1)\middle|s\right]}{u'(c_0)} = (1-\pi^s)Ee^{-\gamma(\mu^s+\varepsilon_0)} + \pi^s Ee^{-\gamma(\mu^{-s}+\varepsilon_0)} = e^{0.5\gamma^2\sigma^2}\left[(1-\pi^s)e^{-\gamma\mu^s} + \pi^s e^{-\gamma\mu^{-s}}\right]. \tag{5.5}$$

Equation (4.1) can then be rewritten as

$$r_1^s = \delta + \gamma m_1^s - 0.5\gamma^2\sigma^2, \tag{5.6}$$

where the exponential of m_1^s is the precautionary equivalent of $(\exp\mu^s, 1-\pi^s; \exp\mu^{-s}, \pi^s)$:

$$e^{-\gamma m_1^s} = (1-\pi^s)e^{-\gamma\mu^s} + \pi^s e^{-\gamma\mu^{-s}}. \tag{5.7}$$

r_1^s is the discount rate for a one-period horizon when the current regime is *s*. Notice that term γm_1^s in (5.6) contains a wealth effect and a precautionary effect, since *m* is the volatility-free component of the precautionary equivalent growth rate of consumption. Because μ^g is larger than μ^b and π^s is smaller than ½, we have that m_1^g is larger than m_1^b. This implies that the short-term discount rate is larger in the good state than in the bad state.

When we explore the possible dynamic evolution of the economy two periods ahead, things become more complex since the economic regime can reverse twice. However, we can proceed as above by using a recursive method. Without going into details, we obtain that

$$r_t^s = \delta + \gamma \frac{m_t^s}{t} - 0.5\gamma^2\sigma^2, \tag{5.8}$$

where m_t^s is defined recursively from m_1^s as follows:

$$e^{-\gamma m_{t+1}^s} = (1-\pi^s)e^{-\gamma(\mu^s + m_t^s)} + \pi^s e^{-\gamma(\mu^{-s} + m_t^{-s})}. \tag{5.9}$$

We thus obtain two term structures r_t^g and r_t^b of efficient discount rates. If the current economic regime is the good one, the short-term discount rate is high because the probability to stay in that high-growth state is larger than ½. However, in the longer run, the probability of a switch to the low-growth state increases, which implies a reduction of the wealth effect. The term structure of the efficient discount rates is thus downward sloping in the good regime. On the contrary, the term structure is upward sloping in the bad regime. In the distant future, the probability distribution of the two regimes becomes independent of the initial state. When t tends to infinity, the probability to be in the good regime tends to its unconditional value $\pi^b/(\pi^b + \pi^g)$.

Numerical illustrations

We hereafter examine two numerical illustrations of this model. The first one is based on an estimation of a two-state regime-switching process for the US economy using the annual per capita consumption data covering the period 1890-1994. The following table reproduces the estimates from Cecchetti, Lam and Mark (2000).

μ^g	μ^b	π^g	π^b	σ
2.25%	*−6.78%*	*2.2%*	*48.4%*	*3.13%*

Table : Estimates of the regime-switching consumption process

Source: Cecchetti et al. (2000, Table 2)

It reveals that the low-growth state is moderately persistent but very bad, with consumption growth in that state being $\mu^b = -6.78\%$. On the contrary, the high-growth state is highly persistent, with consumption growth in that state equalling *2.25%*. The economy spends most of the time in this state with the unconditional probability of being in the good state equalling *96%*. The unconditional expected growth rate is *1.89%*.

The following figures describe the two term structures under these values of the parameters of the Markov process, together with $\delta = 0$ and $\gamma = 2$. The two curves have an asymptote at $r_\infty = 3.26\%$. The short-term rate in the good regime equals $r_1^g = 4.3\%$, whereas it equals $r_1^b = -13.8\%$ in the bad state. The difference between the two wealth effects is the main driver of this result. In the bad state, the recession is expected to be deep in the short run, so that much should be done to transfer consumption for the next few years. Also, the uncertainty about the time at which the economy will switch back to the good state implies a large precautionary effect. This is a situation in which the wealth effect and the precautionary effect go hand-in-hand. The discount rate is negative for time horizons up to 11 years.

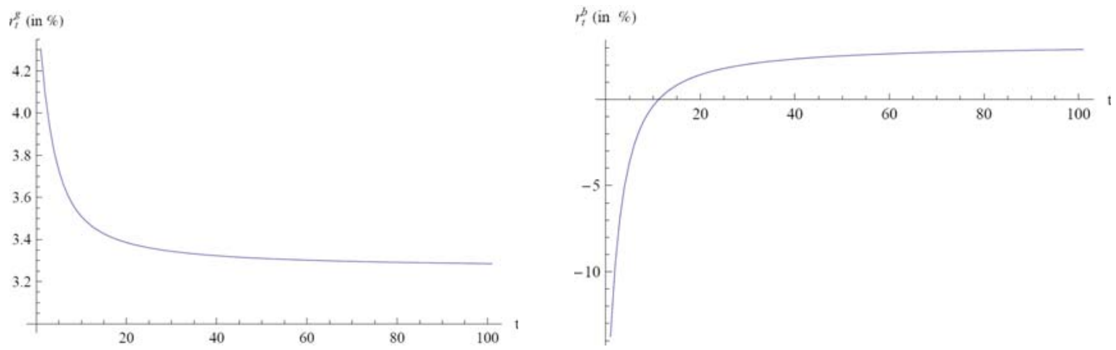


Figure : The term structures of discount rates in the two regimes
under the two-state regime-switching regime estimated by Cecchetti et al. (2000)

The calibration based on data covering the period 1890-1994 fails to recognize a crucial aspect of economic history, as pointed out in the introduction of this chapter. Over at least 6 millenia, the trend of economic growth has been around *0%*, until the end of the XVIIth century, where the western world switched to a trend around *2%*. We use the two-state Markov process presented in this chapter with these two possible trends. We assume that there is a uniform probability of *1%* per year to switch from the current state to the other state. In our next figure, we represented the two term structures for the standard value of the other parameters ($\delta = 0\%$, $\gamma = 2$, and $\sigma = 3.6\%$). In the good state, the discount rate goes down from *3.74%* to *0.77%* from *1* to *500* years. In the bad state, it goes from *-0.26%* to *0.48%* over the same range of time horizons. They both converge to *0.6%* for the far distant future.

μ^g	μ^b	π^g	π^b	σ
2%	*0%*	*1%*	*1%*	*3.6%*

Table : An alternative two-state Markov process

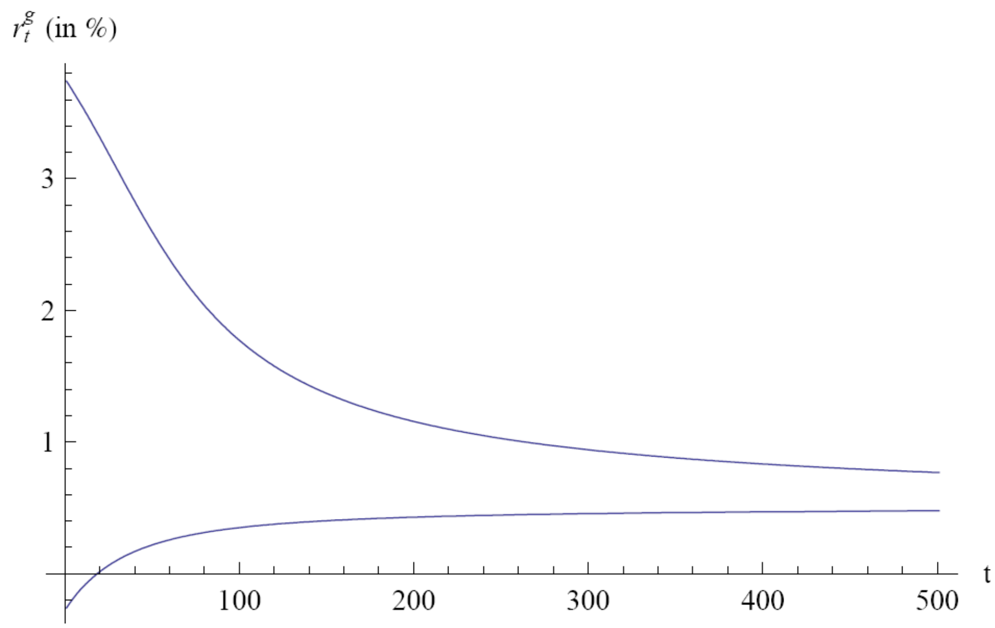


Figure : The term structures of discount rates in the two regimes under the alternative two-state Markov process

This alternative example illustrates the long lasting effects of the uncertainty on the term structure. In the short run, the risk of switching regime adds little to the uncertainty surrounding future consumption. Because the shock on the growth rate is persistent, this risk accumulates over time at a larger pace than when there is no serial correlation. The precautionary effect is thus magnified by the dynamics of regime switches. In the high-growth regime, this first explanation of the long downward-sloping term structure goes hand-in-hand with the wealth effect of the diminishing expectation. In the short run, the expected growth rate is close to *2%,* thereby yielding a wealth effect on the discount rate equalling $\gamma \times 2\%=4\%$. In the longer run, the probability of being in the good state is *50%*, so that the expected trend is only *1%*. So the wealth effect in the distant future amounts to $\gamma \times 1\%=2\%$. In the low-growth regime, the improving expectation tends to shape of term structure upwards, although this is partially countered by the precautionary effect.

Conclusion

All calibration exercises of the term structure of interest rates in the literature rely on macroeconomic data covering a fraction of the last two centuries during which the western world experienced a growth trend around *2%*. This approach is correct when one wants to discount cash flows maturing in the next few years. However, we believe that it is flawed if we want to discount cash flows occurring in the more distant future. The possibility of switching abruptly and persistently to a lower growth regime is an argument in favour of using a smaller rate to discount these cash flows. Both the wealth effect and the precautionary effect concur to this result.

References

Barro, R.J., (2006). "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823-866.

Cecchetti, S.G., P.-S. Lam, and N.C. Mark, (2000), Asset pricing with distorted beliefs: Are equity returns too good to be true, *American Economic Review*, 90, 787-805.

Clark, G., (2007), *Farewell to alms. A brief Economic history of the world,* Princeton University Press.

Parametric uncertainty and fat tails

We started the analysis of the discount rate by considering a sure growth rate of consumption. We extended the analysis by recognizing that economic growth is uncertain. In the previous chapter, we recognized that the parameters governing this uncertainty may be unstable. In this chapter, we go one step further by recognizing that the probability distribution governing economic growth is itself subject to some parametric uncertainty.

The estimation of the parameters governing a stochastic process, as the mean or the volatility, can be performed by collecting a large data set of past realizations of this process. This sample may not contain all possible scenarios that could occur in the future. For example, until the early 1970's, the Mexican currency was pegged to the dollar, so that the estimation of trend and volatility of the exchange rate of the peso were close to zero. It is therefore quite hard to explain the large premium observed between the Mexican and U.S. interest rates. This was called the "peso problem". The sharp devaluation of the peso provided the explanation of the puzzle: the data did not contain this small probability event, although most investors had it in mind.

As invoked in the peso problem, the absence of a sufficiently large data set to estimate the long-term growth process of the economy implies that its parameters are uncertain and subject to learning in the future. This problem is particularly crucial when its parameters are unstable, or when the dynamic process entails low-probability extreme events. The rarer the event, the more ambiguous is our estimate of its frequency. This builds a bridge between the problem of parametric uncertainty, and the one of extreme events.

Uncertain growth

Suppose that the dynamic process $c_0, c_1, c_2, ...$ is a function of a parameter θ. The true value of θ is unknown. For the sake of simplicity, suppose that θ can take n possible values $\theta = 1, ..., n$. Our prior beliefs about θ at date 0 are characterized by a probability distribution

$(q_1,...,q_n)$, $q_\theta > 0, \sum q_\theta = 1$, where q_θ is the probability that the true value of the parameter be θ.By the law of iterated expectations, we have that

$$Eu'(c_t) = \sum_{\theta=1}^{n} q_\theta E\left[u'(c_t)\middle|\theta\right]. \tag{6.1}$$

It implies that the pricing formula (4.1) can be rewritten as

$$r_t = \delta - \frac{1}{t}\ln\sum_{\theta=1}^{n} q_\theta \frac{E\left[u'(c_t)\middle|\theta\right]}{u'(c_0)}. \tag{6.2}$$

Let $r_{t\theta}$ denote the discount rate that would be efficient for horizon *t* if we would know for sure that the true value of the parameter be θ. This means that $r_{t\theta}$ is defined as

$$r_{t\theta} = \delta - \frac{1}{t}\ln\frac{E\left[u'(c_t)\middle|\theta\right]}{u'(c_0)}. \tag{6.3}$$

Combining equations (6.2) and (6.3) yields that

$$e^{-r_t t} = \sum_{\theta=1}^{n} q_\theta e^{-r_{t\theta} t}. \tag{6.4}$$

In other words, the socially efficient discount factor under parametric uncertainty equals the expectation of the conditionally efficient discount factors, i.e., of the discount factors that would be efficient if the value of the parameter would be known with certainty.

Notice that the expectation concerns the discount factors, not the discount rates. In fact, the socially efficient discount rate defined by (6.4) can be interpreted as the certainty equivalent rates $r_{t\theta}$, $\theta = 1,...,n$, under the implicit utility function $h(r) = -\exp(-rt)$. This function is increasing and concave, with an index of concavity measured by *t*. It implies that r_t is smaller than the mean of the $r_{t\theta}$. Moreover, as long as the support of $r_{t\theta}$ remains bounded, r_t tends to the lower bound of this support when *t* tends to infinity. Indeed, using L'Hospital's rule, we have that

$$\lim_{t\to\infty} r_t = -\lim_{t\to\infty}\frac{1}{t}\ln\sum_{\theta=1}^{n} q_\theta e^{-r_{t\theta} t} = \lim_{t\to\infty}\frac{\sum_{\theta=1}^{n} q_\theta r_{t\theta} e^{-r_{t\theta} t}}{\sum_{\theta=1}^{n} q_\theta e^{-r_{t\theta} t}}. \tag{6.5}$$

Let $r_\infty^{\min}$ denote the smallest possible discount rate when *t* tends to infinity: $r_\infty^{\min} = \lim_{t\to\infty} r_{t\theta}$. We get

$$\lim_{t\to\infty} r_t = \lim_{t\to\infty} \frac{\sum_{\theta=1}^{n} q_\theta r_{t\theta} e^{(r_\infty^{\min} - r_{t\theta})t}}{\sum_{\theta=1}^{n} q_\theta e^{(r_\infty^{\min} - r_{t\theta})t}} = r_\infty^{\min}. \tag{6.6}$$

Similarly, we have that

$$\lim_{t\to 0} r_t = \lim_{t\to 0} \frac{\sum_{\theta=1}^{n} q_\theta r_{t\theta} e^{-r_{t\theta}t}}{\sum_{\theta=1}^{n} q_\theta e^{-r_{t\theta}t}} = Er_\theta. \tag{6.7}$$

The rate at which cash flows occurring in the short term should be discounted is equal to the expectation of the conditionally efficient discount rate. In order to get an intuition to these results, let us examine the simplest case when the stochastic process governing $\ln c_t$ is a random walk conditional to θ.

Conditional to θ, the growth process is a random walk

A special case of the above model is as follows:

$$\begin{cases} c_{t+1} = c_t e^{x_t} \\ x_0, x_1, \ldots \mid \theta \ i.i.d. \sim N(\mu_\theta, \sigma_\theta) \ \forall \theta. \end{cases} \tag{6.8}$$

This is a discrete version of an arithmetic Brownian motion with an unknown trend and/or volatility. Although this process is a random walk conditional to θ, x_t exhibits some serial correlation. Using Bayes's rule, the observation of a large x_0 yields an upwards revision of the trend of economic growth.

Conditional to θ, the dynamic process of x_t is a normal random walk. As seen before, equation (6.3) as an analytical solution in that case:

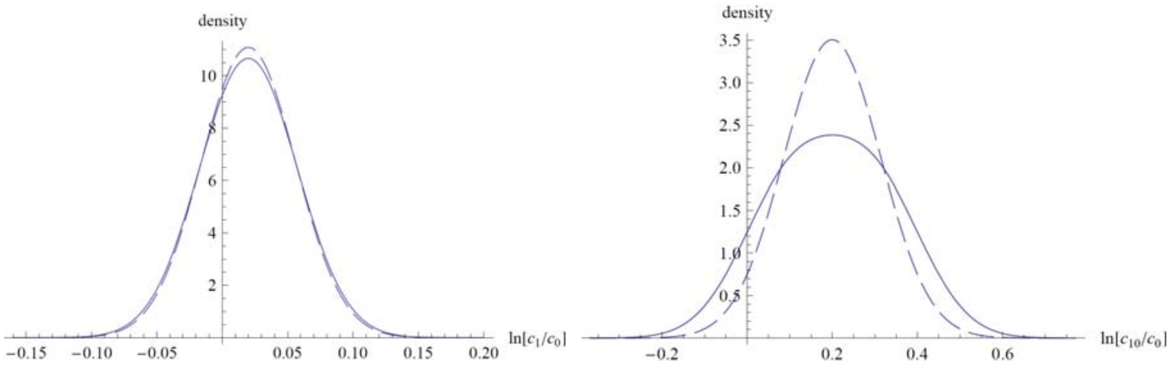
$$r_{t\theta} = \delta + \gamma\mu_\theta - 0.5\gamma^2\sigma_\theta^2. \tag{6.9}$$

In particular, $r_{t\theta}$ is independent of *t*. Under the hidden structure characterized by $(\mu_\theta, \sigma_\theta)$, $\theta = 1,\ldots,n$, the term structure of the socially efficient discount rate is obtained by rewriting equation (6.4) as follows:

$$r_t = \delta - \frac{1}{t} \ln \sum_{\theta=1}^{n} q_\theta e^{(-\gamma\mu_\theta + 0.5\gamma^2\sigma_\theta^2)t}. \tag{6.10}$$

The socially efficient discount rate under this parametric uncertainty is equal to the expected value of $r_\theta = \delta + \gamma\mu_\theta - 0.5\gamma^2\sigma_\theta^2$ for short maturities, is decreasing with *t*, and tends to the smallest possible value of r_θ when *t* tends to infinity.

Following Gollier (2008), the intuition of these results is based on the observation that the parametric uncertainty plays a crucial role in shaping the uncertainty surrounding consumption in the distant future. To understand this, let us assume that the volatility of the growth of log consumption is known and equal to $\sigma = 3.6\%$, but the trend μ is unknown. It can be either *1%* or *3%* with equal probability. In the following figure, we draw the distribution of $\ln c_t / c_0$ for *t=1, 10* and *100.* Ex ante, the distribution of $\ln c_1 / c_0$ is a mixture of two normal densities. However, the uncertainty affecting the trend is a second-order source of uncertainty compared to the volatility of the growth rate. So, in the short-run, assuming a trend of *(1%+3%)/2* to determine the efficient discount rate is a good approximation. On the contrary, the uncertainty affecting consumption in one century is vastly affected by the uncertainty affecting the growth trend. Conditional to the trend of growth $\mu = 1\%$ or $\mu = 3\%$, the expectation of c_{100} / c_0 is $\exp(100(\mu + 0.5 \times 0.036^2))$, which equals *3.5* or *26*. This source of uncertainty should be compared to the one coming from the intrinsic volatility of growth. Assuming $\mu = 2\%$ and $\sigma = 3.6\%$, the interval of confidence at 95% of $\ln c_{100} / c_0$ is $[5.7, 9.6]$. This means that for such a long horizon of one century, the uncertainty about future consumption is mostly affected by the unknown trend of growth, rather than by the natural volatility of growth.



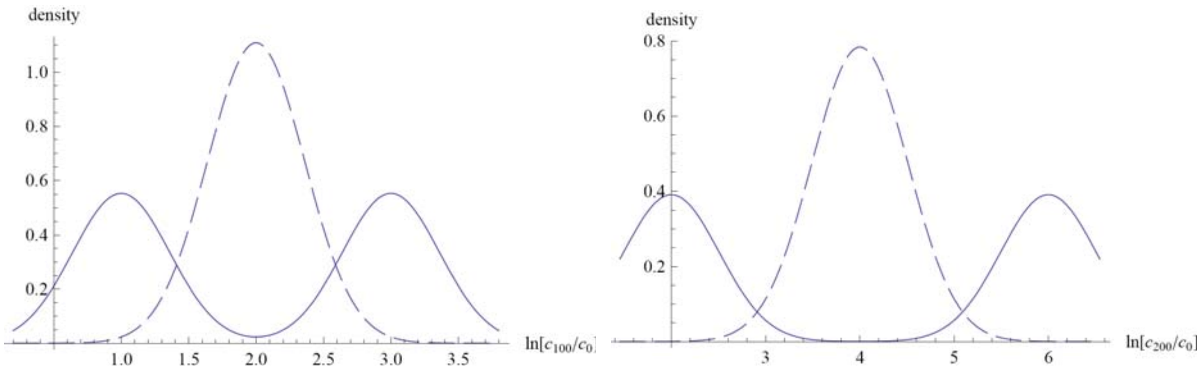


Figure: Density function of $\ln c_t / c_0$ for *t=1, 10, 100* and *200*, under the assumption that $\mu \sim (1\%, 1/2; 3\%, 1/2)$ and $\sigma = 3.6\%$. The dashed curve is the density function without parametric uncertainty and $\mu = 2\%$.

The bottom line is that parametric uncertainty entails fatter tails of the distribution of future consumption. The thickness of the tails increases with the maturity. Integrating out parameter uncertainty by Bayes' rule spreads apart probabilities and thickens the tails of the posterior distribution for predicting the future consumption growth rate. This explains why the term structure of discount rates is decreasing. Indeed, the growing gap of uncertainty compared to the random walk hypothesis in which the term structure is flat magnifies the precautionary effect in the distant future. Because the precautionary effect tends to reduce the discount rate, we get a decreasing term structure. In the long run, the fear of a low economic growth of *1%* dominates all other considerations about how to value the future. Under the assumption that $\delta = 0\%$ and $\gamma = 2$, the discount rate converges to $r_\infty = 2 \times 1\% - 0.5 \times 2^2 \times 0.036^2 = 1.7\%$, as shown in the following figure.

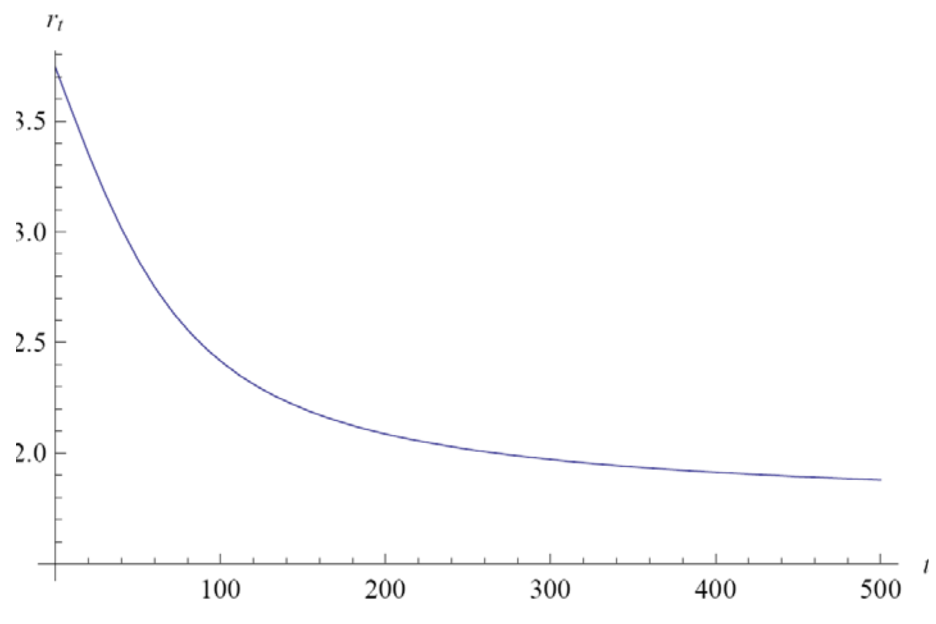


Figure: Efficient term structure with $\mu \sim (1\%,1/2;3\%,1/2)$, $\sigma = 3.6\%$, $\delta = 0\%$ and $\gamma = 2$.

The case of an unknown trend of economic growth

When the growth of log consumption conditional to θ is normally distributed, the term structure of efficient discount rates is characterized by equation (6.10), which is rewritten as follows:

$$r_t = \delta - \frac{1}{t} \ln Ee^{(-\gamma\mu_\theta + 0.5\gamma^2\sigma_\theta^2)t}. \tag{6.11}$$

We hereafter allow θ to have a continuous distribution. In this section, we suppose that the volatility of the growth rate of consumption is known, so that $\sigma_\theta = \sigma$ for all θ. We consider more sophisticated prior distributions for μ_θ than the two-state case considered in the previous section. Suppose that μ_θ is normally distributed with mean μ_0 and variance σ_0^2. We can interpret σ_0 as a measure of the degree of uncertainty about the true growth of log consumption. Observe from (6.11) that we are back once again in a situation to compute the expectation of the exponential of a normally distributed random variable. We obtain that

$$r_t = \delta - \frac{1}{t} \ln e^{(-\gamma\mu_0 + 0.5\gamma^2 t\sigma_0^2 + 0.5\gamma^2\sigma^2)t} = \delta + \gamma\mu_0 - 0.5\gamma^2(\sigma^2 + \sigma_0^2 t). \tag{6.12}$$

This expression can alternatively be derived from the well-known property that if the conditional distribution $\ln c_t / c_0$ given (μ, σ) is normal with mean μt and variance $\sigma^2 t$ and if μt is itself normally distributed with mean $\mu_0 t$ and variance $\sigma_0^2 t^2$, then the unconditional distribution of $\ln c_t / c_0$ is also normal with mean $\mu_0 t$ and variance $\sigma^2 t + \sigma_0^2 t^2$. Define $g = \mu_0 + 0.5(\sigma^2 + \sigma_0^2 t)$ has the growth rate of expected consumption. This allows us to rewrite the above equation as

$$r_t = \delta + \gamma g - 0.5\gamma(\gamma+1)(\sigma^2 + \sigma_0^2 t). \tag{6.13}$$

The term structure of efficient discount rates (6.13) is linearly decreasing in the maturity *t*. It tends to $\min r_\theta = -\infty$ when t tends to infinity. Because the expected growth of log consumption is normally distributed, its support is unbounded below. The plausibility that the economy is facing a true trend of growth that is negative is crucial to value distant cash flows. Although the probability of such an event may be very small, the scenario of a vanishing GDP/cap is much feared by the representative agent. When combining this with the property that $\lim_{c\to 0} u'(c)$ is infinite for power utility functions implies that the social value of transferring wealth to these distant dates where this may happen is highly valued.

One can question the normality of the prior beliefs on the trend of log consumption, or more generally the nature and origin of these prior beliefs. One can approach these questions by using Bayesian inference. Suppose that our current beliefs about the future growth of the economy combines primitive beliefs about it – which may be uninformative – and the observation of a sample of T past growth of log consumption $(x_{-T},...,x_{-1})$. Suppose that these primitive beliefs take the form of three assumptions. First, changes in log consumption are independent and normally distributed. Second, the variance of the change in log consumption is a known constant σ^2. Third, the mean μ of the change in log consumption is normally distributed with mean μ^* and variance σ^{*2}. The observation of the recent changes $(x_{-T},...,x_{-1})$ affects these beliefs. Using Bayes' rule, we have that

$$P\left[\mu \middle| \mu^*,\sigma^*,\sigma,x_{-T},...,x_{-1}\right] = \frac{P\left[x_{-T},...,x_{-1} \middle| \mu,\sigma\right] P\left[\mu \middle| \mu^*,\sigma^*\right]}{P\left[x_{-T},...,x_{-1}\right]}. \tag{6.14}$$

It is well-known that this process of revising beliefs yields a posterior distribution for mean change in $\ln c$ is normally distributed with mean

$$E\left[\mu \middle| \mu^*,\sigma^*,\sigma,x_{-T},...,x_{-1}\right] = \mu_0 = \frac{\mu^*\left(\sigma^*\right)^{-2} + m(\sigma^2/T)^{-1}}{\left(\sigma^*\right)^{-2} + (\sigma^2/T)^{-1}}, \tag{6.15}$$

where $m = T^{-1}\sum_{\tau=-T}^{-1} x_\tau$ is the sample mean of changes in $\ln c$. See for example Leamer (1978, Theorem 2.3). The new expected growth μ_0 is a weighted average of the prior expectation and of the sample mean. A large sample mean pushes beliefs upwards. The sensitiveness of posterior beliefs is an increasing function of the relative precision $(\sigma^2/T)^{-1}$ of the sample information relative to the precision σ^{-2*} of prior beliefs. The posterior variance of μ is equal to

$$Var\left[\mu \middle| \mu^*,\sigma^*,\sigma,x_{-T},...,x_{-1}\right] = \sigma_0^2 = \left(\left(\sigma^*\right)^{-2} + (\sigma^2/T)^{-1}\right)^{-1}. \tag{6.16}$$

The posterior (μ_0,σ_0) can then be considered as the updated mean and standard deviation for the change in log consumption. It can be plugged in equation (6.12) to determine the socially efficient discount rates. A special case is when the prior beliefs are uninformative. This can be approximated by assuming that σ^* is very large. Equations (6.15) and (6.16) then become

$$\mu_0 = m \text{ and } \sigma_0^2 = \frac{\sigma^2}{T}. \tag{6.17}$$

In that case, the beliefs at date *0* are entirely determined by the observation of economic growth. This is the standard way of justifying a normal distribution for the prior beliefs. Notice that this yields a linearly decreasing term structure.

The case of an unknown volatility of economic growth

In a sequence of two recent papers, Weitzman (2007, 2009) considers an alternative model in which the unknown parameter of the distribution of $\ln c_{t+1}/c_t$ is its volatility rather than its mean as in the previous section. So, suppose that $\mu_\theta = \mu$ for all θ. The plausible distribution for the volatility must of course have its support in $\mathbb{R}_+$, which excludes the normal distribution. As already observed in the previous section, it is often more convenient to work with the precision $p_\theta = \sigma_\theta^{-2}$ rather than with the variance. When the precision is unknown, it is standard in the literature to assume that it has a gamma distribution: $p_\theta \sim \Gamma(a,b)$. The gamma distribution has two parameters, a shape parameter *a>0*, and a scale parameter *b>0.* Its density function is

$$f(p;a,b) = p^{a-1}\frac{e^{-p/b}}{b^a\Gamma(a)} \quad \text{for all } p > 0. \tag{6.18}$$

The Gamma function extends the factorial one to non-integer numbers, with $\Gamma(a) = (a-1)!$ when a is a natural integer.

The mean and variance of p_θ are respectively equal to ab and ab^2. In the following figure, we have drawn the Gamma density for various values of the parameters. Remember that the observed volatility of yearly changes in log consumption is around *3.6%*, which gives a precision around $(0.036)^{-2} \approx 800$. We draw four different gamma densities, all with the same mean $ab = 800$.

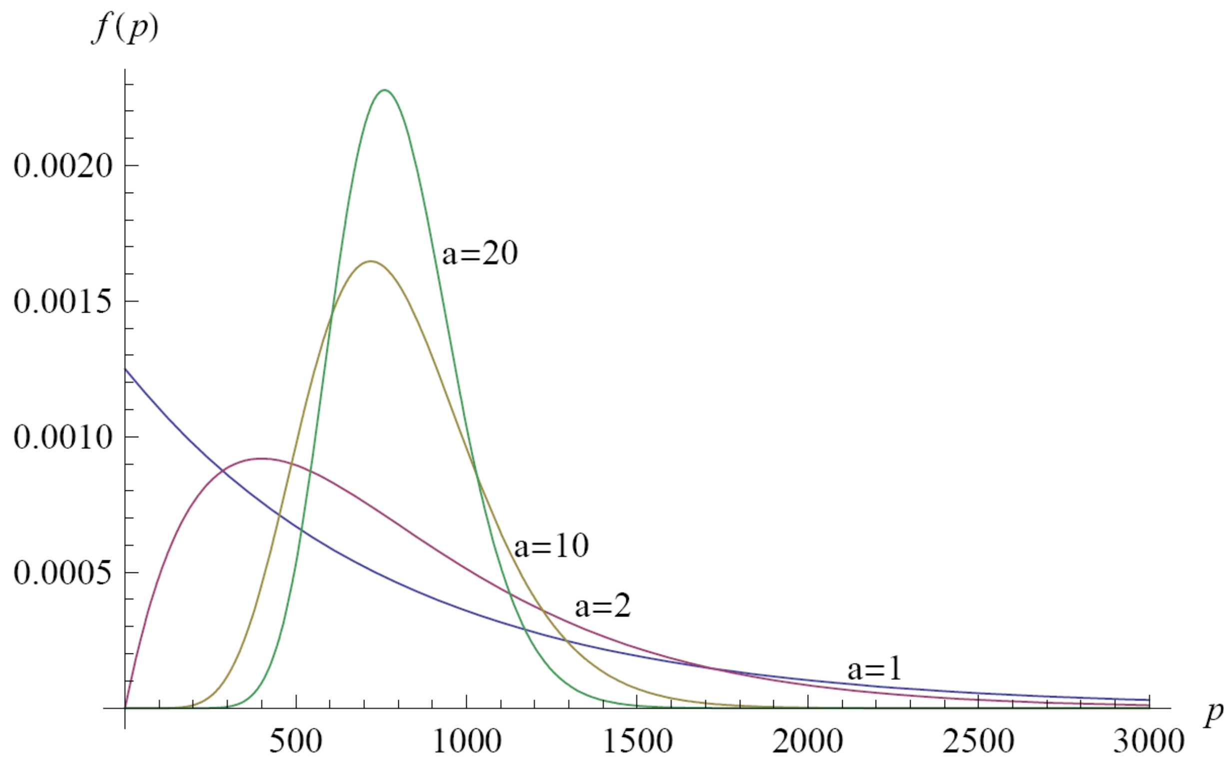


Figure: Gamma densities for different parameters (a,b) with the same $Ep = ab = 800$.

It remains to determine the shape of the term structure under this specification. It is characterized by equation (6.11) which is rewritten as follows:

$$r_t = \delta + \gamma\mu - \frac{1}{t}\ln Ee^{0.5\gamma^2 t/p_\theta} = \delta + \gamma\mu - \frac{1}{t}\ln\int_0^\infty e^{0.5\gamma^2 t/p} f(p;a,b)dp. \qquad (6.19)$$

The integral in this equation -- which is the moment-generating function evaluated at $0.5\gamma^2 t$ of random variable $1/p$ which has an inverted-gamma distribution -- is unbounded. The precautionary effect is here infinite, independent of the degree of parametric uncertainty!

An alternative way to view this problem is obtained by characterizing the unconditional distribution of x_t. Conditionally to σ_θ, it is normal. It is well-known that combining a normal distribution of mean μ with a gamma distribution $\Gamma(a,b)$ for its uncertain precision yields an unconditional distribution that is a Student's *t*-distribution with $v = 2a$ degrees of freedom, with mean μ and variance $1/(a-1)b$:

$$\left.\begin{array}{c} x\big|p \sim N(\mu,\sigma = 1/\sqrt{p}) \\ p \sim \Gamma(a,b) \end{array}\right\} \Rightarrow \frac{x-\mu}{1/\sqrt{ab}} \sim Student(2a) \qquad (6.20)$$

As is well-known also, this Student's *t*-distribution has fatter tails than the corresponding normal distribution with the same mean and variance. In the following figure, we draw different unconditional distributions for the annual change in log consumption by using the same parameters of the gamma distribution as in the previous figure: *(a,b)=(1,800), (2,400), (10,80), and (20,40)*. We assume that x has a mean of $\mu = 2\%$, so that $(x-0.02)\sqrt{800}$ is a Student's t-distribution with $2a$ degrees of freedom. When a tends to infinity, the Student tends to the normal. Finite parameter a has the effect to thicken the tails of the distribution compared to the normal one. As for other sources of parametric uncertainty, the parametric uncertainty about the true volatility of the growth process makes the distribution of the growth rate riskier.

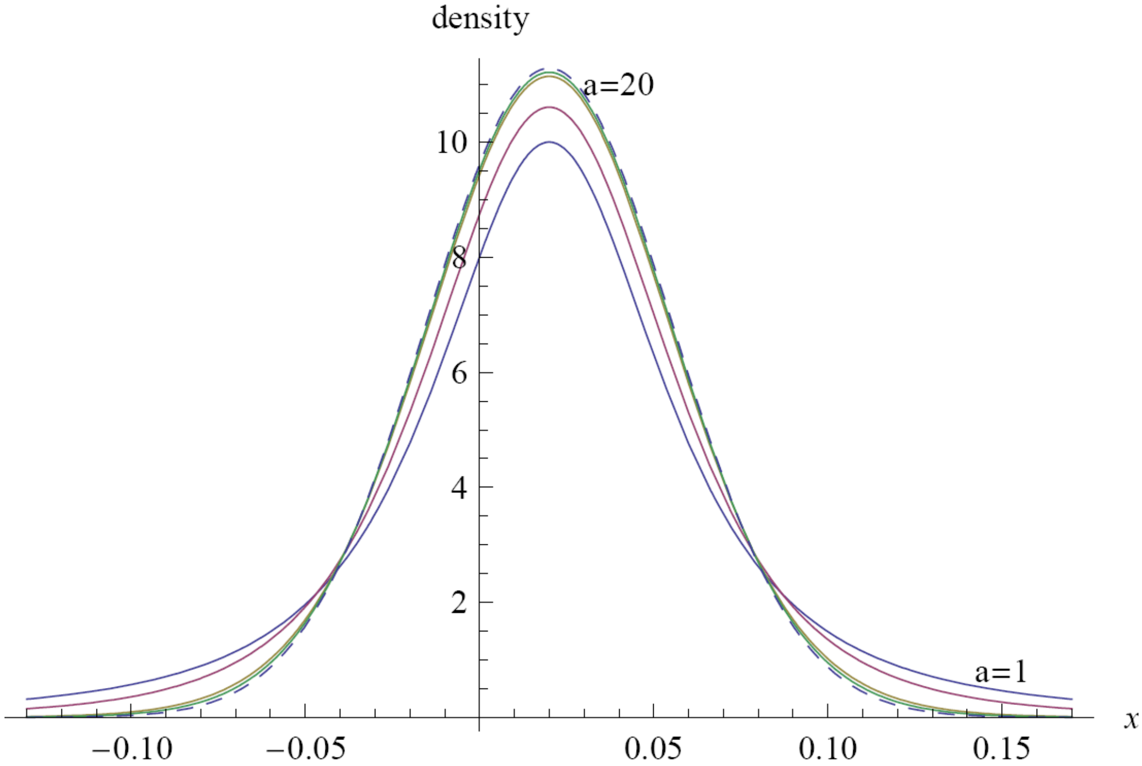


Figure: Density functions for the change in log consumption. We assume that $(x-0.02)\sqrt{800}$ is a Student's t-distribution with $2a$ degrees of freedom, $a = 1, 2, 10$ and 20. The dashed curve is the density of $N(0.02, 1/\sqrt{800})$.

The differences between the normal distribution and the Student's t-distribution may looks quite marginal in the above figure. However, the tails are much different. There is relatively much more probability mass in the second one than in the first. Let us define function $g(t;\nu)$ as the ratio of probabilities that $x_S(\nu)$ and x_N be smaller than t, where $x_S(\nu)$ and x_N are respectively the Student's t-distribution with ν degrees of freedom, and the standardized normal distribution:

$$g(t;\nu) = \frac{P[x_S(\nu) \le t]}{P[x_N \le t]}. \tag{6.21}$$

The table below shows how big *g* can be in the left tail.

	t=-2	*t=-4*	*t=-6*	*t=-8*
$\nu = 1$	6.49	2462.14	5.33×10^7	6.48×10^{13}
$\nu = 10$	1.61	39.76	66952.4	9.64×10^9

Table: Ratio $g(t;\nu)$ of probabilities in the left tail.

What is thus special with this specific parametric uncertainty is that the tails of the unconditional distribution of *x* are here particularly thick. They are so thick that the precautionary effect becomes infinite. This can be checked in the following way. We have that

$$r_1 = \delta - \ln Ee^{-\gamma x_0} = \delta - \ln M_x(-\gamma), \tag{6.22}$$

where $M_x(k) = Ee^{xk}$ is the moment-generating function of random variable *x*. For $x \sim N(\mu,\sigma)$, we know that $M_x(k) = \exp(\mu k + 0.5\sigma^2 k^2)$. But, as is well-known, the Student's t-distribution has an unbounded moment-generating function. Thus, $r_1 = -\infty$.

It may be argued that this result is driven by the fact that "too much" parametric uncertainty is contained in the gamma distribution for the precision *p*. This point raises again the question of the status of our beliefs about the distribution of the uncertain parameter. Suppose that the only source of information is the observation of the past volatility of economic growth. Suppose that the true distribution of x_t is normal. Using Bayes' rule, it can be proved that updating the normal-gamma prior beliefs using the observation of $(x_{-T},...,x_{-1})$ yields a normal-gamma posterior beliefs (see Leamer (1978, Theorem 2.4)). In particular, if μ is known and if the prior on σ is uninformative, the posterior distribution of $p = 1/\sigma^2$ must be a gamma distribution. Thus, the use of an inverse-gamma distribution for the precision is a natural way to model the uncertainty affecting the variance of a Brownian process.

The unboundedness of the efficient discount rate in this case is a consequence of the Inada property $u'(0) = +\infty$ of the utility function, and from the marginalist nature of the economic approach of valuation. The representative agent values a lot any investment that yields a sure positive minimum consumption $\varepsilon > 0$ in the future. Once these investments are implemented, the probability that future consumption will fall below $\varepsilon > 0$ will be zero, and the discount rate will be bounded.

Conclusion

In this chapter, we recognized that the growth process of the economy is not only risky, but it is also surrounded by various parametric uncertainties. After all, who can be sure about the trend and volatility of the economic growth over the next two centuries? We have shown that these parametric uncertainties play a crucial role in shaping the term structure of discount rates. The parametric uncertainty about the trend is a huge source of uncertainty, but only for the distant future. The precautionary effect that it generates provides an intuition for why the term structure should be decreasing in that case. The parametric uncertainty about the volatility of growth makes its unconditional distribution having fatter tails. The fear about the future that it entails induces the representative agent to use a much smaller discount rate for all time horizons.

References

Gollier, C., (2008), Discounting with fat-tailed economic growth, *Journal of Risk and Uncertainty*, 37, 171-186.

Leamer, E. E., (1978), *Specification Searches: Ad Hoc Inference with Non Experimental Data*, John Wiley.

Weitzman, M. L., (2007), Subjective expectations and asset-return puzzle, *American Economic Review*, 97, 1102-1130.

Weitzman, M. L., (2009), On Modeling and Interpreting the Economics of Catastrophic Climate Change, *Review of Economics and Statistics*, 91 (1), 1-19.

The Weitzman's argument

In the first chapter, we have shown that there are basically two methods to determine the socially efficient discount rate. The first one is based on the marginal rate of intertemporal substitution. It leads to the Ramsey rule and to its extension that we have analyzed in details in the previous chapters. The other method is based on the rate of return of capital. At equilibrium, the two methods should lead to the same outcome, which is the equilibrium interest rate.

Let us re-examine the reason why the discount rate should be equalized to the rate of return of risk-free capital in the economy. This reason is based on a simple arbitrage argument. Let r denote this rate of return of capital, which is also the equilibrium interest rate if financial markets are efficient. Consider an investment project that yields a single sure cash flow F in t years per dollar invested today. Because this dollar can alternatively be invested in the economy to yield $\exp(rt)$ dollars in t years, this investment should be implemented only if its future payoff F exceeds $\exp(rt)$, or if the net future value (NFV) $(F-\exp(rt))$ is positive. The NFV is the net future benefit of the investment when compared to the alternative investment in the productive capital of the economy. Behind this positive NFV rule, there is the important notion of the opportunity cost of capital, which tells us that what is spent for one project cannot be spend for other projects. Our efforts in favour of fighting global warming will limit our ability to fight malaria or poverty in developing countries.

The net future value of the project is what the stakeholders get from their investment at date t when financing its initial unit cost by a loan at the interest rate r. An alternative strategy for impatient investors would be to anticipate the future benefit of their investment by borrowing today $F\exp(-rt)$ at rate r, in such a way that the reimbursement F at date t perfectly offsets the cash flow of the project. When doing so, stakeholders get only one benefit, which takes place today and is equal to the net present value (NPV) $-1+F\exp(-rt)$ of the project. It is thus optimal to invest in the project if this NPV is positive. Obviously, because the NPV and the NFV are proportional to each other, they must have the same sign, so that the two decision rules always yields the same decision.

One important difficulty of this approach is that there is no market for risk free assets with very long maturities. Typically government bonds have maturities not exceeding 30 years. Thus, the arbitrage strategies presented above requires a "roll-over" strategy in which the transfer of cash-flows will be made via a sequence of credit contracts scattered in time. It confronts us to a "reinvestment risk", because we don't know what will be the credit market conditions in the future. Alternatively, on could try to guess what will be the rate of return of capital in the future. Although economists have tried for decades to build realistic models of economic growth, we must recognize that the conclusions of the so-called neoclassical growth theory are not impressive. Growth can only marginally explained by the basic driver identified in that theory, which is capital accumulation. Much of the intensity of economic growth is contained in the famous "Solow's residual", which is a mostly exogenous, and which has been interpreted as being fuelled by technological and scientific progresses. The more recent endogenous growth theory tries to model the production of new knowledge, but at this stage, it is not expected to help us very much to characterize the rate of return of capital in the next 200 years. The bottom line is that the above-mentioned arbitrage arguments cannot be applied in the context of sustainable development without more sophistication.

Following Weitzman (1998, 2001), let us take this problem seriously by supposing that the rate of return of capital r will be constant in the future, but is uncertain this morning. It will be observed at the end of the day. To keep it simple, let us consider a numerical example in which r will be either *5%* or *1%* with equal probabilities. Thus, the opportunity cost of capital cannot be evaluated without error today. One dollar invested today in the productive capital of the economy will yield either $\exp(0.05t)$ or $\exp(0.01t)$ dollars at date t. So, it is hard to compare this benefit to the sure benefit F of the investment project. In a word, the NFV of this project is uncertain. One possible decision rule under uncertainty is to require that the sure cash flow of the project be larger than the expected cash flow of the investment in the productive capital of the economy. This is referred to as the expected NFV rule, which requires that the expected NFV of the project be positive to recommend its implementation. This is equivalent to require that the investment has an internal rate of return larger than a critical rate R_t^F which is defined as follows:

$$e^{R_t^F t} = Ee^{rt} \tag{7.1}$$

Weitzman (1998) provides an alternative approach which yields opposite results. He proposes an alternative decision rule under uncertainty: A sure investment project should be implemented if its expected NPV is positive. In spite of the fact that this rule is equivalent to the expected NFV rule when there is no uncertainty, as explained above, this is not the case when there is uncertainty. If the future benefit is offset by borrowing $F\exp(-rt)$ once the rate r will be known, the net present benefit of the investment is equal to $-1+E\left[F\exp(-rt)\right]$, which is equivalent to discounting F at a rate R_t^P defined as

$$e^{-R_t^P t} = Ee^{-rt} \tag{7.2}$$

As observed by Gollier (2004), using the positive expected NFV rule or the positive expected NPV rule leads to opposite results concerning the choice of the discount rate. In particular, we obtained that

$$\min r \le R_t^P \le Er \le R_t^F \le \max r \tag{7.3}$$

for all t. Moreover, the min and max bounds correspond to the asymptotic values of respectively R_t^P and R_t^F when t tends to infinity. The NPV approach is more favourable to the evaluation of sure investment projects than the NFV approach, and this bias is increasing with their maturity.

This analysis has also shown that the two rules differ upon the date at which the risk associated to the alternative investment in the economy is allocated. Under the NFV approach, cash flows and risk are all transferred to the terminal date of the project, whereas they are all transferred to today under the NPV approach. This is a paradox, because of the huge difference in the practical consequences of the two approaches. In the spirit of the Modigliani-Miller's Theorem, the evaluation of an investment project should not depend on the way that it is financed. In the absence of a clear description of the stakeholders' preferences towards risk and time, it is not possible to determine which rule should be preferred, and which discount rate should be selected.

The case of the logarithmic utility function

A surprising result of the expected NFV approach is that the uncertainty affecting the alternative investment project in the productive capital of the economy biases our preferences in favour of this project against the sure one. This suggests that the introduction of risk aversion into the picture should arguably make us to favour the expected NPV rule which goes into the opposite direction.

The ubiquitous idea in this book is that what matters for stakeholders is not the payoff of the project itself, but rather the utility that it generates. We suppose in this section that this utility function is logarithmic, i.e., $u(c) = \ln c$. An important property of this function is that a change in the interest rate does not affect saving. The wealth effect perfectly compensates the substitution effect. This implies that at the end of the day, when r is observed, the level of consumption c_0 is insensitive to this information. This will be shown later in this chapter. On the contrary, consumption in the distant future will be highly sensitive to r. It can be shown that the optimal consumption at date t is proportional to $\exp(rt)$. Thus, at the beginning of the day, there is absolutely no uncertainty about the optimal consumption at the end of the day, but there is a huge uncertainty about consumption in the distant future.

Let us consider the expected NPV approach in this context. Remember that the NPV rule is based on the assumption that all cash flows of the sure project are transformed into additional consumption at the end of the day, and only at that time. This additional consumption is uncertain (it depends upon the unknown r), but it is marginal. Because consumption c_0 at date 0 is risk free, adding this marginal risk to initial wealth increases welfare if and only if the expected NPV is positive, i.e., risk aversion is irrelevant. This is because (independent) risk is a second-order effect in the expected utility model (Segal and Spivak (1990)). When introducing a small lottery in an initially risk free situation, the first-order expectation effect always dominates. This can be seen from observing that, by the Arrow-Pratt approximation (3.3), the risk premium for small risk is proportional to the variance of the payoff, i.e., to the square of the size of the risk. This means that the NPV formula (7.2) is perfectly valid when the representative agent has a logarithmic utility function.

Let us consider alternatively the expected NFV approach, which relies on the assumption that all costs and benefits of the sure project are transferred to the terminal date t. Observe that this NFV is highly negatively related to the interest rate r, since the loan used to finance the initial cost of the project will yield a larger repayment at the terminal date when the interest rate is large. This means that the NFV of the sure project is highly negatively correlated with c_t. In other words, implementing the sure project by this financing strategy provides some hedging against the macroeconomic risk at date t. This is positively valued by consumers; something that the equation (7.1) of the expected NFV approach fails to take into account. Thus, this equation misprices the future.

To sum up, when the sure project is implemented and cash flows are transferred to the present as in the NPV approach, one can do as if the representative agent be risk neutral, because current consumption is risk free. On the contrary, when the sure project is implemented and cash flows are transferred to the terminal date as in the NFV approach, this strategy serves as an insurance against the macroeconomic risk, so that the risk neutrality assumption implicit in equation cannot be sustained. Thus, when the representative agent has a logarithmic utility function, Weitzman's formula (7.2) is right.

When the utility function of the representative agent is not logarithmic, the problem is more complex, because the initial consumption c will optimally react to changes in the rate of return of capital. Therefore, none of the two rules (7.1) and (7.2) will be valid. The next section is devoted to the analysis of the general case.

Taking account of preferences towards risk and time

When considering the expected NFV rule with risk aversion, the marginal additional consumption $F - \exp(rt)$ occurring at date t has differentiated marginal effect on utility because of the different levels of GDP/cap c_t that will be attained in the different states. The underlying strategy of financing the initial cost by a loan at rate r increases the expected utility at date t if

$$Eu'(c_t)\left[F - e^{rt}\right] \geq 0. \tag{7.4}$$

This is equivalent to using a discount rate R_t^F implicitly defined as follows:

$$R_t^F = \frac{1}{t}\ln\frac{E\left[u'(c_t)e^{rt}\right]}{Eu'(c_t)}. \tag{7.5}$$

This formula generalizes equation (7.1) to the case of risk aversion. Because c_t and r are likely to be correlated, the two equations are not equivalent. In fact, because the GDP/cap is expected to be larger when the return of capital is larger, one expects a negative correlation between $u'(c_t)$ and r, which implies that the numerator in equation (7.5) should be smaller than the product of $Eu'(c_t)$ by $E\exp(rt)$. This implies in turn that the right-hand side of this equation should be smaller than the one in equation (7.1). It implies that risk aversion should have a negative impact on the discount rate recommended under the expected NFV approach, and this effect is increasing with the maturity under scrutiny. The intuition of this result is that investing in the productive capital of the economy yields a high risk that has a perfect correlation with the undiversifiable macroeconomic risk. The associated risk premium of this strategy is increasing with the time horizon, which tends to favour the investment in the risk free project.

The same method should also be used under the expected NPV approach. Remember that this approach is based on the assumption that the future cash flow of the risk free project is offset by a loan of $F\exp(-rt)$ at the end of the day. This strategy raises the expected utility of current consumption if

$$Eu'(c_0)\left[Fe^{-rt}-1\right]\geq 0 \tag{7.6}$$

This is equivalent to using a discount rate R_t^P defined as

$$R_t^P=-\frac{1}{t}\ln\frac{E\left[u'(c_0)e^{-rt}\right]}{Eu'(c_0)}. \tag{7.7}$$

Under risk neutrality (*u'* constant), this equation is equivalent to (7.2). The choice of consumption c_0 will in general depend upon the observation of the rate of return of capital at the end of the day. If the substitution effect dominates the wealth effect, c_0 and r are negatively correlated. This means that investing in the economy rather than in the project plays the role of insurance low consumption in the short run. This reduces the relative attractiveness of the sure project under the expected NPV approach. This tends to raise the discount rate R_t^P.

Thus, the introduction of risk aversion tends to reduce the gap between the two discount rates described by inequalities (7.3), by raising the lower rate and reducing the higher one. One can go one step further by showing that the two approaches are in fact equivalent if one recognizes that consumers optimize their consumption plan contingent to the information on the future rate of return of capital. Suppose that r is realized, so that consumers can save and borrow at that interest rate. Consider a marginal increase in saving at date *0* by 1 to increase consumption at date t by $\exp(rt)$. This marginal change in the consumption plan has no effect on welfare if

$$u'(c_0)=e^{-\delta t}e^{rt}u'(c_t). \tag{7.8}$$

This is an optimality condition, which must hold for all possible realizations of r. But then, if one plugs this condition into equation (7.7), one gets

$$R_t^P = -\frac{1}{t}\ln\frac{E\left[u'(c_0)e^{-rt}\right]}{Eu'(c_0)} = \frac{1}{t}\ln\frac{E\left[u'(c_t)e^{rt}\right]}{Eu'(c_t)} = R_t^F. \tag{7.9}$$

This implies that $R_t^P = R_t^F$ for all t! We conclude that once risk and risk aversion are properly combined with intertemporal optimization, the NPV and NFV approaches are equivalent. Moreover, these approaches are equivalent to the one on which the Ramsey rule and our previous chapters are based. Indeed, one also gets

$$R_t^P = -\frac{1}{t}\ln\frac{E\left[u'(c_0)e^{-rt}\right]}{Eu'(c_0)} = \delta - \frac{1}{t}\ln\frac{Eu'(c_t)}{Eu'(c_0)} \tag{7.10}$$

The only difference with respect to what has been presented earlier in this book comes from the possibility that c_0 be random.

The term structure of discount rates

In this model in which shocks on capital productivity are permanent, risk on consumption growth are also permanent, as seen from equation (7.8). This implies that risks are magnified by time, compared to the more standard Brownian motion in which it is known that the term structure of discount rate is flat. The growing importance of the precautionary effect that is entailed by the permanency of the shock on capital productivity yields a decreasing term structure. The property that the term structure must be decreasing in this model can be proved by rewriting equation (7.9) as

$$r_t = R_t^F = R_t^P = -\frac{1}{t}\ln\frac{E\left[u'(c_0)e^{-rt}\right]}{Eu'(c_0)} = -\frac{1}{t}\ln E^*\left[e^{-rt}\right], \tag{7.11}$$

where E^* is the standard risk-neutral expectation operator in which for any function F of r, we have $E^*\left[F(r)\right] = E\left[u'(c_0(r))F(r)\right] / E\left[u'(c_0(r))\right]$. Thus, we see that the efficient term structure under this specification is equivalent to the Weitzman's NPV formula (7.2) up to the risk-neutral transformation of the probability distribution. This implies that we get the same qualitative properties for the term structure than those generated by equation (7.2): it is decreasing and tends to the smallest possible rate of return of capital.

Let us examine this point in more details by characterizing the optimal allocation of risk and consumption through time. Suppose as before that relative risk aversion is a positive constant

γ, so that $u'(c) = c^{-\gamma}$. One can solve equation (7.8) together with the intertemporal budget constraint

$$\int_0^\infty e^{-rt} c_t dt = k_0, \tag{7.12}$$

where k_0 is the initial level of capital in the economy. A solution exists if $r(1-\gamma) < \delta$, which is true in particular when γ is less than unity. The solution is written as

$$c_t = k_0 \left(r - \frac{r-\delta}{\gamma} \right) e^{\frac{r-\delta}{\gamma} t}. \tag{7.13}$$

Observe first that the initial consumption c_0 is independent of the random variable r when γ equals unity. This confirms the property that initial consumption is not sensitive to the interest rate when the utility function is logarithmic. Observe also that, conditional to r, c_t has a constant growth rate $g(r) = (r-\delta)/\gamma$. It is noteworthy that this implies that the ex post equilibrium interest rate is $r = \delta + \gamma g$, which is the Ramsey rule. The problem is to determine the socially efficient discount rate before r is revealed. The fact is that ex post consumption will grow at a constant rate that is unknown ex ante. This simple model is thus equivalent to the following stochastic process for the growth of log consumption:

$$\begin{cases} c_{t+1} = c_t e^{g(\theta)} \\ c_0 = k_0 \left(\theta - g(\theta) \right) \end{cases} \tag{7.14}$$

This is a very special case of the general problem of parametric uncertainty that we examined in the previous chapter, but with an uncertain discrete jump in initial consumption. The arithmetic Brownian motion for the log consumption is degenerated with a zero volatility, so that the uncertainty is fully resolved at date *0*. The riskiness of consumption increases exponentially through time, rather than linearly as in the case of a Brownian motion.

Following Weitzman (2009), let us calibrate this model by assuming that the uncertainty about the future rate of return of capital is governed by a gamma distribution:

$$f(r;a,b) = r^{a-1} \frac{e^{-r/b}}{b^a \Gamma(a)} \quad \text{for all r} > 0, \tag{7.15}$$

where a and b are two positive constant. This implies that the mean rate of return is $Er = \mu = ab$ and its variance is $Var(r) = \sigma^2 = ab^2$. Suppose that $\delta = 0$, which implies that $g(r) = (\gamma - 1) r/\gamma$. The Ramsey pricing formula (7.10) can then be written as follows:

$$r_t = -\frac{1}{t} \ln \frac{Er^{-\gamma} e^{-rt}}{Er^{-\gamma}} = -\frac{1}{t} \ln \frac{\int_0^\infty r^{-\gamma+a-1} e^{-r(t+b^{-1})} dr}{\int_0^\infty r^{-\gamma+a-1} e^{-rb^{-1}} dr}. \tag{7.16}$$

The two integrals in this expression have an analytical solution. Indeed, because the integral of the density $f(r;k,h)$ must be equal to 1, we must have that

$$\int_0^\infty r^{h-1}e^{-r/k}dr = k^h\Gamma(h). \tag{7.17}$$

We apply this property twice in (7.16) for $h = a-\gamma > 0$ and respectively $k = (t+b^{-1})^{-1}$ and $k = b$. It yields

$$r_t = -\frac{1}{t}\ln\frac{(t+b^{-1})^{\gamma-a}\Gamma(a-\gamma)}{b^{a-\gamma}\Gamma(a-\gamma)} = \frac{a-\gamma}{t}\ln(1+tb). \tag{7.18}$$

It is easier to rewrite this equation with parameters (μ,σ) rather than (a,b). This substitution yields the risk-adjusted Weitzman's formula

$$r_t = \frac{(\mu/\sigma)^2-\gamma}{t}\ln\left(1+\frac{t\sigma^2}{\mu}\right). \tag{7.19}$$

As long as γ is smaller than $(\mu/\sigma)^2$, this term structure is decreasing, and tends to zero when t tends to infinity. In the next table, we computed the discount rates for a gamma distribution of the rate of return of capital with mean $\mu = 4\%$ and standard deviation $\sigma = 2\%$, together with $\gamma = 2$. Compared to the expected rate of return of capital of *4%*, we see that the ex ante short term efficient discount rate is only *2%*, thereby illustrating the effect of risk aversion. The additional reduction in the discount rate for longer maturities illustrates the growing precautionary effect.

$t \to 0$	*t=50*	*t=200*	*t=500*	*t=1000*
$r_0 = 2\%$	$r_{50} = 1.62\%$	$r_{200} = 1.10\%$	$r_{500} = 0.72\%$	$r_{1000} = 0.48\%$

Table: Discount rate with $\gamma = 2$ and with a gamma distribution for the shock on the future return of capital. The mean future rate has a mean of *4%* and a standard deviation of *2%.*

Conclusion

We have shown in this chapter that the evaluation of a sure investment project is independent upon how cash flows are allocated through time, as soon as we recognize that economic agents are risk-averse and that they optimize their consumption plans. This Fisher equivalence property is particularly relevant when the rate of return of capital in the economy is uncertain. This reconciles the two approaches for discounting that have been proposed in the literature. The expected net present value rule proposed by Weitzman (1998) has the attractive feature to yield a decreasing term structure of discount rates. We have shown that this property has a simple intuition. When shocks on the productivity of capital are permanent, they yield permanent shocks on the growth rate of the economy. This magnifies the macroeconomic uncertainty associated to distant futures. It is thus precautionary to use smaller discount rates for more distant time horizons, in order to bias our investments towards those with durable positive impacts.

The bottom line is that when shocks on the interest rate have a permanent component, the term structure of discount rates should be decreasing. Newell and Pizer (2003), and Groom, Koundouri, Panopoulou and Pantelidis (2007) have estimated the degree of permanency in the shocks on interest rates, and have shown that it has a crucial role in the shape of the term structure of efficient discount rates.

References

Gollier, C., (2004), Maximizing the expected net future value as an alternative strategy to gamma discounting, *Finance Research Letters*, 1, 85-89.

Gollier, C., and M.L. Weitzman, (2009), How Should the Distant Future be Discounted When Discount Rates are Uncertain?, *Economic Letters*, forthcoming.

Groom, B., P. Koundouri, E. Panopoulou and T. Pantelidis, (2007), An Econometric Approach to Estimating Long-Run Discount Rates. *Journal of Applied Econometrics,* 22, 641-656.

Newell, R. and W. Pizer, 2003. Discounting the Benefits of Climate Change Mitigation: How Much Do Uncertain Rates Increase Valuations? *Journal of Environmental Economics and Management*, 46 (1), 52-71.

Segal, U. and A. Spivak, (1990), First order versus second order risk aversion, *Journal of Economic Theory*, 51, 111-125.

Weitzman, M.L., (1998), Why the far-distant future should be discounted at its lowest possible rate?, *Journal of Environmental Economics and Management*, 36, 201-208.

Weitzman, M.L., (2001), Gamma discounting, *American Economic Review*, 91, 260-271.

Weitzman, M.L., (2009), Risk-adjusted gamma discounting, mimeo, Harvard University.

A theory of the decreasing term structure

The last chapter of Part II of this book is aimed at providing a theoretical foundation to the term structure of discount rates. The benchmark model is based on two assumptions, one on individual preferences toward risk, and the other on the nature of the dynamic uncertainty. We have shown that the constancy of relative risk aversion combined with a random walk for the growth of log consumption yields a flat term structure of efficient discount rates. In this chapter, we relax these two assumptions by using a stochastic dominance approach.

We first show the link between the current long term discount rate and the expectation about what will be the future short term discount rate. Said differently, we first explore how the term structure evolves over time.

Current long discount rate and future short discount rates

We limit the analysis to three equally distant dates, *t=0, 1,* and *2.* We assume that c_0 is known. At date *t=0*, the short and long discount rates are respectively

$$r_1 = \delta - \ln \frac{Eu'(c_1)}{u'(c_0)} \tag{8.1}$$

and

$$r_2 = \delta - \frac{1}{2} \ln \frac{Eu'(c_2)}{u'(c_0)}. \tag{8.2}$$

Suppose now that we are at date *t=1*, with a realized level of consumption c_1. At that date under that state of nature, one should use a short rate denoted $r_{1\to 2} = r_{12}(c_1)$ to discount a sure cash flow occurring one period later at date *t=2*. It is as usual characterized by the following equation:

$$r_{12}(c_1) = \delta - \ln \frac{E\left[u'(c_2) \middle| c_1 \right]}{u'(c_1)}. \tag{8.3}$$

We want to link these three rates r_1, r_2 and r_{12}. This can be done by rewriting equation (8.2) as follows:

$$\begin{aligned} r_2 &= \delta - \frac{1}{2}\ln\frac{Eu'(c_2)}{u'(c_0)} \\ &= \delta - \frac{1}{2}\ln E\left[\frac{E\left[u'(c_2)|c_1\right]}{u'(c_1)}\frac{u'(c_1)}{Eu'(c_1)}\frac{Eu'(c_1)}{u'(c_0)}\right] \\ &= -\frac{1}{2}\ln e^{-r_1} Ee^{-r_{12}(c_1)}\frac{u'(c_1)}{Eu'(c_1)}, \end{aligned} \tag{8.4}$$

This implies that

$$r_2 = 0.5\left(r_1 + R_{12}\right) \tag{8.5}$$

where R_{12} is defined as follows:

$$e^{-R_{12}} = \frac{Eu'(c_1)e^{-r_{12}(c_1)}}{Eu'(c_1)} \tag{8.6}$$

Equation (8.5) tells us that the long rate today is the average of the short rate r_1 today and R_{12}. Observe that the discount factor $\exp(-R_{12})$ is the risk-neutral expectation of the future discount factor $\exp(-r_{12}(c_1))$. Rate R_{12} measured at date *t=0* depends upon the uncertainty about the immediate growth rate and upon its correlation with the interest rate that will prevail in the future. R_{12} as the certainty equivalent of the future short rate r_{12}. To keep terminology simple, let us refer to R_{12} as the forward interest rate. It lies somewhere between the smallest and the largest possible future short rates. Using equations (8.3) and (8.6), it can easily checked that R_{12} can be rewritten as

$$R_{12} = \delta - \ln\frac{Eu'(c_2)}{Eu'(c_1)}. \tag{8.7}$$

This should not be a surprise that the discount factor to be used at date *0* to evaluate a transfer of consumption from date *1* to date *2* is equal to $\exp(-\delta)Eu'(c_2)/Eu'(c_1)$. This is indeed the marginal rate of substitution between c_1 and c_2 when evaluated today. Remember that, by the first theorem of welfare economics, the efficient discount rate is also the equilibrium interest rate in a frictionless economy. In the same spirit, R_{12} is the equilibrium forward interest rate, i.e., the rate of return of a credit contract at date *0* for a loan at date *1* with maturity at date *2*.

Equations (8.5) and (8.6) also describe the relations linking current long rates and the expectation about future shorter rates. It states that the following two investment strategies have the same effect on the expected utility at date *1*. Under both strategies, consumption is reduced by ε at date *2* to increase consumption at date *1*. The first investment strategy is safe. It consist at date *0* to borrow long to invest short. More specifically, one borrows

$\varepsilon \exp(-2r_2)$ today that will yield a reimbursement of ε at date *2*. This loan is used at date zero to invest in a short bond that yields a sure payoff $\varepsilon \exp(-2r_2)\exp(r_1)$ at date *1*. The increase in utility at date *1* is thus equal to that marginal sure increase in consumption multiplied by $Eu'(c_1)$. The alternative investment strategy consists in borrowing $\varepsilon \exp(-r_{12})$ at date *1* that will yield the same reimbursement ε in date *2*. Seen from date *0*, this is a risky strategy because the increased consumption at date *1* will depend upon the short term rate $r_{12}(c_1)$ that will prevail at that time. The increase in expected utility at date *1* is given by $E\exp(-r_{12})u'(c_1)$. At equilibrium, the two strategies must have the same effect on welfare. The following condition must thus be satisfied:

$$\varepsilon e^{-2r_2} e^{r_1} Eu'(c_1) = Ee^{-r_{12}} u'(c_1), \tag{8.8}$$

which is equivalent to equation (8.4) that yields property (8.5). This simple arbitrage argument explains why the long rate today must increase when investors expect the future interest rate to go up. It also explains the role of risk aversion in this relationship.

A vast literature on the term structure of interest rates has examined these interactions. Until seminal works by Vasicek (1977) and Cox, Ingersoll and Ross (1985), economists based their analysis on the so-called “Pure Expectations Hypothesis”, which states that the long rate today is the mean of the sequence of current and future short rates. This is similar to equations (8.5) and (8.6), but with a linear utility function u in (8.6). In spite of its inappropriate assumption of risk neutrality, this theory is compatible with the crucial idea that the current long rate tells us something about the investors’ expectation about the future rates.

Decreasing term structure

We thus have two ways to write the condition for the long rate to be smaller than the short one: $r_2 \le r_1$. First, we clearly see from property (8.5) that this is the case if the current short interest rate r_1 today is larger than the forward rate R_{12}:

$$R_{12} \le r_1. \tag{8.9}$$

Second, we can more directly use conditions (8.1) and (8.2) to get that r_2 is smaller than r_1 if

$$\delta - \frac{1}{2}\ln\frac{Eu'(c_2)}{u'(c_0)} \le \delta - \ln\frac{Eu'(c_1)}{u'(c_0)}, \tag{8.10}$$

i.e., if

$$u'(c_0)Eu'(c_2) \geq \left(Eu'(c_1)\right)^2 \tag{8.11}$$

Of course, given equations (8.1) and (8.7), these two ways yield exactly the same condition for a decreasing term structure.

The case of an i.i.d. dynamic growth process

In this section, we examine the case in which the log of consumption exhibits no serial correlation. We are looking for the condition on *u* that yields a decreasing term structure. Let $x_t = \log c_{t+1} - \log c_t$ denote the change in log consumption between dates *t* and *t+1*. We assume that (x_0, x_1) are i.i.d. It is easier to use variable $y_t = \exp(x_t) = c_{t+1}/c_t$ which is the relative change in consumption between dates *t* and *t+1*. We can thus rewrite condition (8.11) for a decreasing term structure as follows:

$$u'(c_0)Eu'(c_0 y_0 y_1) \geq \left(Eu'(c_0 y_0)\right)^2. \tag{8.12}$$

Let us first consider the special case of power utility functions with $u'(c) = c^{-\gamma}$. The above condition is then equivalent to

$$Ey_0^{-\gamma} y_1^{-\gamma} \geq \left(Ey_0^{-\gamma}\right)^2. \tag{8.13}$$

Because y_0 and y_1 are independent, the left-hand side of this inequality equals $Ey_0^{-\gamma} Ey_1^{-\gamma}$, which in turn is equal to the right-hand side of (8.13) since y_0 and y_1 are identically distributed. We conclude that condition (8.13) holds as an equality, which implies that the term structure of discount rates is flat.

Under constant relative risk aversion, the short term rate r_{12} is independent of c_1. Indeed, from (8.3), we have that

$$r_{12}(c_1) = \delta - \ln \frac{E\left[u'(c_2) \middle| c_1\right]}{u'(c_1)} = \delta - \ln \frac{E\left(c_1 y_1\right)^{-\gamma}}{c_1^{-\gamma}} = \delta - \ln Ey_1^{-\gamma}. \tag{8.14}$$

It is a crucial property of power function that the equilibrium interest rate is independent of the level of economic development. During the XXth century, GDP per capita has been multiplied by a factor around 7, but we did not observe any clear trend to the short term interest rate. This is confirmed in the figure below, in which we draw the series of short term real interest rates between 1900 and 2006 in the United States. This fact is in favour of a constant relative risk aversion. If, in addition, expectations remain stable over time, i.e., if y_0 and y_1 are identically distributed, then comparing (8.14) and (8.1) implies that $r_1 = R_{12} = r_{12}$, which implies in turn that the term structure should be flat.

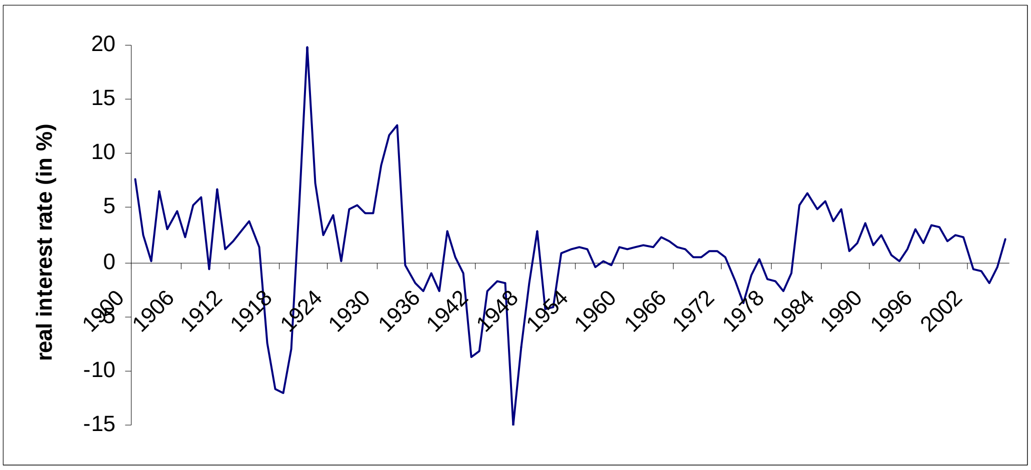


Figure: Real Bill rates in the United States in the XXth century.

Source: Morningstar France.

Let us relax the assumption that relative risk aversion is constant. Let us examine the condition under which r_{12} is decreasing with c_1. From (8.3), this is the case if

$$f(c_1) = \frac{Eu'(c_1 y_1)}{u'(c_1)} \tag{8.15}$$

is increasing in c_1. We have that

$$f'(c) = \frac{u'(c)Ey_1 u''(cy_1) - u''(c)Eu'(cy_1)}{u'(c)^2} \tag{8.16}$$

which is positive if

$$-\frac{Ey_1 u''(cy_1)}{Eu'(cy_1)} \leq -\frac{u''(c)}{u'(c)}. \tag{8.17}$$

This is equivalent

$$E\frac{u'(cy_1)}{Eu'(cy_1)} R(cy_1) \leq R(c), \tag{8.18}$$

where $R(c) = -cu''(c)/u'(c)$ is relative risk aversion. Suppose that consumption never falls, i.e., that y_1 is larger than unity almost surely. If relative risk aversion is decreasing, this implies that $R(cy_1)$ is smaller than $R(c)$ almost surely. This implies that condition (8.18) always holds. Thus, under the assumption that consumption never falls, decreasing relative risk aversion is sufficient for a decreasing term structure. This condition is also necessary if we do not specify the distribution of y_1 with support in $\mathbb{R}_+$. This result is in Gollier (2002a, 2002b).

In the figure below, we draw the term structure of discount rates in the special case of a modified power function with a minimum level of subsistence *k*:

$$u(c) = \frac{(c-k)^{1-\gamma}}{1-\gamma}, \tag{8.19}$$

which is increasing and concave in its domain $]k,+\infty[$. Parameter *k* is interpreted as a minimum level of subsistence in this case since the utility goes to $-\infty$ when consumption goes to this minimum consumption level *k*. It is easily checked that $R(c) = \gamma c/(c-k)$ under this specification. This function is decreasing in its relevant domain. It tends to infinity when consumption approaches the minimum level of subsistence, and it converges to γ for large consumption levels. In this figure we normalize *k* to unity, and we assume that current consumption c_0 is respectively *20%, 50%* or *100%* larger than the minimum level of subsistence. Let us consider $c_0 = 2$ as a benchmark. We also assume that the growth rate of the economy is a sure *2%* per year, and that $\gamma = 1$, so that the relative risk aversion today is $R(1) = 2$. Using the Ramsey rule that states that the interest rate net of the rate of impatience – which is assumed to be *0%* -- must be equal to the product of relative risk aversion by the growth rate of consumption, we obtain a short discount rate of $2 \times 2\% = 4\%$. For very long maturities, the relevant *R* to be used in the Ramsey rule is $R(+\infty) = 1$, which yields a long discount rate equalling $1 \times 2\% = 2\%$.

This figure also depicts the situation for less developed countries whose GDP per capita is closer to the minimum level of subsistence. Let us assume for example that $c_0 = 1.2$, i.e., that current consumption is only *20%* larger than this minimum. The marginal utility of consumption is thus critically large today, which implies that the future gets low priority. This takes the form of a large discount rate $r = R(1) \times 2\% = 12\%$ in the short run. This may explain why poorer countries are short-termist in various investment areas as education or infrastructure building.

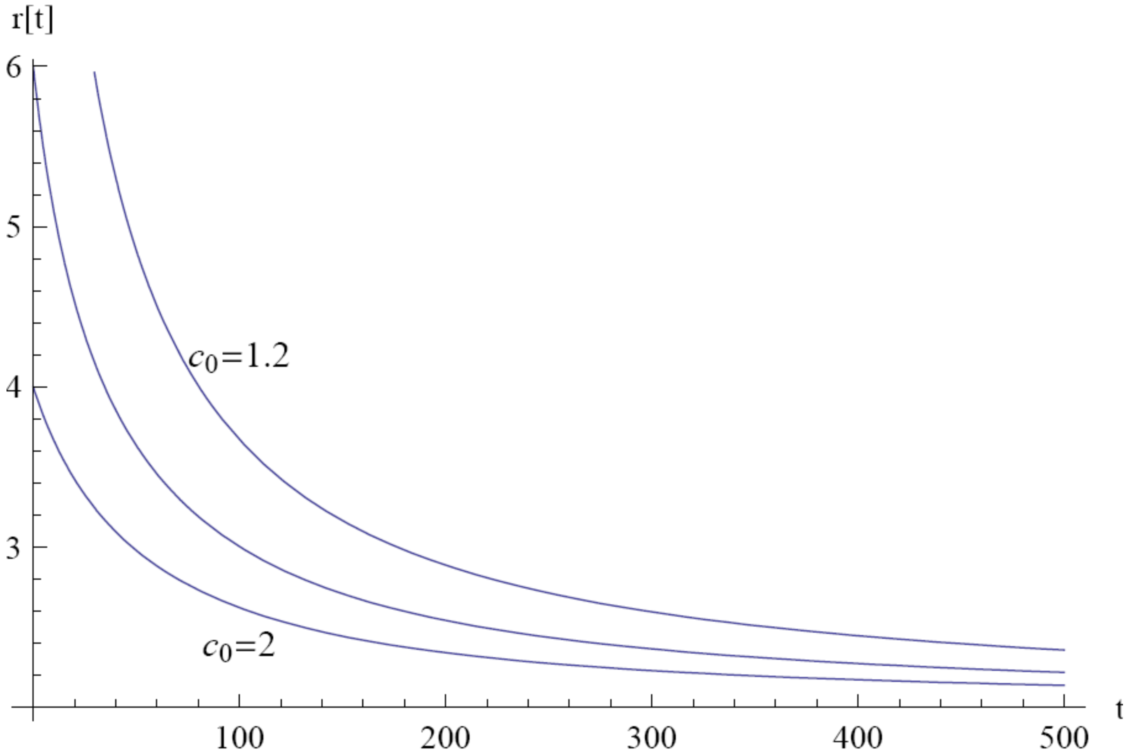


Figure: The term structure of discount rates with

$\delta = 0\%, x_t = 2\%, u'(c) = (c-1)^{-1}, c_0 = 1.2,$ 1.5 and 2.

Under the assumption of a never decreasing consumption, the term structure is decreasing with maturity if and only if relative risk aversion is decreasing with wealth. The intuition of this result is simple. The intensity of the wealth effect is proportional to *R*, which measures the aversion to intertemporal inequality. In a growing economy, this effect decreases over time when *R* is decreasing with wealth. This implies that interest rates will tend to go down in the future, which implies a decreasing term structure of interest rates today. However, this approach is cannot be fruitful because it is based on the implicit assumption that future short term rates tends to go down with time, which is counterfactual. In the next section, we consider another road to justify the downward sloping term structure which is in line with what we have done in this second part of the book.

A concept of concordance: "large values of x_1 go with large values of x_2"

This section is devoted to the analysis of the impact on the forward rate of serial correlation in the growth rate of the economy. The forward rate is characterized by the following equality:

$$R_{12} = \delta - \ln \frac{Eu'(c_0 e^{x_0+x_1})}{Eu'(c_0 e^{x_0})}. \tag{8.20}$$

Thus, the serial correlation in the growth of log consumption matters, as illustrated in the previous chapters. In the special case without serial correlation and constant relative risk aversion, we have that $R_{12} = r_{12} = r_1$, so that, according to condition (8.5), the term structure is flat. We hereafter relax the assumption of serial independence in a framework in which we do not specify a priori the utility function *u*.

In the general expected utility model, the coefficient of correlation is usually notably insufficient to characterize the role of statistical relationship on the term structure. The full joint distribution function is necessary to determine the forward discount rate. Following Tchen (1980) and Epstein and Tanny (1980), we want to formalize the idea that "greater values of x_1 go with greater values of x_2. To do this, consider an initial distribution function F for the pair of random variables (x_1, x_2), with $F(t_1, t_2) = \mathrm{P}[x_1 \le t_1 \cap x_2 \le t_2]$. Consider another pair of random variables $(\hat{x}_1, \hat{x}_2)$ with cdf $\hat{F}$. We define a "marginal-preserving increase in concordance" (MPIC) any transformation of distribution F into distribution $\hat{F}$ that takes the following form: Consider two pairs (t_1, t_2) and (t_1', t_2') such that $t_1' > t_1$ and $t_2' > t_2$. $\hat{F}$ is obtained from F by adding probability mass ε in a small neighbourhood of (t_1, t_2) and (t_1', t_2'), while substracting probability mass ε in a small neighbourhood of (t_1, t_2') and (t_1', t_2'), as depicted in the following Figure.

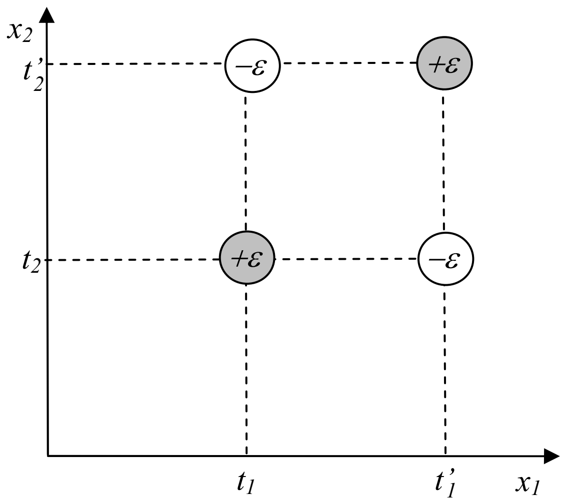


Figure: Transfer of probability mass in a marginal-preserving increase in concordance

This MPIC clearly increases the correlation between the two random variables, without affecting the marginal distributions of the two random variables. Observe also that the new cdf $\hat{F}$ obtained through a MPIC raises the cdf: for all (t_1,t_2), $\hat{F}(t_1,t_2) \geq F(t_1,t_2)$. Following Tchen (1980), this inequality defines the notion of “more concordance” for any two cdfs F and $\hat{F}$ with the same marginals $\hat{F}(t_1,+\infty) = F(t_1,+\infty) \quad \forall t_1 \in \mathbb{R}$, and $\hat{F}(+\infty,t_2) = F(+\infty,t_2)$ $\forall t_2 \in \mathbb{R}$:

$$\hat{F} \succ_c F \quad \Leftrightarrow \quad \forall (t_1,t_2) \in \mathbb{R}^2, \hat{F}(t_1,t_2) \geq F(t_1,t_2). \tag{8.21}$$

A more concordant cdf concentrates more probability mass in any South-East quadrangle of $\mathbb{R}^2$. Tchen (1980, Theorem 1) and Epstein and Tanny (1980) show that two cdfs with the same marginals can be ranked by this notion of increase in concordance, the more concordant cdf can be obtained from the less concordant one through a sequence of MPICs. It is interesting to observe that, by dividing both sides of the inequality in (8.21) by $\hat{F}(t_1,+\infty) = F(t_1,+\infty)$, this definition is equivalent to

$$\hat{F} \succ_c F \quad \Leftrightarrow \quad \forall (t_1,t_2) \in \mathbb{R}^2, P[\hat{x}_2 \leq t_2 \,|\, \hat{x}_1 \leq t_1] \geq P[x_2 \leq t_2 \,|\, x_1 \leq t_1]. \tag{8.22}$$

This is in turn equivalent to the following definition of an increase in concordance, which relies on the notion of First-order Stochastic Dominance (FSD):

$$\hat{F} \succ_c F \quad \Leftrightarrow \quad \forall t_1 \in \mathbb{R},\ \hat{x}_2 \,|\, \hat{x}_1 \leq t_1 \text{ is FSD-dominated by } x_2 \,|\, x_1 \leq t_1. \tag{8.23}$$

This can be seen clearly in the above figure. Suppose that the MPIC represented in this figure is undertaken, and that the information is received that x_1 is smaller than some $t \in]t_1,t_1'[$. What remains visible to the left of t is the downward transfer of probability mass that happens in the neighbourhood of t_1, which is a FSD deterioration in the conditional distribution of x_2. Conditional to the fact that x_1 is smaller than any threshold t_1, the probability distribution of $\hat{x}_2$ is a deterioration of x_2 in the sense of FSD. This means that some probability mass of this conditional distribution is transferred from the high values of x_2 to the lower ones. Under the new distribution, there is always more probability mass in the left-tail of the distribution of $x_2 \,|\, x_1 \leq t_1$.

In words, this means that the present and the future changes in consumption are more correlated after a sequence of MPICs. Bad news in the first period is bad news for the second period’s distribution of consumption. In the statistical literature, this notion is referred to as the "stochastic increasing positive dependence", because x_2 is more likely to take on larger value when x_1 increases (see for example Joe (1997)). It is closely related to the notion of "positive quadrant dependence" proposed by Lehmann (1966).

Suppose that we are interested in the effect of an increase in concordance on the expectation of some function $h:\mathbb{R}^2 \to \mathbb{R}$. Let us first consider the effect of an elementary MPIC defined by pairs (t_1,t_2) and $(t_1^{'},t_2^{'})$ such that $t_1^{'} > t_1$ and $t_2^{'} > t_2$, as in the above figure. Obviously, this MPIC increases the expectation of *h* if and only if

$$h(t_1,t_2)+h(t_1^{'},t_2^{'}) \geq h(t_1^{'},t_2)+h(t_1\,,t_2^{'}). \tag{8.24}$$

Because the two pairs (t_1,t_2) and $(t_1^{'},t_2^{'})$ are arbitrary, this condition must hold for all such pairs such that $t_1^{'} > t_1$ and $t_2^{'} > t_2$. Because any increase in concordance can be expressed as a sequence of MPICs, this condition is necessary and sufficient for an increase in concordance to raise the expectation of *h*. It happens that this condition is well-known in mathematical economics. It is referred to as the supermodularity of *h*.

If *h* represents a von Neumann-Morgenstern utility function in $\mathbb{R}^2$, condition (8.24) in which both sides of the inequality would be divided by 2 states that one would prefer a lottery yielding payoff (t_1,t_2) or $(t_1^{'},t_2^{'})$ with equal probabilities to another lottery yielding payoff $(t_1^{'},t_2)$ and $(t_1,t_2^{'})$ with equal probabilities. This would be the case for example for complement goods where x_1 and x_2 would be respectively the number of left and right shoes in the bundle. Condition (8.24) thus defines a notion of complementarity between x_1 and x_2. Two goods are complement if the marginal utility of the first is increasing in the second, i.e., if the cross derivative of the utility function is positive.

Observe that if *h* is twice differentiable, replacing $(t_1^{'},t_2^{'})$ by (t_1+dx, t_2+dy), inequality (8.24) is equivalent to

$$h_{12}(t_1,t_2)dxdy \geq 0 \tag{8.25}$$

for all $dx > 0$ and $dy > 0$. A simple integration argument implies that when *h* is twice differentiable, the supermodularity of *h* is equivalent to the positiveness of its cross derivative.

Let us summarize what we obtained in the following Lemma, whose formal proof is in Tchen (1980), or Epstein and Tanny (1980).

Lemma 2: *Consider a bivariate function h. The following conditions are equivalent:*

- *For any two pairs of random variables* (x_1, x_2) *and* $(\hat{x}_1, \hat{x}_2)$ *such that* $(\hat{x}_1, \hat{x}_2)$ *is more concordant than* (x_1, x_2), $Eh(\hat{x}_1, \hat{x}_2) \geq Eh(x_1, x_2)$.
- *h is supermodular.*

Moreover, assuming that *h* is twice differentiable, Tchen (1980, Theorem 2) shows that

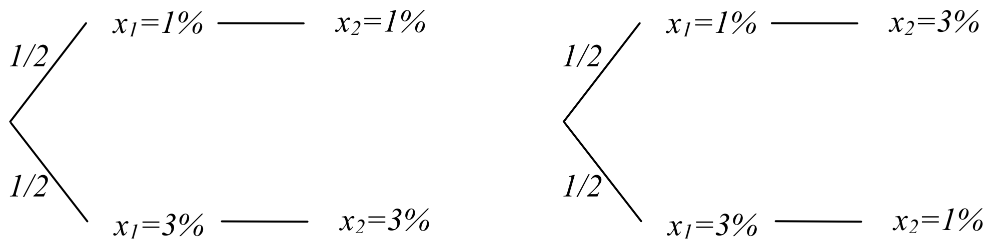
$$Eh(\hat{x}_1, \hat{x}_2) - Eh(x_1, x_2) = \int\int h_{12}(t_1, t_2)\left[\hat{F}(t_1, t_2) - F(t_1, t_2)\right]dt_1 dt_2. \tag{8.26}$$

This can be obtained by a double integration by parts. By the definition (8.21) of an increase in concordance, we see that equation (8.26) provides a simple proof to the above Lemma.

An immediate application of this Lemma is obtained by considering function $h(x_1, x_2) = x_1 x_2$, which is supermodular. Lemma 2 tells us that an increase in concordance raises the expectation of *h*. Because $E\hat{x}_i = Ex_i$ since the marginal distributions are preserved, this shows that an increase in concordance necessarily raises the covariance between the two random variables.

The effect of an increase in concordance of economic growth on the forward discount rate

The link between the notions of supermodularity and of an increase in concordance is clear. Consider two dynamic processes for the growth of consumption:



(a) Perfect positive concordance (b) Perfect negative concordance

The perfect positive concordant pair of random variables in (a) is obtained from the perfect negative concordant pair in (b) through a MPIC transferring all the probability mass from the upward diagonal of the rectangle to the downward one. In the two cases, the marginal distributions of x_1 and x_2 are the same: $x_t \sim (1\%,1/2;3\%,1/2)$, but they are perfectly positively correlated in case (a), whereas they are perfectly negatively correlated in case (b).

Define

$$h(x_1,x_2) = u'(c_0 e^{x_0+x_1}) \tag{8.27}$$

Equation (8.20) tells us that the forward discount rate R_{12} is negatively affected by an increase in concordance if *Eh* is positively affected by it. Using Lemma 2, this requires that *h* be supermodular. We have that

$$h_{12}(x_1,x_2) = c_2 u''(c_2)\left[1-P(c_2)\right], \tag{8.28}$$

where $c_2 = c_0 \exp(x_1 + x_2)$ is consumption at date *t=2* and $P(c) = -cu'''(c)/u''(c)$ is the index of relative prudence. This proves the following proposition.

Proposition: *Any increase in correspondence in the growth of log consumption reduces the forward discount rate if and only if relative prudence is uniformly larger than unity.*

By equation (8.5), $P \geq 1$ is also necessary and sufficient to reduce the long discount rate. Now, remember that combining the assumption of i.i.d. (x_1,x_2) with the constancy of relative risk aversion implies a flat term structure. Remember also that constant relative risk aversion implies that relative prudence is also constant and is equal to relative risk aversion plus one. Thus, when relative risk aversion is constant, it must be that relative prudence is larger than unity. When relative risk aversion is constant, the term structure of discount rates is decreasing if the growth process exhibits more concordance than in the the case of independence and the marginal cdf are the same.

The intuition of this result is based on the observation that the second moment of c_2 is supermodular in (x_1,x_2). Indeed, function

$$h(x_1,x_2) = \left(c_0 e^{x_1+x_2}\right)^2 \tag{8.29}$$

is supermodular. It implies that the increase in concordance in the change in log consumption tends to raise the variance of c_2. This reduces the forward discount rate under prudence. However, observe also that the expectation of c_2 is increased by the positive FSD dependence in (x_1,x_2), since $h(x_1,x_2) = c_0 \exp(x_1 + x_2)$ is supermodular. This wealth effect goes against the precautionary effect. This explains why positive prudence is not sufficient to sign the effect of an increase in concordance of *log consumption*. Using the above Lemma, it is easy to check that positive prudence is necessary and sufficient when the dynamic process of *consumption* exhibits more concordance than in the case of independence.

Unified explanation for a decreasing term structure of discount rates

The stochastic processes that we examined in chapters 4 (mean-reversion), 5 (Markov switches) and 6 (parametric uncertainty) exhibited some forms of stochastic dependence in serial changes of log consumption. Their common feature is the increased concordance of successive changes in log consumption compared to the case of a random walk, thereby providing a common ground for a decreasing term structure. The simplest illustration of this is obtained in the case of Markov switches. Suppose that there are two regimes, one with a sure growth rate of 2%, and one with a sure growth rate of 0%. There is a 1% probability to switch from one regime to the other every year. The Figure below on the left describes the probability distribution for the growth rate in the first two years, assuming that one experienced a good state in the previous year. The Figure on the right describes the probability distribution that yields no serial correlation, and with the same marginal probabilities as in the original distribution on the left. We see that the Markov-switch process to the left is more concordant than in the right case of independence, since it is obtained from the latter through a MPIC of a probability mass of *0.97%*.

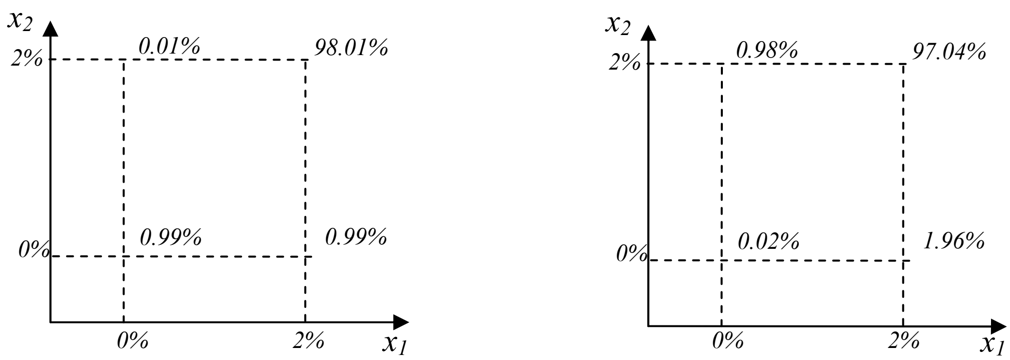


Figure: A two-state Markov process (left) that is more concordant than in the case of independence (right). The switching probability in each period is 1%.

Consider alternatively the mean-reverting process $x_{t+1} = \phi x_t + (1-\phi)\mu + \varepsilon_t$, with $\phi \in [0,1]$ and where ε_t is normally distributed with mean 0 and volatility σ. We have seen in chapter 4 that this yields a decreasing term structure under CRRA when $x_0 = \mu$, which guarantees that $Ex_1 = Ex_2$. In the Figure below, we depicted the iso-density curves of (x_1, x_2) together with the ones of the pair of independent random variables with the same marginal distributions ($x_1 \sim N(\mu, \sigma^2)$ and $x_2 \sim N(\mu, (1+\phi^2)\sigma^2)$). We clearly see that the pair exhibiting mean-reversion exhibits more concordance than the corresponding independent pair. A similar observation can be made for the case of parametric uncertainty.

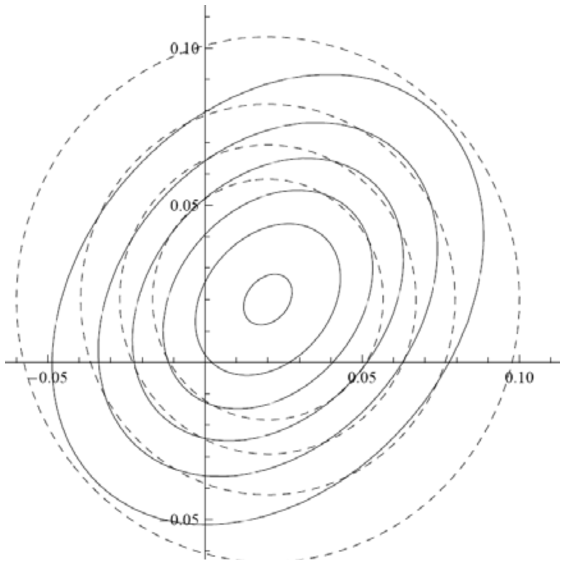


Figure: Iso-density curves in the case of mean-reversion with *μ=2%*, *σ=3.6%* and *ϕ=0.3*. The dashed curves correpond to the iso-density curves of the pair of random variables with the same marginal distributions.

Conclusion

This chapter was devoted to more technical analyses of the term structure of discount rates. We have developed a theory of this term structure based on concepts of stochastic dominance. In the benchmark case of a random walk for changes in log consumption, the growth in the first period yields no information about the growth in subsequent periods. Under constant relative risk aversion, this typically yields a flat term structure. We considered alternatively the case in which a larger growth rate in the first period improves the distribution of the growth rate in the second period in the sense of first-degree stochastic dominance. We have shown that most stochastic processes that we examined in the second part of this book exhibit this property. This positive statistical dependence in the growth process raises the uncertainty about consumption in the distant future, thereby reducing the long discount rate under prudence. Because it also increases the expected future consumption, this is formally true only if relative prudence is larger than unity rather than zero.

We also explored the possibility to depart from constant relative risk aversion. When relative risk aversion is decreasing, the wealth effect tends to fade away in growing economy, thereby reducing the forward discount rate. This tends to favour a downward-sloping term structure. This may explain the relatively important intensity of short-termism in public investments observed in developing countries, in which their citizens are close to their minimum level of subsistence.

References

Cox, J., Ingersoll, J., and S. Ross, (1985), A theory of the term structure of interest rates, *Econometrica*, 53, 385-403.

Epstein, L.G. and S.M. Tanny, (1980), Increasing Generalized Correlation: A Definition and Some Economic Consequences, *Canadian Journal of Economics*, 13, 16-34.

Gollier, C., (2002a), Discounting an uncertain future, *Journal of Public Economics*, 85, 149-166.

Gollier, C., (2002b), Time horizon and the discount rate, *Journal of Economic Theory*, 107, 463-473.

Joe, H., (1997), *Multivariate models and dependence concepts*, Chapman and Hall/CRC.

Lehmann, E.L., (1966), Some concepts of dependence, *Annals of Mathematical Statistics*, 37, 1153-1173.

Tchen, A.H., (1980), Inequalities for distributions with given marginals, *The Annals of Probability*, 8, 814-827.

Vasicek, 0., (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-188.

PART III

Extensions

Inequalities

In the canonical models of the term structure presented earlier in this book, we assumed that a single agent benefits from the cash flow of the project in each period. Another way to interprete this model is that there are several people that share equally both the GDP of the economy and the cash flow of the project. The real world is quite different from this model. In particular, our societies are unequal, and people are unequally affected by macroeconomic shocks. Moreover, costs and benefits of most public policies are not spread equally among various categories of citizens. To illustrate, consider our effort to curb our emissions of greenhouse gas. It is likely that most of these efforts will be borned by the western world, whereas the biggest beneficiaries will be the populations most vulnerable to climate change, many of them being in the third world, as predicted by various integrated models. Thus, fighting climate change in this way has some additional value coming from its virtue to reduce wealth inequality around the world. But even if one abstracts oneself from this heterogeneous allocation of costs and benefits, the existence of huge wealth inequalities across different countries and within them necessitates an adaptation of the canonical model which does not recognize this crucial feature of our world. This is the aim of this chapter.

Two models will be considered in this chapter. In the first model, we recognize that people are borne and live unequal in our society, but we also assume that they are able to share risk efficiently, and that they can implement mutually improving long term credit contracts. In the second model, we relax these assumptions.

Description of the economy

Suppose that the economy is composed of inifitely living N agents. They can be interpretated as family dynasties, or countries. They are indexed by $i=1, 2,...,N$. To keep the model simple, we assume that they have the same preference, which is classically represented by the rate of pure preference for the present δ,[3] and the increasing and concave utility function u. We first focus the analysis on the rate to be used at date 0 to discount a sure cash flow at date t. At date

[3] Gollier and Zeckhauser (2005) examine the effect of heterogeneous rates of impatience.

0, there is some inequality of the endowment $(z_{10},...,z_{N0})$, where z_{i0} is agent *i*'s endowment of the single consumption good at that date. One does not know at date *0* the distribution of the endowment occuring at date *t*. This uncertainty is characterized by *S* possible states of nature, *s=1, 2,..., S*, and by the associated state probabilities $(p_1,...,p_S)$, with $\Sigma_s p_s = 1$. Let z_{is} denote the endowment of agent *i* at date *t* in state *s*. Observe that *s=0* designates date *0* rather than a possible state to occur at date *t*. We define the income per capita in state *s* (or in date *0*) as

$$z_s = \frac{1}{N}\sum_{i=1}^{N} z_{is}. \tag{9.1}$$

Let us first assume that there exists in date *0* a complete market of insurance and credit contracts. In other words, we hereafter assume that for each *s=1,...,S,* there exists a contract for the delivery of one unit of the consumption good at date *t* if and only if state *s* is realized. Moreover, there exists a competitive market for each of these so-called Arrow-Debreu securities. An Arrow-Debreu security can be interpreted as an insurance contract, in which an indemnity is paid by the counterpart of the contract if a specific event occurs. Any risky asset can be interpreted as a bundle of Arrow-Debreu securities. A special case is the risk free asset, which is characterized as a bundle containing exactly one unit of each of the Arrow-Debreu securities. Let Π_s denote the equilibrium price of the Arrow-Debreu security associated to state *s*. It is useful at this stage to also define the state price per unit of probability $\pi_s = \Pi_s / p_s$, and $\pi_0 = \Pi_0$.

A competitive equilibrium is characterized by the vector $(\Pi_0,...,\Pi_S)$ of Arrow-Debreu securities at date *0*, and by a matrix (c_{is}), *i=1,...,N, s=0,1,...,S*, of actual consumption levels in the economy. Observe that $c_{is} - z_{is}$ is the demand for the Arrow-Debreu security *s* by agent *i.* The equilibrium must satisfy two sets of conditions:

- Each agent maximizes his welfare under the intertemporal budget constraint: $\forall i = 1,...,N$:

$$\max_{c_{is}} u(c_{i0}) + e^{-\delta t}\sum_{s=1}^{S} p_s u(c_{is}) \quad s.t. \quad \Pi_0\left(c_{i0} - z_{i0}\right) + \sum_{s=1}^{S}\Pi_s\left(c_{is} - z_{is}\right) = 0. \tag{9.2}$$

- Markets clear: $\forall s = 0,1,...,S$:

$$\sum_{i=1}^{N}\left(c_{is}-z_{is}\right)=0. \tag{9.3}$$

Observe that this condition can be rewritten as

$$\frac{1}{N}\sum_{i=1}^{N}c_{is}=z_s, \tag{9.4}$$

which is a feasibility condition.

Of course, if agents have all the same preferences and the same endowments ($z_{is}=z_s$ for all *s=0,1,…,S*), no trade must be an equilibrium, and the canonical model described earlier in this book applies. If the endowment is unequally allocated at date *0* or in some states in date *1*, some additional work is required to define a "representative agent" in this economy.

Existence of a representative agent

The first-order condition associated to program (9.2) can be written as

$$\begin{cases} u'(c_{i0})=\lambda_i\pi_0 \\ u'(c_{is})=\lambda_i e^{-\delta t}\pi_s, \quad s=1,...,S, \end{cases} \tag{9.5}$$

where λ_i is the lagrangian multiplier associated to agent *i*'s budget constraint. The competitive equilibrium is the solution of this set of *N(S+1)* first-order conditions (9.5) combined with the *S+1* market-clearing conditions (9.4). Standard theorems from General Equilibrium Theory can be used to prove the existence and the unicity (up to a normalization of the vector of prices) of the competitive equilibrium, and to prove that it is Pareto-efficient.

An important property of the competitive equilibrium is the so-called mutuality principle, which states that if there are two states at date *t*, say *s=a* and *s=b*, such that the wealth per capita are the same, i.e., $z_a=z_b$, then all agents will enjoy the same consumption level in the two states, i.e., $c_{ia}=c_{ib}$ for all *i=1,…,N*. It also implies that the two states per unit of probability must be the same, i.e., $\pi_a=\pi_b$. The simplest way to prove this is to check that the set of equations corresponding to the two states are equivalent. More intuitively, the mutuality principle states that all diversifiable risks are diversified at equilibrium. Suppose for example that there are only two states, and that the wealth levels per capita are the same in the two states. This later assumption means that there is no aggregate risk in the economy. This must imply that all agents are fully insured at equilibrium. Departing from this rule would force people to face zero-mean risks, which is a Pareto-inferior allocation by risk aversion.

The mutuality principle means that state-dependent variables c_{is} and π_s depend upon the state only through the wealth level per capita z_s: There exist functions C_i and v' such that $c_{is} = C_i(z_s)$ and $\pi_s = v'(z_s)$ for all *s=1,...,S.* Equation (9.5) can thus be rewritten as

$$\frac{u'(c_{is})}{u'(c_{is'})} = \frac{v'(z_s)}{v'(z_{s'})}, \tag{9.6}$$

for all *s* and *s'* in $\{1,...,S\}$, and for all *i.* As is well-known, the equilibrium is characterized by the equalization across all agents of their marginal rate of substitution of consumption in any pair of states. Equation (9.6) tells us that the equilibrium marginal rate of substitution is the same as in an economy in which all agents would consume the income per capita z_s, but the utility function u would be replaced by function v when computing the ratio of marginal utility.

Suppose without loss of generality that there exists a state *s'* such that $z_0 = z_{s'}$. Equation (9.5) implies that $c_{i0} = c_{is'} = C_i(z_0)$ for all *i,* and $\pi_0 = e^{-\delta t}\pi_{s'} = e^{-\delta t} v'(z_0)$. Thus, we also have that

$$\frac{u'(c_{is})}{u'(c_{i0})} = e^{\delta t}\frac{v'(z_s)}{v'(z_0)}, \tag{9.7}$$

for all s in $\{1,...,S\}$, and for all *i.* At equilibrium, the marginal rates of substitution between consumption at date *0* and in any specific state at date *t* are equalized across agents. They are equal to the one of an agent consuming the income per capita at date *0* and in any state at date *t*, but where the original utility function *u* would be replaced by function *v*. This function is hereafter refered to as the utility function of the representative agent who consumes the income per capita in all states and at all dates. An egalitarian economy composed by *N* identical agents with this utility function *v* would price all assets in this economy in exactly the same way as in the unequal economy described in the previous section. We have shown in this section that the existence of a complete set of competitive markets for Arrow-Debreu securities implies the existence of such representative agent, as initially shown by Wilson (1968). In the next section, we characterize the preferences of the representative agent.

Characterization of the representative agent

We have seen in the previous section that the utility function v of the representative agent can be derived from the original utility function by solving the following set of equalities:

$$u'(C_i(z)) = \lambda_i v'(z) \quad i = 1,...,N,$$
$$\frac{1}{N}\sum_{i=1}^{N} C_i(z) = z \tag{9.8}$$

for all z. Notice that this set of equations characterizes the solution of the following cake-sharing problem:

$$v(z) = \max_{(C_1,...,C_N)} \frac{1}{N}\sum_{i=1}^{N} \lambda_i^{-1} u(C_i) \quad \text{s.t.} \quad \frac{1}{N}\sum_{i=1}^{N} C_i = z. \tag{9.9}$$

The competitive allocation of risk maximizes the social welfare in each state of nature, where the social welfare function is the sum of individual utilities weighted by λ_i^{-1}.

The unequal distribution of wealth in the economy is entirely concentrated in the vector of lagrange multipliers $(\lambda_1,...,\lambda_N)$. If all agents have the same market value of their endowment, the λ_i would be all the same, thereby trivially yielding solution $v \equiv u$ and $C_i(z) = z$ for all z. Suppose alternatively that the market values of the individual endowment are unequal, so that the lagrange multipliers are heterogeneous. Fully differentiating the above equations with respect to z yields

$$u''(C_i(z))\frac{dC_i}{dz} = \lambda_i v''(z) \quad i = 1,...,N,$$
$$\frac{1}{N}\sum_{i=1}^{N} \frac{dC_i}{dz} = 1 \tag{9.10}$$

Let $T(c) = -u'(c)/u''(c)$ and $T_v(z) = -v'(z)/v''(z)$ denote the degree of absolute risk tolerance of the utility function of respectively the original agent and of the representative agent. The first equality in (9.10) can be rewritten as

$$\frac{dC_i}{dz} = \frac{T(C_i(z))}{T_v(z)} \quad i = 1,...,N. \tag{9.11}$$

This formula is intuitive. It states that the share of the aggregate risk borne by agent i – which is measured by the sensitiveness of his own consumption to income per capita -- is proportional to his degree of absolute risk tolerance. Using the second equality in (9.10) implies that we must have

$$T_v(z) = \frac{1}{N}\sum_{i=1}^{N} T(C_i(z)). \tag{9.12}$$

This equation first derived by Wilson (1968) tells us that the degree of risk tolerance of the representative agent is the mean of the absolute risk tolerance of the original agents evaluated at their actual level of consumption. This equation fully characterizes the utility function v of the representative agent in this inequal economy. Once v is obtained, one can determine the socially efficient discount rate by using the standard pricing formula in the canonical model:

$$r_t = \delta - \frac{1}{t}\ln\frac{Ev'(z_t)}{v'(z_0)}, \tag{9.13}$$

where z_t is the random variable which is distributed as $(z_1, p_1;...;z_S, p_S)$. It is obtained as usual by considering a marginal investment project in which the income per capita at date *0* is reduced by ε to increase the income per capita in all states at date *t* by $\varepsilon\exp(r_t t)$. The r_t defined in (9.13) is the one for which this investment project has no effect on the intertemporal social welfare $v(z_0) + e^{-\delta t}Ev(z_t)$ at the margin. We assume that benefits and costs are added and substracted to aggregate wealth, and are reallocated in the population according to the cake-sharing rule derived from program (9.9) and described by rule (9.11). In other words, this means that markets for Arrow-Debreu securities remain active after the investment decision is made.

The impact of wealth inequality on the efficient discount rate

In order to explore the effect of wealth inequality on the efficient discount rate, let us first examine the special case of an economy in which agents have the same classical power utility function with $u'(c) = c^{-\gamma}$. This implies that $T(c) = c/\gamma$. This implies in turn that

$$T_v(z) = \frac{1}{N}\sum_{i=1}^{N} T(C_i(z)) = \frac{1}{N}\sum_{i=1}^{N}\frac{C_i(z)}{\gamma} = \frac{z}{\gamma}, \tag{9.14}$$

for all *z*. It implies that the utility function of the representative agent is also a power function, with the same constant relative risk aversion than *u*. This proves that, under this specification, wealth inequality has absolutely no effect on the shape of the utility function of the representative agent, and therefore on the efficient discount rate. Because this utility function is widely used by economists, we can conclude that the sole presence of (large) wealth inequalities around the world is not enough to justify a departure from the extended Ramsey rule, which itself relies on a power utility function.

More generally, if the utility function *u* exhibits linear risk tolerance, the representative agent will have the same utility function *u*, whatever the degree of wealth inequality in the economy. On the contrary, if the utility function *u* exhibits a convex risk tolerance *T*, the Jensen's inequality implies that

$$T_v(z) = \frac{1}{N}\sum_{i=1}^{N} T(C_i(z)) \geq T\left(\frac{1}{N}\sum_{i=1}^{N} C_i(z)\right) = T(z) \tag{9.15}$$

for all z. The opposite result holds if risk tolerance is concave. A simple application is obtained in the special case of a safe growth between dates *0* and t. Suppose that T is convex, so that $T_v(z)$ is larger than $T(z)$ for all z. This means that v is less concave than u in the sense of Arrow-Pratt, or that there exists an increasing and convex function ψ such that $v(z) = \psi(u(z))$ for all z. This implies in turn that if $z_t \geq z_0$,

$$r_t = \delta - \frac{1}{t}\ln\frac{v'(z_t)}{v'(z_0)} = \delta - \frac{1}{t}\ln\frac{\psi'(u(z_t))u'(z_t)}{\psi'(u(z_0))u'(z_0)} \leq \delta - \frac{1}{t}\ln\frac{u'(z_t)}{u'(z_0)}, \tag{9.16}$$

because $\psi'(u(z_t)) \geq \psi'(u(z_0))$. This means that if the sure growth of the economy is positive, and if risk tolerance is convex, then wealth inequality reduces the efficient discount rate. Assuming that economic growth is uncertain makes the problem quite complex, because it requires describing the degree of prudence of the representative agent in addition to his risk tolerance.[4]

Epitaph for long-term risk-sharing allocations

Up to now in this chapter, we assumed that agents can credibly commit to share risk efficiently over long time horizons. This assumption would fit quite well the real western world over time horizons corresponding to life expectancies, in which real people write insurance and credit contracts that can be certified by a judge. This view is however limited by the existence of transaction costs and asymmetric information that yields no-borrowing constraints faced by households. If time horizon t exceeds the lifetime of the current generation, these risk-sharing arrangements described earlier can only be implicit, which raises a commitment problem. An alternative view is that the agents described above are governments that commit their citizens to intergenerational risk sharing contracts. This is quite unrealistic. Even within the European Union, countries have a limited commitment to assist other countries in economic distress, as illustrated for example by the absence of solidarity within the EU during the great recession in 2008-2010, in particular in the face of the Greek financial distress in the spring of 2010.

[4] For more details, see Gollier (2001).

The potential social value of international risk sharing is enormous, in particular when we take a long term perspective. Imagine for a second a Marco Polo as a plenipotential ambassador of the western world going to China to sign a multisecular treaty of risk sharing with the eastern world, each party accepting to compensate financially the other in case of a durable divergence in their respective growth rates. Imagine for another second that we would be able today to create a global "Commonwealth" for a progressist mutual assistance scheme where, prospectively, unlucky countries would get positive transfers from the lucky ones over the next two centuries. In both cases, there exists a large set of mutually-improving risk-sharing contracts, but which are hardly implemented – even at the margin -- because the huge commitment and agency problems that they would generate.

This means that the model that we presented earlier in this chapter is irrealistic, in particular for the time horizons that correspond to the global investment projects that we have in mind. It is a useful benchmark however, since it is the classical model used in the modern theory of finance, which heavily relies on the existence of a representative agent.

The case of inefficient risk sharing

We hereafter take dynastic or country-specific dynamic processes of consumption as completely exogeneous. An extreme interpretation of this model is that there is no transfer at all among this community, and that each agent consumes at each date its exogeneous endowment of the consumption good. Let c_{it} denote the consumption of agent *i* at date *t*. The dynamic stochastic process of $(z_1,...,z_N)$ is not specified at this stage, but it may exhibit temporal and geographical correlations. Let us consider an investment project that allocates costs and benefits in a non-discriminatory way. More specifically, consider a safe investment project that reduces consumption of all agents by ε at date *0*, and that raises consumption of all agents by $\varepsilon\exp(r_t t)$ at date *t*. We look for the critical rate of internal return of this project that has no effect at the margin on the intertemporal social welfare. As before, we define it as the discounted sum of the flow of temporal welfare. The welfare at date *t* is arbitrarily defined as the sum of the individual felicities weighted by Pareto-weights $(q_1,...,q_N)$, with $\Sigma_i q_i = 1$. Thus, the objective function is defined as

$$W_0 = \sum_{t=0} e^{-\delta t} \sum_{i=1}^{N} q_i Eu(c_{it}). \tag{9.17}$$

Following the same road as in chapter 1, we get that the critical rate of internal return of the safe project is characterized by the following rule:

$$r_t = \delta - \frac{1}{t}\ln\frac{\sum_{i=1}^{N} q_i Eu'(c_{it})}{\sum_{i=1}^{N} q_i u'(c_{i0})}. \tag{9.18}$$

Consider the discount rate that should be used by agent *i* who would bear alone the full costs and benefits of the project:

$$r_{it} = \delta - \frac{1}{t}\ln\frac{Eu'(c_{it})}{u'(c_{i0})}. \tag{9.19}$$

Let us also define the date-*0* inequality-neutral Pareto-weights $(\hat{q}_1,...,\hat{q}_N)$ such that

$$\hat{q}_i = q_i \frac{u'(c_{i0})}{\sum_{j=1}^{N} q_j u'(c_{j0})}. \tag{9.20}$$

Using equation (9.18), it is then easy to check that the efficient rate r_t to discount an egalitarian cash-flow at date *t* is linked to the individual discount rates $(r_{1t},...,r_{Nt})$ as follows:

$$e^{-r_t t} = \sum_{i=1}^{N} \hat{q}_i e^{-r_{it} t}. \tag{9.21}$$

The efficient discount factor is a weighted mean of the individual discount factors. This is reminiscent of equation (6.4) that describes the term structure of efficient discount rates when there is a single representative agent with utility function *u*, but in which there is some uncertainty about the true stochastic process of the growth of consumption per capita. With probability $\hat{q}_i$, the true stochastic process will be c_{it}. In this model, there is thus a formal equivalence between the fact that different agents may face different destinies, and the fact that all agents face the same uncertain destiny. This equivalence is nothing else than an illustration of John Rawls' concept of the veil of ignorance. We can thus limit the analysis in this section by a copy-pasting of the results presented in Chapter 6. For example, for distant time horizons, the efficient discount rate will tends to the smallest individual long-term discount rate. The equivalence with the model of parametric uncertainty is perfect only when there is no inequality of consumption at date *0*, otherwise the Pareto-weights need to be biased.

To illustrate, consider a specification similar to what we examined in Chapter 6:

$$\begin{cases} c_{it+1} = c_{it} e^{x_{it}} \\ x_{i0}, x_{i1},... \;\; i.i.d. \sim N(\mu_i, \sigma_i) \quad \forall i = 1,...,N. \end{cases} \tag{9.22}$$

Under constant relative risk aversion γ, it implies that

$$r_{it} = r_i = \delta + \gamma\mu_i - 0.5\gamma^2\sigma_i^2. \tag{9.23}$$

Using equation (9.21), it implies that the efficient discount rate r_t is decreasing and tends to the smallest r_i when t tends to infinity. Moreover, it satisfies the following property:

$$\lim_{t\to 0} r_t = \sum_{i=1}^{N} \hat{q}_i r_i. \tag{9.24}$$

The short-term discount rate is the weighted mean of the individual discount rates. The intuition is as in the framework of parametric uncertainty. For the distant future, what really matter when evaluating the project is whether it can improve the welfare of the poorest agent. The true shape of the term structure depends upon the distorted Pareto-weights $\hat{q}_i$, which depends upon our ethical values $(q_1,...,q_N)$, the initial degree of inequality $(c_{10},...,c_{N0})$, and its correlation with the distribution of economic growth.

Economic convergence and the discount rate

In order to have a more precise description of the term structure, one needs to specify the degree of convergence of economic development. Let us first consider an economy without any convergence, i.e., in which the current level of development of a country is uninformative about its future economic growth. More precisely, suppose that $\log c_{i0}$ is independent of the distribution of the growth rate x_{it}. Let y_{it} be the cumulated growth of log consumption between *0* and *t*. Under constant relative risk aversion γ, this implies that

$$\begin{aligned}\frac{\sum_{i=1}^{N} q_i Eu'(c_{it})}{\sum_{i=1}^{N} q_i u'(c_{i0})} &= \frac{\sum_{i=1}^{N} q_i E\exp\left(-\gamma \ln c_{i0} - \gamma y_{it}\right)}{\sum_{i=1}^{N} q_i \exp(-\gamma \ln c_{i0})} \\ &= \frac{\left(\sum_{i=1}^{N} q_i \exp\left(-\gamma \ln c_{i0}\right)\right)\left(\sum_{i=1}^{N} q_i E\exp\left(-\gamma y_{it}\right)\right)}{\sum_{i=1}^{N} q_i \exp(-\gamma \ln c_{i0})}.\end{aligned} \tag{9.25}$$

The second equality is a direct consequence of the no-convergence hypothesis. This implies in turn that

$$r_t = \delta - \frac{1}{t}\ln \sum_{i=1}^{N} q_i E\exp\left(-\gamma y_{it}\right). \tag{9.26}$$

This means that the term structure of discount rates is independent of the initial distribution of wealth in this framework. Only the inequal expectation about future growth matters.

Let us now consider the case of economic convergence. We say that the global growth process at horizon t exhibits some degree of economic convergence if agents with a larger initial level of development face less optimistic growth prospects. Let random variables c_0 and y_t be distributed respectively as $(c_{10},q_0;...;c_{N0},q_N)$ and $(y_{1t},q_1;...;y_{Nt},q_t)$. We compare this global economy exhibiting some degree of economic convergence to another economy where c_0 and y_t are statistically independent, but with the same marginal distributions. Technically, following the same technique as in Chapter 8, we say that this economy exhibits economic convergence if (c_0,y_t) is less concordant than the corresponding economy without statistical dependence between c_0 and y_t. This alternative global economy exhibits no tendency of convergence, and its term structure is governed by equation (9.26) under CRRA.

Lemma 2 in chapter 8 is useful to evaluate the impact of economic convergence on the efficient discount rate. Observe that the numerator in the right-hand side of equation (9.18) can be expressed as

$$\sum_{i=1}^{N} q_i Eu'(c_{it}) = Eu'(c_0 e^{y_t}) = Eu'(\exp(\ln c_0 + y_0)). \tag{9.27}$$

Lemma 2 in chapter 8 tells us that the reduction in concordance of $\ln c_0$ and y_t, i.e., economic convergence, reduces this numerator if and only if $h(x_1,x_2)=u'(\exp(x_1+x_2))$ is supermodular. This is true if and only if relative prudence is uniformly larger than unity. Thus, we conclude that economic convergence raises the efficient discount rate if relative prudence is larger than unity. This is the case for example in the case of constant relative risk aversion. In that case, equation (9.26) underestimates the true discount rate at all time horizons. Symmetrically, economic divergence tends to reduce the discount rate.

To illustrate, consider a global economy with two countries. County *i=1* has a GDP/cap at date *0* that is normalized to unity. Country *i=2* has a GDP/cap at date *0* that is 50 times larger. Our ethical values impose $q_1=q_2=1/2$. Suppose that the two countries converge, with country *1* enjoying a constant growth rate of *3%*, whereas the growth rate of the wealthier country *2* is only *1%*. It implies that the two countries will have the same consumption level per capita in around 195 years. Consider alternatively an economy with the same $(c_{10},c_{20})=(1,50)$, but the same uncertain growth rate for the two countries, which will be either *1%* or *3%* with equal probabilities. In the figure below, we have drawn the term structure of efficient discount rates in these two economies. As expected, the two curves are decreasing, and the discount rates are larger when there is convergence.

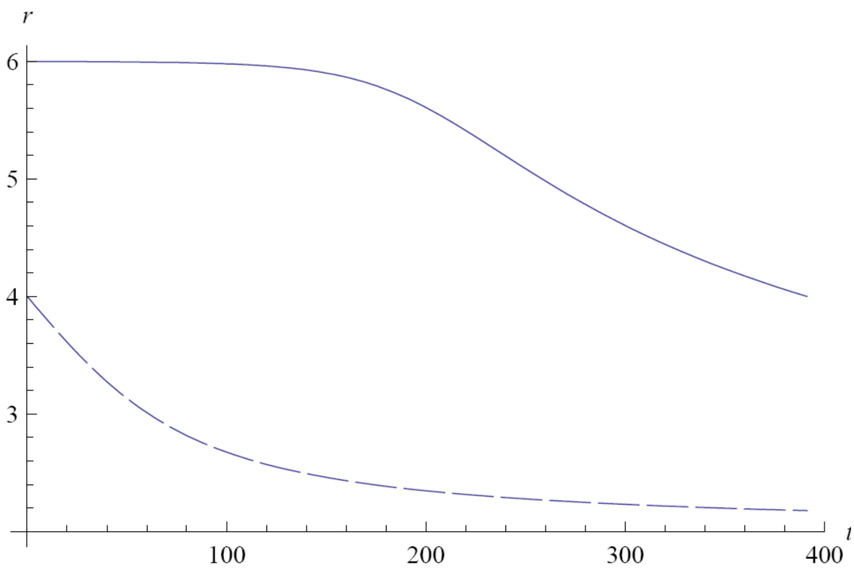


Figure: Term structure in a two-country model with $(c_{10}, c_{20}) = (1,50)$ and $(x_{1t}, x_{2t}) = (3\%, 1\%)$. We assume that $\gamma = 2$ and $\delta = 0\%$. The dashed curve corresponds to the case where the two countries face an uncertain constant growth rate of either *1%* or *3%* with equal probabilities.

What do we know about economic convergence? The classical economic theory of economic growth provides an argument for economic convergence, since decreasing marginal productivity of capital implies that wealthier countries should grow at a smaller rate. Furthermore, poorer countries can replicate succesful production methods, technologies and institutions implemented earlier by more developed countries. However, in spite of some the existence of some succesful newly developed countries as India, Singapur, Korea, China or Brazil, many poor countries seem to be permanently underdeveloped, whereas some others become ever poorer (Haïti, Zimbabwe). According to Clark (2007), the industrial revolution has reduced inequalities within societies, but it has increased them between societies, in a process labeled the Great Divergence following Pomeranz (2000).

Conclusion

This chapter has decribed two discounting models with wealth inequality. In the first model, we assume that our modern society has developed efficient risk-sharing schemes: insurance markets, markets for futures and options, social security. Under this view, there is no loss of generality to assume that there is a representative agent who aggregates the preferences towards risk and time into a single utility function. Wealth inequality is irrelevant for the determination of the term structure of discount rates, under CRRA.

If we alternatively recognize that risk-sharing schemes do not work efficiently, in particular towards risk occurring in the distant future, this can justify a decreasing term structure in a way similar to the case of parametric uncertainty. Moreover, the plausibility of economic convergence tends to raise the efficient discount rate.

References

Clark, G., (2007), *A farewell to alms: A brief history of the world*, Princeton University Press.

Gollier, C., (2001), Wealth inequality and asset pricing, *Review of Economic Studies*, 68, 181-203.

Gollier, C., and R.J. Zeckhauser (2005), Aggregation of heterogeneous time preferences, *Journal of Political Economy*, 113, 878-898.

Pomeranz, K., (2000), *The great divergence: China, Europe, and the making of the modern world economy*, Princeton University Press.

Wilson, R., (1968), "The theory of syndicates", *Econometrica* 36, 113-132.

Discounting non-monetary benefits

Humans extract utility from a large set of determinants, from the consumption of goods and services to the quality of their environment, their health, or of their social interactions. Earlier in this book, we simplified much the analysis by assuming that utility is derived from a univariate variable that we called consumption, or income. This approach relies on the well-known notion of the *indirect* utility function. This function characterizes the maximum utility that can be extracted from a given income through a clever selection among the bundles of these determinants that satisfy the budget constraint. This approach is often not satisfactory for our purpose for at least two reasons. First, many of these determinants are not tradable, such as various environmental assets for example. Second, this indirect utility function depends upon the vector of prices of the tradable goods and services. Because of the evolution of their relative scarcity, these prices fluctuate over time. So does the indirect utility function. Think for example of the relative price of oil, of land, of masterpieces of art, or more prosaically of the services of a plumber. When valuing a project that generates multidimensional impacts scattered over a long time span, it is crucial to take into account of these transformations of the indirect utility function, and of the relative value of these impacts.

The main economic justification of discounting is based on the wealth effect. If one believes that future generations will be wealthier than us, one more unit of consumption is more valuable to us than to them, under decreasing marginal utility of consumption. However, a large fraction of the impacts of our actions, for example towards climate change, affects the quality of the environment (increased temperature, reduced biodiversity, or destruction of environmental assets for example) rather than consumption. In this chapter, we address the question of how one should discount future changes in the quality of the environment. If we believe that the environment is deteriorating over time, and if we assume that the marginal utility of the quality of the environment is decreasing, then increasing the environmental quality is more valuable to future generations than to us. This argument, which is symmetric to the Ramsey's wealth effect, is in favour of using a smaller discount rate for changes in the environment than for changes in consumption. The full characterization of this "ecological" discount rate should also take into account of the substitutability between environmental assets and consumption, and of the uncertainty that affects the dynamics of consumption and of the environment. This chapter is based on Gollier (2010).

Two methods to evaluate future non-financial benefits

There are two possible methods to evaluate the present monetary value of a sure future environmental impact. The classical one consists in first measuring the future monetary value of the impact, and second discounting this monetary equivalent impact to the present. This involves a pricing formula to value future changes in environmental quality, and an economic discount rate to discount these monetarized impacts. As first suggested by Malinvaud (1953), the second approach consists in first discounting the future environmental impact to transform it into an immediate equivalent environmental impact, and then measuring the monetary value of this immediate impact. This involves an ecological discount rate, to discount environmental impacts. Of course, these two methods are strictly equivalent. As shown by Guesnerie (2004), Weikard and Zhu (2005) and Hoel and Sterner (2007) in the case of certainty, the two discount rates differ if the monetary value of environmental assets evolves over time.

The classical method is not well adapted to the case of uncertainty. Indeed, the value of environmental assets in the future depends upon their relative scarcity, which is unknown. This is a problem because the economic discount rate is useful to discount sure future monetary benefits. Because the monetary value of environmental impacts is uncertain, one needs to compute its certainty equivalent. This requires the use of a stochastic discount factor, which determines at the same time the risk premium and the economic discount rate. Standard pricing formulas exist that can be borrowed from the theory of finance, but they are seldom used in cost-benefit analyses of environmental projects because of their complexity. In this chapter, we describe in details the alternative methods based on the ecological discount rate. The ecological discount factor associated to date t is the immediate sure environmental impact that has the same impact on intergenerational welfare than a unit environmental impact at date t. The (shadow) price of an immediate environmental impact can then be used to value environmental projects. This alternative method is simpler because one does not need to compute certainty equivalent future values.

A simple model of the ecological discount rate

To keep the notation simple, we assume that the representative agent's felicity is affected by two determinants or "goods", available in quantities (x_{1t}, x_{2t}) at date *t*. It is notationally demanding but easy to extend this model to more than two dimensions. Determinant *1* is hereafter assumed to be an aggregated consumption good, whereas x_{2t} is an index of the quality of the environment, which includes the comfort generated from the climate, the services extracted from the biodiversity, the human morbidity due to various pollutions, or the life expectancy for example. The felicity at any date *t* is a function *U* of the available quantities (x_{1t}, x_{2t}) of the two goods. *U* is assumed to be increasing and concave. The intertemporal social welfare is measured by the discounted value of the flow of temporal expected felicity:

$$W = \sum_{t=0} e^{-\delta t} EU(x_{1t}, x_{2t}). \tag{10.1}$$

The expectation is linked to the fact that, seen from date *0*, the evolution of the availability of the consumption good and of the quality of the environment is uncertain.

We first examine the economic discount rate. Let us consider a simple marginal project that would reduce consumption by $\varepsilon \exp(-r_t t)$ today, and that would raise consumption by a sure amount ε at date *t*, leaving the environment unaffected by the action. The internal rate of return r_t that is such that implementing the project has no effect on *W* at the margin is called the *economic* discount rate, and is denoted r_{1t}:

$$r_{1t} = \delta - \frac{1}{t} \ln \frac{EU_1(x_{1t}, x_{2t})}{U_1(x_{10}, x_{20})}, \tag{10.2}$$

where $U_i(x_1, x_2)$ is the partial derivative of *U* with respect to x_i. This economic discount rate allows for the comparison of the value of different consumption increments at different dates.

Consider alternatively an investment project that increases the environmental quality by ε at date *t*. The standard way to include this environmental impact in the cost-benefit analysis would be to first express this impact in future monetary terms. The instantaneous value v_t of the environment at date *t* is measured by the marginal rate of substitution between consumption and the environment:

$$v_t = -\left.\frac{dx_{1t}}{dx_{2t}}\right|_U = \frac{U_2(x_{1t}, x_{2t})}{U_1(x_{1t}, x_{2t})}. \tag{10.3}$$

If the quality of the environment would be traded, v_t would be its equilibrium price, taking the aggregate consumption good as the numeraire. More generally, v_t is the instantaneous willingness to pay for improving environmental quality. Its evolution over time is uncertain, i.e., v_t is a random variable seen from $t=0$. So is the future monetary benefit εv_t of the sure improvement of the environmental quality. It implies that in spite of the fact that we are considering an investment project with a sure ecological benefit, its moneraty benefit is uncertain. Up to now in this book, we focused the analysis on the valuation of sure cash flows, and we leave to the next chapters the exploration of the valuation of uncertain projects.

A much simpler approach is obtained by defining an ecological discount rate. Consider a marginal project that would increase the environmental quality by a sure amount ε at date t, and that would reduce the environmental quality by $\varepsilon \exp(-r_t t)$ today. Implementing this project would be socially efficient if

$$r_t \geq r_{2t} = \delta - \frac{1}{t} \ln \frac{EU_2(x_{1t}, x_{2t})}{U_2(x_{10}, x_{20})}. \tag{10.4}$$

This equation defines the *ecological* discount rate r_{2t} associated to time horizon t. It allows us to compare sure changes in the environment quality at different dates. Namely, an increase in environmental quality by ε at date t has an effect on intertemporal welfare that is equivalent to an increase in current environmental quality by $\varepsilon \exp(-r_{2t} t)$. In monetary terms, this is equal to $v_0 \varepsilon \exp(-r_{2t} t)$, where v_0 is the current value of one unit of environmental quality.

To sum up, the benefit of a unit increment in environmental quality at date t should be accounted for in the evaluation of a project as equivalent to an immediate increase in consumption by $v_0 \exp(-r_{2t} t)$. This really means that environmental costs and benefits should be discounted at the ecological rate r_{2t}, which needs not to be the same as the economic discount rate r_{1t}. The potential discrepancy between the economic discount rate and the ecological discount rate takes into account the stochastic changes in the relative social valuation of the environment.

Determinants of the ecological discount rate

In this section, we examine the determinant of the rate r_{2t} at which one should discount a sure increase in environmental quality at date t, which is characterized by equation (10.4). Let us first focus on the role of the level and uncertainty surrounding x_{2t}. Because U is concave in

x_2, a larger future environmental quality raises the ecological discount rate, ceteris paribus. This effect is symmetric to the wealth effect presented in chapter 2. Because of the decreasing marginal utility of environmental quality, one is ready to sacrifice less today if the future quality of the environment is larger. We call this the ecological growth effect.

If we assume that U_2 is convex in x_2, then the uncertainty surrounding x_{2t} reduces the ecological discount rate. This effect, that we call the ecological prudence effect, is parallel to the precautionary effect for monetary cash flows described in chapter 3. The basic idea is that one should do more to improve the future quality of the environment if it is more uncertain.

One should also take into account of changes in the GDP/cap x_{1t} on the level of the ecological discount rate. Suppose for example that the two goods are substitutes, i.e. that the marginal utility of x_2 is decreasing in x_1. In other words, suppose that U_{12} is negative. Then, an increase in the GDP/cap in *t* reduces the marginal utility of environmental quality at that date. Therefore, it raises the ecological discount rate. We call this the substitution effect. One difficulty is to determine whether consumption and the environment are substitute ($U_{12} \leq 0$) or complement ($U_{12} \geq 0$). There is a simple way to solve this question. Consider an arbitrary situation characterized by (x_1, x_2), an arbitrary reduction in consumption $l_1 > 0$, and an arbitrary reduction in environmental quality $l_2 > 0$. Consider two lotteries. Lottery A is a fifty-fifty chance to face the monetary loss or the environmental loss. Lottery B is a fifty-fifty chance to face the two losses simultaneously, or to loose nothing. If one prefers A to B, it must be that U_{12} is negative. Indeed, it means that

$$\frac{1}{2}U(x_1 - l_1, x_2) + \frac{1}{2}U(x_1, x_2 - l_2) \geq \frac{1}{2}U(x_1 - l_1, x_2 - l_2) + \frac{1}{2}U(x_1, x_2), \tag{10.5}$$

which is equivalent to

$$\int_{x_2 - l_2}^{x_2} U_2(x_1 - l_1, y)dy \geq \int_{x_2 - l_2}^{x_2} U_2(x_1, y)dy. \tag{10.6}$$

This requires that U_{12} be negative. Eeckhoudt, Rey and Schlesinger (2007) call this notion "correlation aversion", which is another way to say that the two goods are substitutes.

A more complex problem is to evaluate the effect of the uncertainty affecting the economy on the ecological discount rate. Obviously, a zero-mean risk on x_{1t} raises $EU_2(x_{1t}, x_{2t})$ if U_2 is convex in its first argument. We call this effect the "cross-prudence in consumption" effect. In order to evaluate whether condition $U_{211} \geq 0$ is reasonable, let us follow Eeckhoudt, Rey and Schlesinger (2007) who used a multidimensional version of equation (3.10). Consider again an arbitrary initial situation (x_1, x_2), an arbitrary zero-mean risk in consumption ε_1, and an arbitrary reduction in environmental quality $l_2 > 0$. Consider two lotteries. Lottery A is a fifty-fifty chance to face the monetary risk or the environmental loss. Lottery B is a fifty-fifty chance to face the monetary risk and the environmental loss simultaneously, or to loose nothing. If one prefers A to B, it must be that U_{211} is positive. Indeed, this preference implies that

$$\frac{1}{2}EU(x_1+\varepsilon_1, x_2)+\frac{1}{2}U(x_1, x_2-l_2) \geq \frac{1}{2}EU(x_1+\varepsilon_1, x_2-l_2)+\frac{1}{2}U(x_1, x_2), \tag{10.7}$$

which is equivalent to

$$\int_{x_2-l_2}^{x_2} EU_2(x_1+\varepsilon_1, y)dy \geq \int_{x_2-l_2}^{x_2} U_2(x_1, y)dy. \tag{10.8}$$

This requires that U_2 be convex in x_1. Because the preference of lottery A over lottery B, it provides an economic justification to reduce the ecological discount rate when the economic growth becomes more uncertain.

Finally, the existence of a positive correlation between the economic growth and the improvement of the quality of the environment provides a last determinant of the ecological discount rate. As many readers may now anticipate, we formalize the positive statistical dependence of (x_{1t}, x_{2t}) through the notion of positive first-degree stochastic dependence, in which an increase in x_{1t} improves the distribution of $x_{2t} | x_{1t}$ in the sense of first-degree stochastic dominance. Using Lemma 2 of chapter 8, this positive correlation raises $EU_2(x_{1t}, x_{2t})$ if and only if U_2 is supermodular, i.e., if U_{221} is positive. By symmetry to the notion of cross-prudence in consumption, this means that the representative agent is cross-prudent in the environment. In a fifty-fifty lottery, he prefers to face a sure monetary loss or a zero-mean environmental risk in isolation than altogether in one state, and nothing in the other state. Under this assumption, the existence of a positive correlation in the economic and ecological growth rates raises $EU_2(x_{1t}, x_{2t})$, thereby reducing the efficient ecological discount rate. Intuitively, one wants to do more about the future when the economic and ecological risks are positively correlated than when they are independent. This is the correlation effect.

In this section, we assumed the sign of the cross-derivatives of the utility function to guarantee that the representative agent always prefer to incur one of the two harms for certain, with the only uncertainty being about which one will be received, as opposed to a 50-50 gamble of receiving the two harms simultaneously, or receiving neither. Following a terminology introduced by Eeckhoudt and Schlesinger (2006), pairs of harms are "mutually aggravating". Under this set of assumptions, it implies that the following factors raise the ecological discount rate:

- An increase in the future environmental quality;
- An increase in the future GDP/cap.

On the contrary, the following factors reduce it:

- An increase in the uncertainty affecting the future quality of the environment;
- An increase in the uncertainty affecting the future GDP/cap;
- An increase in the correlation in the two risks.

A symmetric analysis can be made for the determinants of the economic discount rate r_{1t}.

An analytical solution

The integral $EU_2(x_{1t}, x_{2t})$ has an analytical solution in the special case of a bivariate geometric Brownian motion for (x_{1t}, x_{2t}) and a Cobb-Douglas utility function. Suppose that

$$U(x_1, x_2) = kx_1^{1-\gamma_1} x_2^{1-\gamma_2} \tag{10.9}$$

in the domain $(x_1, x_2) \in \mathbb{R}_+^2$. We suppose that

$$k = sign(1-\gamma_1) = sign(1-\gamma_2) \tag{10.10}$$

in order to guarantee that U is increasing in its two arguments. The concavity of U with respect to its two arguments requires that γ_1 and γ_2 be positive. If they are both larger than unity, it is easy to check that this utility function satisfies the assumptions made in the previous section, i.e., pairs of harms are "mutually aggravating":

$$U_{12} < 0; U_{222} > 0; U_{112} > 0; U_{122} > 0; U_{111} > 0. \tag{10.11}$$

The two goods are substitutes, and the agent is (cross-)prudent in consumption and in the quality of the environment.

In the same fashion that the benchmark univariate model presented in chapter 3, let us assume that $(\ln x_{1t}, \ln x_{2t})$ be normaly distributed with mean $(\ln x_{10} + \mu_1 t, \ln x_{20} + \mu_2 t)$ and variance-covariance matrix $\Sigma = (\sigma_{ij} t)_{i,j=1,2}$. We have that

$$EU_2(x_{1t}, x_{2t}) = k(1-\gamma_2)E\exp(z_t), \tag{10.12}$$

where $z_t = (1-\gamma_1)x_{1t} - \gamma_2 x_{2t}$ is normally distributed with mean

$$Ez_t = (1-\gamma_1)\left(\log x_{10} + \mu_1 t\right) - \gamma_2\left(\log x_{20} + \mu_2 t\right), \tag{10.13}$$

and variance

$$Var(z_t) = \left((1-\gamma_1)^2\sigma_{11} + \gamma_2^2\sigma_{22} - 2(1-\gamma_1)\gamma_2\sigma_{12}\right)t. \tag{10.14}$$

Using Lemma 1 yields

$$\frac{EU_2(x_{1t}, x_{2t})}{U_2(x_{10}, x_{20})} = \exp\left((1-\gamma_1)\mu_1 - \gamma_2\mu_2 + 0.5\left((1-\gamma_1)^2\sigma_{11} + \gamma_2^2\sigma_{22} - 2(1-\gamma_1)\gamma_2\sigma_{12}\right)\right)t. \tag{10.15}$$

By equation (10.4), we obtain that

$$r_{2t} = \delta + \gamma_2\mu_2 - 0.5\gamma_2^2\sigma_{22} - (1-\gamma_1)\mu_1 - 0.5(1-\gamma_1)^2\sigma_{11} + (1-\gamma_1)\gamma_2\sigma_{12}. \tag{10.16}$$

Finally, let $g_i = t^{-1}\log(Ex_{it}/x_{i0}) = \mu_i + 0.5\sigma_{ii}$ be the growth rate of Ex_{it}. The above equation can thus be rewritten as

$$r_{2t} = \delta + \gamma_2 g_2 - 0.5\gamma_2(\gamma_2+1)\sigma_{22} + (\gamma_1-1)g_1 - 0.5\gamma_1(\gamma_1-1)\sigma_{11} - (\gamma_1-1)\gamma_2\sigma_{12}. \tag{10.17}$$

The term structure of the ecological discount rate is flat. In such an economy, the random evolution of aggregate consumption and of the environmental quality does not justify to use a smaller rate to discount benefits occurring in a more distant future.

We recognize in the right-hand side of equality (10.17) the 5 determinants of the ecological discount rate that we described in the previous section, in addition to the rate of pure preference for the present:

- $\gamma_2 g_2$ is the positive ecological growth effect, assuming an improving quality of the environment;
- $-0.5\gamma_2(\gamma_2+1)\sigma_{22}$ is the negative ecological prudence effect;
- $(\gamma_1-1)g_1$ is the positive substitution effect, assuming a growing economy;
- $-0.5\gamma_1(\gamma_1-1)\sigma_{11}$ is the negative cross-prudence in consumption effect;
- $-(\gamma_1-1)\gamma_2\sigma_{12}$ is the negative correlation effect, assuming a positive correlation between the economic and ecological growth rates.

Symmetrically, we can compute the economic discount rate:

$$r_{1t} = \delta + \gamma_1 g_1 - 0.5\gamma_1(\gamma_1+1)\sigma_{11} + (\gamma_2-1)g_2 - 0.5\gamma_2(\gamma_2-1)\sigma_{22} - (\gamma_2-1)\gamma_1\sigma_{12}. \tag{10.18}$$

We can also determine the difference between the two discount rates :

$$r_{2t} - r_{1t} = g_2 - g_1 + \left(\gamma_1\sigma_{11} - \gamma_2\sigma_{22}\right) + \left(\gamma_2 - \gamma_1\right)\sigma_{12}. \tag{10.19}$$

Interestingly enough, under certainty, the difference between the two discount rates is independent of the parameters of the Cobb-Douglas utility function. This equation provides two arguments in favor of using an ecological discount rate smaller than the economic discount rate. First, it is often suggested that the growth rate of environmental quality is smaller than the economic growth rate ($g_2 < g_1$), the first being potentially negative. Second, it seems that there is much more uncertainty surrounding the evolution of the environmental quality than the evolution of the economy itself ($\sigma_{22} > \sigma_{11}$). If the degrees aversion to risk on x_1 and on x_2 are not too heterogeneous, this would imply that $\left(\gamma_1\sigma_{11} - \gamma_2\sigma_{22}\right)$ be negative. The last term of the right-hand side of equation (10.19) is more difficult to sign.

A calibration exercice

Because of the lack of time-series data about environmental quality, calibrating the specification above is problematic. Various authors have argued in favor of a closer link between the environmental quality and economic growth. Following this idea, let us alternatively assume that the environmental quality is a deterministic function of economic achievement: $x_{2t} = f(x_{1t})$. Common wisdom suggests that the environmental quality is a decreasing function of GDP per capita, but this is heavily debated in scientific circles. The environmental Kuznets curve hypothesizes that the relationship between per capita income and the environmental quality has an inverted U-shaped, but there is no consensus about it (see for example Millimet, List and Stengos (2003)). We hereafter hypothesize a monotone relationship by assuming that there exists $\rho \in \mathbb{R}$ such that $x_{2t} = x_{1t}^{\rho}$, where ρ can be either positive or negative. If we assume that x_{1t} follows a geometric brownian motion, x_{2t} also follows a geometric Brownian motion, so that we obtain an analytical solution for the discount rates. Using the standard trick of Lemma 1, we obtain

$$r_{2t} = \delta + \left(\rho\gamma_2 + \gamma_1 - 1\right)\left(g_1 - 0.5(\rho\gamma_2 + \gamma_1)\sigma_{11}\right), \qquad (10.20)$$

and

$$r_{1t} = \delta + \left(\gamma_1 + \rho(\gamma_2 - 1)\right)\left(g_1 - 0.5(1 + \gamma_1 + \rho(\gamma_2 - 1))\sigma_{11}\right). \qquad (10.21)$$

The interested reader can recover from these equations the different determinants of these two rates that we exhibited earlier in this chapter.

In order to calibrate this model, let us assume that the rate of pure preference for the present δ is zero. We also assume as before that the relative aversion to risk on consumption is a constant $\gamma_1 = 2$. The parameter γ_2 of aversion to environmental risk is not easy to calibrate. Observe however that

$$\gamma^* = \frac{\gamma_2 - 1}{\gamma_1 + \gamma_2 - 2} \tag{10.22}$$

is the share of total consumption expenditures that the representative agent would use on environmental quality if environmental quality would be a tradable good. Hoel and Sterner (2007) and Sterner and Persson (2008) suggested γ^* somewhere *10%* and *50%*, which implies that γ_2 should be somewhere between *1.1* and *2* under our specification. We hereafter assume $\gamma^* = 0.3$, which implies that $\gamma_2 = 1.4$. Suppose also that $g_1 = 2\%$, and $\sqrt{\sigma_{11}} = 3.6\%$.

The last parameter to calibrate is the elasticity ρ of the environmental quality to changes in GDP/cap. It depends upon how we define the environmental quality. In order to estimate ρ, we considered the SYS_LAN indicator contained in the Environmental Sustainability Index (ESI2005, Yale Center for Environmental Law and Policy, (2005)), which measures for 146 countries in 2005 the percentage of total land area (including inland waters) having very high anthropogenic impact. Let x_1 be the 2005 GDP/cap from the World Economic Outlook Database of IMF (April 2008), and x_2 be defined as $3 + SYS\_LAN$ from ESI2005. In the figure below, we have represented this database and the associated OLS regression line which is

$$\ln x_2 = 1.93 - 0.10 \ln x_1 + \varepsilon \tag{10.23}$$

The *p*-value for the slope-coefficient is *-4.69*, whereas the $R2$ coefficient equals *0.13*. Plugging $\rho = -0.10$ in equations (10.20) and (10.21) yields

$$r_{2t} = 1.6\% \text{ and } r_{1t} = 3.5\%. \tag{10.24}$$

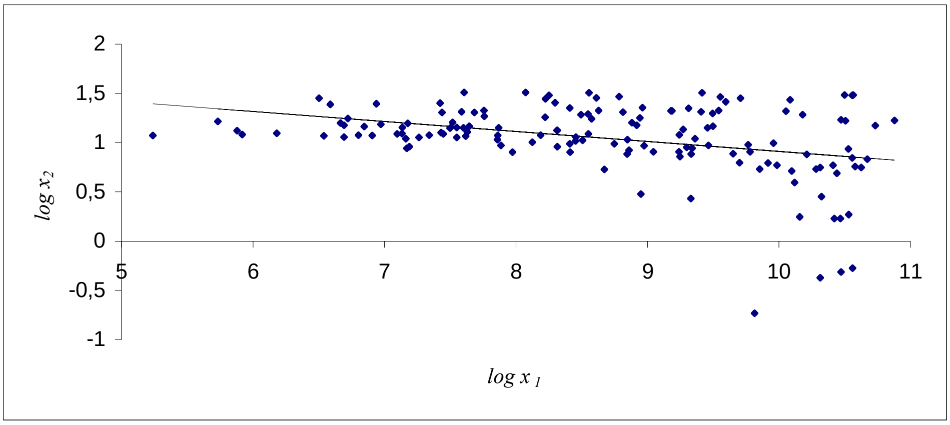


Figure : OLS regression using a panel of 146 countries in 2005, with x_1 being the GDP/cap (World Economic Outlook Database of IMF), and $x_2 = 3 + SYS\_LAN$ (Environmental Sustainability Index 2005).

It is useful to provide a few comments on this result. First, the difference between the ecological rate and the economic rate comes mostly from the large expected economic growth rate ($g_1 = 2\%$) compared to the expected environmental growth rate ($g_2 = \rho g_1 = -0.2\%$). Second, the level of the ecological rate is mostly determined by the substitution effect. Because ρ is small in absolute value, the (negative) ecological growth effect $\gamma_2 \rho g_1 = -0.28\%$ is indeed small. This needs to be compared to the substitution effect $(\gamma_1 - 1)g_1 = 2\%$. Third, the effect of the uncertainty (prudence, cross-prudence and correlation effects) is marginal because of the low volatility of x_1 and x_2, and because we assume that shocks are not serially correlated. Finally, one should compare the economic discount rate obtained here to the one that we estimated around *3.6%* in chapter 3 in the absence of the environment. The diminishing expectation about the quality of the environment and its associated substitution effect $(\gamma_2 - 1)\rho g_1 = -0.08\%$ explains most of the discrepancy between the benchmark *3.6%* to the *3.5%* obtained here.

Extension to parametric uncertainty

In the previous two specifications of the bivariate model, we heavily relied on a geometric Brownian motion with known parameters. Without surprise, we obtained flat term structures under this framework. One easy extension is obtained when recognizing that some of the parameters governing the stochastic economic and ecological growth are uncertain. Consider for example the model that we calbrated in the previous section, and suppose that the parameters (g_1, σ_{11}, ρ) depend upon a variable θ that is not perfectly known. Suppose as in chapter 6 that θ can take interger values *1* to *n*, respectively with probabilities $q_1,...,q_n$. Then, as before, it is easy to derive from equation (10.4) that

$$e^{-r_{2t}t} = \sum_{\theta=1}^{n} q_\theta e^{-r_2(\theta)t}, \tag{10.25}$$

where $r_2(\theta)$ is the ecological discount rate that would prevail if the true value of the unknown parameter would be θ, i.e.,

$$r_2(\theta) = \delta + \big(\rho(\theta)\gamma_2 + \gamma_1 - 1\big)\big(g_1(\theta) - 0.5(\rho(\theta)\gamma_2 + \gamma_1)\sigma_{11}(\theta)\big). \tag{10.26}$$

The reader is now acustomed to the fact that this model yields a decreasing term structure that converges to the smallest $r_2(\theta)$. A symmetric result holds for the economic discount rate.

Suppose for example that g_1 and σ_{11} are known, but the elasticity ρ of environmental quality to changes in GDP is not. Rather than assuming that $\rho = -0.1$ that we estimated in the previous section (with a small *R2* of the OLS estimation), let us suppose that ρ is either *-0.6* or *+0.4* with equal probabilities. All other parameters remain unchanged compared to the previous section. We draw the term structure of r_{1t} and r_{2t} in the next figure. Since the economic growth follows a Brownian motion, the economic discount rate is almost independent of the time horizon. It goes down to a lower *3.2%* for distant cash flows, which would be the efficient economic discount rate if the elasticity ρ would be *-0.6*. In that case, the negative substitution effect would be stronger than in the benchmark with ρ=*-0.1*. The ecological discount rate goes from *1.6%* to *0.3%* when *t* goes from *0* to infinity. The high uncertainty affecting the long-term evolution of the environment in this specification explains why the term structure of the ecological discount rate is decreasing. Another way to interprete this result is obtained by examining the worst-case scenario. In the case in which ρ would be *-0.6*, the large economic growth rate would have a strong negative impact on the quality of the environment. This would generate a strong negative ecological growth effect $\gamma_2 \rho g_1 = -1.68\%$, which offsets most of the substitution effect $(\gamma_1 - 1)g_1 = 2\%$.

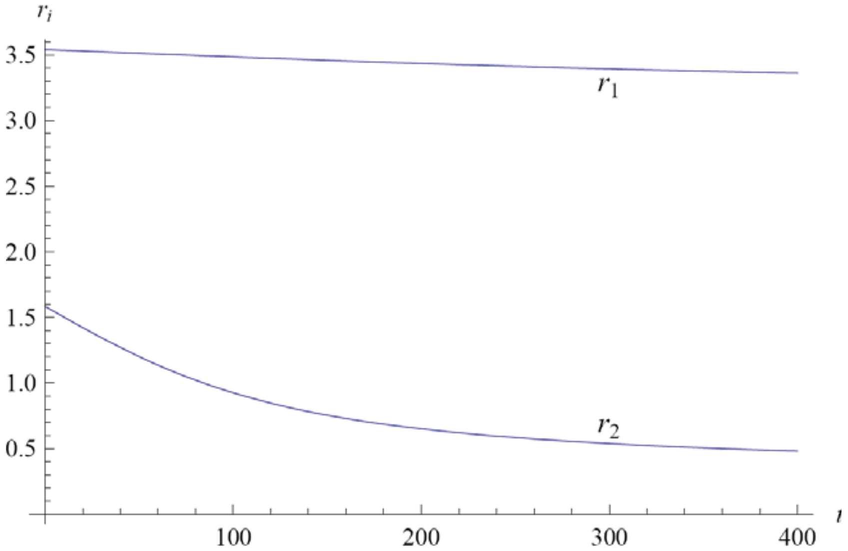


Figure: The economic and ecological discount rates with $U(x_1,x_2)=-x_1^{-1}x_2^{-0.4}$, $\delta=0\%$, $g_1=2\%$, $\sqrt{\sigma_{11}}=3.6\%$ and $x_{2t}=x_{1t}^{\rho}$ with $\rho\sim(-0.6,1/2;0.4,1/2)$.

CES utility functions

Guesnerie (2004), Hoel and Sterner (2007), Sterner and Persson (2008) and Traeger (2007) considered the case of certainty, which implies that the only determinants at play for the ecological discount rate are the ecological growth effect and the substitution effect. In echange for this simplification, they examined a family of utility functions that are more general than the Cobb-Douglas. Namely, they just assumed that U has constant elasticity of substitution $\sigma>0$:

$$U(x_1,x_2)=\frac{y^{1-\alpha}}{1-\alpha} \text{ with } y=\left[(1-\gamma)x_1^{\frac{\sigma-1}{\sigma}}+\gamma x_2^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{10.27}$$

where $\alpha>0$ is relative aversion towards the risk on "aggregate good" y, and $\gamma\in[0,1]$ is a preference weight in favor of the environment. Parameter σ is the percentage rate at which the demand for x_2 declines when the relative price of x_2 is increased by *1%*. When σ tends to unity, y tends to $x_1^{1-\gamma}x_2^{\gamma}$, so that we get the Cobb-Douglas specification as a special case. When $\sigma\neq 1$, the additive nature in y implies that it can never be lognormally distributed, thereby prohibiting the possibility to get an analytical solution under uncertainty. We have that

$$U_2(x_1,x_2)=\gamma y^{\frac{1}{\sigma}-\alpha}x_2^{-\frac{1}{\sigma}}. \qquad (10.28)$$

We can check that the two goods are substitutes if $\alpha\sigma-1$ is positive. Under this condition, an increase in the economic growth raises the ecological discount rate. Under the same condition, U_{22} is negative, so that an anticipated deterioration in the quality of the environment reduces the ecological discount rate. To make this more explicit, suppose that growth rates are constant, which means that $x_{it}=\exp(g_i t)$. The following equation is a direct rewriting of equation (10.4) under this specification:

$$r_{2t}=\delta+\frac{g_2}{\sigma}+\left(\alpha-\frac{1}{\sigma}\right)G(t), \qquad (10.29)$$

with

$$G(t)=\frac{\sigma}{\sigma-1}\frac{1}{t}\ln\left[(1-\gamma)e^{g_1\frac{\sigma-1}{\sigma}t}+\gamma e^{g_2\frac{\sigma-1}{\sigma}t}\right]. \qquad (10.30)$$

Observe that $\exp G(t)$ is the certainty equivalent of $(\exp g_1,1-\gamma;\exp g_2,\gamma)$ under utility function $v_t(G)=(\sigma/(\sigma-1))G^{((\sigma-1)/\sigma)t}$, which is increasing, and whose Arrow-Pratt coefficient of risk aversion is increasing (decreasing) in *t* when σ is smaller (larger) than unity. This implies that the certainty equivalent *G(t)* is decreasing (increasing) in *t* when σ is smaller (larger) than unity. This implies in turn that the term structure of the ecological discount rate is decreasing if $(\alpha\sigma-1)(1-\sigma)$ is positive. More details are given in Guesnerie (2004) and Gollier (2010).

Conclusion

Environmentalists are often quite skeptical about using standard cost-benefit analysis to shape environmental policies because environmental damages incurred in the distant future are claimed to receive insufficient weights in the economic evaluation. This may be due either because future environmental assets are undervalued, or because the economic discount rate is too large. In this chapter, we addressed these two questions altogether by defining an ecological discount rate compatible with social welfare when the representative agent cares about both the economic and ecological environment faced by future generations. This ecological rate at which future environmental damages are discounted may be much smaller than the economic rate at which economic damages are discounted, because of the integration of the potentially increasing willingness to pay for the environment into the ecological discount rate. This increased interest in environmental assets is modelled in this chapter by the potential increased scarcity of these assets, which drives its upward through time. We have also shown that the uncertainties surrounding the evolutions of the quality of the environment and of the economy tend to reduce the discount rates, in particular if they are positively correlated.

References

Eeckhoudt, L., and H. Schlesinger, (2006), Putting risk in its proper place, *American Economic Review*, 96(1), 280-289.

Eeckhoudt, L., B. Rey, and H. Schlesinger, (2007), A good sign for multivariate risk taking, *Management Science*, Vol. 53 (1), 117-124.

Gollier, C., (2010), Ecological discounting, *Journal of Economic Theory*, forthcoming.

Guesnerie, R., (2004), Calcul économique et développement durable, *Revue Economique*.

Hoel, M., and T. Sterner, (2007), Discounting and relative prices, *Climatic Change*, DOI 10.1007/s10584-007-9255-2, March 2007.

Malinvaud, E., (1953), Capital accumulation and efficient allocation of resources, *Econometrica*, 21 (2), 233-268.

Millimet, D. L., J. A. List and T. Stengos (2003), "The environmental Kuznets curve: Real progress or misspecified models?", *The Review of Economics and Statistics*, 85, 1038-1047.

Sterner, T. and M. Persson, (2008),"An Even Sterner Report": Introducing Relative Prices into the Discounting Debate, *Review of Environmental Economics and Policy*, vol 2, issue 1.

Traeger, C.P., (2007), Sustainability, limited substituability and non-constant social discount rates, Dpt of Agricultural & Resource Economics DP 1045, Berkeley.

Weikard, H.-P., and X. Zhu, (2005), Discounting and environmental quality: When should dual rates be used?, *Economic Modelling*, 22, 868-878.

Yale Center for Environmental Law and Policy, (2005), *2005 Environmental Sustainability Index: Benchmarking national environmental stewardship*, Yale University.

Alternative decision criteria

The discounted expected utility (DEU) model that is used in this book is not exempt of critics. Since Allais (1953), many researchers have found certain contexts in which human beings behave in a way that is not compatible with the DEU model. It is clear that this model is violated in many contexts by many people, which implies that it is not a good model to explain actual behaviours under uncertainty. However, our aim in this book is not positive; it is normative. We are not directly interested here by what people actually do. Rather, we tried to determine what they should do. Some of the violations of the DEU model are informative about intrinsic features of actual preferences, whereas others are generated by errors, biases in beliefs, lack of information, or lack of time and effort to find the optimal strategy.

For example, many experiments stress weaknesses of the independence axiom (IA), which is the cornerstone of the von Neumann-Morgenstern expected utility theory. The IA is illustrates as follows. Suppose that you are offered to go to the theatre or to the restaurant tonight. What do you prefer? Suppose that you prefer going to the restaurant. Now, you are explained that the theater and the restaurant are downtown. Because you live in the suburb, the only way to go there is to take the subway. The problem is that there is a 10% probability that the subway will be on strike tonight. Therefore, the actual decision choice that you face is not whether you prefer to go to the restaurant with certainty, or to go to the theatre with certainty. The actual choice is whether you prefer lottery R to lottery T, where lottery R is a good diner to the restaurant with probability *0.9*, or stay at home with probability *0.1*, and lottery T is a nice evening at the theater with probability *0.9*, or stay at home with probability *0.1*. The IA claims that it is natural to assume that the mere fact that there is a *0.1* probability to face the same alternative (stay at home) does not reverse the initial preference. If one prefers the restaurant to the theater, one should also prefer lottery R to lottery T. This is intuitive, and we believe that it is desirable that our collective preference satisfies this axiom.

The last 3 decades have witnessed the blossoming of various interesting decision criteria alternative to the expected utility model. Most of them violate the independence axiom, and will not be examined here. Our aim in this chapter is to describe decision criteria that have features that are normatively attractive.

Recursive expected utility

The concavity of the utility function plays two roles in the DEU model. The indice $-u''/u'$ measures the aversion to consumption inequalities across time and across states of nature. The first feature yields the crucial wealth effect in the Ramsey rule, whereas the second is minked to risk aversion and to prudence. One may question the logic for decreasing marginal utility of consumption to generate at the same time an aversion to risk in each period and an aversion to non-random fluctuations of consumption over time . If the marginal gain of one more unit of consumption is less than the marginal loss due to one unit reduction of consumption, agents will reject the opportunity to gamble on a fifty-fifty chance to gain or lose one unit of consumption. For the same reason, if their current consumption plan is smooth, patient consumers will reject the opportunity to exchange one unit of consumption today against one unit of consumption tomorrow. Kreps and Porteus (1978), Selden (1978) and Epstein and Zin (1991) claimed that there is no logical reason to impose that one use the same utility function for these two psychological processes. They proposed an alternative model to disentangle the attitudes toward consumption smoothing over time and across states. Following Gollier (2002), this section summarizes the application of this model to the problem of evaluating a safe investment project.

We limit the analysis to a model with two dates. As before, let c_0 and c_1 denote consumption per capita respectively in the present and in the future. Welfare at date *0* is evaluated “recursively” in two steps, by backward induction. One first evaluates the certainty equivalent *m* of future consumption c_1 by using an increasing and concave von-Neumann-Morgenstern utility function *v*:

$$v(m) = Ev(c_1). \tag{11.1}$$

We then use a time-aggregating utility function *u* to evaluate the intertemporal welfare *W*:

$$W = u(c_0) + e^{-\delta} u(m). \tag{11.2}$$

The reader can easily check that we recover the standard DEU model if functions *u* and *v* are identical. The utility function *v* characterizes the attitude towards risk, whereas function *u* characterizes the attitude towards time. If *v* is linear, we have $m = Ec_1$ and the agent is risk neutral. This is compatible with a positive wealth effect in the Ramsey rule if *u* is concave and $Ec_1 > c_0$. In other words, one can be risk neutral and have a preference for the reduction of consumption fluctuations over time. Symmetrically, one can be risk aversion and, at the same time, be neutral to consumption fluctuations over time. This would be the case if *v* is concave and u is linear. To sum up, *-v''/v'* measures risk aversion, whereas *–u''/u'* measures the aversion to intertemporal inequality of consumption.

Consider a safe investment project that generates at date *1* $\exp(r)$ euros per euro invested at date *0*. A marginal investment in this project has no effect on intertemporal welfare *W* if

$$-u'(c_0)+e^{r-\delta}u'(m)\left.\frac{\partial m}{\partial s}\right|_{s=0}=0, \tag{11.3}$$

Where *m=m(0)* and *m(s)* is defined as follows:

$$v(m(s))=Ev(s+c_1). \tag{11.4}$$

It yields

$$\left.\frac{\partial m}{\partial s}\right|_{s=0}=\frac{Ev'(c_1)}{v'(m)}. \tag{11.5}$$

All this implies that the efficient discount rate equals

$$r_1=\delta-\ln\frac{u'(m)Ev'(c_1)}{u'(c_0)v'(m)}. \tag{11.6}$$

When $u\equiv v$, we recover the standard pricing formula used in this book. Let us first examine the wealth effect as in chapter 2. Suppose that c_1 is safe, so that $m=c_1$. In that case, equation (11.6) simplifies to (2.9) with *t=1*.

The analysis of the precautionary effect is more complex than in chapter 3. We hereafter determine the condition under which adding a zero-mean risk to c_1 reduces the efficient discount rate. Using (11.6), this is the case if and only if

$$\frac{u'(m)Ev'(c_1)}{u'(c_0)v'(m)}\geq\frac{u'(Ec_1)}{u'(c_0)}, \tag{11.7}$$

or equivalently, if

$$\frac{Ev'(c_1)}{v'(m)}\geq\frac{u'(Ec_1)}{u'(m)}. \tag{11.8}$$

Observe first that the right-hand side of this inequality is less than unity, because m is larger than Ec_1 under risk aversion. This upper bound is attained when the representative agent is neutral to consumption inequalities over time. Thus, inequality (11.8) will surely holds if its left-hand side is larger than unity, i.e., if $Ev'(c_1)$ is larger than $v'(m)$. Let $x=v(c_1)$ and $g(x)=v'(v^{-1}(x))$. With this notation, this condition can be rewritten as

$$Eg(x)\geq g(Ex). \tag{11.9}$$

By Jensen's inequality, this is the case if and only if g is concave. Because $g(v(c)) = v'(c)$, we obtain that

$$-g'(v(x)) = -\frac{v''(x)}{v'(x)}, \tag{11.10}$$

which immediately implies that g is concave if and only if v exhibits decreasing absolute risk aversion (DARA).

We conclude that the precautionary effect on the discount rate is negative as in the standard DEU model if v exhibits DARA. This condition is necessary and sufficient if u is linear. It is noteworthy that DARA is a condition that is stronger prudence since

$$\frac{\partial}{\partial c}\left(\frac{-v''(c)}{v'(c)}\right) = \frac{v''(c)^2 - v'''(c)v'(c)}{v'(c)^2} = \left(\frac{-v''(c)}{v'(c)}\right)\left(\left(\frac{-v''(c)}{v'(c)}\right) - \left(\frac{-v'''(c)}{v''(c)}\right)\right). \tag{11.11}$$

It implies that DARA holds if and only if prudence is larger than risk aversion. It should not be a surprise that a more general model than DEU is more demanding to generate a specific comparative static property.

We can calibrate this model by using power functions and a lognormal distribution for c_1:

$$v'(c) = c^{-\gamma_v}, \quad u'(c) = c^{-\gamma_u}, \quad and \quad \ln c_1 \sim N(\ln c_0 + \mu, \sigma). \tag{11.12}$$

Observe that function v exhibits DARA, so we must expect a negative precuationary effect. Using Lemma 1, we get that

$$\ln m = \ln c_0 + \mu + 0.5\left(1-\gamma_v\right)\sigma^2 = \ln c_0 + g - 0.5\gamma_v\sigma^2, \tag{11.13}$$

with $g = \mu + 0.5\sigma^2$. It implies in turn that

$$\begin{aligned}\frac{u'(m)Ev'(c_1)}{u'(c_0)v'(m)} &= \frac{\exp(-\gamma_u(g-0.5\gamma_v\sigma^2))\exp(-\gamma_v(\mu-0.5\gamma_v\sigma^2))}{\exp(-\gamma_v(g-0.5\gamma_v\sigma^2))} \\ &= \exp(-\gamma_u(g-0.5\gamma_v\sigma^2))\exp(0.5\gamma_v\sigma^2).\end{aligned} \tag{11.14}$$

It implies that the socially efficient discount rate equals

$$r_1 = \delta + \gamma_u g - 0.5\gamma_v(\gamma_u + 1)\sigma^2. \tag{11.15}$$

In the DEU case with $\gamma_u = \gamma_v$, this formula is equivalent to equation (3.21). This shows that this model does not radically modify our understanding of the determinants of the efficient discount rate. In the short run, the driving force of the discount rate is the wealth effect, which is the same as in the DEU case. Because σ^2 is small, changing the precautionary effect from $0.5\gamma_u(\gamma_u+1)\sigma^2$ to $0.5\gamma_v(\gamma_u+1)\sigma^2$ does not impact r_1 very much. An appraisal of the effect of $\gamma_v \neq \gamma_u$ for the long term discount rate remains to be done.

Maxmin ambiguity aversion

In chapter 6, we examined models in which the true probability distribution of future consumption c_1 is uncertain. We used the DEU model to evaluate safe projects under this 2-stage risk context, with stage 1 being about the random selection of the true distribution, and stage 2 being about the random draw of the realization of c_1 from this distribution. Since Ellsberg (1961), we know that many people do not evaluate this 2-stage risk in that way. Let us consider a simplified version of the Ellsberg game. Consider an urn that contains 100 balls, some are black, and the others are white. The two games that will be considered have the same basic structure. The player must pay an entry fee to play the game. The player bets on one of the two colors. The experimenter randomly extracts a ball from the urn, and pays 1000 euros to the player if the color of the ball corresponds to the one on which he bet. In the first game, wich we call the "risky game", there are exactly 50 black balls and 50 white balls. Because betting on any of the two colors yields the same lottery to win 1000 euros with probability ½, most people are indifferent on the color to bet. Because of risk aversion, the maximum entry fee that people are ready to pay is less than the expected gain of 500 euros.

Consider alternatively the "ambiguous game", in which the player gets no information about the proportion black and white balls in the urn. The closed ambiguous urn is brought in front of the player before he selects the color. What is usually observed in this second experiment is that most people are still indifferent between betting on white or on black, but that they are ready to pay much less to play this ambiguous game than the risky game. This cannot be explained under the DEU model. Indeed, if the player is indifferent between white or black, this must mean that they believe that their chance to win by betting white is the same as by betting black. Because probabilities must sum up to unity, this means that their expected probability to win is ½, independent of the color on which the player bets. But then, the player faces a lottery to win 1000 euros with probability ½, which is the same lottery as in the risky game. The player should thus be ready to pay the same in two games. The fact that most people are ready to pay much less for the ambiguous game than for the risky game tells us that people are ambiguity-averse, a psychological trait that cannot be explained by the DEU model. Ambiguity aversion just means that people prefer a lottery to win a widget with a sure probability p than another lottery to win the same widget with an ambiguous probability with mean p.

The first attempt to produce a decision criterion that entails ambiguity aversion was made by Gilboa and Schmeidler (1989). Suppose that people form expectation about the set of plausible distributions of the random variable x that they face. Gilboa and Schmeidler claim that agents evaluate their welfare ex ante by the minimum expeted utility over the set of plausible probability distributions. This would explain the behaviour observed in the Ellsberg game. Indeed, suppose that people believe that the probability of a white draw is either *0.25* or *0.75*. Their welfare will be measured by the expected utility of 1000 euros with the minimum plausible probability, which is *0.25* whatever they bet on white or on black! The certainty equivalent of that lottery is indeed much smaller than in the risky game in which the probability to win is *0.5*.

Let us apply this idea to the discounting problem. To retain our earlier notation, suppose that the distribution of c_1 depends upon an unknown parameter θ that can take *n* possible values $\theta = 1,...,n$. Let $\theta = 1$ denote the value of the parameter that yields the smallest expected utility at date *1*. The efficient discount rate would then satisfy the standard pricing formula (3.14), but in which the distribution of c_1 would be $c_1|\theta_1$ rather than the unconditional distribution of c_1. What would be the consequence for the short-term efficient discount rate r_1? Suppose that the uncertainty is about the mean growth rate. In that case, ambiguity aversion would replace the mean growth rate by the minimum growth rate in the Ramsey rule. Suppose alternatively that the uncertainty is about the volatility. In that case, ambiguity aversion would replace the mean volatility by the maximum volatility in the Ramsey rule. In the two cases, the problem becomes equivalent to computing the discount rate that would be efficient conditional to each realization of θ, and then to selecting the smallest of these rates as the efficient discount rate r_1. Interestingly enough, this *short-term* discount rate that is efficient under Gilboa-Schmeidler's maxmin theory is the discount rate that is efficient for the *distant future* in the DEU model examined in chapter 6!

Smooth ambiguity aversion

This maxmin model is difficult to implement to provide normative recommendation because it does not explain how to determine the set of plausible distributions that is part of the preferences of the representative agent. This is problematic because this model is very sensitive to the characteristics of the worse probability distribution, which could be arbitrarily catastrophist. Klibanoff, Marinacci and Mukerji (KMM, 2005, 2010) have recently proposed a model that is easier to implement, and is less sensitive to the extreme plausible distribution. They define ambiguity aversion as the aversion to any mean-preserving spread in the space of probabilities. Remember that risk aversion is an aversion to any mean-preserving spread in the space of payoffs. For example, risk aversion means that one prefers to get 500 in two equally probable states, than to receive 1000 in state 1, and 0 in state 2. Taking this risky lottery as a benchmark, ambiguity aversion means that one prefers this lottery to another one in which the true probability of state 1 is not *0.5* with certainty but rather *0.25* or *0.75* with equal probabilities.

KMM have proposed the following decision criterion under ambiguity. For each possible value of θ, one computes the conditional expected utility $E\left[u(c_1)\middle|\theta\right]$. Then, rather than taking its mean under the subjective distribution $(q_1,...,q_n)$ of θ, one computes its certainty equivalent by using some increasing and concave function ϕ:

$$W = u(c_0) + e^{-\delta}M \quad with \quad \phi(M) = \sum_{\theta=1}^{n} q_\theta \phi\left(E\left[u(c_1)\middle|\theta\right]\right). \tag{11.16}$$

Because ϕ is concave, M is smaller than the unconditional expected utility, which means that this welfare functional exhibits ambiguity aversion. Two special cases are useful to examine. First, if function ϕ is the identity function, then this welfare function is as in the standard DEU case, in which agents are neutral to mean-preserving spreads in probabilities. The expected utility criterion is linear in probabilities. In fact, function $-\phi''/\phi'$ is an index of absolute ambiguity aversion. The other special case is obtained by assuming that $\phi(u) = -A_\phi^{-1}\exp(-A_\phi u)$, where the index of absolute ambiguity aversion A_ϕ tends to infinity. We have demonstrated in chapter 6 that $E\phi(u)$ tends to the min of u in that case, so that we get the maxmin criterion as another special case.

As usual, consider a safe investment project that yields $\exp(r)$ euros at date *1* per euro invested at date *0*. At the margin, this project has no effect on the intertemporal welfare W if

$$-u'(c_0) + e^{r-\delta}\frac{\sum_{\theta=1}^{n} q_\theta \phi'\left(E\left[u(c_1)\middle|\theta\right]\right)E\left[u'(c_1)\middle|\theta\right]}{\phi'(M)} = 0. \tag{11.17}$$

It yields the following efficient discount rate:

$$r_1 = \delta - \ln\frac{\sum_{\theta=1}^{n} q_\theta \phi'\left(E\left[u(c_1)\middle|\theta\right]\right)E\left[u'(c_1)\middle|\theta\right]}{u'(c_0)\phi'(M)}. \tag{11.18}$$

Gierlinger and Gollier (2009) exhibit two effect of ambiguity aversion in this model : an ambiguity prudence effect and a pessimism effect. The ambiguity prudence effect is easiest to explain if we assume that the representative agent is risk-neutral, i.e., if u is the identity function. This switches off both the wealth effect and the precautionary effect of the standard model. In that case, equation (11.18) simplifies to

$$r_1 = \delta - \ln \frac{\sum_{\theta=1}^{n} q_\theta \phi'\left(\overline{c}_{1\theta}\right)}{\phi'(M)} \quad with \quad \phi(M) = \sum_{\theta=1}^{n} q_\theta \phi\left(\overline{c}_{1\theta}\right), \tag{11.19}$$

Where $\overline{c}_{1\theta}$ is the conditional expected consumption at date *1*. Thus, the ambiguous distribution of economic growth reduces the efficient discount rate if

$$\sum_{\theta=1}^{n} q_\theta \phi'\left(\overline{c}_{1\theta}\right) \geq \phi'(M) \quad whenever \quad \sum_{\theta=1}^{n} q_\theta \phi\left(\overline{c}_{1\theta}\right) = \phi(M). \tag{11.20}$$

We have encoutered excatly the same technical condition in the section on recursive expected utility, and we have shown that it requires that the ϕ function exhibits decreasing absolute aversion: $(-\phi''/\phi')' \leq 0$. This is more demanding than the prudence of ϕ ($\phi''' \geq 0$). This ambiguity prudence condition guarantees that, under risk-neutrality, the existence of some ambiguity on the distribution of future consumption reduces the discount rate.

The pessimism effect is similar to the one that is obtained under the maxmin criterion. It is easiest illustrated by switching off the ambiguity prudence effect, that is, by assuming that absolute ambiguity aversion $-\phi''/\phi'$ is constant. If we assume that $\phi(u) = -A\exp(-Au)$, we have that $\phi'(M)$ equals $\Sigma_\theta q_\theta \phi'\left(Eu_\theta\right)$. It implies that we can rewrite equation (11.18) as follows:

$$r_1 = \delta - \ln \frac{\sum_{\theta=1}^{n} \hat{q}_\theta E\left[u'(c_1)\middle|\theta\right]}{u'(c_0)} \quad with \quad \hat{q}_\theta = q_\theta \frac{\phi'\left(E\left[u(c_1)\middle|\theta\right]\right)}{\sum_{\tau=1}^{n} q_\tau \phi'\left(E\left[u(c_1)\middle|\tau\right]\right)}. \tag{11.21}$$

If we compare this to the discount rate that we obtain under the standard DEU criterion, which is equation (6.2) with *t=1*, we observe that the only difference is that we distorted the beliefs described by $(q_1,...,q_n)$ into $(\hat{q}_1,...,\hat{q}_n)$ defined in (11.21). Because ϕ' is decreasing, these distorted beliefs put more probability weight to the θ that yields smaller conditional expected utility. This is a clear expression of pessimism, whose extreme version was illustrated by the maxmin model. If we suppose for example that the uncertainty is about the expected growth rate, the probabilities will be distorted in favour of the θ with the smallest expected growth, for which the expected marginal utility is larger. That will tend to reduce the discount rate r_1.

To sum up, ambiguity aversion tends to reduce the discount rate. One can illustrate this intuituive idea[5] by considering the following specification suggested in Gierlinger and Gollier (2009). Suppose as in chapter 6 that $\ln c_t \,|\theta$ is normally distributed with mean $\ln c_0 + \mu_\theta t$ and variance $\sigma^2 t$. Suppose that the mean of the change μ_θ in the log of consumption is itself normally distributed with mean μ_0 and variance σ_0^2. Consider the case of a power utility function with constant relative risk aversion γ. This model is exactly the benchmark case that we considered in chapter 6. The only new dimension is ambiguity aversion. Suppose that ϕ exhibits constant relative ambiguity aversion $\eta = -|u|\phi''(u)/\phi'(u)$. Using Lemma 1 twice, Gollier and Gierlinger (2009) obtained the following formula:

$$r_t = \delta + \gamma g - 0.5\gamma(1+\gamma)(\sigma^2 + \sigma_0^2 t) - 0.5\eta\left|1-\gamma^2\right|\sigma_0^2 t, \tag{11.22}$$

where $g = \mu_0 + 0.5(\sigma^2 + \sigma_0^2 t)$ is the growth rate of expected consumption. This equation should be compared to equation (6.13), which is a special case of (11.22) with $\eta = 0$. This observation allows us to conclude that ambiguity aversion yields a fourth determinant to the discount rate, which is negative and linear with the time horizon under the specification considered here. This is because the degree of ambiguity is magnified by the time horizon in this framework with an uncertain trend in economic growth.

Intergenerational habit formation

Although we consume much more goods and services than our parents, we are not really happier than them. This is a paradox that the indices of happiness do not parallel those of the GDP per capita (see for example Layard (2005)). One possible explanation is that people do evaluate their well-being in relative rather than in absolute terms. In particular, their felicity at date *t* is not a function of their consumption at date *t* alone. In the literature on external habit formation, it is assumed that the agent's felicity at date *t* is a function of c_t and of a weighted average of past consumption $(c_{t-1}, c_{t-2}, ...)$. This breaks down the time-additivity property of the DEU model. Constantinides (1990) has argued for a positive effect of past consumption

[5] Gierlinger and Gollier (2009) show that this is not always the case, even under decreasing absolute ambiguity aversion.

on today's marginal utility of consumption, which is a simple definition of consumption habit. A large consumption level in the past raises the marginal utility of current consumption, thereby creating some form of addiction to consumption.

A simple specification is the multiplicative habit in which the felicity at date t is measured by $u(c_t / c_{t-1}^{\alpha})$, for some positive constant $\alpha \leq 1$. A special case is $\alpha = 1$, in which case the felicity is a function of the growth rate of consumption rather than of the level of consumption. For example, if the growth rate of consumption is a positive constant, the felicity will remain constant over time in this model. Under these preferences, increasing temporarily consumption at any time above its historical trend is beneficial in the short run, but it generates a negative externality on future welfare because of the consumption habit that this transitory increase generates. When α is less than unity, this negative externality is reduced. Therefore, α is a measure of habit formation.

To keep the model very simple, let us assume that $u(x) = x^{1-\gamma}/(1-\gamma)$ with $\gamma > 1$. Suppose also that that the growth rate of consumption is a positive constant g. Observe now that

$$u\left(\frac{c_t}{c_{t-1}^{\alpha}}\right) = \frac{1}{1-\gamma} c_t^{(1-\alpha)(1-\gamma)} g^{\alpha(1-\gamma)} = k c_t^{1-\gamma'}, \tag{11.23}$$

with $\gamma' = \alpha + (1-\alpha)\gamma$. This shows that the existence of a multiplicative internal consumption habit transforms the intertemporal welfare function in a very simple way. First, it multiplies the felicity by a common positive constant $g^{\alpha(1-\gamma)}$. Second, it modifies the degree of relative risk aversion from γ to γ', which is a mean of γ and *1*, weighted respectively by $(1-\alpha)$ and α. Since we usually assume that γ is larger than unity, we have that this model of habit formation just reduces the degree of concavity of the felicity function. The Ramsey rule (2.11) thus still holds, but with γ being replaced by the smaller γ':

$$r_t = \delta + \gamma' g. \tag{11.24}$$

Because consumption habit downsizes the wealth effect, it yields a smaller discount rate. The intuition is that investing for the future is a good way to impose self-control in consumption, thereby limiting the formation of consumption habits that have adverse effects on future welfare. Gollier, Johansson-Stenman and Sterner (2010) extend this result to the case of uncertainty.

The internal habit formation model brievely described above has some interesting features to explain real behaviours of human beings. For example, it can contribute to solve the equity premium puzzle (Constantinides (1990)). However, it is still an open question to determine whether this model should be used or not for normative analysis of public policies spanning several generations. It is clear that parents transfer consumption habits to their children, so that the habit formation is not strictly speaking an intra-individual feature. But is it enough to justify more sacrifices for the current generation?

Conclusion

In this chapter, we have illustrated the recent blossoming of new decision criteria in the face of risk and time, focusing on their applications to the selection of the discount rate. We examined successively the recursive expected utility model, the maxmin and the smooth ambiguity aversion models, and we provided a short introduction to the internal habit formation model. Many other models could have been considered for inclusion in this chapter. For the sake of conciseness, we had to make choices. For example, we could have considered the cumulative prospect theory introduced by Tversky and Kahneman (1992). This model shares with the habit formation model the idea that future consumption will be evaluated in relation to some reference point that may be related to past consumption. But prospect theory has other features, as the assumption that agents are risk-lovers in the range of losses (below the reference point). It is also assumed that they distort the decumulative distribution function by using some specific nonlinear function that plays a role symmetric to the utility function that transforms payoffs into utility in a nonlinear way. This transformation raises the subjective probability of extreme events, which will have the effect to raise the precautionary term in the extended Ramsey rule, thereby reducing the discount rate. It is still too early to determine which of these innovations will survive the long term tests of the scientific validation process.

References

Allais, M., (1953), Le comportement de l'homme rationnel devant le risque, Critique des postulats et axiomes de l'école américaine, *Econometrica*, 21, 503-46.

Constantinides, G. (1990), “Habit formation: a resolution of the equity premium puzzle”, *Journal of Political Economy*, 98, 519−543.

Epstein, L.G., and S. Zin, (1991), Substitution, Risk aversion and the temporal behavior of consumption and asset returns: An empirical framework, *Journal of Political Economy*, 99, 263-286.

Giergingler, J., and C. Gollier, (2009), Socially efficient discounting under ambiguity aversion, mimeo, Toulouse School of Economics.

Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with a non-unique prior, *Journal of Mathematical Economics*, 18, 141--153.

Gollier, C., (2002), Discounting an uncertain future, *Journal of Public Economics*, 85, 149-166.

Gollier, C., O. Johansson-Stenman and Th. Sterner, (2010), Ramsey Discounting when Relative Consumption Matters, unpublished manuscript, Toulouse School of Economics.

Kreps, D.M., and E.L. Porteus, (1978), Temporal resolution of uncertainty and dynamic choice theory, *Econometrica*, 46, 185-200.

Klibanoff, P., M. Marinacci, and S. Mukerji, (2005), A smooth model of decision making under ambiguity, *Econometrica*, 73(6), 1849—1892.

Klibanoff, P., M. Marinacci, and S. Mukerji, (2010), Recursive smooth ambiguity preferences. *Journal of Economic Theory*, forthcoming.

Layard, Richard (2005), *Happiness: Lessons from a new Science*, Penguin Press.

Selden, L., (1979), An OCE analysis of the effect of uncertainty on saving under risk independence, *Review of Economic Studies*, 73-82.

Tversky, A., and D. Kahneman, (1992), Advances in prospect theory - Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5, 297-323.

PART IV

Evaluation of risky and uncertain projects

Evaluation of risky projects

This book is mostly devoted to the evaluation of safe investment projects. However, most real projects are not safe, and many of them are very risky, in particular those yielding cash flows in the distant future. The last part of this book is devoted to the exploration of adaptation of the rules presented earlier in this book to risky and uncertain projects. The evaluation of risky projects and of risky assets has been the Holy Grail of the theory of asset pricing, which is the main branch of the modern theory of finance. This chapter provides a short overview of the main concepts, ideas and tools that have been produced by more than fifty years of research in that field.

The equity premium

It is easy to have a crude estimation of the effect of risk on the evaluation of projects and assets in the economy. Investors on financial markets have the opportunity to invest in a large set of projects. Their optimal assets allocation is such that they are indifferent at the margin to transfer wealth from one asset to any other one. This is why two safe assets with the same maturity must have the same return. By risk aversion, if an asset has a cash flow that correlates positively with the aggregate risk in the economy, its equilibrium price will be smaller than the corresponding safe asset with the same expected payoff at the same maturity. In other words, the expected return of the risky asset will be larger than the return of the safe asset. This means that investors discount the expected cash flows of the risky asset at a larger rate. The social planner should do the same to evaluate risky public investments. This chapter is devoted to the analysis of the risk premium that should be added to the discount rate for safe projects.

Dimson, Marsh and Staunton (2002) have computed the annualized return of bonds and equity for different countries over the 20th century. Using extended data from the same authors over the period 1900-2006, we summarized the main facts in the following figure. In the United States, the return on 10-year Treasury bonds, which are probably the safest assets in the world, yield a real return around *1.9%*, whereas equity delivered an average real return of *6.6%* per year. This yields an equity premium around *4.7%*. The real return of bonds varies much across different countries during the period. In particular, the real return of bonds was negative in countries confronted with a world war on their soil, as Japan, France and Italy. However, the equity premium is surprisingly stable across countries, in the range of *3-5%*.

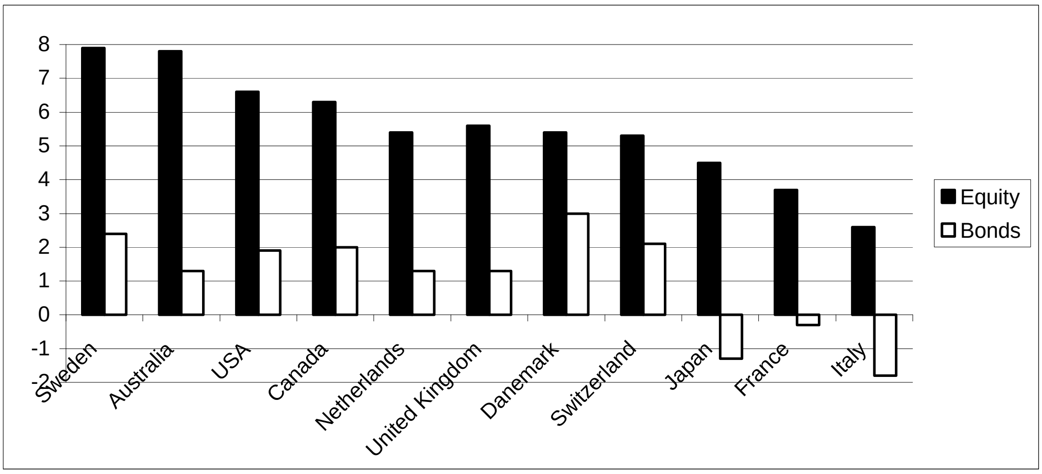


Figure : Average annual real returns of equity and bonds from 1900 to 2006.

Sources: Morningstar and Dimson, Marsh and Staunton, (2002)

We have reproduced in the following figure the same exercise over the shorter period of time 1971-2006. It is noteworthy that the safe return has been much larger than over the entire century, whereas the return on equity remained stable. A possible explanation of this fact comes from the successful fight of inflation by central banks over the period. It implies a smaller equity premium. For example, in the United States, the annualized real return on bonds has been *4%*, whereas the annualized real return on equity has been *6.6%*, yielding an equity premium of *2.6%*.

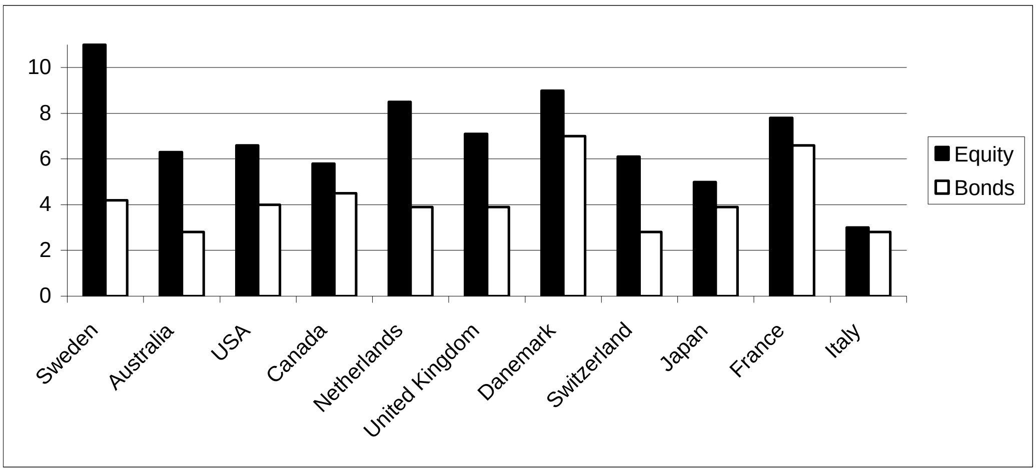


Figure : Average annual real returns of equity and bonds from 1971 to 2006.

Sources: Morningstar and Dimson, Marsh and Staunton, (2002) By the standard arbitrage argument, these numbers justifies a discount rate of *4%* to evaluate safe projects in the United States. At the same time, if the project under scrutiny has a risk profile similar to the one of U.S. equity, a discount rate of *6.6%* should be used. This is not far from the *7%* that is recommended by the OMB in 1992. However, it would be inefficient to use that discount rate to evaluate a safe project. This numbers give us some orders of magnitude of the effect of risk on the evaluation of risky projects.

Certainty equivalent and risk premium

Let us consider an investment project that yields B_t euros per capita at date t per euro invested today. We allow for B_t to be random and potentially correlated to consumption c_t. Investing ε in the project yields the following intertemporal welfare:

$$W(\varepsilon) = u(c_0 - \varepsilon) + e^{-\delta t} Eu(c_t + \varepsilon B_t). \tag{12.1}$$

A marginal investment in that project has a positive effect on the intertemporal welfare if

$$-u'(c_0) + e^{-\delta t} EB_t u'(c_t) \geq 0. \tag{12.2}$$

This can be rewritten as

$$-1 + e^{-\delta t} \frac{Eu'(c_t)}{u'(c_0)} \frac{EB_t u'(c_t)}{Eu'(c_t)} \geq 0. \tag{12.3}$$

It is easier to write this condition as

$$NPV = -1 + e^{-r_t t} F_t \geq 0, \tag{12.4}$$

with

$$r_t = \delta - \frac{1}{t} \ln \frac{Eu'(c_t)}{u'(c_0)}, \tag{12.5}$$

and

$$F_t = \frac{EB_t u'(c_t)}{Eu'(c_t)}. \tag{12.6}$$

When the future cash flow is uncertain, its evaluation requires a two-step procedure. First, the risky cash flow B_t is replaced by its certainty equivalent F_t defined by (12.6). This first operation simplifies the problem to the one of valuing a safe project. Therefore, the second step is obvious: this certainty equivalent must be discounted by using the discount rate r_t defined by (12.5). The reader recognizes the rate that is efficient for safe projects, and that we have been describing all along this book. The project should be implemented if and only if its net present value computed in that way is positive. This procedure is very useful, because it tells us that what we have done in this book to characterize the efficient discount rate can also be used to evaluate risky projects.

The only new element to be examined in this chapter is the transformation of a risky cash-flow B_t into its certainty equivalent F_t. If this project can be traded on frictionless financial markets, its equilibrium price should be equal to F_t. Equation (12.6) is in fact the classical equilibrium asset pricing formula that can be found in any textbook on the theory of finance. It happens that F_t is a weighted mean of the different possible realizations of B_t. For example, if B_t is sure, then $F_t = B_t$. If it is risky, let us define the "risk-neutral expectation" operator $\hat{E}$ as follows:

$$\hat{E}f(b) = \frac{Ef(b)u'(c_1)}{Eu'(c_1)}. \tag{12.7}$$

It corresponds to the notion of the « risk-neutral probability" of a state, which is the true state probability multiplied by the marginal utility of consumption in that state, and divided by $Eu'(c_1)$ in order to guarantee that the risk-neutral probabilities sum up to one. We thus get that $F_t = \hat{E}B_t$. The certainty equivalent of a cash flow is equal to its risk-neutral expectation. We hereafter describe the implications of this observation. It is natural to define the risk premium in the valuation of the cash flow B_t as the difference between the expected cash flow EB_t and its certainty equivalent $F_t = \hat{E}B_t$.

The Arrow-Lind Theorem

The simplest case arises when the cash flow B_t is risky, but this risk is independent from the systematic risk corresponding to c_t. In that case, applying equation (12.6) immediately implies that $F_t = EB_t$. The equilibrium price – and the efficient valuation – of that asset is actuarially fair, in the sense that the risk premium vanishes. There is no risk premium associated to idiosyncratic risk. This result is usually refered to as the Arrow-Lind Theorem in the literature in public economics (Arrow and Lind (1970)).

It is important to get the intuition of this result. Put it simply, risks that are uncorrelated with the aggregate risk are in fact fully diversified away in the portfolio of the representative agent. Adding this risk to the portfolio does not increase the portfolio riskiness. This is due to the fact that the risk premium for small risk is proportional to its variance. This comes from the Arrow-Pratt approximation (3.3). Thus, when the size k of the risk goes to zero, its risk premium goes to zero as k^2, whereas its expected value goes to zero as k. This means that when the size of the risk is small, only the mean matters to value it. The risk premium is of the second order. In the DEU model, risk aversion is a second-order phenomenon, following Segal and Spivak (1990). This is not the case for other decision criteria under uncertainty, as for example the prospect theory.

The consumption-based capital asset pricing model

Suppose alternatively that the cash flow of the project and the GDP per capita be positively stochastically dependent, in the sense that an increase in c_t improves the conditional distribution of B_t in the sense of first-degree stochastic dominance. Using Lemma 2 in chapter 8, this statistic dependence of (B_t, c_t) raises the value F_t of the cash flow if $h(B_t, c_t) = B_t u'(c_t)$ is supermodular, i.e., if u is concave. In other words, the risk premium is positive if the cash flow is positively correlated with the systemic risk, and the risk premium is negative if they negatively correlated. We get the Arrow-Lind theorem in the limit case of independence. In case of a negative correlation, implementing the project reduces the global risk. It has therefore an insurance value, which takes the form of a negative risk premium.

Suppose that $(\ln B_t, \ln c_t)$ follows an arithmetic Brownian motion. Their trends and volatilities are denoted respectively (μ_B, μ_c) and (σ_B, σ_c). Their indice of correlation is denoted ρ. It implies that $(\ln B_t, \ln c_t)$ are jointly normal. Suppose that $u'(c) = c^{-\gamma}$. We can then use Lemma 1 twice to compute the two expectations in (12.6):

$$Eu'(c_t) = \exp\left(-\gamma\left(\ln c_0 + \mu_c t - 0.5\gamma\sigma_c^2\right)\right). \tag{12.8}$$

$$\begin{aligned} EB_t u'(c_t) &= E\left(\exp\left(\ln B_t - \gamma \ln c_t\right)\right) \\ &= \exp\left(\ln B_0 + \mu_B t - \gamma \ln c_0 - \gamma\mu_c t + 0.5t\left(\sigma_B^2 + \gamma^2\sigma_c^2 - 2\gamma\sigma_B\sigma_c\rho\right)\right). \end{aligned} \tag{12.9}$$

Using (12.6), we obtain that

$$F_t = B_0 \exp\left(\mu_B t + 0.5t\left(\sigma_B^2 - 2\gamma\sigma_B\sigma_c\rho\right)\right). \tag{12.10}$$

Now, observe that $EB_t = \exp(\ln B_0 + \mu_B t + 0.5\sigma_B^2 t)$, so that the above equation can finally be rewritten as

$$F_t = \left(EB_t\right)e^{-\gamma\beta\sigma_c^2 t} = \left(EB_t\right)e^{-\pi(\beta)t}. \tag{12.11}$$

where the "consumption β" of the project is defined as

$$\beta = \frac{\rho\sigma_B}{\sigma_c} = \frac{\text{cov}(\ln B_t / B_{t-1}, \ln c_t / c_{t-1})}{\sigma_c^2}, \tag{12.12}$$

and where $\pi(\beta) = \gamma\beta\sigma_c^2$ is defined as the risk premium of the project. Equation (12.11) confirms that the signs of the risk premium and of the covariance of (B_t, c_t) coincide. Under this specification, the risk premium is proportional to the expected cash flow of the project.

Computing this risk premium thus requires information about the volatility σ_B of the cash flows and about its correlation ρ with the growth of GDP/cap. If similar investment projects have been implemented in the past, one can use these observations to estimate these parameters by using standard regression methods. If these data are not available, the Monte-Carlo methodology is a good alternative. What remains crucial to keep in mind is that the idiosyncratic risk of the project has no value, because agents diversify them away. As stated by the Arrow-Lind Theorem, only the correlation with the macro risk is relevant.

In this chapter, we have adviced the reader to disentangle the problem of time (discounting) and the problem of risk (certainty equivalent). However, under the joint lognormal specification considered in this section, a nice simplication occurs. Observe from equation (12.11) that the certainty equivalents of the cash flows expressed as a fraction of their expected values varies exponentially with time. Therefore, taking into account of this treatment of risk is equivalent to adapting the discount rate to the riskiness of the project in the following way. As explained in chapter 4, the rate r_t to discount safe projects is constant. Let us denote it $r_f = \delta + \gamma\mu_c - 0.5\gamma^2\sigma_c^2$. Combining equations (12.4) and (12.11) yields

$$NPV = -1 + e^{-r_f t}F_t = -1 + e^{-rt}EB_t, \tag{12.13}$$

with

$$r = r_f + \gamma\beta\sigma_c^2. \tag{12.14}$$

Equation (12.13) tells us that the two-step evaluation procedure that we presented earlier in this chapter is equivalent to an alternative procedure in which one discounts the expected cash flows at a rate that takes into account of the riskiness of the project. This risk-adjusted rate r defined by equation (12.14) is the sum of the risk free discount rate r_f examined in this book and a risk premium $\pi(\beta) = \gamma\beta\sigma_c^2$. This risk-adjusted discount rate is specific to each project through the estimation of its β. Equation (12.14) is usually refer to as the "consumption-based capital asset pricing" formula (CCAPM) à la Lucas (1978).

This alternative evaluation procedure is very specific to the joint lognormal specification considered above. In general, the certainty equivalent cash flows are not proportional to their expected values, and when they are, they do not vary exponentially with time, as in (12.11). Consider for example the case of the nuclear sector. Producing electricity with the nuclear technology has a lifecycle with different phases yielding very different risk. During the building phase, risks on cash flows come mostly from the uncertainty surrounding costs of labour and physical inputs. During the long duration of the production phase, the uncertainty is mostly about the price of the electricity on the markets. In the demantling phase, the uncertainty is about the cost of recycling or storing nuclear wastes. Clearly, the correlations of these cash flows with the macro risk are much different across these three phases, and this alternative evaluation procedure needs to be fixed. This can be done by estimating the beta of the cash flows in each phase separately, and by using different discount rates for them accoding to the CCAPM formula (12.14).

Valuation of the macroeconomic risk and the equity premium

In this section, we examine an investment project whose risk profile exactly duplicates the macroeconomic risk. This project has a cash flow that duplicates the GDP per capita. When c_t is increased or decreased by *1%*, so does B_t. This project has a consumption β equalling *1*. Under the geometric Brownian specification, the riskiness of such a project should be taken into account by raising the discount rate above r_f by $\gamma\sigma_c^2$. Earlier in this book, we have estimated the risk aversion γ to be around *2*, whereas the volatility of the growth of GDP/cap σ_c was estimated around *3.6%.* Thus, we get a macroeconomic risk premium around $\pi_m = \pi(1)$=*0.26%*. Because we estimated the safe discount rate r_f at *3.6%,* this means that one should discount such an investment project with a discount rate of *3.86%.*

Suppose alternatively a project whose cash flows increase by β% when GDP/cap increases by *1%*. Observe that it implies that $\mathrm{cov}(\ln B_t / B_{t-1}, \ln c_t / c_{t-1}) / \sigma_c^2 = \beta$, so that we are indeed refereeing here to the consumption β. Following the CCAPM equation (12.14), such a project should be evaluated by using the following discount rate:

$$r(\beta) = r_f + \beta\pi_m = 3.6\% + \beta \times 0.26\%. \tag{12.15}$$

Suppose that this investment corresponds to a traded asset. At equilibrium, agents should be indifferent to a marginal increase in this investment, so that its price must be such that NPV of buying the asset is zero. This is the case if its the equilibrium expected return is $r(\beta)$.

Let us now consider an asset that duplicates the equity market. Kocherlakota (1996) used the Standard & Poors 500 annual data for the U.S. equity market over the period 1889-1978. He obtained a consumption β for this equity portfolio around $\beta_{SP500} = 1.72$. Applying equation (12.15), it must be that that the expected excess return of the SP500 be around $1.72 \times 0.26\% = 0.44\%$. However, as shown earlier in this chapter, the excess return of equity in the U.S. during the 20th century was rather around *4-5%* per year. This large discrepancy between the observed equity premium and its prediction by the CCAPM is called the equity premium puzzle.

This puzzle has raised muc attention in our profession, in hundreds of papers have been published to try to solve it. The main difficulty comes from the low level of the macroeconomic risk premium $\pi_m = \gamma\sigma_c^2$, and of the risk on economic growth that lies behind it. As seen earlier in this book, there are reasons to believe that this is underestimated. To solve this problem, one can reverse the method that led to equation (12.15) to evaluate the efficient risk-adjusted discount rate. Suppose that markets estimate correctly the macroeconomic risk and the equity's consumption $\beta_{SP500} = 1.72$**.** The average real return of the equity market in the United States has been r_{SP500}=*6.6%.* Combining this with an observed risk free rate of $r_f = 1.9\%$ yields an estimation of the macroeconomic risk premium $\pi_m = \gamma\sigma_c^2$ by using equation (12.14):

$$\pi_m = \frac{r_{SP500} - r_f}{\beta_{SP500}} = \frac{6.6\% - 1.9\%}{1.72} = 2.73\%. \tag{12.16}$$

It implies the following alternative formula to the risk-adjusted discount rate:

$$r(\beta) = 1.9\% + \beta \times 2.73\%. \tag{12.17}$$

For example, a project whose risk profile duplicates the macroeconomic risk ($\beta = 1$) should be discounted at a rate of *3.63%*. An investment whose risk profile is similar to the riskiness of the SP500 ($\beta = 1.72$) should be discounted at *6.6%.*

The CCAPM discount rate r defined by (12.17) is linked to the "weighted average cost of capital" (WACC) used by firms to evaluate the NPV of their investment projects. At equilibrium, the cost of capital of corporation having a portfolio of investments with different β must be the capital-weighted average of the discount rates $r(\beta)$ of these investments. Each new project should be evaluated with its own $r(\beta)$ rather than with the firm's WACC, however.

A solution to the equity premium puzzle

At this stage, an important question emerges about pricing risky investment projects. Which of the two rules (12.15)(12.17) should we use for the risk-adjusted discount rate? Compare to observed prices on the market, the calibration of the CCAPM suggest a larger risk free rate (*3.6%* vs *1.9%*) and a smaller macroeconomic risk premium (*0.26%* vs *2.73%*). These two discrepancies can be explained by the hypothesis that the markets assume a larger macro risk σ_c than what is in the data. Indeed, a larger uncertainty on the economic growth reduces the risk free rate because of the magnified precautionary effect, in particular in the long run. We have discussed in Part II various arguments for why the macro risk could be underestimated in the long term, and we have shown there that reducing the interest rate from *4%* to *2%* is within the range of reasonable values. Observe now that raising the perceived macroeconomic risk σ_c also raise the macroeconomic risk premium $\pi_m = \gamma\sigma_c^2$. Thus, what we have done in Part II may be useful to solve the equity premium puzzle.

One possible road is to recognize that our calibration can be affected by the Peso problem that we illustrated in chapter 6 for the interest rate. It may just be the case that the data set does not contain the deep potential recessions and economic catastrophes that investors have in mind when determining their assets allocation. Barro (2006) shows that this can explain the puzzle. Weitzman (2007) proposes an alternative explanation based on the presence of uncertainty surrounding the stochastic dynamics of the economy. Let us brievely describe the idea, which follows the lines developed in chapter 6.

Suppose that the growth process of the economy be lognormal with parameters (μ_c, σ_c), but the true values of these parameters are uncertain. As usual, let us describe this parametric uncertainty by assuming that they are functions of parameter θ, which can take interger values *1* to *n*, respectively with probability q_1 to q_n. Let us reconsider the macroeconomic risk premium $\pi_m = \pi(1)$. Without parametric uncertainty, by using equations (12.6) and (12.11), it is equal to

$$\pi_m = -\frac{1}{t}\ln\frac{F_t}{Ec_t} = -\frac{1}{t}\ln\frac{Ec_t u'(c_t)}{Ec_t Eu'(c_t)}. \tag{12.18}$$

With the parametric uncertainty described above, this equation must be rewritten as follows :

$$\pi_m = -\frac{1}{t}\ln\frac{\sum_{\theta=1}^{n} q_\theta E\left[c_t u'(c_t)\middle|\theta\right]}{\left(\sum_{\theta=1}^{n} q_\theta E\left[c_t\middle|\theta\right]\right)\left(\sum_{\theta=1}^{n} q_\theta E\left[u'(c_t)\middle|\theta\right]\right)}. \tag{12.19}$$

Assume constant relative risk aversion γ. Using Lemma 1, one can rewrite this in the following way:

$$\pi_m = -\frac{1}{t}\ln\frac{\sum_{\theta=1}^{n} q_\theta e^{(1-\gamma)\left(\mu_c+0.5(1-\gamma)\sigma_c^2\right)t}}{\left(\sum_{\theta=1}^{n} q_\theta e^{\left(\mu_c+0.5\sigma_c^2\right)t}\right)\left(\sum_{\theta=1}^{n} q_\theta e^{-\gamma\left(\mu_c-0.5\gamma\sigma_c^2\right)t}\right)}. \tag{12.20}$$

In the sepcial case of no parametric uncertainty, this simplifies to $\pi_m = \gamma\sigma_c^2$. Otherwise, it can be shown that the macro risk premium is increasing with the time horizon. Weitzman (2007) shows that if the uncertainty is about σ_c^2 whose inverse would be distributed according to a gamma distribution as described in chapter 6, then π_m becomes infinite, thereby reversing the equity premium puzzle. Let us consider alternatively a model in which $\sigma_c = 3.6\%$ is known, but the growth of log consumption is either *1%* or *3%* with equal propabilities, as in our simple calibration exercise in chapter 6. Taking $\gamma = 2$ as usual, we obtain a term structure of the macro risk premium as described in the following figure. Because this parametric uncertainty magnifies the long term risk, it raises the equilibrium risk premium. The long term risk premium enters in the range of the equity premium observed on financial markets over the last century.

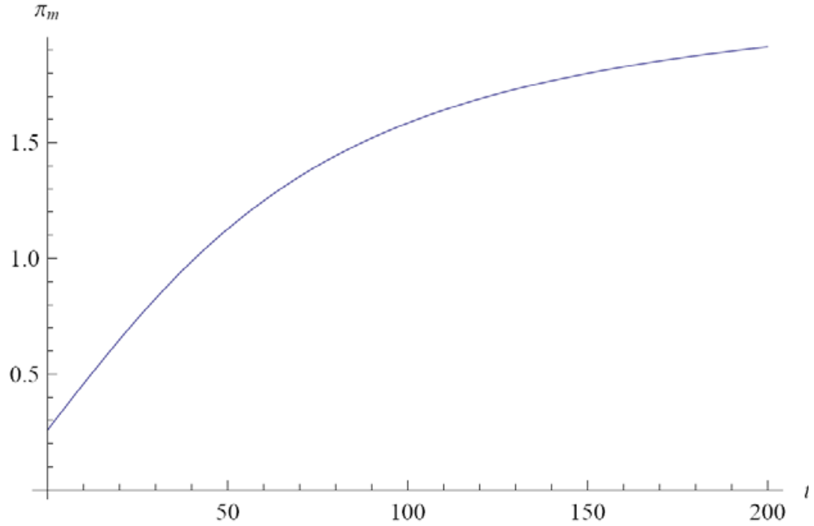


Figure: The term structure of the macro risk premium with $\delta = 0\%$, $\gamma = 2$, $\sigma_c = 3.6\%$ and $\mu_c \sim (1\%, 1/2; 3\%, 1/2)$.

A simple picture emerges from this analysis. For short horizons, the safe discount rate should be relatively large, and the risk premium should be relatively small. However, for longer horizons, one should use a smaller safe discount rate r_f following the methods that we developed in Part II. At the same time, a larger macro risk premium π_m should be used, as justified by argument like the one developed above. This is line with the intuition that if the macro risk increases with time at a rate faster than the one assumed by the standard Brownian model used in finance, then one should do two things. First, it should induce us more effort for the future in general (reduction of the discount rate). Second, it should distort our investment in favour of safer projects.

The capital asset pricing model

In chapter 9, we have justified the use of a representative agent through the existence of efficient risk-sharing schemes in the economy. Real people may have very different von Neumann-Morgenstern preferences, and very heterogeneous income risks or investment projects. Still, if insurance markets are complete, one can assume the existence of a representative agent who consumes the income per capita in the economy, and who gets a fair share of the cash flows of the investment project under consideration. The efficiency of the allocation of risk in the economy implies that all agents will value collective investment projects in the same way. People will unanimously accept or reject them. We systematically used this property of competitive complete markets in this manuscript.

Since Townsend (1994), economists have tested the efficiency of risk sharing in our economies. The general tone of the results obtained in this literature is that risk are not shared efficiently, even in small rural villages of developing countries where stronger informal incentive devices exist to control risk transfers. It implies that different people exposed to different risks will value collective investment differently. Consider for example an investor who is fully invested in a diversified portfolio of risky assets, and has no other source of income than this investment. Therefore, the income of this investor is the return of that stocks portfolio, which is denoted r_t^p. This can well be representative of the community of large investors on financial markets. From their specific point of view, how will they value an investment project? Their intertemporal welfare will be written as

$$W^p(\varepsilon) = u(r_0^p - \varepsilon) + e^{-\delta t} Eu(r_1^p + \varepsilon B_1), \tag{12.21}$$

where the investment project consists in investing ε today for a risky payoff εB_1 at date *1*. The same methodology as above can be used to get a symmetric result. These investors will use a risk-adjusted discount rate

$$r(\beta^p) = r_f + \beta^p \pi^p, \tag{12.22}$$

where β^p measures the correlation of the return of the project with the investor's portfolio rather than with the macro risk, and $\pi^p = \gamma\left(\sigma^p\right)^2$ is the risk premium associated to that portfolio.

The capital asset pricing model developed in the 60's used the capital market as the representative portfolio of investors to price assets. Other reference portfolios or income profiles may be used. The fact that people facing different risks will evaluate collective project in different ways confronts collective decision makers to a difficult challenge. This tells us that the process of valuing an investment project can in general not be disentangled from the question of who will bear the risk.

Valuing the reduction of inequalities

A side product of the analysis presented in this chapter is related to evaluate projects that reduce (or increase) inequalities in our society. Suppose that the economy is composed of *N* agents, indexed by *i=1,…,N*. Let q_i be the Pareto-weight of agent *i* in the social welfare function, with $\Sigma_i q_i = 1$, and let c_{it} denote his consumption at date *t*. Consider an investment project whose sure payoffs are not distributed homogeneously in the population, yielding potentially an increase or a reduction of income inequalities. Let B_{it} be the benefit accruing to agent *i* at date *t*. One can define a inequality-neutral payoff F_t à la Dalton-Atkinson:

$$\sum_{i=1}^{N} q_i u(c_{it} + \varepsilon B_{it}) = \sum_{i=1}^{N} q_i u(c_{it} + \varepsilon F_t) \tag{12.23}$$

For a marginal investment, we obtain

$$F_t = \frac{\sum_{i=1}^{N} q_i B_{it} u'(c_{it})}{\sum_{i=1}^{N} q_i u'(c_{it})} = \frac{EB_t u'(c_t)}{Eu'(c_t)}, \tag{12.24}$$

Where the expectation operator is with respect to (B,c) which takes value (B_{it}, c_{it}) with probability q_i, under the veil of ignorance. Equation (12.24) is formally equivalent to (12.6), and the same methodology as the one developed to evaluate the risk premium can be used to evaluate the "inequality premium". In particular, if (B,c) exhibits positive stochastically dependence, i.e., if the project raises the income inequality at date *t*, the inequality-neutral payoff will be smaller than the Pareto-weighted average payoff.

References

Arrow, K.J., and R.C. Lind, (1970), Uncertainty and the evaluation of public investment decision, *American Economic Review*, 60, 364-378.

Barro, R.J., (2006). "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823-866.

Dimson, E., P. Marsh and M. Staunton, (2002), *Triumph of the Optimists: 101 Years of lobal Investment Returns*, Princeton University Press, Princeton.

Kocherlakota, N.R., (1996), The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature*, 34, 42-71.

Lucas, R., (1978), Asset prices in an exchange economy, *Econometrica*, 46, 1429-46.

Mehra, R. and E. Prescott, (1985), The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 10, 335-339.

Segal, U. and A. Spivak, (1990), First order versus second order risk aversion, *Journal of Economic Theory*, 51, 111-125.

Townsend, R.M., (1994), Risk and insurance in village India, *Econometrica*, 62, 539-592.

Weitzman, M. L., (2007), Subjective expectations and asset-return puzzle, *American Economic Review*, 97, 1102-1130.

Evaluation of uncertain projects

Discounting non-marginal projects

Ph. Trainar

Appendix IV

Slides of the talk by Christian Gollier

Séance inaugurale de la Chaire Financière de la Cité - TSE

L’évaluation à court terme
des stratégies d’investissement à long terme :

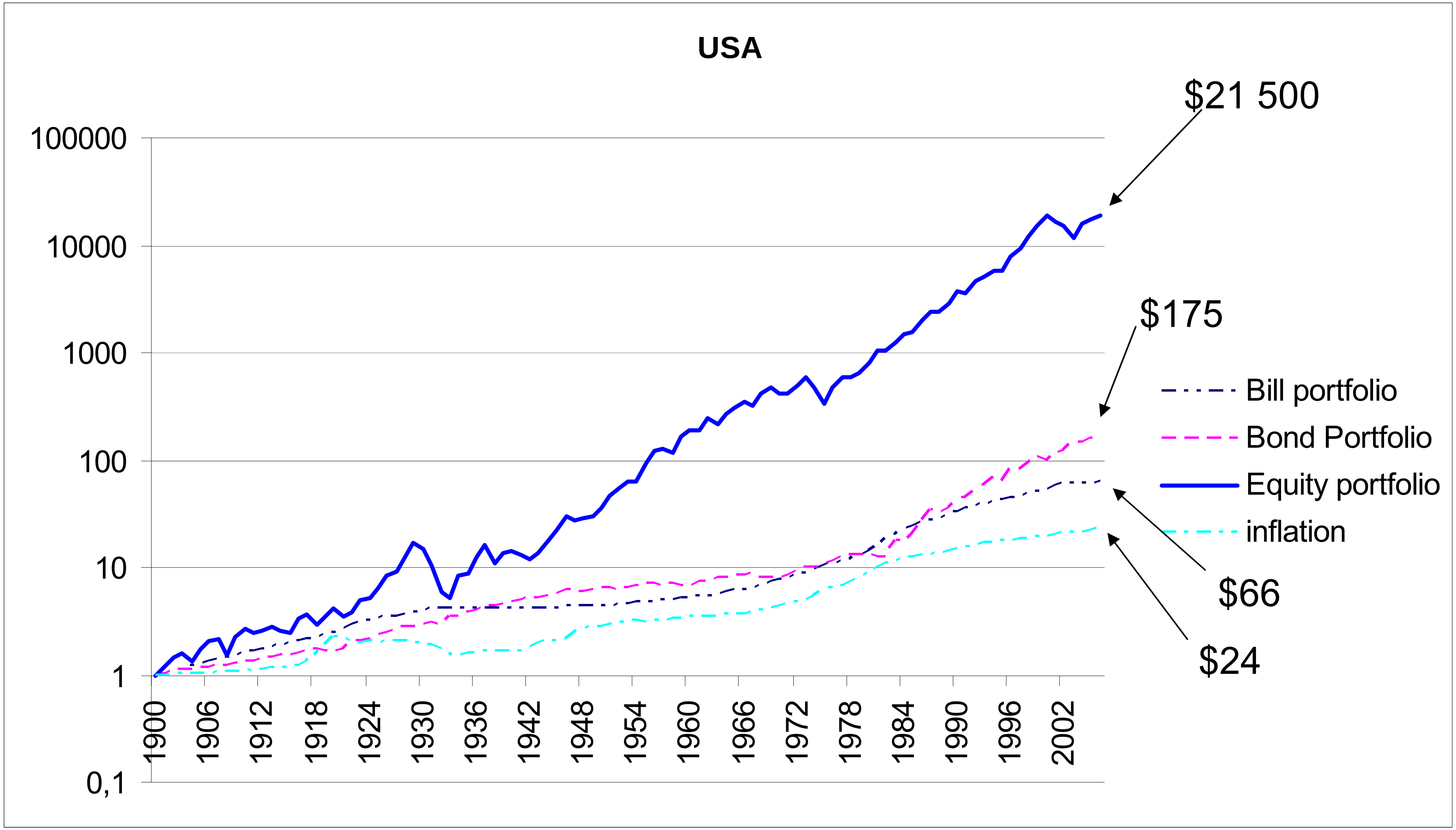
L’allocation d’actifs au risque des nouvelles
normes prudentielles

FINANCIĒRE DE LA CITĒ

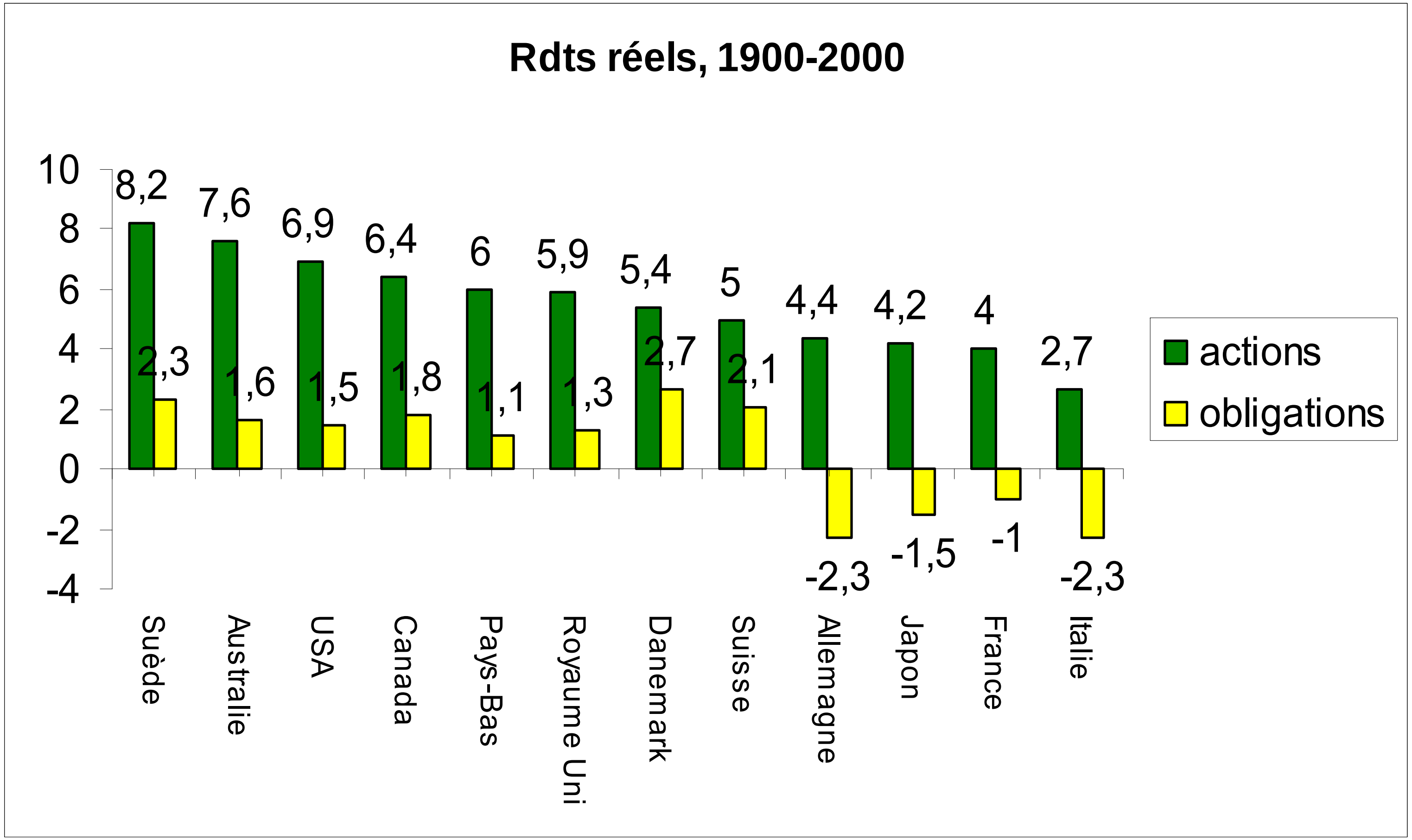
Christian Gollier

Toulouse School of Economics

Le décor: Le paradoxe de la prime de risque



Les autres marchés



Estimez votre propre aversion au risque !

- Supposons que votre richesse soit sujette à un risque de gain ou de perte de α%.
- Quel pourcentage de votre richesse êtes-vous prêt à payer pour éliminer ce risque?

RRA	α=10%	α=30%
γ=0.5	π=0.3%	π=2.3%
γ=1	π=0.5%	π=4.6%
γ=4	π=2.0%	π=16.0%
γ=10	π=4.4%	π=24.4%
γ=40	π=8.4%	π=28.7%

Niveau d'aversion au risque estimé → (γ=1 à γ=4)

Niveau d'aversion au risque nécessaire pour expliquer les prix → (γ=40)

Lien entre risque et temps

- Histoire de P. Samuelson (1963): Est-ce que l'option de prendre des risques à l'avenir augmente la tolérance au risque aujourd'hui?
- Une interprétation fallacieuse de la loi des grands nombres.
- Diversifier, c'est subdiviser des risques.
- Un assureur qui couvre deux fois plus de risques indépendants a une variance de son résultat deux fois plus élevée.
- Donc: le lien entre allocation de portefeuille et horizon de détention est ambigu.

Une illusion d'optique

- Loi des Grands Nombres: lorsque l'horizon devient grand, le *rendement annuel moyen* du portefeuille d 'actions tend vers sa moyenne.
- L 'incertitude sur le *rendement moyen* se réduit de façon exponentielle, mais sur un capital qui croît de façon exponentielle.
- Ce qui importe pour les ménages, c'est le capital global accumulé, pas le rendement annuel moyen de l 'investissement.
- Le risque sur le *capital* ne se réduit pas avec le temps, bien au contraire!

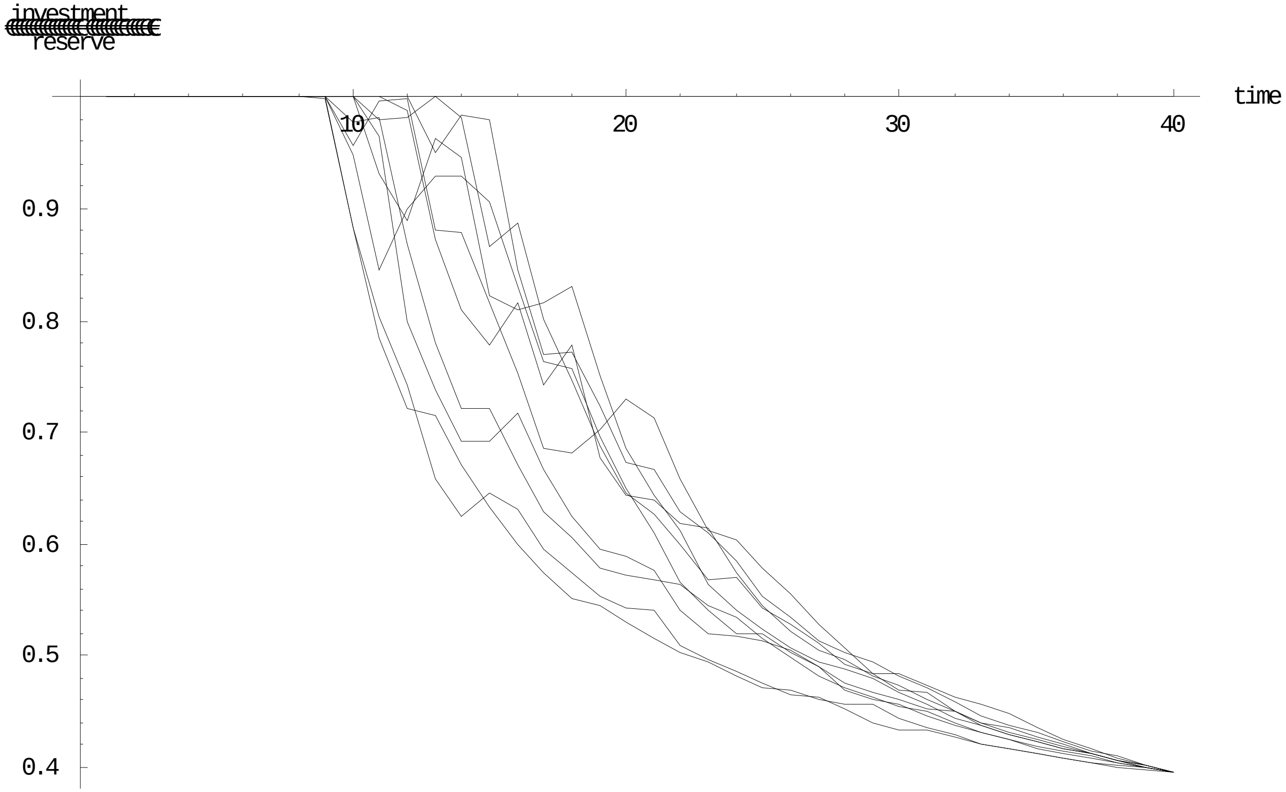
Le résultat de Merton-Samuelson

- Si
  - votre aversion au risque est indépendante de votre richesse;
  - le rendement de l'actif risqué n'a pas de corrélation sérielle;
  - votre objectif d'épargne est pour une dépense à une date précise;
- Alors votre stratégie dynamique optimale consiste à investir une part fixe de votre richesse dans l'actif risqué.
- Gollier-Zeckhauser (2002): Extension à d'autres préférences.

Rôle du capital humain

- Prendre en compte le capital humain dans la mesure de la richesse.
- La richesse des jeunes est souvent essentiellement composée du capital humain. Il baisse avec l'âge, et se transforme partiellement en capital financier.
- Prendre en compte le béta de son capital humain.
- Prendre en compte la flexibilité ex-post:
  - Offre de travail;
  - Adaptation du niveau de vie (immobilier).

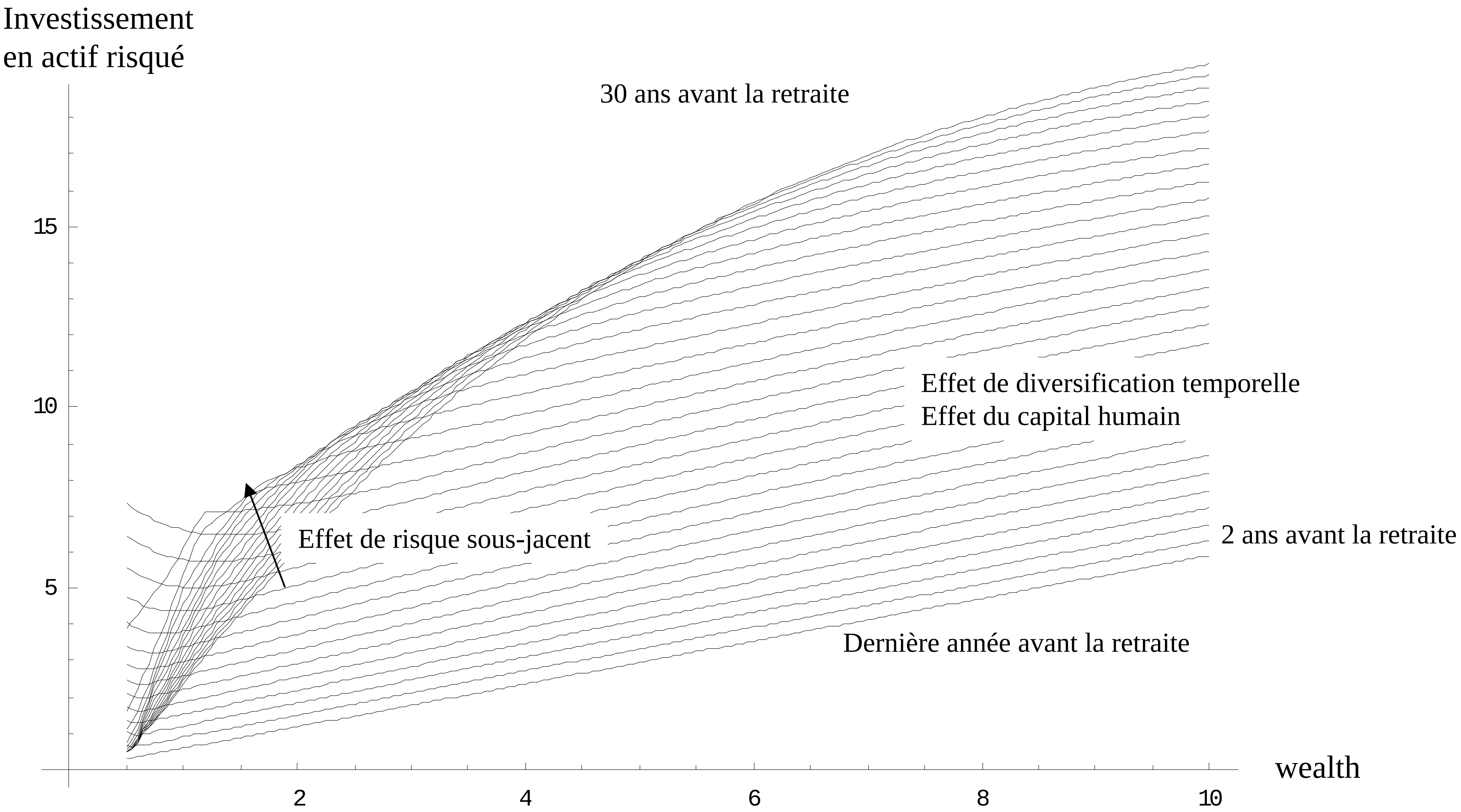
Portefeuille optimal



Diversification temporelle

- Idéalement, un risque sur la richesse se traduit en de petites variations de la consommation sur le cycle de vie.
- Stratégie de lissage des chocs financiers par l'accumulation/réduction d'épargne.
- Quand ce processus est actif, la part investie en risque est proportionnelle à la durée résiduelle de vie.
- Mais l'épargne liquide est faible, ce qui limite l'efficacité de ce système de diversification temporelle.
- Aversion décroissante avec la richesse.

Portefeuille optimal en période d'emploi



Que savons-nous des stratégies individuelles?

Table: Premium Allocation Patterns by Age, June 2000, TIAA-CREF

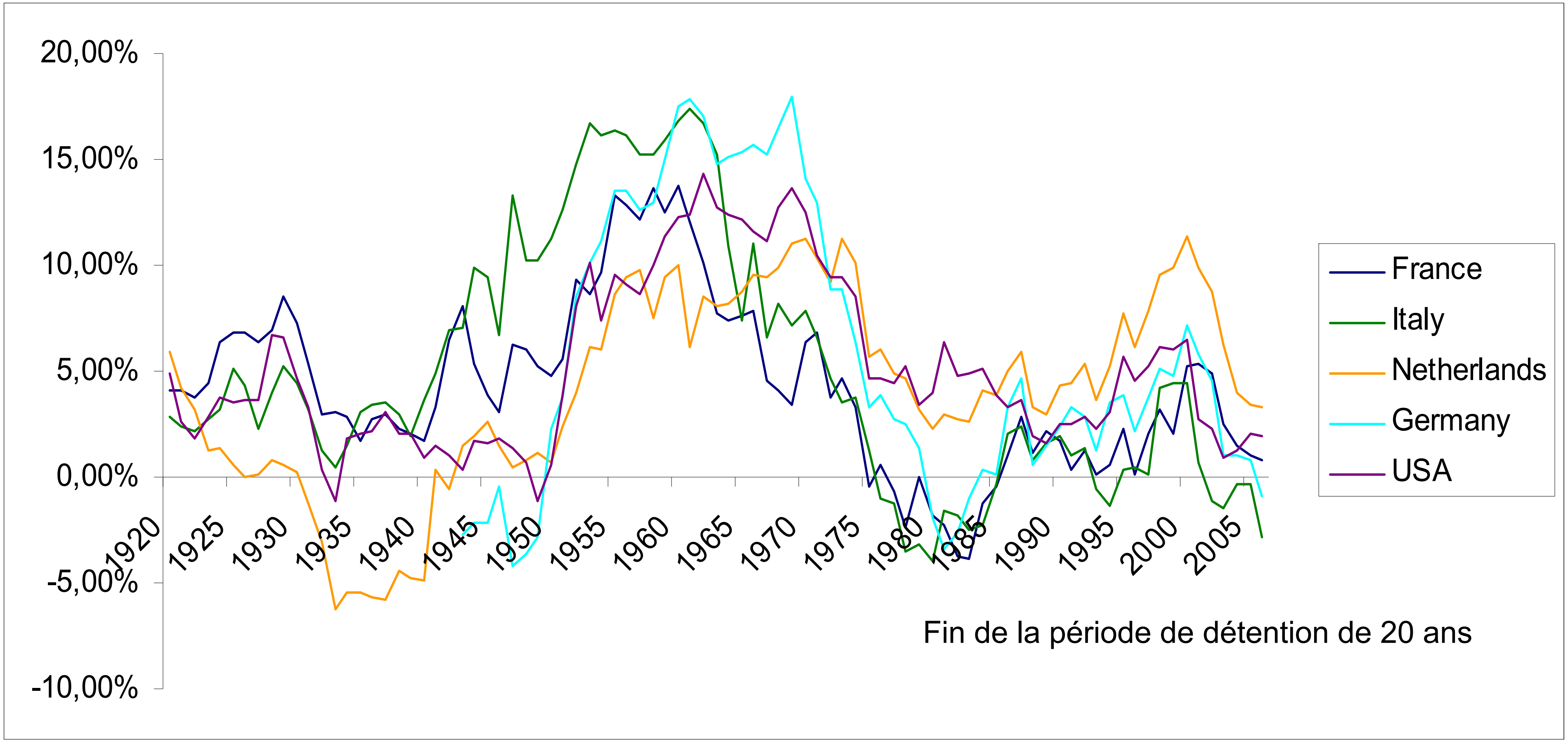
	June 2000			
Allocation Pattern	Under 35	35-44	45-54	55+
100% Guaranteed (« Gtd. »)	4.0%	5.2%	6.5%	9.4%
100% Equity	34.9	32.6	30.1	28.1
100% Fixed Income	7.4	4.1	3.2	3.9
50% Equity, 50% Gtd.	3.8	8.3	14.1	18.7
Mostly Guaranteed				
75.1%-99.9% Gtd.	0.5	0.8	1.0	1.2
50.1%-75% Gtd.	2.2	3.6	5.2	6.7
Mostly Equity				
75.1%-99.9% Equity	17.1	13.8	10.3	6.6
50.1%-75% Equity	19.5	20.4	19.5	17.2
Mostly Fixed Income				
75.1%-99.9% Fixed Income	0.3	0.2	0.2	0.2
50%-75% Fixed Income	3.2	3.2	2.8	2.6
Other combinations	7.1	7.8	7.2	5.4
Total	100.0	100.0	100.0	100.0
Any Real Estate	18.2	10.9	7.8	5.1

Sources: TIAA-CREF Institute Research (2000), and TIAA-CREF Actuarial Technical (1986)

Un premier éclairage sur Solvency II

- Les épargnants ont un horizon (largement) supérieur à un an. Il est socialement désirable qu'ils prennent une part importante du risque collectif.
- Le régulateur impose aux intermédiaires une évaluation des risques à 12 mois, indépendamment de l'horizon temporel des porteurs de risque.
- Seuls les court-termistes investiront dans de tels systèmes.

Ecarts de rendements annualisés entre les actions et les obligations sur des périodes de détention de 20 ans



Diversification intergénérationnelle

- L'assurance-vie et les systèmes de retraite ont la capacité de mutualiser/diversifier les risques entre générations d'épargnants.
- Nécessite une gouvernance et un contrôle intelligent/flexible de l'institution sur le long terme.
- Le bénéfice de ce partage de risque entre générations est équivalent en termes de bien-être à une augmentation de 1% du rendement annuel de l'épargne.

Un deuxième éclairage sur Solvency II

- La communauté des assurés, présents et à venir, offre un horizon très long à l'intermédiaire qui les représente.
- Cette mutualisation intergénérationnelle implique une prise de risque désirable plus importante.
- Solvency II ne reconnaît pas cette raison d'être des assureurs-vie.

Un troisième éclairage sur Solvency II

- Les systèmes d'épargne longue collective offrent souvent une assurance de portefeuille.
- Cela limite la solidarité intergénérationnelle, détruit de la valeur, force les assureurs à réduire leur prise de risque, impose une régulation de la solvabilité.
- Quelle contrepartie, quel coût à l'assurance de portefeuille de masse?
- Vers la fin de l'assurance-vie de bon-papa?
- …sans aller jusqu'à l'extrême des Unités de compte!

Cyclicité des marchés

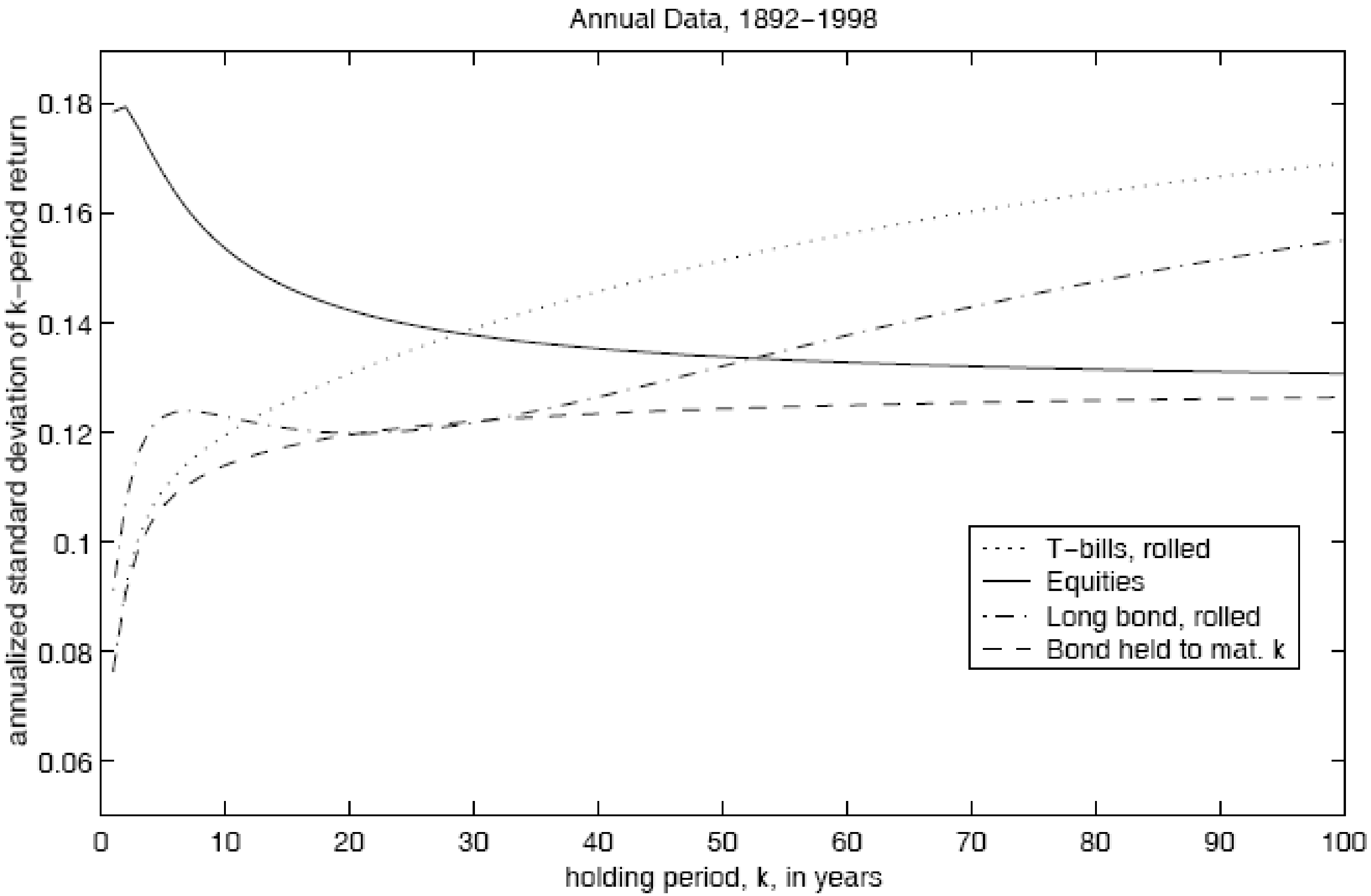
- Les rendements des actifs sont partiellement prévisibles.
- Dès lors, les investisseurs de long terme devraient
  - Biaiser leur portefeuille vers les actifs dont les rendements sont « mean-reverting »;
  - Avoir une allocation contra-cyclique.
- Rôle de la régulation de la solvabilité des intermédiaires financiers.

Une illustration : Cochrane 2008
Données annuelles, USA, 1926-2004

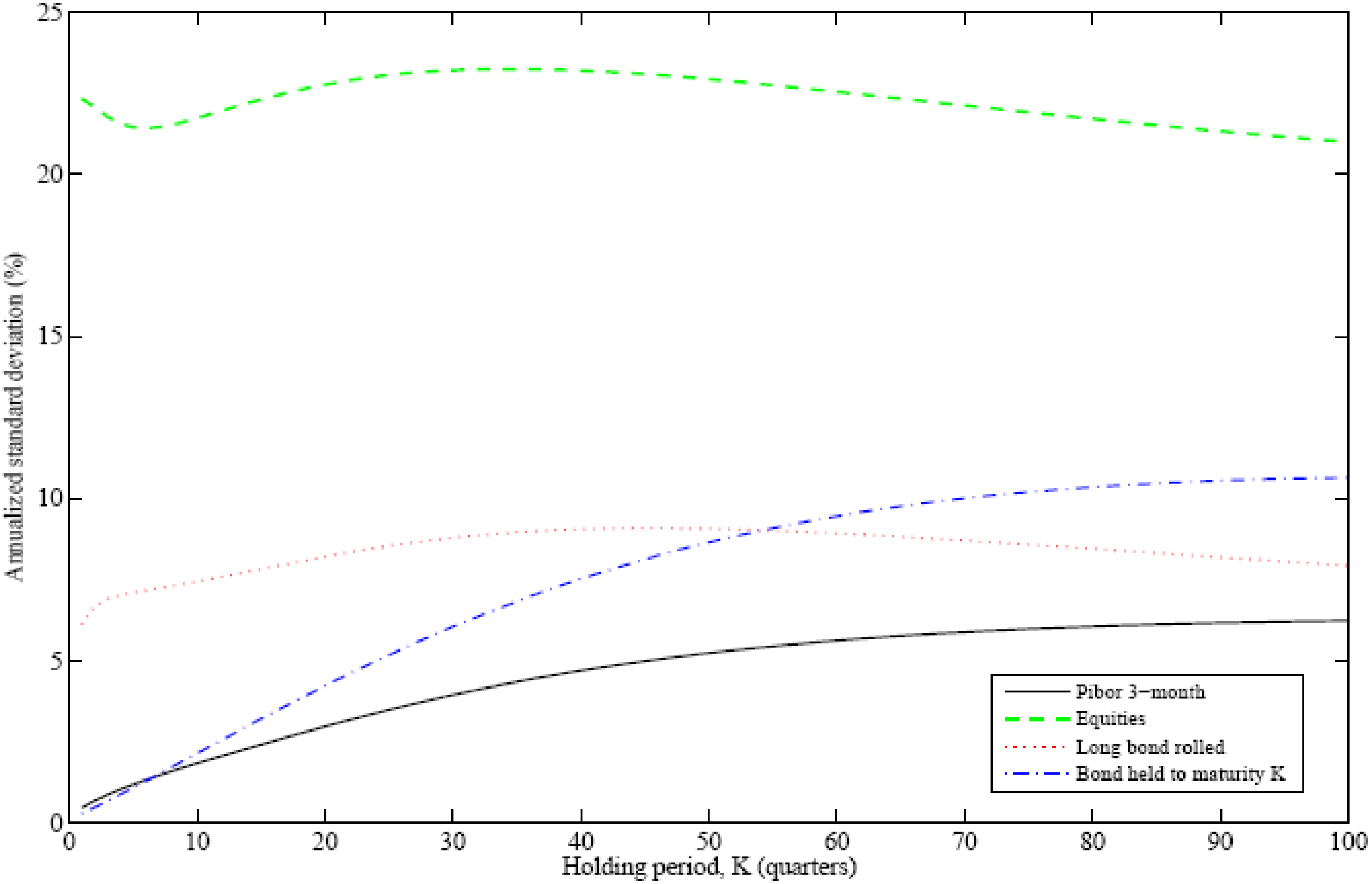
$$R_{t+1} - R_t^f = a + 3.83\left(\frac{D_t}{P_t}\right) + \varepsilon_{t+1}.$$

- *t*-statistic du coefficient D/P : 2.61
- R^2: 7.4%

Retour à la moyenne (Campbell, USA)



Retour à la moyenne (Bec-Gollier, France)



Estimation de la prévisibilité pour la France 1970-2006 Bec and Gollier (2008)

Le modèle prédictif

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + v_t,$$

$r_{0t} = \log(R_{0t})$: log (or continuously compounded) real short term rate;

$x_{et} = r_{et} - r_{0t}$: excess log return of stocks;

$x_{bt} = r_{bt} - r_{0t}$: excess log return of long bonds;

r_{0t}^{nom} : log nominal short term rate;

$ldmp_t$: log dividend-price ratio;

spr_t : difference between the 10-year bond yield and the 3-month rate.

L'estimation du modèle prédictif

	$r_{0,t}$	$x_{e,t}$	$x_{b,t}$	$r_{0,t}^{nom}$	$ldmp_t$	spr_t
$r_{0,t-1}$	**0.931**	**6.444**	**1.747**	-0.040	**-0.109**	-0.156
	(0.058)	(2.833)	(0.793)	(0.058)	(0.033)	(0.211)
	[15.91]	[2.27]	[2.20]	[-0.69]	[-3.33]	[-0.74]
$x_{e,t-1}$	0.001	-0.016	-0.024	0.001	-0.000	-0.002
	(0.002)	(0.091)	(0.026)	(0.002)	(0.001)	(0.007)
	[0.42]	[-0.17]	[-0.94]	[0.59]	[-0.38]	[-0.33]
$x_{b,t-1}$	-0.005	**0.620**	**0.243**	**-0.021**	-0.006	**0.054**
	(0.006)	(0.307)	(0.086)	(0.006)	(0.003)	(0.023)
	[-0.79]	[2.02]	[2.82]	[-3.29]	[-1.74]	[2.35]
$r_{0,t-1}^{nom}$	0.078	-4.217	-0.238	**1.014**	**0.053**	-0.027
	(0.047)	(2.261)	(0.633)	(0.046)	(0.026)	(0.168)
	[1.68]	[-1.86]	[-0.37]	[21.89]	[2.04]	[-0.16]
$ldmp_{t-1}$	-0.146	**8.617**	0.933	-0.074	**0.850**	0.153
	(0.083)	(3.997)	(1.118)	(0.082)	(0.046)	(0.297)
	[-1.76]	[2.15]	[0.83]	[-0.91]	[18.39]	[0.52]
spr_{t-1}	**0.070**	1.537	**0.618**	**0.043**	**-0.021**	**0.729**
	(0.018)	(0.893)	(0.250)	(0.018)	(0.010)	(0.066)
	[3.78]	[1.72]	[2.47]	[2.35]	[-2.09]	[10.98]
c	-0.899	**46.024**	3.697	-0.414	**-0.757**	1.155
	(0.460)	(22.278)	(6.235)	(0.456)	(0.258)	(1.656)
	[-1.95]	[2.07]	[0.59]	[-0.91]	[-2.94]	[0.70]
R-squared	0.88	0.10	0.12	0.93	0.95	0.66

Standard errors in ( ) and t-statistics in [ ].

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + v_t,$$

	r_0	x_e	x_b	r_0^{nom}	$ldmp_t$	spr_t
r_0	0.232	-0.267	-0.377	0.806	0.145	-0.742
x_e	—	11.230	0.243	-0.301	-0.804	0.198
x_b	—	—	3.142	-0.571	-0.231	0.071
r_0^{nom}	—	—	—	0.230	0.250	-0.857
$ldmp$	—	—	—	—	0.130	-0.150
spr	—	—	—	—	—	0.835

Stratégie dynamique optimale

$$U_t = U(C_t, U_{t+1}, z_t) = \left[(1-\delta) C_t^{1-\psi^{-1}} + \delta \left(E_t U_{t+1}^{1-\gamma} \right)^{\frac{1-\psi^{-1}}{1-\gamma}} \right]^{\frac{1}{1-\psi^{-1}}},$$

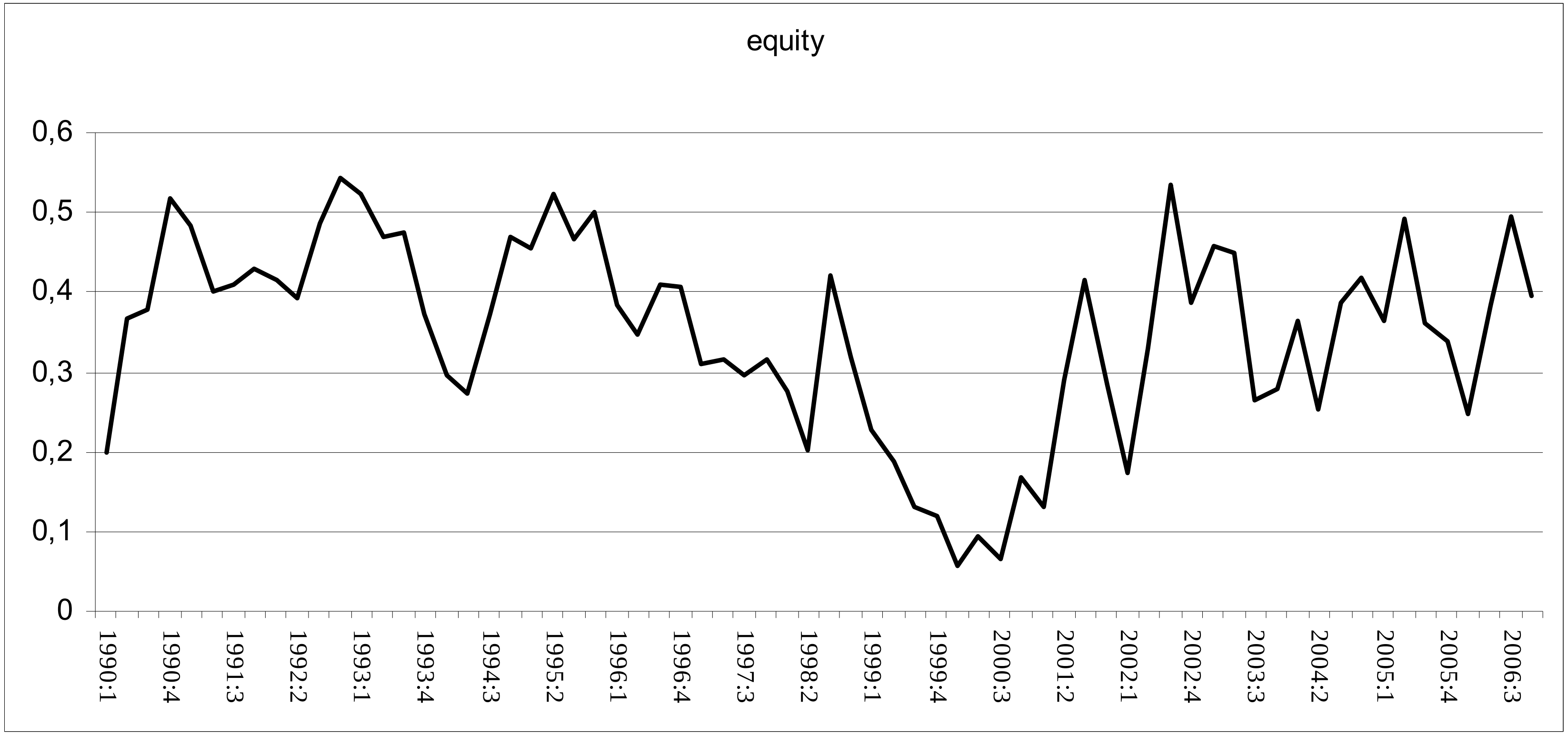
$$\alpha_t = A_0 + A_1 z_t,$$

$$\log\left(\frac{C_t}{W_t}\right) = b_0 + B_1' z_t + z_t' B_2 z_t,$$

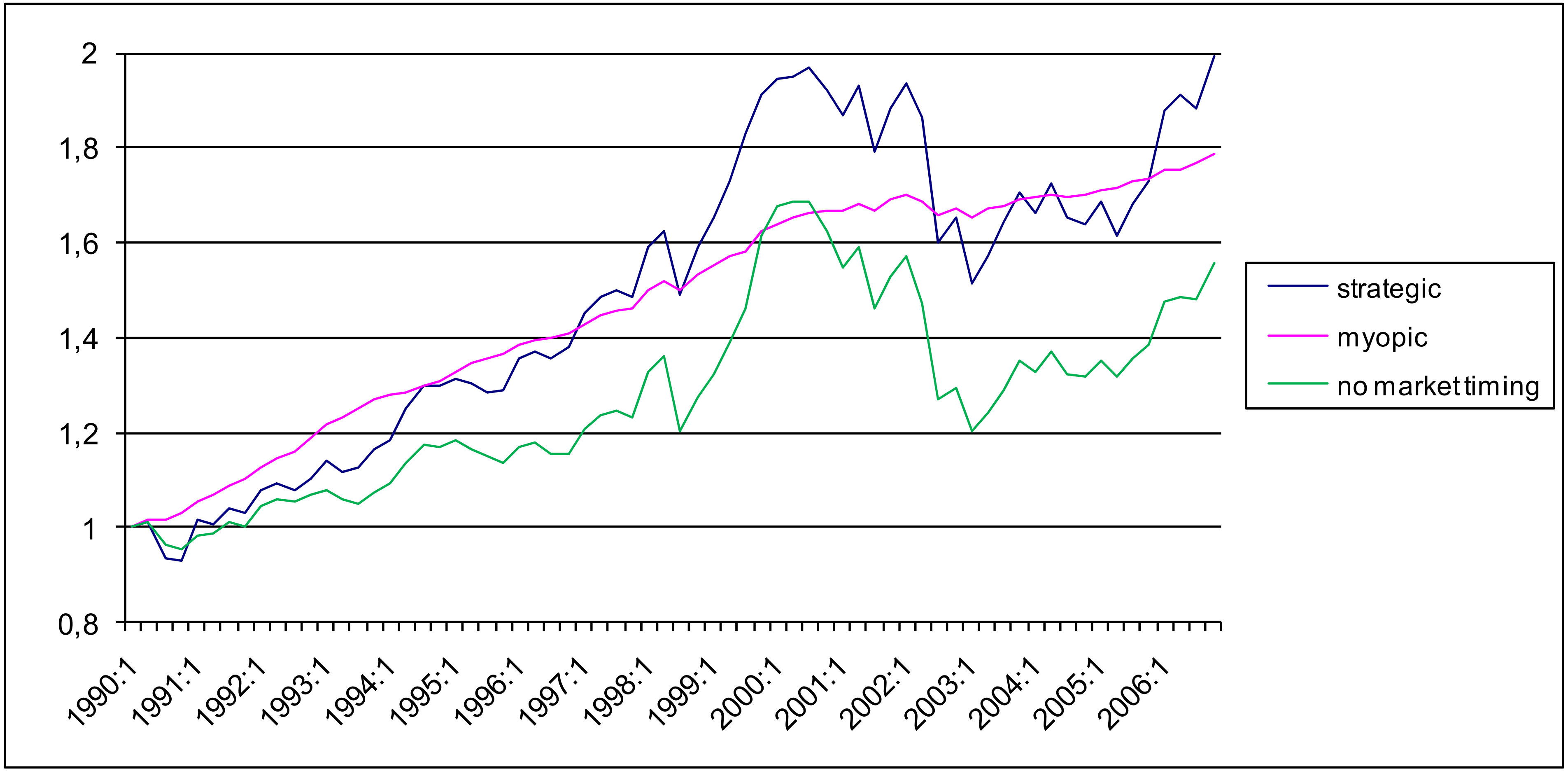
$$\alpha_e = \alpha_{e0} + 25.6\, r_0 + 0.006\, x_e + 1.89\, x_b - 19.61\, r^{nom} + 0.39\, ldmp + 0.055\, spr$$

$$\alpha_b = \alpha_{b0} + 46.42\, r_0 - 1.15\, x_e + 9.90\, x_b + 14.23\, r^{nom} - 0.06\, ldmp + 0.209\, spr$$

Stratégie d'allocation



Performance



Un quatrième éclairage sur Solvency II

- La demande d'actifs risqués des investisseurs de long terme est, de façon optimale, anticyclique.
- Le régulateur SII ne reconnaît pas que la VAR est anticyclique.
  - Il favorise la prise de risque quand il ne le faut pas.
  - Il force à la réduction de risque quand il faudrait l'inciter.
- Il faudrait une exigence en capital basée sur une VAR dépendante de l'état des variables prédictives, calculée sur base d'un modèle économétrique certifié (modèle interne?).

Crises et événements extrêmes

Toulouse School of Economics

Evénements rares

Table 1 Declines of 15% or More in Real Per Capita GDP
Part A: 20 OECD Countries in Maddison [2003]

Event	Country	Years	% fall in real per capita GDP
World War I	Austria	1913-19	35
	Belgium	1916-18	30
	Denmark	1914-18	16
	Finland	1913-18	35
	France	1916-18	31
	Germany	1913-19	29
	Netherlands	1913-18	17
	Sweden	1913-18	18
Great Depression	Australia	1928-31	20
	Austria	1929-33	23
	Canada	1929-33	33
	France	1929-32	16
	Germany	1928-32	18
	Netherlands	1929-34	16
	New Zealand	1929-32	18
	United States	1929-33	31
Spanish Civil War	Portugal	1934-36	15
	Spain	1935-38	31
World War II	Austria	1944-45	58
	Belgium	1939-43	24
	Denmark	1939-41	24
	France	1939-44	49
	Germany	1944-46	64
	Greece	1939-45	64
	Italy	1940-45	45
	Japan	1943-45	52
	Netherlands	1939-45	52
	Norway	1939-44	20
Aftermaths of wars	Canada	1917-21	30
	Italy	1918-21	25
	United Kingdom	1918-21	19
	United Kingdom	1943-47	15
	United States	1944-47	28

Table 2 Stock and Bill Returns during Economic Crises

Event	real stock return (% per year)	real bill return (% per year)
World War I		
Austria, 1914-18	--	-4.1
Denmark, 1914-18	--	-6.9
France, 1914-18	-5.7	-9.3
Germany, 1914-18	-26.4	-15.6
Netherlands, 1914-18	--	-5.2
Sweden, 1914-18	-15.9*	-13.1
Great Depression		
Australia, 1928-30	-3.6	8.2
Austria, 1929-32	-17.3*	7.1
Canada, 1929-32	-23.1*	7.1
Chile, 1929-31	-22.3*	--
France, 1929-31	-20.5	1.4
Germany, 1928-31	-14.8	9.3
Netherlands, 1929-33	-14.2*	5.7
New Zealand, 1929-31	-5.6*	11.9
United States, 1929-32	-16.5	9.3
Spanish Civil War		
Portugal, 1934-36	13.4*	3.8
World War II		
Denmark, 1939-45	-3.7*	-2.2
France, 1943-45	-29.3	-22.1
Italy, 1943-45	-33.9	-52.6
Japan, 1939-45	-2.3	-8.7
Norway, 1939-45	1.7*	-4.5
Post-WWII Depressions		
Argentina, 1998-01	-3.6	9.0
Chile, 1981-82	-37.0*	14.0
Indonesia, 1997-98	-44.5	9.6
Philippines, 1982-84	-24.3	-5.0
Thailand, 1996-97**	-48.9	6.0
Venezuela, 1976-84	-8.6*	--

Risques non normaux

- La théorie de la finance est basée sur des hypothèses simplificatrices, dont souvent la normalité des rendements.
- Cygnes noirs et allocation optimale de portefeuille.
- Ambigüité sur les distributions de probabilités et aversion à l'ambigüité.
- Cygnes noirs et aversion à l'ambigüité peuvent expliquer la prime de risque.
- Erreur de deux types: focalisation sur le danger de sous-estimer le risque.
- Faut-il donner le pouvoir aux pessimistes?

Conclusion

Toulouse School of Economics

Messages à retenir

- La myopie est rarement optimale:
  - Capacité des individus à réagir aux chocs: FLEXIBILITE;
  - Information contenue dans les chocs: PREVISIBILITE.
- Incertitude sur les modèles d'évaluation des risques:
  - aversion à l'ambigüité;
  - sensibilités des stratégies aux événements extrêmes.
- Evaluation des risques: Une affaire dépendante de l'horizon temporel.