# Three essays on exchange rate dynamics and model uncertainty

by

**Edouard Tsague Djeutem**

MA, Simon Fraser University, 2011
MA, Economics School of Louvain (UCL & FUNDP), 2010
B.Eng. Statistics, ISSEA, Yaounde, Cameroon, 2006

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts & Social Sciences


**SIMON FRASER UNIVERSITY**
**Springl 2016**

# Approval

**Name:** **Edouard Tsague Djeutem**

**Degree:** **Doctor of Philosophy (Economics)**

**Title:** ***Three essays on exchange rate dynamics and model uncertainty***

**Examining Committee:** **Chair:** Simon Woodcock
Professor Associate
Department of Economics, SFU

**Kenneth Kasa**
Senior Supervisor
Professor
Department of Economics, SFU

**Nicolas Schmitt**
Supervisor
Professor
Department of Economics, SFU

**Bertille Antoine**
Supervisor
Associate Professor
Department of Economics, SFU

**Chris Muris**
Supervisor
Assistant Professor
Department of Economics, SFU

**David Campbell**
Internal Examiner
Associate Professor
Department of Statistics and Actuarial Science

**Yulei Luo**
External Examiner
Associate Professor
Department of Economics, SFU
The University of Hong Kong

**Date Defended:** 19 May 2016

# Abstract

At least since Knight (1921), economists have suspected that the distinction between risk and 'uncertainty' might be important in economics. However, Savage (1954) showed this distinction is meaningless if agents adhere to certain axioms, which seem to be normatively compelling. Savage's Subjective Expected Utility (SEU) model became the dominant paradigm in economics, and remains so to this very day. Still, suspicions that the distinction matters never really died. The Ellsberg Paradox (1961) first raised doubts about the SEU model. Then, Gilboa and Schmeidler (1989) showed how to modify Savage's axioms so that the distinction *does* matter. In their model, agents entertain a *set* of priors, and optimize against the worst-case prior. Finally, Hansen and Sargent (2008) operationalized this new approach by linking it to the engineering literature on 'robust control'. My dissertation applies the Hansen-Sargent framework to the foreign exchange market. I show that if we think of market participants as confronting both uncertainty and risk, then we can easily explain several well known empirical puzzles in the foreign exchange market.

The second chapter of my dissertation, entitled **Robustness and Exchange Rate Volatility**, was published in the *Journal of International Economics* in 2013, and is coauthored with my supervisor, Prof. Kenneth Kasa. This paper uses the monetary model of exchange rates. It assumes investors are aware of their own lack of knowledge about the economy. They respond to their ignorance strategically, by constructing forecasts that are robust to model misspecification. We show that revisions of robust forecasts are more sensitive to new information, and can easily explain observed violations of Shiller's variance bound inequality.

The third chapter, entitled **Model Uncertainty and the Forward Premium Puzzle**, was published in the *Journal of International Money and Finance* in 2014. It studies a standard two-country Lucas (1982) asset-pricing model. The main objective is to understand the determinants of observed excess return in the foreign exchange market. The paper shows that Hansen-Jagannathan (1991) volatility bounds can be attained with both reasonable degrees of risk aversion and empirically plausible detection error probabilities. Hence, excess returns in the foreign exchange market appear to be primarily driven by a 'model uncertainty premium' rather than a risk premium.

The fourth chaper, entitled **Robust Learning in the Foreign Exchange Market**, was recently revised and resubmitted to the *Canadian Journal of Economics*. Following Hansen and Sargent (2010), it assumes agents cope with uncertainty by both *learning* and by formulating robust decision rules. Agents entertain two competing models, differing by the *persistence* of consumption growth. As in my previous paper, agents continue to doubt the specification of each model. It shows that robust learning can not only explain unconditional risk premia in the foreign exchange market, but can also explain the cyclical *dynamics* of risk premia. In particular, an empirically plausible concern for model misspecification and model uncertainty generates a stochastic discount factor that uniformly satisfies the spectral Hansen-Jagannathan bound of Otrok et. al. (2007).

**Keywords:** Volatility, Robustness, Exchange rates, Model uncertainty, Forward premium puzzle, Detection error probability, Robust learning, state dependent risk, persistence

# Dedication

*To my Godchildren Nanan Antolio Edouard and Noumedem Edouard*
*To the memory of my father Tsague Mathieu and to my mother Gnitedem Monique*
*To my stepparents Sobgue Edouard and Sobgue Marie for teaching me the taste of hard work*

*"Above the clouds the sky is always blue."* St Therese of Lisieux

# Acknowledgements

The journey towards my PhD was both intellectually and personally challenging, and could not have been achieved without the help of everyone who supported me - students, professors, housemates and friends. I would like to take a moment to acknowledge all these persons, and a few in particular who greatly contributed to the success of my academic adventure.

With patience, attention and kindness, my senior supervisor, Kenneth Kasa, guided me through the entire process. He shared his theoretical and empirical "encyclopedic knowledge" of economics with me. I learned from him the importance of completing research work efficiently and through prioritization of activities.

I also benefited from the exceptional support of my other supervisors. Nicolas Schmitt reinforced my passion for international trade and supply chain through his classes. His joviality made every office visit memorable. Teaching through example, Bertille Antoine showed me how to be meticulous and to make level-headed decisions. I am also grateful for the openness, availability and pragmatism of Chris Muris.

Many thanks also go to Christoph Luelfesmann who gave me a chance at SFU, and to Simon Woodcock for all his advice, guidance and understanding. I am most grateful to Pierre Nguimkeu and Annick Nembot for their unconditional support. Pierre shepherded me through my Masters and PhD programs at SFU. Many thanks also to my housemate and friend Denis Dogah for all the "Joloff" and words of encouragment. I am particularly humbled and privileged to have met many classmates - among them are: Xiaowen Lei, Victor Hadad, Kevin Chen, Johannes Strobel, Dirk Assman, and Matthew Lewis.

A special note of thanks goes to my family (Djeutem , Sobgue and Tsague), and to my friends : Chidom, Willibrod, Valery, Ornella, Justin, Olive, Elvire, Vincent de Paul, Loudine, Chuma, Fr Fernando, Clemence, Beribelle, Landry, Marie-Therese and Kim Huynh. My gratitude also goes to my stepparents in Vancouver: Berenice and Ron Elliot.

I gratefully acknowledge the financial support of the Peter Kennedy Memorial fund, and thank all the staff and faculty of the department of economics at SFU, especially Gwen Wild and Kathleen Vieira-Ribeiro.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Robustness and Exchange Rate Volatility

This chapter[1] studies exchange rate volatility within the context of the monetary model of exchange rates. We assume agents regard this model as merely a benchmark, or reference model, and attempt to construct forecasts that are robust to model misspecification. We show that revisions of robust forecasts are more volatile than revisions of nonrobust forecasts, and that empirically plausible concerns for model misspecification can explain observed exchange rate volatility. We also briefly discuss the implications of robust forecasts for a number of other exchange rate puzzles.

## 1.1 Introduction

Exchange rate volatility remains a mystery. Over the years, many explanations have been offered - bubbles, sunspots, 'unobserved fundamentals', noise traders, etcetera. Our paper offers a new explanation. Our explanation is based on a disciplined retreat from the Rational Expectations Hypothesis. The Rational Expectations Hypothesis involves two assumptions: (1) Agents know the correct model of the economy (at least up to a small handful of unknown parameters, which can be learned about using Bayes rule), and (2) Given their knowledge of the model, agents make statistically optimal forecasts. In this paper, we try to retain the idea that agents process information efficiently, while at the same time relaxing what we view as the more controversial assumption, namely, that agents know the correct model up to a finite dimensional parameterization.

Of course, if agents don't know the model, and do not have conventional finite-dimensional priors about it, the obvious question becomes - How are they supposed to forecast the future? Our answer is to suppose that agents possess a simple benchmark model of the

[1] This chapter is a joint work with Kenneth Kasa (Dept of Economics, Simon Fraser University, Email: kkasa@sfu.ca) and was published in the Journal of International Economics, 91 (2013) 27-39

economy, containing a few key macroeconomic variables. We further suppose that agents are aware of their own ignorance, and respond to it strategically by constructing forecasts from the benchmark model that are robust to a wide spectrum of potential misspecifications. We show that revisions of robust forecasts are quite sensitive to new information, and in the case of exchange rates, can easily account for observed exchange rate volatility.

Our paper is closely related to prior work by Hansen and Sargent (2008), Kasa (2001), and Lewis and Whiteman (2008). Hansen and Sargent have pioneered the application of robust control methods in economics. This literature formalizes the idea of a robust policy or forecast by viewing agents as solving dynamic zero sum games, in which a so-called 'evil agent' attempts to subvert the control or forecasting efforts of the decisionmaker. Hansen and Sargent show that concerns for robustness and model misspecification shed light on a wide variety of asset market phenomena, although they do not focus on exchange rate volatility. Kasa (2001) used frequency domain methods to derive a robust version of the well known Hansen and Sargent (1980) prediction formula. This formula is a key input to all present value asset pricing models. Lewis and Whiteman (2008) use this formula to study stock market volatility. They show that concerns for model misspecification can explain observed violations of Shiller's variance bound. They also apply a version of Hansen and Sargent's detection error probabilities to gauge the empirical plausibility of the agent's fear of model misspecification. Since robust forecasts are the outcome of a minmax control problem, one needs to make sure that agents are not being excessively pessimistic, by hedging against models that could have been easily rejected on the basis of observed historical time series. Lewis and Whiteman's results suggest that explaining stock market volatility solely on the basis of a concern for robustness requires an excessive degree of pessimism on the part of market participants. Interestingly, when we modify their detection error calculations slightly, we find that robust forecasts *can* explain observed exchange rate volatility.[2]

Since there are already many explanations of exchange rate volatility, a fair question at this point is - Why do we need another one? We claim that our approach enjoys several advantages compared to existing explanations. Although bubbles and sunspots can obviously generate a lot of volatility, these models require an extreme degree of expectation coordination. So far, no one has provided a convincing story for how bubbles or sunspots emerge in the first place. Our approach requires a more modest degree of coordination. Agents must merely agree on a simple benchmark model, and be aware of the fact that this model may be misspecified.[3] It is also clear that noise traders can generate a lot of volatility. However,

[2]This is not the first paper to apply robust control methods to the foreign exchange market. Li and Tornell (2008) show that a particular type of structured uncertainty can explain the forward premium puzzle. However, they do not calculate detection error probabilities. Colacito and Croce (2011a) develop a dynamic general equilibrium model with time-varying risk premia, and study its implications for exchange rate volatility. They adopt a 'dual' perspective, by focusing on a risk-sensitivity interpretation of robust control. However, they do not focus on Shiller bounds or detection error probabilities, as we do here.

[3]On bubbles, see *inter alia* Meese (1986) and Evans (1986). On sunspots, see Manuelli and Peck (1990) and King, Weber, and Wallace (1992). It should be noted that there *are* ways to motivate the emergence

as with bubbles and sunspots, there is not yet a convincing story for where these noise traders come from, and why they aren't driven from the market. An attractive feature of our approach is that, if anything, agents in our model are *smarter* than usual, since they are aware of their own lack of knowledge about the economy.[4]

Our approach is perhaps most closely related to the 'unobserved fundamentals' arguments in West (1987), Engel and West (2004), and Engel, Mark, and West (2007). These papers all point out that volatility tests aren't very informative unless one is confident that the full array of macroeconomic fundamentals are captured by a model.[5] As a result, they argue that rather than test whether markets are 'excessively volatile', it is more informative to simply compute the fraction of observed exchange rate volatility that *can* be accounted for by innovations in observed fundamentals. Our perspective is similar, yet subtlely different. In West, Engel-West, and Engel-Mark-West, fundamentals are only unobserved by the outside econometrician. Agents within the (Rational Expectations) model are presumed to observe them. In contrast, in our model it is the agents themselves who suspect there might be missing fundamentals, in the form of unobserved shocks that are correlated both over time and with the observed fundamentals. In fact, however, their benchmark model could be perfectly well specified. (In the words of Hansen and Sargent, their doubts are only 'in their heads'). It is simply the prudent belief that they *could be* wrong that makes agents aggressively revise forecasts in response to new information.

In contrast to 'unobserved fundamentals' explanations, which are obviously untestable, there is a sense in which our model *is* testable. Since specification doubts are only 'in their heads', we can ask whether an empirically plausible degree of doubt can rationalize observed exchange rate volatility. That is, we only permit agents to worry about alternative models that could have plausibly generated the observed time series of exchange rates and fundamentals, where plausible is defined as an acceptable detection error probability, in close analogy to a significance level in a traditional hypothesis test. We find that given a sample size in the range of 100-150 quarterly observations, detection error probabilities in the range of 10-20% can explain observed exchange rate volatility.

The remainder of the paper is organized as follows. Section 1.2 briefly outlines the monetary model of exchange rates. We assume agents regard this model as merely a benchmark, and so construct forecasts that are robust to a diffuse array of unstructured alternatives. Section 1.3 briefly summarizes the data. We examine quarterly data from 1973:1-2011:3 on

of sunspots via an adaptive learning process (Woodford (1990)), but then this just changes the question to how agents coordinated on a very particular learning rule.

[4] On the role of noise traders in fx markets, see Jeanne and Rose (2002). A more subtle way noise traders can generate volatility is to prevent prices from revealing other traders' private information. This can produce a hierarchy of higher order beliefs about other traders' expectations. Kasa, Walker, and Whiteman (2010) show that these higher order belief dynamics can explain observed violations of Shiller bounds in the US stock market.

[5] Remember, there is an important difference between unobserved fundamentals and unobserved information about observed fundamentals. The latter can easily be accommodated using the methods of Campbell and Shiller (1987) or West (1988).

six US dollar bilateral exchange rates: the Australian dollar, the Canadian dollar, the Danish kroner, the Japanese yen, the Swiss franc, and the British pound. Section 1.4 contains the results of a battery of traditional excess volatility tests: Shiller's original bound applied to linearly detrended data, the bounds of West (1988) and Campbell-Shiller (1987), which are robust to inside information and unit roots, and finally, a couple of more recent tests proposed by Engel and West (2004) and Engel (2005). Although the results differ somewhat by test and currency, a fairly consistent picture of excess volatility emerges. Section 1.5 contains the results of our robust volatility bounds. We first apply Kasa's (2001) robust Hansen-Sargent prediction formula, based on a so-called $H^{\infty}$ approach to robustness, and show that in this case the model actually predicts exchange rates should be far *more* volatile than observed exchange rate volatility. We then follow Lewis and Whiteman (2008) and solve a frequency domain version of Hansen and Sargent's evil agent game, which allows us to calibrate the degree of robustness to detection error probabilities. This is accomplished by assigning a penalty parameter to the evil agent's actions. We find that observed exchange rate volatility can be explained if agents are hedging against models that have a 10-20% chance of being the true data-generating process. Section 1.6 relates robustness to other puzzles in the foreign exchange market. In particular, we show that robust forecasts can explain the forward premium puzzle. In fact, explaining the forward premium puzzle is *easier* than explaining the volatility puzzle, since the associated detection error probabilities are larger. Section 1.7 contains a few concluding remarks.

## 1.2 The Monetary Model of Exchange Rates

The monetary model has been a workhorse model in open-economy macroeconomics. It is a linear, partial equilibrium model, which combines Purchasing Power Parity (PPP), Uncovered Interest Parity (UIP), and reduced-form money demand equations to derive a simple first-order expectational difference equation for the exchange rate. It presumes monetary policy and other fundamentals are exogenous. Of course, there is evidence against each of these underlying ingredients. An outside econometrician would have reasons to doubt the specification of the model. Unlike previous variance bounds tests using this model, we assume the agents within the model share these specification doubts.

Since the model is well known, we shall not go into details. (See, e.g., Mark (2001) for a detailed exposition). Combining PPP, UIP, and identical log-linear money demands yields the following exchange rate equation:

$$s_t = (1 - \beta) f_t + \beta E_t s_{t+1} \tag{1.2.1}$$

where $s_t$ is the log of the spot exchange rate, defined as the price of foreign currency. The variable $f_t$ represents the underlying macroeconomic fundamentals. In the monetary model, it is just

$$f_t = (m_t - m_t^*) - \lambda(y_t - y_t^*)$$

where $m_t$ is the log of the money supply, $y_t$ is the log of output, and asterisks denote foreign variables. In what follows, we assume $\lambda = 1$, where $\lambda$ is the income elasticity of money demand. The key feature of equation (1.2.1) is that it views the exchange rate as a an *asset price*. It's current value is a convex combination of current fundamentals, $f_t$, and expectations of next period's value. In traditional applications employing the Rational Expectations Hypothesis, $E_t$ is defined to be the mathematical expectations operator. We relax this assumption. Perhaps not surprisingly, $\beta$ turns out to be an important parameter, as it governs the weight placed on expectations in determining today's value. In the monetary model, this parameter is given by, $\beta = \alpha/(1+\alpha)$, where $\alpha$ is the interest rate semi-elasticity of money demand. Following Engel and West (2005), we assume $.95 < \beta < .99$.

By imposing the no bubbles condition, $\lim_{j\to\infty} E_t \beta^j s_{t+j} = 0$, and iterating eq. (1.2.1) forward, we obtain the following present value model for the exchange rate

$$s_t = (1-\beta)E_t \sum_{j=0}^{\infty} \beta^j f_{t+j} \tag{1.2.2}$$

which expresses the current exchange rate as the expected present discounted value of future fundamentals. If $f_t$ is covariance stationary with Wold representation $f_t = A(L)\varepsilon_t$, then application of the Hansen-Sargent prediction formula yields the following closed-form expression for the exchange rate,

$$s_t = (1-\beta)\left[\frac{LA(L) - \beta A(\beta)}{L-\beta}\right]\varepsilon_t \tag{1.2.3}$$

We shall have occasion to refer back to this in Section 1.5, when discussing robust forecasts. In practice, the assumption that $f_t$ is covariance stationary is questionable. Instead, evidence suggests that for all six countries $f_t$ contains a unit root. In this case, we need to reformulate equation (1.2.2) in terms of stationary variables. Following Campbell and Shiller (1987), we can do this by defining the 'spread', $\phi_t = s_t - f_t$, and expressing it as a function of expected future *changes* in fundamentals, $\Delta f_{t+1} = f_{t+1} - f_t$[6]

$$s_t - f_t = E_t \sum_{j=1}^{\infty} \beta^j \Delta f_{t+j} \tag{1.2.4}$$

Applying the Hansen-Sargent formula to this expression yields the following closed form expression for the spread,

$$s_t - f_t = \beta\left[\frac{A(L) - A(\beta)}{L-\beta}\right]\varepsilon_t \tag{1.2.5}$$

[6]This suggests that $s_t - f_t$ should be stationary, i.e., $s_t$ and $f_t$ are cointegrated. As we show in the next section, the evidence here is more mixed.

where we now assume that $\Delta f_t = A(L)\varepsilon_t$.

## 1.3 The Data

We study six US dollar exchange rates: the Australian dollar, the Canadian dollar, the Danish kroner, the Japanese yen, the Swiss franc, and the British pound. The data are taken from the IFS, and are quarterly end-of-period observations, expressed as the dollar price of foreign currency. Money supply and income data are from the OECD Main Economic Indicators. Money supply data are seasonally adjusted M1 (except for Britain, which is M4). Income data are seasonally adjusted real GDP, expressed in national currency units. All data are expressed in natural logs.

Table 1.1 contains summary statistics for $\Delta s_t$ and $\Delta f_t$. There are three noteworthy fea-

Table 1.1: Summary Statistics: Quarterly Data (1973:1-2011:3)

| | Aus | Can | Den | Jap | Swz | UK |
|---|---|---|---|---|---|---|
| $\underline{\Delta s}$ | | | | | | |
| Mean | $-.0024$ | $-.0002$ | .0007 | .0080 | .0083 | $-.0030$ |
| StDev | .0547 | .0311 | .0576 | .0590 | .0649 | .0529 |
| $\rho_1$ | .067 | .178 | .067 | .060 | $-.008$ | .151 |
| $\underline{\Delta f}$ | | | | | | |
| Mean | $-.0091$ | $-.0071$ | $-.0105$ | $-.0035$ | $-.0053$ | $-.0136$ |
| StDev | .0307 | .0197 | .0277 | .0273 | .0455 | .0236 |
| $\rho_1$ | .197 | .367 | .277 | .394 | .092 | .461 |

Notes: (1) USA is reference currency, with $s$ = price of fx in US dollars.
(2) Both $s$ and $f$ are in natural logs, with $f = m - m^* - (y - y^*)$.
(3) $\rho_1$ is the first-order autocorrelation coefficient.

tures. First, it is apparent that exchange rates are close to random walks, while at the same time changes in fundamentals are predictable. This apparent contradiction was explained by Engel and West (2005). They show that if the discount factor is close to one, as it is with quarterly data, then exchange rates *should* be close to random walks. Second, it is apparent that the mean of $\Delta s_t$ does not always match up well with the mean of $\Delta f_t$. This simply reflects the fact that some of these currencies have experienced long-run *real* appreciations or depreciations. This constitutes prima facie evidence against the monetary model, independent of its volatility implications. Our empirical work accounts for these missing trends by either detrending the data, or by including trends in posited cointegrating relationships. This gives the model the best chance possible to explain exchange rate volatility. Third, and most importantly for the purposes of our paper, note that the standard deviations of $\Delta s_t$ are around twice as large as the standard deviations of $\Delta f_t$. Given the mild persistence in $\Delta f_t$, this is a major problem when it comes to explaining observed exchange rate volatility.

As we discuss in more detail in the following section, variance bounds tests require assumptions about the nature of trends in the data. The original tests of Shiller (1981) presumed stationarity around a deterministic trend. Subsequent work showed that this causes a bias toward rejection if in fact the data contain unit roots, which then led to the development of tests that are robust to the presence of unit roots (Campbell and Shiller (1987), West (1988)). Table 1.2 contains the usual tests for unit roots and cointegration. Clearly, there is little evidence against the unit root null for both the exchange rate and

Table 1.2: Unit Root and Cointegration Tests

| | Aus | Can | Den | Jap | Swz | UK |
|---|---|---|---|---|---|---|
| Dickey-Fuller | | | | | | |
| $s$ | $-1.43$ | $-1.16$ | $-1.96$ | $-2.08$ | $-2.56$ | $-2.70$ |
| $f$ | $-1.03$ | $-2.09$ | $-1.57$ | $-1.24$ | $-2.27$ | $-1.43$ |
| $(s-f)$ | $-2.05$ | $-1.91$ | $-2.44$ | $-2.37$ | $-3.52$ | $-3.25$ |
| Engle-Granger | | | | | | |
| $(s,f)$ | $-1.44$ | $-1.60$ | $-2.49$ | $-4.16$ | $-3.62$ | $-3.30$ |
| Johansen (Trace) | | | | | | |
| $(s,f)$ | 4.73 | 11.18 | 11.33 | 20.12 | 23.51 | 19.21 |

Notes: (1) DF and EG regressions include constant, trend, and four lags.
(2) $(5\%, 10\%)$ critical values for $DF = (-3.44, -3.14)$.
(3) $(5\%, 10\%)$ critical values for $EG = (-3.84, -3.54)$.
(4) Johansen based on VAR(4) with deterministic trend. 5% critical value = 18.40.

fundamentals. This casts doubt on the applicability of the original Shiller bound. One can also see that the cointegration evidence is more mixed. There is some evidence in favor of cointegration for Japan, Switzerland, and the UK (especially when using the Johansen test), but little or no evidence for Australia, Canada, or Denmark. Later, when implementing tests based on the unit root specification, we simply assume the implied cointegration restrictions hold.

## 1.4 Traditional Volatility Tests

As motivation for our paper, this section briefly presents results from applying traditional volatility tests. We present them roughly in order of their historical development. Given our purposes, we do not dwell on the (many) statistical issues that arise when implementing these tests.[7]

### 1.4.1 Shiller (1981)

The attraction of Shiller's original variance bound test is that it is based on such a simple and compelling logic. Shiller noted that since asset prices are the conditional expectation

[7] See Gilles and LeRoy (1991) for an excellent summary of these issues.

of the present value of future fundamentals, they should be *less* volatile than the ex post realized values of these fundamentals. More formally, if we define $s_t^* = (1-\beta)\sum_{j=0}^{\infty}\beta^j f_{t+j}$ as the ex post realized path of future fundamentals, then the monetary model is equivalent to the statement that $s_t = E_t s_t^*$. Next, we can always decompose a random variable into the sum of its conditional mean and an orthogonal forecast error,

$$\begin{aligned} s_t^* &= E_t s_t^* + u_t \\ &= s_t + u_t \end{aligned}$$

Since by construction the two terms on the right-hand side are orthogonal, we can take the variance of both sides and get

$$\sigma_{s^*}^2 = \sigma_s^2 + \sigma_u^2 \qquad \Rightarrow \qquad \sigma_s^2 \leq \sigma_{s^*}^2$$

It's as simple as that.[8] Two practical issues arise when implementing the bound. First, prices and fundamentals clearly trend up over time. As a result, neither possesses a well defined variance, so it is meaningless to apply the bounds test to the raw data. Shiller dealt with this by linearly detrending the data. To maintain comparability, we do the same, although later we present results that are robust to the presence of unit roots. Second, future fundamentals are obviously unobserved beyond the end of the sample. Strictly speaking then, we cannot compute $\sigma_{s^*}^2$, and therefore, Shiller's bound is untestable. One can always argue that agents within any given finite sample are acting on the basis of some as yet unobserved event. Of course, this kind of explanation is the last refuge of a scoundrel, and moreover, we show that it is unnecessary. Shiller handled the finite sample problem by assuming that $s_T^*$, the end-of-sample forecast for the discounted present value of future fundamentals, was simply given by the sample average. Unfortunately, subsequent researchers were quick to point out that this produces a bias toward rejection. So in this case, we depart from Shiller by using the unbiased procedure recommended by Mankiw, Romer, and Shapiro (1985). This involves iterating on the backward recursion $s_t^* = (1-\beta)f_t + \beta s_{t+1}^*$, with the boundary condition, $s_T^* = s_T$.

The first row of Table 1.3 reports results from applying Shiller's bound. Evidently, rather than being less volatile than ex post fundamentals, exchange rates are actually between 3 and 32 times as volatile as fundamentals.[9] Perhaps the most striking result that Shiller

[8] As emphasized by Kasa, Walker, and Whiteman (2010), matters aren't quiet so simple in models featuring heterogeneous beliefs. The law of iterated expectations does not apply to the average beliefs operator. They show conventional applications of Shiller's bound can easily generate false rejections when there are heterogeneous beliefs.

[9] By way of comparison, Shiller (1981) found that U.S. stock prices were 5 times as volatile as fundamentals. In the first application to exchange rates, Huang (1981) found that the pound and deutschemark were between 3 and 10 times too volatile. However, Diba (1987) pointed out that Huang's results were tainted by a miscalibrated discount factor. With empirically plausible values of $\beta$, Huang's tests showed no signs of excess volatility.

Table 1.3: Traditional Volatility Tests

| | Aus | Can | Den | Jap | Swz | UK |
|---|---|---|---|---|---|---|
| Shiller<br>$\text{var}(s)/\text{var}(s^*)$ | 3.13 | 7.85 | 25.7 | 2.93 | 14.2 | 32.5 |
| Campbell-Shiller<br>$\text{var}(\phi)/\text{var}(\phi^*)$ | 3.41 | 10.2 | 7.98 | 211.2 | 178.2 | 9.00 |
| West<br>$\text{var}(\varepsilon)/\text{var}(\hat{\varepsilon})$ | 2.36 | 1.81 | 2.33 | 1.34 | 1.73 | 1.77 |
| Engel-West<br>$\text{var}(\Delta s)/\text{var}(\Delta x_H)$ | 1.83 | 0.99 | 1.66 | 1.30 | 1.46 | 1.15 |
| Engel<br>$\text{var}(\Delta s)/\text{var}(\Delta \hat{s})$ | 5.98 | 10.8 | 3.35 | 7.42 | 20.7 | 14.6 |

Notes: (1) Shiller bound based on detrended data, assuming $s_T^* = s_T$ and $\beta = .98$.
(2) Campbell-Shiller bound based on a VAR(2) for $(\Delta f, \phi)$, assuming $\beta = .98$.
(3) Engel-West bound based on AR(2) for $\Delta f$, assuming $\beta = .98 \approx 1$.
(4) Engel bound based on AR(1) for detrended $f_t$, assuming $\beta = .98$.

presented was a simple time series plot of $s_t$ versus $s_t^*$. This, more than anything else, is what convinced many readers that stock prices are excessively volatile.[10] Figure 1 reproduces the Shiller plot for each of our six currencies. These particular plots are based on the assumption $\beta = .98$, but similar plots are obtained for the empirically plausible range, $.95 \leq \beta \leq .99$.

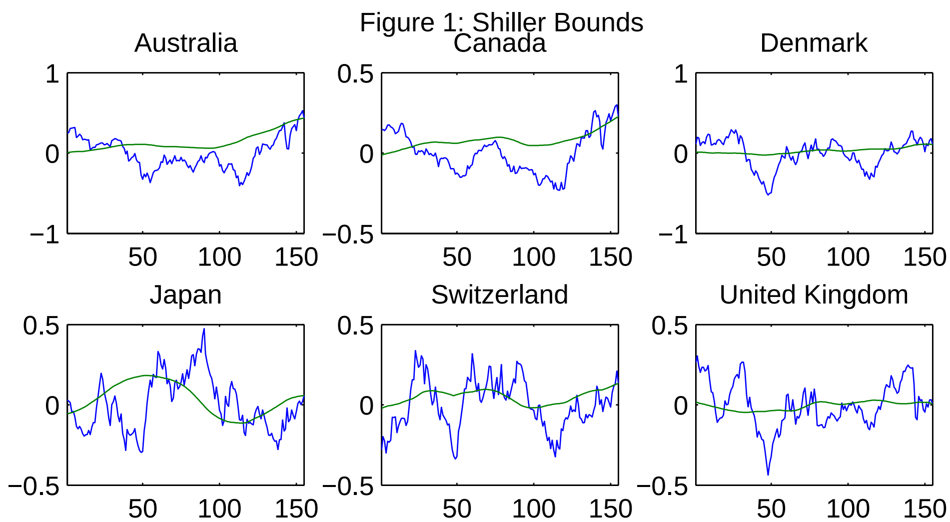
As with Shiller, these plots paint a clear picture of 'excess volatility'. Unfortunately, there are enough statistical caveats and pitfalls associated with these results, that it is worthwhile to consider results from applying some of the more recently proposed bounds tests, starting with the influential work of Campbell and Shiller (1987).

### 1.4.2 Campbell and Shiller (1987)

The results of Shiller (1981) are sensitive to the presence of unit roots in the data. As we saw in Table 1.2, there appear to be unit roots in both $s_t$ and $f_t$ (even after the removal of a deterministic trend). Campbell and Shiller (1987) devised a volatility test that is valid when the data are nonstationary. In addition, they devised a clever way of capturing potential information that market participants might have about future fundamentals that is unobserved by outside econometricians. To do this, one simply needs to include current and lagged values of the exchange rate when forecasting future fundamentals. Under the null, the current exchange rate is a sufficient statistic for the present value of future fundamentals. Forecasting future fundamentals with a VAR that includes the exchange rate converts

[10] Not everyone was convinced. In particular, Kleidon (1986) warned that these plots could be quite misleading if the underlying data are nonstationary.

Figure 1.1: Shiller Bounds




Shiller's variance bound inequality to a variance *equality*. The model-implied forecast of the present value of fundamentals should be identically equal to the actual exchange rate!

To handle unit roots, Campbell and Shiller (1987) presume that $s_t$ and $f_t$ are cointegrated, and then define the 'spread' variable $\phi_t = s_t - f_t$. The model implies that $\phi_t$ is the expected present value of future values of $\Delta f_t$ (see eq. (1.2.4)). If we define the vector $(\Delta f_t, \phi_t)'$ we can then estimate the following VAR,

$$x_t = \Psi(L)x_{t-1} + \epsilon_t$$

By adding lags to the state, this can always be expressed as a VAR(1)

$$\hat{x}_t = \hat{\Psi}\hat{x}_{t-1} + \epsilon_t$$

where $\hat{\Psi}$ is a double companion matrix, and the first element of $\hat{x}_t$ is $\Delta f_t$. The model-implied spread, $\phi_t^*$, is given by the expected present discounted value of $\{\Delta f_{t+j}\}$, which can be expressed in terms of observables as follows,

$$\phi_t^* = e1'\beta\hat{\Psi}(I - \beta\hat{\Psi})^{-1}\hat{x}_t \tag{1.4.6}$$

where $e1$ is a selection vector that picks off the first element of $\hat{x}_t$. The model therefore implies $\text{var}(\phi_t) = \text{var}(\phi_t^*)$.

The second row of Table 1.3 reports values of $\text{var}(\phi_t)/\text{var}(\phi_t^*)$ based on a VAR(2) model (including a trend and intercept). We fix $\beta = .98$, although the results are similar for values in the range $.95 \leq \beta \leq .99$. Although Campbell and Shiller's approach is quite different, the results are quite similar. All ratios substantially exceed unity. Although the Campbell-Shiller method points to relatively less volatility in Denmark and the UK, it actually suggests greater excess volatility for Japan and Switzerland.

### 1.4.3 West (1988)

West (1988) proposed an alternative test that is also robust to the presence of unit roots. Rather than look directly at the volatility of observed prices and fundamentals, West's test is based on comparison of two innovation variances. These innovations can be interpreted as one-period holding returns, and so are stationary even when the underlying price and fundamentals processes are nonstationary. Following West (1988), let $I_t$ be the market information set at time-$t$, and let $H_t \subseteq I_t$ be some subset that is observable by the outside econometrician. In practice, $H_t$ is often assumed to just contain the history of past (observable) fundamentals. Next, define the following two present value forecasts:

$$x_{tH} = \sum_{j=0}^{\infty} \beta^j E(f_{t+j}|H_t) \qquad x_{tI} = \sum_{j=0}^{\infty} \beta^j E(f_{t+j}|I_t)$$

West then derived the following variance bound,

$$E(x_{t+1,H} - E[x_{t+1,H}|H_t])^2 \geq E(x_{t+1,I} - E[x_{t+1,I}|I_t])^2$$

This says that if market participants have more information than $H_t$, their forecasts should have a smaller innovation variance. Intuitively, when forecasts are based on a coarser information set, there are more things that can produce a surprise. We apply this bound to forecasts of the spread, $\phi_t = s_t - f_t$, with $H_t$ assumed to contain both the history of fundamentals *and* exchange rates. From the Campbell-Shiller logic, the inclusion of $s_t$ in $H_t$ converts West's variance bound inequality to a variance *equality*.

To derive the market's innovation, we can write the present value model recursively as follows

$$\phi_t = \beta E_t(\phi_{t+1} + \Delta f_{t+1})$$

Exploiting the decomposition, $\phi_{t+1} + \Delta f_{t+1} = E_t(\phi_{t+1} + \Delta f_{t+1}) + \varepsilon_{t+1}$, we can then derive the ex post observable expression for the market's innovation

$$\varepsilon_{t+1} = \phi_{t+1} + \Delta f_{t+1} - \frac{1}{\beta}\phi_t$$

Note that $\varepsilon_{t+1}$ is just a one-period excess return. We can do the same for the predicted spread, $\hat{\phi}_t$, estimated from the same VAR(2) model used to compute the Campbell-Shiller test. This yields the fitted excess return,

$$\hat{\varepsilon}_{t+1} = \hat{\phi}_{t+1} + \Delta f_{t+1} - \frac{1}{\beta}\hat{\phi}_t$$

Under the null, $\mathrm{var}(\varepsilon) = \mathrm{var}(\hat{\varepsilon})$. The third row of Table 1.3 contains the results. Interestingly, the West test indicates a much less dramatic degree of excess volatility. Although the point estimates continue to exceed unity, most are in the range 1.5-2.0.

### 1.4.4 Engel and West (2004)

Engel and West (2004) use results from Engel and West (2005) to derive a variance bound for the limiting case, $\beta \to 1$. We implement this bound assuming $\beta = .98$. Two things happen as $\beta \to 1$. First, from West (1988) we know

$$\mathrm{var}(\varepsilon_{t,H}) = \frac{1-\beta^2}{\beta^2}\mathrm{var}(x_{t,H} - x_{t,I}) + \mathrm{var}(\varepsilon_{t,I})$$

Since $\mathrm{var}(x_{t,H} - x_{t,I})$ is bounded, it is clear that $\mathrm{var}(\varepsilon_{t,H}) \approx \mathrm{var}(\varepsilon_{t,I})$ as $\beta \to 1$. Second, from Engel and West (2005) we know that $s_t$ converges to a random walk as $\beta \to 1$. Therefore, $\mathrm{var}(\varepsilon_{t,I}) \approx \mathrm{var}(\Delta s_t)$ as $\beta \to 1$. Combining, we conclude that $\mathrm{var}(\varepsilon_{t,H}) \approx \mathrm{var}(\Delta s_t)$ under the null as $\beta \to 1$. To estimate $\mathrm{var}(\varepsilon_{t,H})$ we first estimate a univariate VAR(2) for $\Delta f_t$,

$$\Delta f_t = \hat{\gamma}_0 + \hat{\gamma}_1 \Delta f_{t-1} + \hat{\gamma}_2 \Delta f_{t-2} + u_t$$

We can then estimate $\mathrm{var}(\varepsilon_{t,H})$ as follows

$$\mathrm{var}(\varepsilon_{t,H}) = (1 - \beta\hat{\gamma}_1 - \beta^2\hat{\gamma}_2)^{-2}\mathrm{var}(u)$$

The fourth row of Table 1.3 reports the ratio $\mathrm{var}(\Delta s_t)/\mathrm{var}(\varepsilon_{t,H})$ for each of our six currencies. Overall, the results are quite similar to the results from the West test, the only significant difference being Canada, which has a point estimate (slightly) below unity. The results are also quite similar to those reported by Engel and West (2004), although as noted earlier, they interpret the results reciprocally, as the share of exchange rate volatility that can be accounted for by observed fundamentals.

### 1.4.5 Engel (2005)

Engel (2005) derives a variance bound that is closely related to the West (1988) bound. Like the West bound, it is robust to the presence of unit roots. Let $\hat{s}_t$ be the forecast of the present discounted value of future fundamentals based on a subset, $H_t$, of the market's information. Engel shows that the model, along with the Rational Expectations Hypothesis, implies the following inequality[11]

$$\text{var}(s_t - s_{t-1}) \leq \text{var}(\hat{s}_t - \hat{s}_{t-1})$$

For simplicity, we test this bound by assuming that $f_t$ follows an AR(1) around a deterministic trend. That is, letting $\tilde{f}_t$ denote the demeaned and detrended $f_t$ process, we assume $\tilde{f}_t = \rho \tilde{f}_{t-1} + u_t$. In this case, we have

$$\hat{s}_t = \alpha_0 + \alpha_1 \cdot t + \left(\frac{1-\beta}{1-\rho\beta}\right)\tilde{f}_t \qquad \Rightarrow \qquad \Delta\hat{s}_t = \left(\frac{1-\beta}{1-\rho\beta}\right)\Delta\tilde{f}_t$$

The bottom row reports values of $\text{var}(\Delta s_t)/\text{var}(\Delta \hat{s}_t)$. Not surprisingly, given our previous results, they all exceed unity by a substantial margin, once again pointing to excess volatility in the foreign exchange market.

## 1.5 Robust Volatility Tests

The results reported in the previous section are based on tests that make different assumptions about information, trends, and the underlying data generating process. Despite this, a consistent picture emerges - exchange rates appear to exhibit 'excess volatility'. Of course, we haven't reported standard errors, so it is *possible* these results lack statistical significance (although we doubt it). However, we agree with Shiller (1989). Pinning your hopes on a lack of significance does not really provide much support for the model. It merely says there isn't enough evidence yet to reject it.

Although the previous tests differ along several dimensions, there is one assumption they all share, namely, the Rational Expectations Hypothesis (REH). As noted earlier, this is a joint hypothesis, based on two assumptions: (1) Agents have common knowledge of the correct model, and (2) Agents make statistically optimal forecasts. Of course, many previous researchers have interpreted Shiller's work as evidence against the Rational Expectations Hypothesis (including Shiller himself!). However, there is an important difference between our response and previous responses. Previous responses have focused on the second part of the REH, and so have studied the consequences of various kinds of ad hoc forecasting strategies. In contrast, we try to retain the idea that agents make statistically optimal forecasts, and instead relax the first part of the REH. We further assume that the sort of model uncertainty that agents confront cannot be captured by a conventional (finite-dimensional)

[11] Engel (2005) also derives a second bound. As before, let $s_t^*$ denote the ex post realized present value of future fundamentals. Engel shows that the model implies $\text{var}(s_t^* - s_{t-1}^*) \leq \text{var}(s_t - s_{t-1})$. Note, this is like the Shiller bound, but in first differences. However, note that the direction of the bound is reversed! Although we haven't formally checked it, given our previous results, we suspect this bound would be easily satisfied.

Bayesian prior. As a result, we define an optimal forecast in terms of 'robustness'.[12] Of course, it is possible to abandon *both* parts of the REH at the same time. This has been the approach of the adaptive learning literature. For example, Kim (2009) shows that (discounted) Least Squares learning about the parameters of an otherwise known fundamentals process can generate significant exchange rate volatility (although he does not focus on violation of variance bounds inequalities). Markiewicz (2012) assumes agents entertain a set of competing models, each of which is adaptively estimated. She shows that endogenous model switching can generate time-varying exchange rate volatility.

### 1.5.1 A Robust Hansen-Sargent Prediction Formula

In contrast to the previous section, where agents *knew* the monetary model was the true data-generating process, here we suppose agents entertain the possibility that they are wrong. Although the monetary model is still regarded as a useful benchmark, agents suspect the model could be misspecified in ways that are difficult to capture with a standard Bayesian prior. To operationalize the idea of a robust forecast, agents employ the device of a hypothetical 'evil agent', who picks the benchmark model's disturbances so as to *maximize* the agent's mean-squared forecast errors. Since the sequence of error-maximizing disturbances obviously depends on the agent's forecasts, agents view themselves as being immersed in a dynamic zero-sum game. A robust forecast is a Nash equilibrium of this game.

We provide two solutions to this game. The first makes the agent's present value forecasts maximally robust, in the sense that the (population) mean-squared error remains totally invariant to a wide spectrum of potential dynamic misspecifications. As stressed by Hansen and Sargent (2008), this may not be an empirically plausible solution, as it may reflect concerns about alternative models that could be easily rejected given the historically generated data. We want our agents to be prudent, not paranoid. It will also turn out to be the case that maximally robust forecasts deliver *too much* exchange rate volatility. Hence, we also construct a solution that limits the freedom of the evil agent to subvert the agent's forecasts. The more we penalize the actions of evil agent, the closer we get to the conventional minimum mean-squared error forecast. Our empirical strategy is to first select a penalty parameter that replicates observed exchange rate volatility. We then calculate the detection error probability associated with this parameter value. We find that detection error probabilities in the range of 10-20% can explain observed exchange rate volatility.[13]

To handle general forms of dynamic misspecification, it is convenient to solve the problem in the frequency domain. As a first step, let's return to the problem of forecasting the present

[12] It is possible to provide axiomatic foundations that formalize the connection between robustness and optimality. See, e.g., Strzalecki (2011).

[13] Note, our focus on detection error probabilities differentiates our approach from the the 'rare disasters' literature. In that literature, agents are hedging against very low probability events. See, e.g., Farhi and Gabaix (2011).

value of fundamentals (eq. (1.2.2)) under the assumption that the data-generating process is known. In particular, suppose agents know that the Wold representation for fundamentals is $f_t = A(L)\varepsilon_t$. Transformed to the frequency domain our problem becomes (omitting the $(1-\beta)$ constant for simplicity),

$$\min_{g(z)\in H^2} \frac{1}{2\pi i}\oint \left|\frac{A(z)}{1-\beta z^{-1}} - g(z)\right| \frac{dz}{z} \tag{1.5.7}$$

where $\oint$ denotes contour integration around the unit circle, and $H^2$ denotes the Hardy space of square-integrable analytic functions on the unit disk. Once we find $g(z)$, the time-domain solution for optimal forecast is $s_t = (1-\beta)g(L)\varepsilon_t$. Restricting the $z$-transform of $g(L)$ to lie in $H^2$ guarantees the forecast is 'causal', i.e., based on a square-summable linear combination of current and past values of the underlying shocks, $\varepsilon_t$. The solution of the optimization problem in (1.5.7) is a classic result in Hilbert space theory (see, e.g., Young (1988), p. 188). It is given by,

$$g(z) = \left[\frac{A(z)}{1-\beta z^{-1}}\right]_+ \tag{1.5.8}$$

where $[\cdot]_+$ denotes an 'annihilation operator', meaning 'ignore negative powers of $z$'. From (1.5.8) it is a short step to the Hansen-Sargent prediction formula. One simply subtracts off the principal part of the Laurent series expansion of $A(z)$ around the point $\beta$ (see, e.g., Hansen and Sargent (1980, Appendix A)). This yields

$$g(z) = \frac{A(z) - \beta A(\beta) z^{-1}}{1-\beta z^{-1}}$$

which is the well know Hansen-Sargent formula. (Note that $g(z) \in H^2$ by construction, since the pole at $z = \beta$ is cancelled by a zero at $z = \beta$).

Now, what if the agent *doesn't* know the true Wold representation of $f_t$? In particular, suppose the $z$-transform of the actual process is $A^a(z) = A^n(z) + \Delta(z)$, where $A^n(z)$ is the agent's original benchmark (or nominal) model, and $\Delta(z)$ is an unknown (one-sided) perturbation function. Applying (1.5.8) in this case yields the following mean-squared forecast error:

$$\begin{aligned}
\mathcal{L}^a &= \mathcal{L}^n + ||\Delta(z)||_2^2 + \frac{2}{2\pi i}\oint \Delta(z)\left[\frac{A(z)}{1-\beta z^{-1}}\right]_- \frac{dz}{z} \\
&= \mathcal{L}^n + ||\Delta(z)||_2^2 + \frac{2}{2\pi i}\oint \Delta(z)\left(\frac{\beta A(\beta)}{z-\beta}\right) \frac{dz}{z}
\end{aligned} \tag{1.5.9}$$

where $[\cdot]_-$ is an annihilation operator that retains only negative powers of $z$, and $\mathcal{L}^a$ and $\mathcal{L}^n$ denote actual and nominal mean-squared forecast errors. The point to notice is that $\mathcal{L}^a$ could be much greater than $\mathcal{L}^n$, even when $||\Delta(z)||_2^2$ is small, depending on how $\Delta(z)$ interacts with $\beta$ and $A(z)$. To see this, apply Cauchy's Residue Theorem to (5.9), which yields

$$\mathcal{L}^a = \mathcal{L}^n + ||\Delta(z)||_2^2 + 2A(\beta)[\Delta(\beta) - \Delta(0)]$$

Notice that the last term is scaled by $A(\beta)$ which, though bounded, could be quite large. It turns out that the key to achieving greater robustness to model misspecification is to switch norms. Rather than evaluate forecast errors in the conventional $H^2$ sum-of-squares norm, we are now going to evaluate them in the $H^\infty$ supremum norm. In the $H^\infty$-norm the optimal forecasting problem becomes[14]

$$\min_{g(z)\in H^\infty} \max_{|z|=1} \left| \frac{A(z)}{1-\beta z^{-1}} - g(z) \right|^2 \tag{1.5.10}$$

where $H^\infty$ denotes the Hardy space of essentially bounded analytic functions on the unit disk.[15] Problem (1.5.10) is an example of a wide class of problems known as 'Nehari's Approximation Problem', which involves minimizing the $H^\infty$ distance between a two-sided $L^\infty$ function and a one-sided $H^\infty$ function. For this particular case, Kasa (2001) proves that the solution takes the following form

$$g(z) = \frac{zA(z) - \beta A(\beta)}{z-\beta} + \frac{\beta^2}{1-\beta^2} A(\beta) \tag{1.5.11}$$

Notice that the first term is just the conventional Hansen-Sargent formula. The new element here comes from the second term. It shows that a concern for robustness causes the agent to revise his forecasts more aggressively in response to new information. Note that this vigilance is an increasing function of $\beta$. As $\beta \to 1$ the agent becomes more and more concerned about low frequency misspecifications.[16] It is here that the agent is most exposed to the machinations of the evil agent, as even small misjudgments about the persistent component of fundamentals can inflict large losses on the agent. Of course, the agent pays a price for this vigilance, in the form of extra noise introduced at the high end of the frequency spectrum. An agent concerned about robustness is happy to pay this price.

To illustrate the role of robust forecasts in generating exchange rate volatility, we consider the spread, $\phi_t = s_t - f_t$, as a present value forecast of future values of $\Delta f_t$, and assume $\Delta f_t$ is an AR(1), so that $A(L) = 1/(1-\rho L)$ in the above formulas. Because the forecasts begin at $j = 1$ in this case, the Hansen-Sargent formula changes slightly (see eq. (1.2.5)). One can readily verify that the second term in the robust forecast remains unaltered.

[14] Using the sup norm to attain robustness is a time-honored strategy in both statistics (Huber (1981)) and control theory (Zames (1981)).

[15] In principle, we should write 'inf' and 'sup' in (1.5.10), but in our case it turns out that the extrema are attained.

[16] This intuition is also the key to the Engel-West (2005) Theorem, which shows that exchange rates converge to random walks as $\beta \to 1$. This happens because agents become increasingly preoccupied by the permanent component of fundamentals as $\beta \to 1$.

Table 1.4 reports ratios of actual to predicted spread variances for both traditional and robust forecasts.

Table 1.4: Robust Volatility Tests

| | Aus | Can | Den | Jap | Swz | UK |
|---|---|---|---|---|---|---|
| NonRobust AR(1) Forecast $\text{var}(\phi)/\text{var}(\hat{\phi})$ | 3486. | 1091. | 3482. | 1162. | 15661. | 740.1 |
| Robust AR(1) Forecasts $\text{var}(\phi)/\text{var}(\hat{\phi})$ | .227 | .291 | .478 | .358 | .218 | .363 |
| AR coef | .197 | .380 | .281 | .404 | .092 | .488 |

Notes: (1) Both bounds based on predictions of $\phi = s - f$ using an AR(1) model for $\Delta f$.
(2) When computing present values, it is assumed $\beta = .98$.

Here traditional forecasts look even worse than before, partly because we are failing to account for potential inside information, as we did in the Campbell-Shiller tests, and also partly because an AR(1) benchmark model is probably a bit too simple. Note, however, that this only makes our job harder when it comes to explaining excess volatility. The key results are contained in the middle row of Table 1.4, which reports ratios of actual to predicted spread variances with robust forecasts. Notice that all are well below one. If anything, robust forecasts generate too much volatility! This is illustrated by Figure 2, which plots of the actual spread (blue dashed line) against the nonrobust predicted spread (solid red line) and robust predicted spread (black dotted line). The Shiller plot is inverted![17]

Perhaps not surprisingly in light of eq. (1.5.11), the results here are rather sensitive to the value of $\beta$. For example, the above results assume $\beta = .98$. If instead we set $\beta = .92$, then all variance ratios exceed one and the model continues to generate excess volatility. This is not a serious problem in our view, since with quarterly data $\beta = .98$ *is* a more plausible value. A potentially more serious issue arises from the possibility that the agent is being excessively pessimistic here, and is worrying about alternative models that have little likelihood of being the true data-generating process. Fortunately, it is easy to avoid this problem by parameterizing the agent's concern for robustness. We do this by penalizing the actions of the evil agent.

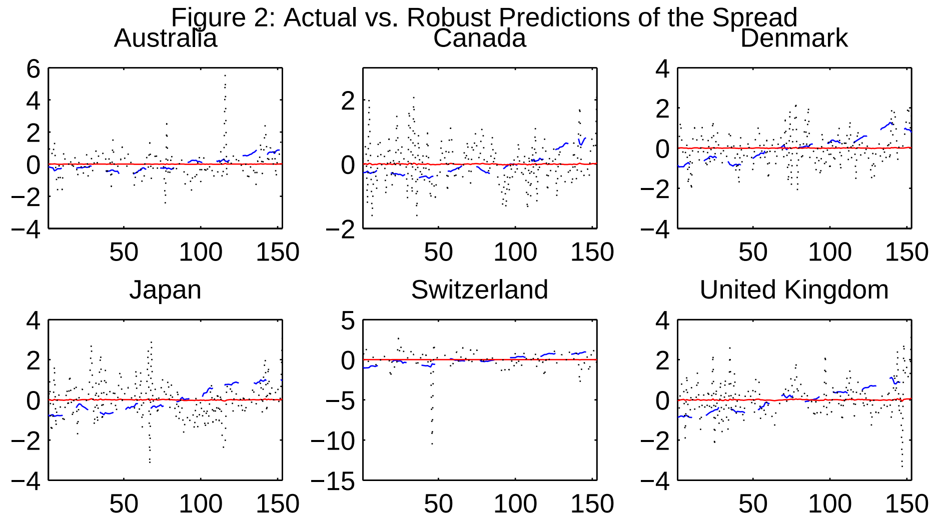
### 1.5.2 The Evil Agent Game

The previous results incorporated robustness by evaluating forecast errors in the $H^{\infty}$-norm. This delivers the maximal degree of robustness, but may entail unduly pessimistic beliefs. Following Lewis and Whiteman (2008), we now go back to the $H^2$-norm, and instead model

[17] We use eq. (1.5.11) to generate the time path of the robust spread. Expressed in terms of observables, this implies

$$\hat{\phi}_t^r = \frac{1}{1-\beta\rho}\left[\left(\beta\rho + \frac{\beta^2}{1-\beta^2}\right)\Delta f_t - \frac{\rho\beta^2}{1-\beta^2}\Delta f_{t-1}\right]$$

Figure 1.2: Actual vs. Robust Predictions of the Spread




robustness by penalizing the actions of the hypothetical evil agent. In particular, we assume the evil agent picks shocks subject to a quadratic penalty. As emphasized by Hansen and Sargent (2008), a quadratic penalty is convenient, since in Gaussian environments it can be related to entropy distortions and the Kullback-Leibler Information Criterion, which then opens to door to an interpretation of the agent's worst-case model in terms of detection error probabilities. This allows us to place empirically plausible bounds on the evil agent's actions.

One new issue arises here. Normally, the choice between modeling the level of the exchange rate (eq. (1.2.2)) and modeling the spread (eq. (1.2.4)) is based solely on statistical considerations. However, here there is a substantive issue at stake. If we view the agent as constructing a robust forecast of the spread, we are effectively conditioning on the current value of $f_t$, meaning that we are not permitting the evil agent to maliciously select current shock realizations. In contrast, most applications of robust control view the agent and his evil twin as making simultaneous (Nash) choices. Although it might seem natural to condition forecasts on current fundamentals, remember that here we want the agent to worry about potential unobserved fundamentals. If there are unobserved fundamentals, it would actually be *less* robust to condition on the current level of $f_t$. Hence, even if it makes more sense statistically to examine the spread, we want to first compute a robust forecast of $s_t$, and then use standard manipulations to write $s_t - f_t$ in terms of $\Delta f_t$. (Note, due to discounting, levels forecasts are still well defined even when there are unit roots in $s_t$ and $f_t$).

To begin, we assume the agent has a benchmark (nominal) model, $f_t = A^n(L)\varepsilon_t$. At the same time, he realizes this model may be subject to unstructured perturbations of the form $\Delta(L)\varepsilon_t$, so that the actual model becomes, $f_t = A^a(L)\varepsilon_t = A^n(L)\varepsilon_t + \Delta(L)\varepsilon_t$. The Evil Agent Game involves the agent selecting a forecast function, $g(L)$, to minimize mean-squared forecast errors, while at the same time the evil agent picks the distortion function, $\Delta(L)$. Hence, both agents must solve a calculus of variations problem. These problems are related to each other, and so we must solve for a Nash equilibrium. Following Lewis and Whiteman (2008), we can express the problem in the frequency domain as follows

$$\min_{g(z)} \max_{A^a(z)} \frac{1}{2\pi i} \oint \left\{ \left| \frac{A^a(z)}{1-\beta z^{-1}} - g(z) \right|^2 - \theta \left| \frac{A^a(z) - A^n(z)}{1-\beta z^{-1}} \right|^2 \right\} \frac{dz}{z} \tag{1.5.12}$$

where for convenience we assume the evil agent picks $A^a(z)$ rather than $\Delta(z)$. The key element here is the parameter $\theta$. It penalizes the actions of the evil agent. By increasing $\theta$ we get closer to the conventional minimum mean-squared error forecast. Conversely, the smallest value of $\theta$ that is consistent with the concavity of the evil agent's problem delivers the maximally robust $H^\infty$ solution.

The Wiener-Hopf first-order condition for the agent's problem is

$$g(z) - \frac{1}{1-\beta z^{-1}} A^a(z) = \sum_{\infty}^{-1} \tag{1.5.13}$$

where $\sum_{\infty}^{-1}$ denotes an arbitrary function in negative powers of $z$. The evil agent's Wiener-Hopf equation can be written

$$\frac{(1-\theta)A^a(z)}{1-\beta z} - \frac{1-\beta z^{-1}}{1-\beta z} g(z) + \frac{\theta A^n(z)}{1-\beta z} = \sum_{\infty}^{-1} \tag{1.5.14}$$

Applying the annihilation operator to both sides of eq. (1.5.13), we can then solve for the agent's policy in terms of the policy of the evil agent

$$g(z) = \frac{zA^a(z) - \beta A^a(\beta)}{z - \beta}$$

Then, if we substitute this into the evil agent's first-order condition and apply the annihilation operator we get

$$\frac{-\theta A^a(z)}{1-\beta z} + \left[ \frac{\beta z^{-1} A^a(\beta)}{1-\beta z} \right]_+ + \frac{\theta A^n(z)}{1-\beta z} = 0$$

which then implies

$$A^a(z) = A^n(z) + \frac{\beta^2}{\theta} A^a(\beta) \tag{1.5.15}$$

To determine $A^a(\beta)$ we can evaluate (1.5.15) at $z = \beta$, which implies $A^a(\beta) = \frac{\theta A^n(\beta)}{\theta - \beta^2}$. Substituting this back into (1.5.15) yields

$$A^a(z) = A^n(z) + \frac{\beta^2}{\theta - \beta^2} A^n(\beta) \tag{1.5.16}$$

This gives the worst-case model associated with any given benchmark model. In game-theoretic terms, it represents the evil agent's reaction function. Finally, substituting (1.5.16) into the above solution for $g(z)$ delivers the following robust present value forecast

$$g(z) = \frac{zA^n(z) - \beta A^n(\beta)}{z - \beta} + \frac{\beta^2}{\theta - \beta^2} A^n(\beta)$$

Notice that as $\theta \to \infty$ we recover the conventional minimum mean-squared error solution. Conversely, notice the close correspondence to the previous $H^\infty$ solution when $\theta = 1$.

### 1.5.3 Detection Error Calibration

The idea here is that the agent believes $A^n(L)\varepsilon_t$ is a useful first-approximation to the actual data-generating process. However, he also recognizes that if he is wrong, he could suffer large losses. To minimize his exposure to these losses, he acts as if $f_t = A^a(L)\varepsilon_t$ is the data-generating process when formulating his present value forecasts. Reducing $\theta$ makes his forecasts more robust, but produces unnecessary noise if in fact the benchmark model is correct. To gauge whether observed exchange rate volatility might simply reflect a reasonable response to model uncertainty, we now ask the following question - Suppose we calibrate $\theta$ to match the observed volatility of the spread, $s_t - f_t$. (We know this is possible from the $H^\infty$ results). Given this implied value of $\theta$, how easy would it be for the agent to statistically distinguish the worst-case model, $A^a(L)$, from the benchmark monetary model, $A^n(L)$? If the two would be easy to distinguish, then our explanation doesn't carry much force. However, if the probability of a detection error is reasonably large, then we claim that model uncertainty provides a reasonable explanation of observed exchange rate volatility.

To begin, we need to take a stand on the benchmark model. For simplicity, we assume $\Delta f_t$ follows an AR(1), so that $A^n(L) = 1/[(1 - \rho L)(1 - L)]$. (We de-mean the data). Given this, note that eq. (1.5.16) then implies the worst case model, $A^a(L)$, is an ARMA(1,1), with the same AR root. To facilitate notation, we write this model as follows,

$$\Delta f_t = \frac{1 + \kappa - \rho\kappa L}{1 - \rho L} \varepsilon_t$$

where $\kappa \equiv \frac{\beta^2}{\theta-\beta^2}\frac{1}{1-\rho\beta}$. Notice that $\kappa \to 0$ as $\theta \to \infty$, and the two models become increasingly difficult to distinguish. As always, there are two kinds of inferential errors the agent could make: (1) He could believe $A^n$ is the true model, when in fact $A^a$ is the true model, or (2) He could believe $A^a$ is the true model, when in fact $A^n$ is the true model. Denote the probability of the first error by $P(A^n|A^a)$, and the probability of the second error by $P(A^a|A^n)$. Following Hansen and Sargent (2008), we treat the two errors symmetrically, and define the detection error probability, $\mathcal{E}$, to be $\frac{1}{2}[P(A^n|A^a) + P(A^a|A^n)]$. From Taniguchi and Kakizawa (2000) (pgs. 500-503), we have the following frequency domain approximation to this detection error probability,

$$\mathcal{E} = \frac{1}{2}\left\{\Phi\left[-\sqrt{T}\frac{I(A^n, A^a)}{V(A^n, A^a)}\right] + \Phi\left[-\sqrt{T}\frac{I(A^a, A^n)}{V(A^a, A^n)}\right]\right\} \tag{1.5.17}$$

where $T$ is the sample size and $\Phi$ denotes the Gaussian cdf.[18] Note that detection error probabilities decrease with $T$. The $I$ functions in (1.5.17) are the KLIC 'distances' between the two models, given by

$$I(A^n, A^a) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[-\log\frac{|A^n(\omega)|}{|A^a(\omega)|} + \frac{A^n(\omega)}{A^a(\omega)} - 1\right]d\omega \tag{1.5.18}$$

$$I(A^a, A^n) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[-\log\frac{|A^a(\omega)|}{|A^n(\omega)|} + \frac{A^a(\omega)}{A^n(\omega)} - 1\right]d\omega \tag{1.5.19}$$

The $V$ functions in (1.5.17) can be interpreted as standard errors. They are given by the square roots of the following variance functions

$$V^2(A^n, A^a) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[A^n(\omega)\left(\frac{1}{A^n(\omega)} - \frac{1}{A^a(\omega)}\right)\right]^2 d\omega \tag{1.5.20}$$

$$V^2(A^a, A^n) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[A^a(\omega)\left(\frac{1}{A^a(\omega)} - \frac{1}{A^n(\omega)}\right)\right]^2 d\omega \tag{1.5.21}$$

[18] The same formula applies even when the underlying data are non-Gaussian. All that changes is that higher-order cumulants must be added to the variance terms in (5.20)-(5.21). However, for simplicity, we suppose the agent knows the data are Gaussian. Intuitively, we suspect that relaxing this assumption would only strengthen our results.

By substituting in the expressions for $A^a$ and $A^n$ and performing the integrations, we obtain the following expressions for $I$ functions[19]

$$
\begin{aligned}
I(A^n, A^a) &= \frac{1}{2}\left[\log(1+\kappa) + \frac{1}{(1+\kappa)^2(1-\psi^2)} - 1\right] \\
I(A^a, A^n) &= \frac{1}{2}\left[-\log(1+\kappa) + (1+\kappa)^2(1+\psi^2) - 1\right]
\end{aligned}
$$

where $\psi \equiv \rho\kappa/(1+\kappa)$. As a simple reality check, note that $\theta \to \infty \Rightarrow \kappa \to 0 \Rightarrow I \to 0$, which we know must be the case. Now, doing the same thing for the $V$ functions gives us

$$
\begin{aligned}
V^2(A^n, A^a) &= \frac{1}{2}\frac{1}{(1+\kappa)^2(1-\psi^2)}\left[(1+\kappa)^4 - 2(1+\kappa)^2 + \frac{1+\psi^2}{(1-\psi^2)^2}\right] \\
V^2(A^a, A^n) &= \frac{1}{2}\left[\frac{1}{(1+\kappa)^2} - 2(1+\psi^2) + (1+4\psi^2+\psi^4)(1+\kappa)^2\right]
\end{aligned}
$$

We use these formulas as follows. First, we estimate country-specific values of $\rho$. These are given in the bottom row of Table 1.4. Then we calibrate $\theta$ for each country to match the observed variance of its spread, $s_t - f_t$. The resulting values are reported in the first row of Table 1.5. Since the earlier $H^\infty$ forecasts generated *too much* volatility, it is not surprising that the implied values of $\theta$ all exceed one (but not by much). Finally, as we have done throughout, we set $\beta = .98$. We can then calculate the $\kappa$ and $\psi$ parameters that appear in the above detection error formulas. The results are contained in the second and third rows of Table 1.5.

We report detection error probabilities for two different sample sizes. The first assumes $T = 150$, which is (approximately) the total number of quarterly observations available in our post-Bretton Woods sample. In this case, the agent could distinguish the two models at approximately the 10% significance level.[20] Although one could argue this entails excessive pessimism, keep in mind that the data are generated in real-time, and so agents did not have access to the full sample when the data were actually being generated. As an informal

[19] We employed the following trick when evaluating these integrals. First, if $A^n(L) = \frac{1}{1-\rho L}$, we can then write $A^a = (1+\kappa)\frac{1-\psi L}{1-\rho L}$, where $\kappa$ is defined above and $\psi = \rho\kappa/(1+\kappa)$. We then have (omitting inessential constants)

$$
\oint -\log\frac{|A^n|}{|A^a|}\frac{dz}{z} = \oint \log(1+\kappa)\frac{dz}{z} + \oint \log\left|\frac{1-\psi z}{1-\rho z}\right|\frac{dz}{z} - \oint \log\left|\frac{1}{1-\rho z}\right|\frac{dz}{z} = \log(1+\kappa)
$$

where the last equality follows from the well known innovation formula, $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\omega)d\omega = \log(\sigma^2)$, where $f(\omega)$ is the spectral density of a process, and $\sigma^2$ is its innovation variance. This cancels the last two terms since they have the same innovation variance. The same trick can be used to evaluate the other log integral.

[20] It is a little misleading to describe these results in the language of hypothesis testing. The agent is not conducting a traditional hypothesis test, since both models are treated symmetrically. It is more accurate to think of the agent as conducting a *pair* of hypothesis tests with two different nulls, or even better, as a Bayesian who is selecting between models with a uniform prior. The detection error probability is then expected loss under a 0-1 loss function.

Table 1.5: Evil Agent Game: Calibrated Detection Errors

| | Aus | Can | Den | Jap | Swz | UK |
|---|---|---|---|---|---|---|
| $\theta$ | 1.0442 | 1.0349 | 1.0178 | 1.0273 | 1.0454 | 1.0220 |
| Det Error Prob($T = 150$) | .083 | .109 | .112 | .119 | .075 | .131 |
| Det Error Prob ($T = 100$) | .107 | .131 | .134 | .140 | .099 | .151 |

Notes: (1) $\theta$ calibrated to match observed variance of $\phi = s - f$.
(2) When computing present values, it is assumed $\beta = .98$.

correction for this, we also report detection error probabilities assuming $T = 100$. Now the detection error probabilities lie between 10-15%.

## 1.6 Other Puzzles

Excess volatility is not the only puzzle plaguing the foreign exchange market. Undoubtedly, the most widely studied puzzle is the forward premium puzzle, based on the observation that high interest rate currencies appreciate on average. Many proposed explanations of this puzzle link it to another puzzle, ie., the 'delayed overshooting puzzle'.[21] More recently, Engel (2012) has identified a new puzzle. One way to explain the forward premium puzzle is to argue that high interest rate currencies are relatively risky. Engel notes that a problem with standard risk premium theories is that the *level* of the exchange rate in high interest rate countries tends to be quite strong, stronger than can be accounted for by Uncovered Interest Parity. This suggests that high interest rate currencies are *less* risky.

Given these other puzzles, one natural concern is that our proposed explanation of excess volatility comes at the expense of exacerbating one or more of these other puzzles. In this section we show that robustness can readily account for the forward premium puzzle, but is less successful in resolving the delayed overshooting and Engel puzzles.

### 1.6.1 The Forward Premium Puzzle

Efforts to explain the forward premium puzzle generally fall into one of two categories: (1) Models with time-varying risk premia, or (2) Models with distorted expectations. Our model clearly falls into the second category.[22] Table 6 documents the forward premium

[21]See *inter alia* Eichenbaum and Evans (1995),Gourinchas and Tornell (2004), Li and Tornell (2008), Bacchetta and van Wincoop (2010).

[22]In ongoing work, Djeutem (2012) studies the implications of model uncertainty and robustness in a setting with stochastic discount factors and time-varying risk premia. He finds that Hansen-Jagannathan bounds can be satisfied with low degrees of risk aversion as long as investors are hedging against a worst-case consumption process with detection error probability in the same range as this paper (ie, 10-20%).

puzzle for the case of the Canadian dollar, the Japanese yen, and the British pound. The data are quarterly, for the period 1978:3-2011:1, and were downloaded from Engel's website.

Table 1.6: UIP Regressions: 1978:3 - 2011:1

$\Delta s_{t+1} = \alpha + \beta(i_t - i_t^*) + \varepsilon_{t+1}$

| | $\alpha$ | $\beta$ | $R^2$ |
|---|---|---|---|
| Canada | .001 | $-.059$ | .00 |
| | (.003) | (.702) | |
| Japan | .027 | $-2.60$ | .07 |
| | (.008) | (.816) | |
| UK | $-.007$ | $-1.28$ | .02 |
| | (.006) | (.810) | |

The puzzle is that not only is $\beta$ not equal to one, its point estimate is actually negative, and significantly less than one for Japan and the UK. These results are quite typical and quite robust (at least for the major currencies). On the surface, it would seem that our model is incapable of addressing this puzzle. After all, the monetary model *assumes* Uncovered Interest Parity! Remember, however, that we are *not* assuming Rational Expectations. Although by construction our model would satisfy UIP in the event the agent's worst-case model is the true data-generating process, if instead his benchmark model is the true model, so that his doubts are only 'in his head', then the agent's forecasts are biased, and this bias produces systematic deviations from UIP. Following Hansen and Sargent (2008), these deviations can be interpreted as a 'model uncertainty premium', as opposed to a risk premium.

Since our primary goal here is to address excess volatility, we make no pretense to providing a full treatment of the links between robustness and the forward premium puzzle. Instead, we merely show that a reasonably parameterized example can generate the sort of deviations from UIP that are seen in the data. As noted earlier, an empirically plausible specification for fundamentals is the ARIMA(1,1,0) process, $\Delta f_t = \rho \Delta f_{t-1} + \varepsilon_t$. Given this, the robust spread can be written as follows:

$$s_t - f_t = \frac{1}{1-\beta\rho}\left[\beta\rho\Delta f_t + \left(\frac{\beta^2}{\theta-\beta^2}\right)(\Delta f_t - \rho\Delta f_{t-1})\right]$$

which then implies:

$$\Delta s_{t+1} = \frac{1}{1-\beta\rho}\left[\rho(1-\beta)\Delta f_t - \left(\frac{\beta^2}{\theta-\beta^2}\right)\varepsilon_t + \left(1+\frac{\beta^2}{\theta-\beta^2}\right)\varepsilon_{t+1}\right]$$

There are two new features here. First, notice that the middle term generates an omitted variables bias in conventional UIP regressions. Evidently, the bias is downward. Second, notice from the third term that robust expectations produce a more volatile error term, which potentially explains why UIP regressions typically have rather low $R^2$'s.

Now, PPP and money demand imply the following relationship between the nominal interest differential and the spread: $s_t - f_t = \beta(i_t - i_t^*)/(1-\beta)$. With Rational Expectations (i.e, as $\theta \to \infty$) we would also have $s_t - f_t = \frac{\rho\beta}{1-\rho\beta}\Delta f_t$, which would imply UIP holds in the above equation. However, if we instead relate the nominal interest differential to the robust spread, we obtain the following relationship between exchange rate changes and nominal interest rate differentials:

$$\Delta s_{t+1} = \psi(i_t - i_t^*) + \frac{\rho\beta^2}{(1-\beta\rho)(\theta-\beta^2)}\left[1 + \psi\frac{1-\beta}{\beta}\right]\Delta f_{t-1} + \left(1 + \frac{\beta^2}{\theta-\beta^2}\right)\varepsilon_{t+1}$$

where the UIP regression slope coefficient is given by

$$\psi = \frac{1-\tau/(1-\beta)}{1+\tau/\beta} \qquad \tau = \frac{\beta^2}{\rho(\theta-\beta^2)}$$

Note that $\psi \to 1$ as $\theta \to \infty$, as it should. More importantly, notice that $\psi < 0$ for $\theta < \beta\sqrt{1+[\rho(1-\beta)]^{-1}}$, which offers some hope of resolving the forward premium puzzle. To gauge this, we take parameter values from our above resolution of the excess volatility puzzle, and see what they would imply about regressions of $\Delta s_{t+1}$ on lagged interest rate differentials. In principle, the entire history of $i_t - i_t^*$ should appear, but it is clear that for modest values of $\rho$ these lags damp out quite quickly. To simplify, we just report the implied coefficients on the first two lags.

Table 1.7: Robust UIP Regressions

$\Delta s_{t+1} = \beta_1(i_t - i_t^*) + \beta_2(i_{t-1} - i_{t-1}^*) + \varepsilon_{t+1}$

| | $\beta_1$ | $\beta_2$ | Implied $\theta$ |
|---|---|---|---|
| Canada | $-47.6$ | 0.20 | 120.1 |
| Japan | $-47.7$ | 0.22 | 32.2 |
| UK | $-47.5$ | 0.35 | 43.0 |

Notes: (1) $\beta_i$ coefficients based on country-specific values of $\theta$ and $\rho$ from Tables 4 and 5.
(2) Implied $\theta$ calibrated to match observed UIP slope estimates.

Evidently, we have a case of *coefficient* overshooting! Not only do we get negative implied coefficients, but they are far too negative. What's happening is that initial exchange rate reactions greatly exceed their final equilibrium, but revert to it quite quickly. At the same time, positive persistence in $\Delta f_t$ means that interest differentials approach their new equilibrium level more gradually. These movements generate a negative correlation between $\Delta s_{t+1}$ and $i_t - i_t^*$. Of course, traditional UIP regressions only include the first lag, and given the positive persistence in interest differentials, this omitted variable bias would moderate the estimates on the first lag somewhat. However, it seems unlikely this would bring them down enough. An alternative reconciliation is to consider less extreme values of $\theta$. In fact, we can pursue the exact same calibration strategy that we used for the volatility puzzle, and simply pick a value of $\theta$ that replicates empirical findings. These implied values of $\theta$ are reported in the final column of Table 1.7. Naturally, they are much larger. Interestingly, this suggests that the forward premium puzzle is actually *easier* to explain than the excess volatility puzzle, in the sense that detection error probabilities are larger in this case. It would be interesting to consider both puzzles simultaneously, and attempt to find a single value of $\theta$ that achieves some optimal compromise between the two puzzles. However, we leave this for future work.

That overreaction can explain the forward premium puzzle is not too surprising in light of the recent work of Burnside, Han, Hirshleifer, and Wang (2011). However, there is an important difference between their model and ours. They attribute overreaction to 'irrationality', which produces overconfidence. They appeal to psychological evidence in support of this posited irrationality. In contrast, we argue that it can be quite rational to overreact to news if you are unsure (in a Knightian sense) about the underlying model.

Finally, it should be noted that whether overreaction explains the forward premium puzzle depends on the underlying source of model uncertainty.[23] Our monetary model presumes money and income are the exogenous determinants of exchange rates, which agents then attempt to forecast. For the sake of argument, suppose these fundamentals are stationary. Then, without model uncertainty, a positive money shock drives down the nominal interest rate and causes the exchange rate to depreciate. The interest rate falls because the price level rises less than the money supply (due to known mean reversion). During the transition back to the steady state, we find negative interest rates accompanied by currency appreciation, in accordance with Uncovered Interest Parity. With model uncertainty, however, nominal interest rates can *rise* in response to a positive money shock, since the price level (and exchange rate) can rise more than the money supply. In this case, one would find positive interest rates accompanied by subsequent currency appreciation, as seen in the data. However, what if instead we suppose *interest rates* are the exogenous determinants of exchange rates? This might be a reasonable assumption if Central Banks follow an interest rate rule. Now the present value model is $s_t = -E_t \sum_{j=0}^{\infty}(i_{t+j} - i^*_{t+j})$. Overreaction to an exogenous interest rate shock now produces an *upward* bias in the estimated UIP slope coefficient.[24]

[23] We thank an anonymous referee for pointing this out.

[24] Ilut (2012) assumes interest rates are exogenous, and is still able to explain the forward premium puzzle by appealing to robust, ambiguity averse, forecasts. Interestingly, his model predicts a state dependent UIP slope coefficient. However, we suspect his model exacerbates the excess volatility puzzle.

### 1.6.2 The Delayed Overshooting and Engel Puzzles

Although robustness seems to offer a promising route toward resolving two of the leading puzzles in foreign exchange markets, it seems less promising as an explanation of the delayed overshooting and Engel puzzles. Since both these puzzles relate *real* exchange rates to *real* interest differentials, we must obviously modify our benchmark monetary model in some way if we are to even get started, since it imposes PPP, and PPP does not depend on expectations in this model. A natural way of doing this is to relax the assumption of exogenous monetary policy, and to instead follow the recent literature on Taylor Rules and exchange rates. In particular, we can simply import the following real exchange rate equation from Engel and West (2006) (see their paper for motivation and details):

$$q_t = (1-\beta)f_t + \beta E_t q_{t+1}$$

where $q_t$ is the real exchange rate, defined as the relative price of foreign goods. Notice that this has the exact same form as eq. (1.2.1). Now, however, $f_t$ refers to the following Taylor rule fundamentals:

$$f_t = \frac{1}{\chi}\left[\psi_\pi E_t(\pi^*_{t+1} - \pi_{t+1}) + \psi_y(y^*_t - y_t)\right]$$

where $\chi > 0$ is the coefficient on the foreign central bank's reaction to real exchange rates. It turns out that $\beta = 1/(1+\chi)$, which given empirically plausible values of $\chi$, once again implies that $\beta$ should be quite close to unity. The $\psi_\pi$ and $\psi_y$ coefficients can be inferred from standard Taylor rule estimates.

Engel and West (2006) apply this equation to the case of the deutschmark-dollar real exchange rate. They find some support for it. Interestingly, however, they find that it does not generate sufficient volatility. We instead suppose that the Central Bank (and the public) formulate *robust* expectations. Superficially, the analysis can proceed exactly as before. There is one important difference, however. Now fundamentals are *endogenous*. In particular, the model implies the following relationship between $f_t$ and $q_t$:

$$f_t = \frac{\beta}{1-\beta}(r^*_t - r_t) + q_t$$

where $r^*_t - r_t$ is the real interest rate differential. If one were to specify a process for $f_t$, use it to evaluate the present value, and then take account of endogeneity, one would find that *real* interest parity holds. That is, one would find:

$$q_{t+1} - q_t = r_t - r^*_t + \varepsilon_{t+1} \tag{1.6.22}$$

where $\varepsilon_{t+1}$ is an i.i.d, mean zero projection error. Engel (2012) shows that in fact the same sort of anomalous results are obtained with real exchange rates, i.e., estimated coefficients on $r_t - r^*_t$ are usually negative. Moreover, he points to a new puzzle confronting conventional risk premium explanations of these results. In particular, if we define $\lambda_t = r_t^* - r_t + E_t q_{t+1} - q_t$ as the period-$t$ risk premium on the foreign currency, we can iterate eq. (1.6.22) forward (assuming stationarity of the real interest differential), and obtain the following expression for the *level* of the real exchange rate

$$q_t = \bar{q} - R_t - \Lambda_t$$

where $\bar{q} = \lim_{j\to\infty}(E_t q_{t+j})$ is the long-run mean of the real exchange rate, and $R_t = E_t \sum_{j=0}^{\infty}(r_{t+j} - r_{t+j}^* - \bar{r})$ and $\Lambda_t = E_t \sum_{j=0}^{\infty}(\lambda_{t+j} - \bar{\lambda})$. Engel refers to $\Lambda_t$ as the 'level risk premium', since it relates risk to the *level* of the real exchange rate. Using VARs to compute present values, Engel shows that $\text{cov}(\Lambda_t, r_t - r_t^*) > 0$. That is, when domestic real interest rates are relatively high, the domestic currency is *less* risky. Of course, there is no *necessary* conflict between having a currency be relatively risky in the short-run and relatively safe in the long-run, but Engel goes on to show that conventional risk premium models fail to generate the necessary sign reversal.

As was the case with nominal interest parity, robust forecasts will generate systematic deviations from real interest parity if in fact the benchmark Taylor Rule model is the true DGP. These deviations can be interpreted as a 'model uncertainty premium'. Can model uncertainty premia explain the Engel (2012) puzzle? Following the exact same procedures as before, we can derive the following expression for the real exchange rate (assuming, as in Engel (2012), the real interest differential is stationary):

$$q_t = (1-\beta)\left\{\left[\frac{LA(L) - \beta A(\beta)}{L-\beta}\right] v_t + \left(\frac{\beta^2}{\theta - \beta^2}\right) A(\beta) v_t\right\} \tag{1.6.23}$$

where $f_t$ follows the process, $f_t = A(L)v_t$. A subtlety arises here from the combination of robust forecasts and the endogeneity of $f_t$. For example, suppose we adopt the empirically plausible assumption that the real interest differential follows the AR(1) process, $(r_t^* - r_t) = \rho(r_{t-1}^* - r_{t-1}) + v_t$. Then, because $f_t = \frac{\beta}{1-\beta}(r_t^* - r_t) + q_t$, the robust real exchange rate equation in (1.6.23) implies that $f_t$ is actually an ARMA(1,1). One can readily verify that $q_t$ inherits the same AR root as the real interest differential, but the MA root must be found via a fixed point procedure. That is, we guess $f_t = A_0\left(\frac{1-\gamma L}{1-\rho L}\right)$, use this guess to evaluate (1.6.23), and then match coefficients to find $A_0$ and $\gamma$. It turns out $A_0 > 0$ and $0 < \gamma\beta < 1$ for feasible values of $\theta$, so that $A(\beta) > 0$. To see what this implies about real interest parity and the Engel puzzle, we can use (1.6.23) to derive the following equation for real exchange rate changes:

$$\Delta q_{t+1} = r_t - r_t^* - (1-\beta)A(\beta)\left(\frac{\beta^2}{\theta - \beta^2}\right) v_t + (1-\beta)A(\beta)\left[1 + \beta\left(\frac{\beta^2}{\theta-\beta^2}\right)\right] v_{t+1} \tag{1.6.24}$$

Not surprisingly, we see that real interest parity holds as $\theta \to \infty$. However, as before, notice that for $\theta < \infty$ the middle term generates an omitted variables bias in conventional regression tests. Substituting for $v_t$ in terms of $r_t - r_t^*$ and $r_{t-1} - r_{t-1}^*$, one finds two offsetting effects. First, the 'true' coefficient (ie, with $r_{t-1} - r_{t-1}^*$ included) exceeds unity. Second, if the lag is excluded, the omitted variable bias is downward when $\rho > 0$. One can show that for the knife-edge case, $\rho = 1$, the two effects exactly offset, and one would find that real interest parity holds. If $\rho > 1$, one would find estimated coefficients below one, and perhaps even negative for small enough values of $\theta$. In this case, the Engel puzzle would be resolved. However, for the more plausible case that $\rho < 1$, we would find estimates in *excess* of unity. Although the model would generate needed volatility, and would explain why high interest rate currencies are strong, it would not explain observed deviations from real interest parity. The Engel puzzle would remain.

Engel (2012) conjectures that a resolution might be found by combining over-reaction with 'momentum', and cites Hong and Stein (1999) as a promising lead. The model in Hong and Stein generates short-term underreaction and long-term overreaction by positing investor heterogeneity, and by assuming that dispersed private information diffuses slowly within the market. Unfortunately, given the initial underreaction, it's not clear whether such a model can explain our main objective here, i.e., excess volatility. However, introducing robustness into a model with gradual information diffusion might be a way to generate both initial overreaction and momentum.

## 1.7 Concluding Remarks

This paper has proposed a solution to the excess volatility puzzle in foreign exchange markets. Our solution is based on a disciplined retreat from the Rational Expectations Hypothesis. We abandon the assumption that agents know the correct model of the economy, while retaining a revised notion of statistically optimal forecasts. We show that an empirically plausible concern for robustness can explain observed exchange rate volatility, even in a relatively simple environment like the constant discount rate/flexible-price monetary model.

Of course, there are many competing explanations already out there, so why is ours better? We think our approach represents a nice compromise between the two usual routes taken toward explaining exchange rate volatility. One obvious way to generate volatility is to assume the existence of bubbles, herds, or sunspots. Although these models retain the idea that agents make rational (self-fulfilling) forecasts, in our opinion they rely on an implausible degree of expectation coordination. Moreover, they are often not robust to minor changes in market structure or information. At the other end of the spectrum, many so-called 'behavioral' explanations have the virtue of not relying on strong coordination assumptions, but only resolve the puzzle by introducing rather drastic departures from conventional notions of optimality.

As noted at the outset, our paper is closely related to Lewis and Whiteman (2008). They argue that robustness can explain observed US stock market volatility. However, they also find that if detection errors are based only on the agent's ability to discriminate between alternative models for the economy's exogenous dividend process, then implausibly small detection error probabilities are required. If instead detection errors are based on the agent's ability to discriminate between bivariate models of dividends *and* prices, then stock market volatility can be accounted for with reasonable detection errors. This is not at all surprising, since robustness delivers a substantially improved fit for prices. Interestingly, we find that even if detection errors are only based on the exogenous fundamentals process, exchange rate volatility can be accounted for with reasonable detection error probabilities. Still, one could argue that they are a bit on the low side, so it might be a useful extension to apply the bivariate Lewis-Whiteman approach to computing detection error probabilities. We conjecture that this would only strengthen our results. A second useful extension would be to consider in more detail the links between robustness and other exchange rate puzzles. For example, while we have shown that robust forecasts can also explain the forward premium puzzle, it would interesting to see to what extent both puzzles can be explained simultaneously.

# Chapter 2

# Model Uncertainty and the Forward Premium Puzzle

This chapter[1] studies the Forward Premium Puzzle in a setting where investors doubt the specification of their models, and thus engage in *robust* portfolio strategies ( Hansen and Sargent (2008)). It shows that an empirically plausible concern for model misspecification can explain the Forward Premium Puzzle. In particular, the paper shows that Hansen and Jagannathan (1991) volatility bounds can be attained with both reasonable degrees of risk aversion and reasonable detection error probabilities. Hence, observed excess returns in the foreign exchange market appear to be primarily driven by a *model uncertainty premium.*

## 2.1 Introduction

It is commonly found in empirical international finance that high interest rate currencies tend to appreciate on average, while Uncovered Interest Parity (UIP) predicts they should depreciate. That is, under the joint hypotheses of Rational Expectations and risk neutrality, the regression of realized exchange rate changes on interest rate differentials should give a coefficient of one. Most studies find that not only is this coefficient statistically different from one, but it is often negative.[2] One common explanation attributes this failure to time varying risk premia. However, empirical tests using standard utility models require implausibly high degrees of risk aversion to account for observed excess returns in foreign exchanges markets.[3]

This paper revisits the puzzle in a setting where investors are both risk and ambiguity averse. Following Hansen and Sargent (2008), I use the notion of a preference for *robustness*[4]

[1] This chapter was published in the Journal of International Money and Finance, 46 (2014) 16-40

[2] Prominent examples include Hansen and Hodrick (1980), Fama (1984), Engel (1996) and Bansal and Dahlquist (2000).

[3] See *inter alia* Engel (1996) and Mark (2001) for a review of conventional risk premium explanations.

[4] I use interchangeably the expressions model uncertainty, robustness, knightian uncertainty and ambiguity aversion.

to distinguish between risk and ambiguity. More specifically, I consider an ambiguity averse investor who decides how much to consume and how much to invest in domestic and foreign bonds. The overall portfolio is therefore risky due to exchange rate risk, and the investor is assumed to be uncertain about the low frequency covariance between consumption growth and the exchange rate.

In my model, fears of misspecification pertain to the equilibrium consumption growth process. In response, the agent constructs a set of unstructured alternative consumption growth models surrounding a benchmark approximating model. Each model in this set is difficult to distinguish statistically from the benchmark model. I use model detection theory (Anderson, Hansen, and Sargent (2003)) to calibrate the robustness parameter, and show that there is strong empirical evidence supporting the ambiguity aversion interpretation. In fact, the paper shows that a *model uncertainty premium* is more important than a risk premium in explaining the forward premium puzzle.

Behavioural foundations for robustness go back to Knight (1921), who tried to distinguish between risk and uncertainty. For Knight, *uncertainty* refers to situations where a decision-maker does not know the probability distribution of an event, while *risk* corresponds to the case where this probability distribution is known or can be constructed from past data. Although most economists found Knight's arguments intuitively plausible, Savage (1954) showed that this distinction is irrelevant when individuals can formulate subjective probabilities. However, Ellsberg (1961) urn experiments suggest that, empirically, individuals seem to have a preference for knowing the probability distribution rather than having to form it subjectively. One of the most attractive approaches that takes account of the Ellsberg paradox is the multiple priors (or maxmin) approach developed by Gilboa and Schmeidler (1989). They show that in the presence of Knightian uncertainty, agents cannot form a unique probability distribution over states of the world. As a result, they proposed an approach where agents formulate multiple priors, and then base decisions on the worst probability measure. This approach has been extended to dynamic recursive environments by Hansen and Sargent (2008). It is inspired by robust control theory, widely used in engineering, and gives rise to so called *multiplier preferences*.[5]

This paper is not the first to use ambiguity as a potential solution to the forward premium puzzle. Li and Tornell (2008) use an overlapping generations framework with an ambiguity averse investor and an exogenous interest rate differential. They assume interest rate differentials are governed by temporary and persistent components that are unobserved by investors. Rational investors thus engage in robust filtering, and systematically distort their forecasts. A negative UIP coefficient is then the result of forecast distortion in response to interest rate differential shocks. Recently, Ilut (2012) used a similar assumption and

[5]Axiomatic characterization of these preferences are provided by Strzalecki (2011) and Maccheroni, Marinacci, and Rustichini (2006)

showed that ambiguity averse investors systematically underestimate the hidden state of interest rate differentials, underreacting to good news and overacting to bad news.

There are several features that distinguish this paper from Li and Tornell (2008) and Ilut (2012). First, both these papers use a partial equilibrium model where interest rate differentials are exogenously specified. Their solutions depend sensitively on the specification of the interest differential. For instance, Ilut (2012, p53) shows that, with less persistence, his model cannot account for the puzzle, while a robust decision against the higher variance of temporary component generates a negative UIP slope. My test strategy relies on testing the volatility implications of the Euler equations of a general equilibrium model. This strategy is due to Hansen and Jagannathan (1991), and I examine the restrictions imposed by their volatility bound using multiplier preferences. In particular, my paper addresses the following question: can model uncertainty be an alternative to the implausibly high degrees of risk aversion found with standard preference specifications?[6] Second, these two papers do not focus on volatility, and it is likely that the combination of underaction to good news and overaction to bad news exacerbates the excess volatility puzzle (Djeutem and Kasa (2013)). Third, my choice of the stochastic discount factor environment is justified by the need to generate a time varying risk premium. As pointed out by Cochrane (2001, p. 451), and emphasized by Alvarez, Atkeson, and Kehoe (2009), variation in risk is essential for understanding movements in assets prices. In addition, Fama (1984), Benigno, Benigno, and Nisticò (2011), Verdelhan (2010), Menkhoff, Sarno, Schmeling, and Schrimpf (2012) and Engel (2012) have provided evidence for a time varying risk component in the foreign exchange markets. Fourth, these two papers do not calibrate empirically the degree of model uncertainty. In contrast, I use detection error probabilities to investigate whether implied model uncertainty premia are empirically plausible. I find that observed excess returns can be explained if investors are hedging against models that have a 30-40% probability of being the true model.

The remainder of the paper is organized as follows. Section 2.2 describes the framework. Section 2.3 presents the data and the empirical results. Section 2.4 concludes, and an appendix provides details of some mathematical derivations.

## 2.2 The Framework

### 2.2.1 The Forward Premium Puzzle

The uncovered interest parity relation states that:

$$(1 + i_t) = (1 + i_t^*)\frac{E_t S_{t+1}}{S_t} \tag{2.2.1}$$

[6] This paper is methodologically closely related to Tallarini (2000), Barillas, Hansen, and Sargent (2009) and Bidder and Smith (2011).

where $S$ is the spot exchange rate defined as the price of foreign currency, $i$ and $i^*$ are domestic and foreign one period nominal interest rates, and $E_t$ denotes the market expectation. Taking logarithms on both sides of the relation yields the linear approximation:

$$E_t s_{t+1} - s_t = \alpha_0 + i_t - i_t^* \tag{2.2.2}$$

where $s = \ln(S)$ and $\alpha_0$ is a linearization constant. Assuming Rational Expectations,

$$s_{t+1} = E_t s_{t+1} + \varepsilon_{t+1} \;\Rightarrow\; s_{t+1} - s_t = i_t - i_t^* + \varepsilon_{t+1} \tag{2.2.3}$$

where $\varepsilon_{t+1}$ is an i.i.d forecast error that is orthogonal to the time-$t$ information set. The last expression implies the following testable implication, often called a Fama regression:

$$\Delta s_{t+1} = \alpha_0 + \alpha_1 (i_t - i_t^*) + \varepsilon_{t+1} \qquad H_0 : \alpha_1 = 1 \tag{2.2.4}$$

where $\Delta s_{t+1} = s_{t+1} - s_t$. As noted in the Introduction, empirical evidence suggests that for the major currencies $\alpha_1$ is actually negative. Existing explanations of these results fall into one of two broad categories: (1) Relaxations of the risk neutrality assumption, and (2) relaxations of the Rational Expectations Hypothesis. In what follows, I propose an explanation that involves a sort of *interaction* between these two. We shall see that ambiguity aversion can be interpreted as producing pessimistically distorted beliefs.[7]

### 2.2.2 Stochastic Discount Factors Without Robustness

#### Basic Set up

The setup is a simplified version of the typical consumption-based asset pricing model due to Lucas (1978). This is a two-country flexible exchange rate model with an exogenous nonstorable consumption good. There are two assets: a one period nominal bond in domestic currency, and a one period nominal bond in foreign currency, with interest rates $i_t$ and $i_t^*$ respectively. The domestic nominal bond is a (nominally) risk free asset, while the foreign bond is a risky asset due to exchange rate risk. Given an initial level of wealth $x_0$, a representative consumer wants to maximize his lifetime expected utility by choosing how much to consume each period $C_t$, and how much to save $B_t$. A share $\omega_t$ of this saving is invested in the domestic bond, and $1 - \omega_t$ is invested in the foreign bond with return $(1 + i_t^*)\frac{S_{t+1}}{S_t}$. The consumption-saving problem is therefore:

[7] In this paper I am using quarterly data.

$$
\begin{aligned}
&\underset{\{C_t,\omega_t\}_{t=0}^{\infty}}{\text{Max}} && E_0 \sum_{t=0}^{\infty} \beta^t U(C_t) \\
&\text{s.t.} && P_t C_t + B_t = x_t, \\
& && x_{t+1} = B_t \left[ \omega_t (1+i_t) + (1-\omega_t)(1+i_t^*) \frac{S_{t+1}}{S_t} \right]
\end{aligned}
$$

where $\beta$ is the subjective discount factor. Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ denote a sequence of increasing conditioning information sets available to the agent. Then $E[.|\mathcal{F}_t]$ is the conditional expectation with respect to the information set available at date $t$, denoted alternatively $E_t$. The Bellman equation for this problem is:

$$
\begin{aligned}
V(x_t) = &\underset{C_t,\omega_t}{\text{Max}}\, U(C_t) + \beta E_t V(x_{t+1}) \\
&\text{s.t.} x_{t+1} = \left[ \omega_t (1+i_t) + (1-\omega_t)(1+i_t^*) \frac{S_{t+1}}{S_t} \right] [x_t - P_t C_t]
\end{aligned}
$$

The first-order condition with respect to $\omega_t$ is given by:

$$
E_t \left[ \beta \frac{U'(C_{t+1})}{P_{t+1}} \left( (1+i_t) - (1+i_t^*) \frac{S_{t+1}}{S_t} \right) \right] = 0 \tag{2.2.5}
$$

Multiplying both sides of 2.2.5 by $\frac{P_t}{U'(C_t)}$ yields:

$$
E_t \left[ \beta \frac{U'(C_{t+1})}{U'(C_t)} \left( (1+i_t) - (1+i_t^*) \frac{S_{t+1}}{S_t} \right) \frac{P_t}{P_{t+1}} \right] = 0 \tag{2.2.6}
$$

Equivalently the Euler equation can be written in a compact form as:

$$
E_t(m_{t+1} R_{t+1}^e) = 0 \tag{2.2.7}
$$

where $R_{t+1}^e = \left( (1+i_t) - (1+i_t^*) \frac{S_{t+1}}{S_t} \right) \frac{P_t}{P_{t+1}}$ is the real return differential, and $m_{t+1} = \beta \frac{U'(C_{t+1})}{U'(C_t)}$ is the agent's stochastic discount factor. The Euler equation states that if investors are optimizing, the expected marginal utility from a leveraged position in the foreign currency is zero.

Many empirical studies have tested this Euler equation. One approach consists of specifying a general equilibrium model and then using Generalized Method of Moments to test the Euler equation's moment conditions (Mark 1985, Modjtahedi 1991). Using a CRRA utility function, both studies found that a degree of risk aversion ranging from 40 to 65 was required for the Euler restrictions in the forward foreign exchange market to hold. However, instead of testing directly that relationship Hansen and Jagannathan (1991) use a nonparametric reverse engineering approach to infer the properties that any valid discount factor must satisfy. In particular, they derived a lower bound for the variance of the discount factor using the first two moments of returns.

**Hansen Jagannathan Lower Bound**

The Euler equation of consumption-based asset pricing is expressed in general as :

$$E_t(m_{t+1}R_{t+1}) = 1_N \qquad (2.2.8)$$

where $R_{t+1}$ is a N-dimensional vector of one period returns from date $t$ to date $t+1$. A valid discount factor satisfying the previous Euler equation could be $m^* = R^T \left[E(RR^T)\right]^{-1} 1_N$, and thus $m = m^* + \epsilon$ is another valid discount factor, with $\epsilon$ being any random variable orthogonal to $R$. Using this insight, the Hansen Jagannathan procedure consists of deriving the SDF with the lowest variance. The trick is to use the property of Ordinary Least Squares together with the Cauchy Schwarz inequality.

More formally, let $\mu_m \equiv E(m_t)$, $\mu_R \equiv E(R_t)$, $\Sigma_R \equiv E(R_t - \mu_R)(R_t - \mu_R)^T$ and then project $m - \mu_m$ onto $R - \mu_R$. That is :

$$m - \mu_m = (R - \mu_R)^T \beta + u \qquad (2.2.9)$$

An estimate of $\beta$ is given by:

$$\begin{aligned} \hat{\beta} &= \Sigma^{-1} E(R - \mu_R)(m - \mu_m) \\ &= \Sigma^{-1} \left[E(mR) - \mu_R \mu_m\right] \\ &= \Sigma^{-1} \left[1_N - \mu_R \mu_m\right] \text{ By using the Euler equation} \end{aligned}$$

where $\Sigma$ is the variance-covariance matrix of the random vector $R - \mu_R$. Since by construction $u$ is orthogonal to $(R - \mu_R)^T \beta$, we can derive the variance of $m$ as:

$$\begin{aligned} \sigma_m^2 &= E\left[\hat{\beta}^T (R - \mu_R)(R - \mu_R)^T \hat{\beta}\right] + \sigma_u^2 \\ &= (1_N - \mu_R \mu_m)^T \Sigma^{-1} (1_N - \mu_R \mu_m) + \sigma_u^2 \end{aligned}$$

Implying the following lower bound for the volatility of the stochastic discount factor:

$$\sigma_m \geq \sqrt{(1_N - \mu_R \mu_m)^T \Sigma^{-1} (1_N - \mu_R \mu_m)} \qquad (2.2.10)$$

In foreign exchange markets, where all the assets are either forward contracts or zero net investment positions, the inequality becomes:

$$\sigma(m) \geq E(m) \sqrt{\bar{R}^{e\prime} \Sigma^{-1} \bar{R}^e} \qquad (2.2.11)$$

where $\bar{R}^e$ is a vector of mean return differentials and $\Sigma$ is their variance-covariance matrix. Mark (2001, p143) reports that a coefficient of relative risk aversion of at least 30 is needed to attain the Hansen-Jagannathan bound.

Standard expected utility models have a hard time pricing risk in general, and give rise to several well known paradoxes or puzzles, such as the Ellsberg paradox, Allais paradox, Equity premium puzzle or risk-free rate puzzle The response to these puzzles has been to develop more sophsticated utility models, which differ from expected utility in various ways: (i) the aggregation procedure of utility, (ii) time of resolution of uncertainty and, (iii) the persistence of risk Ljungqvist and Sargent (2012, p 532-535). Prominent theoretical contributions include Kreps and Porteus (1978), Epstein and Zin (1989), Weil (1990), and Tallarini (2000).

### 2.2.3 Stochastic Discount Factors With Robustness

An agent with multiplier preferences is aware of his own ignorance, and would like to make decisions that are robust to a set of unstructured alternative models expressed in terms of distortions to the shocks in his benchmark model. The desire for robustness is implemented using Gilboa and Schmeidler (1989) insight, implying a robust decision is the outcome of a dynamic zero sum game between the agent and a hypothetical evil agent who chooses shocks to minimize the agent's payoff.

Let $x_{t+1} = h(x_t, u_t, \varepsilon_{t+1})$ be the agent's benchmark approximating model, where $x_t$ is the state vector, $u_t = [log\, C_t, \omega_t]$ is the control vector, and $\varepsilon_t$ a sequence of i.i.d shocks to the state vector. Denote $r(x_t, u_t)$ the payoff function and let $p(x_t|x_{t-1})$ be the benchmark conditional probability describing the state transition equation.

A preference for robustness can be modelled as follows:

$$\begin{aligned} W(x_0) = \underset{\{g_{t+1}\}}{\text{Min}} \quad & \sum_{t=0}^{\infty} E\left\{\beta^t G_t \left[r(x_t, u_t) + \beta\theta E(g_{t+1} \log g_{t+1} | \mathcal{F}_t)\right] | \mathcal{F}_0\right\} \\ \text{s.t.} \quad & x_{t+1} = h(x_t, u_t, \varepsilon_{t+1}) \\ & G_{t+1} = g_{t+1} G_t \\ & E[g_{t+1}|\mathcal{F}_t] = 1 \ , \ g_{t+1} \geq 0 \ , \ G_0 = 1 \\ & x_0 \text{ given} \end{aligned}$$

Where $\theta$ is a penalty parameter on the conditional relative entropy associated with $g_{t+1}$. When $\theta$ increases, the evil agent pays a bigger cost to distort the agent's model. Essentially, the agent cares less about robustness. $\{G_t; t \geq 0\}$ is a nonnegative martingale that is used to construct a distorted probability measure generated by the approximating model.

The robust optimization problem now takes the form of a zero-sum dynamic game:

$$
\begin{aligned}
\underset{\{u_t\}}{\text{Max}}\ \underset{\{g_{t+1}\}}{\text{Min}} \quad & \sum_{t=0}^{\infty} E\left\{\beta^t G_t\left[r(x_t,u_t)+\beta\theta E(g_{t+1}\log g_{t+1}|\mathcal{F}_t)\right]|\mathcal{F}_0\right\} \\
\text{s.t.} \quad & x_{t+1}=h(x_t,u_t,\varepsilon_{t+1}) \\
& G_{t+1}=g_{t+1}G_t \\
& E[g_{t+1}|\mathcal{F}_t]=1\ ,\ g_{t+1}\geq 0\ ,\ G_0=1 \\
& x_0 \text{ given}
\end{aligned}
$$

And the corresponding Bellman equation is given by:

$$
\begin{aligned}
V(x_t)=\underset{u_t}{\text{Max}}\ \underset{g(x_{t+1},x_t)}{\text{Min}} \quad & r(x_t,u_t)+\beta\int\Big\{g(x_{t+1},x_t)V(x_{t+1})+ \\
& \theta g(x_{t+1},x_t)\log\ g(x_{t+1},x_t)\Big\}p(x_{t+1}|x_t)dx_{t+1} \\
\text{s.t.} \quad & x_{t+1}=h(x_t,u_t,\varepsilon_{t+1}) \\
& G_{t+1}=g_{t+1}G_t \\
& E[g_{t+1}|\mathcal{F}_t]=1\ ,\ g_{t+1}\geq 0\ ,\ G_0=1 \\
& x_0 \text{ given}
\end{aligned}
$$

Under mild regularity conditions[8], the order of optimization in the previous problem does not matter. Solving the Evil agent's problem yields the following martingale increment:

$$
g(x_{t+1},x_t)=\frac{e^{-\frac{V(x_{t+1})}{\theta}}}{E(e^{-\frac{V(x_{t+1})}{\theta}}|\mathcal{F}_t)} \tag{2.2.12}
$$

Plugging the worst-case martingale increment back into the Bellman equation gives:

$$
\begin{aligned}
V(x_t)=\underset{u_t}{\text{Max}} \quad & r(x_t,u_t)-\beta\theta\log\ E\left[exp\left(-\frac{V(x_{t+1})}{\theta}\right)|x_t\right] \\
\text{s.t.} \quad & x_{t+1}=h(x_t,u_t,\varepsilon_{t+1}) \\
& x_0 \text{ given}
\end{aligned} \tag{2.2.13}
$$

This recursive representation corresponds toHansen and Sargent (1995) *risk sensitive preferences* [9], and thus shows a connection between risk sensitivity and robustness.[10] Moreover,

[8] These conditions are known as the Isaac conditions, and require the objective function to be globally concave in $u_t$ and globally convex in $g_{t+1}$

[9] With $\theta=-\frac{1}{(1-\beta)(1-\gamma)}$ and $\gamma>1$ denotes the agent's risk sensitivity

[10] More formally, robustness and risk sensitivity are mathematical duals, in the sense that they are Legendre transforms of each other.

with multiplier preferences, the stochastic discount factor becomes[11]:

$$\hat{m}_{t+1} = \left(\beta \frac{C_t}{C_{t+1}}\right) \left(\frac{e^{-\frac{V_{t+1}}{\theta}}}{E(e^{-\frac{V_{t+1}}{\theta}}|\mathcal{F}_t)}\right)$$

And the Euler equation eq. (2.2.7) takes the form:

$$\begin{aligned} E_t\left(\hat{m}_{t+1} R^e_{t+1}\right) &= E_t\left(g_{t+1} m_{t+1} R^e_{t+1}\right) = 0 \\ &= \hat{E}_t\left(m_{t+1} R^e_{t+1}\right) = 0 \end{aligned} \tag{2.2.14}$$

Where $m_{t+1} = \beta \frac{C_t}{C_{t+1}}$ corresponds to the traditional risk averse SDF (with a risk aversion coefficient equal to one). Note that $g_{t+1}$ can be interpreted either as a multiplicative shock to the conventional discount factor, or as a distortion to the traditional expectations operator (where $\hat{E}$ represent expectations taken with respect to this pessimistically distorted probabability measure).

### 2.2.4 Consumption growth dynamics

Evidently, the specification of the consumption process and the utility function are essential to study the forward premium. In fact, the value function of the investor's problem, the stochastic discount factor and mean-variance frontier of the SDF all depend tightly on the consumption time series. The previous maxmin problem can be very difficult to estimate, due to the recursive nature of the preferences and the nonlinearity of the transition law. Without loss of generality, the following three simplifying assumptions are made. Let $log\, C_t \equiv c_t$.

**Assumption 2.2.1.** *Consumption follows one of the following two processes:*

$$\begin{aligned} \textit{Random Walk (RW):} \quad & c_{t+1} = \mu + c_t + \sigma_\epsilon \varepsilon_{t+1} \\ \textit{Trend Stationary (TS):} \quad & c_t = \lambda + \mu t + z_t \\ & z_t = \rho z_{t-1} + \sigma_\epsilon \varepsilon_t \end{aligned}$$

*where $\varepsilon_t$ is i.i.d. $N(0,1)$.*

The specification of the consumption process is identical to Tallarini (2000), Barillas, Hansen, and Sargent (2009), and Bidder and Smith (2011). These two specifications can also be classified within the broader class of univariate consumption growth specifications encountered in the literature (Mehra and Prescott 1985, Cecchetti, Lam, and Mark 1994, Bansal and Yaron 2004, Hansen, Heaton, Lee, and Roussanov 2007, e.g.).

[11] To ease notation, let $g_{t+1} \equiv g(x_{t+1}, x_t)$ and $V_t \equiv V(x_t)$

**Assumption 2.2.2.** *The one period payoff function (utility) of the investor is log linear:*

$$r(x_t, u_t) = c_t$$

This requirement on the one period utility is solely used for tractability purposes.

**Conjecture 2.2.3.** *Using the multiplier preferences, we conjecture that the value function is also linear in $c_t$:*

$$\begin{aligned} \textit{RW: } V_t &= A + Bc_t \\ \textit{TS: } V_t &= D + Et + Fc_t \end{aligned}$$

Assumptions 2.2.1 and 2.2.2 together with conjecture 2.2.3 delivers a closed form solution for the stochastic discount factor. The following three propositions summarize these results. [12]

**Proposition 2.2.4.** *Under multiplier preferences 2.2.13, assumptions 2.2.1 and 2.2.2, and conjecture 2.2.3; the agent's problem gives the following optimal distortion:*

$$\begin{aligned} \textit{RW:} \quad & g_{t+1} = exp\left\{-\tfrac{B\sigma_\epsilon}{\theta}\varepsilon_{t+1} - \tfrac{B^2\sigma_\epsilon^2}{2\theta^2}\right\} \\ \textit{TS:} \quad & g_{t+1} = exp\left\{-\tfrac{F\sigma_\epsilon}{\theta}\varepsilon_{t+1} - \tfrac{F^2\sigma_\epsilon^2}{2\theta^2}\right\} \end{aligned}$$

*where* $B = \frac{1}{1-\beta}$ *and* $F = \frac{1}{1-\beta\rho}$

*Proof.* See A.2 □

The results from the above proposition are central to our analysis in the sense that, first, the martingal increment is essential to find the corresponding distorted probability distribution $\tilde{p}(x_{t+1}|x_t)$ by using its likelihood ratio interpretation. That is,

$$\tilde{p}(x_{t+1}|x_t) = g(x_{t+1}, x_t)p(x_{t+1}|x_t)$$

In our environment where the approximating distribution of consumption growth is posited to follow either a random walk with drift or trend stationary process, the equilibrium worst case distribution distorts only the mean. These distortion are respectively given by $w_{t+1} = -\sigma_\epsilon(\gamma - 1)$ and $w_{t+1} = -\sigma_\epsilon\frac{(1-\beta)(\gamma-1)}{1-\beta\rho}$.

Second, it will also be useful in determining the unconditional distribution of the stochastic discount factor summarize in proposition below.

[12] See Barillas, Hansen, and Sargent (2009) for alternative proofs of proposition 2.2.4 and 2.2.5 .

**Proposition 2.2.5.** *Under multiplier preference 2.2.13, assumptions 2.2.1, 2.2.2 and conjecture 2.2.3; The stochastic discount factor has the following unconditional distributions:*

$$\begin{aligned}
&\text{RW:} \qquad \log m_{t+1} \sim \mathcal{N}\left(\log\beta - \mu - \tfrac{B^2}{2\theta^2}\sigma_\varepsilon^2,\ \sigma_\varepsilon^2(1+\tfrac{B}{\theta})^2\right) \\
&\text{TS:} \quad \log m_{t+1} \sim \mathcal{N}\left(\log\beta - \mu - \tfrac{F^2}{2\theta^2},\ \sigma_\varepsilon^2(\tfrac{2}{1+\rho} + 2\tfrac{F}{\theta} + \tfrac{F^2}{\theta^2})\right)
\end{aligned}$$

*Proof.* See A.3 □

The unconditional distribution of the SDF are necessary to describe the loci $(E(m), \sigma(m))$ spanned by their first two moments. In this case, $\gamma$ can be interpreted as a model uncertainty premium rather than risk aversion coefficient Hansen, Sargent, and Tallarini (1999).

### 2.2.5 Detection Error Probability

So far, the model uncertainty parameter $\gamma$ is assumed to be a free parameter. I will now use the procedure outlined by Hansen and Sargent (2008, ch 9) to calibrate this parameter. It is calibrated to match the difficulty of statistically discriminating the approximating model from the worst case model, given the observed data. This is called a *detection error probability*, and is computed using the likelihood ratio principle.

Letting $A$ and $B$ denote respectively the approximating and worst case models, assigning a prior probability 0.50 to each model, the detection error probability is given by:

$$p(\gamma) = \frac{p_A + p_B}{2}$$

where $p_A = \mathbb{P}(log\frac{L_A}{L_B} < 0|A)$, $p_B = \mathbb{P}(log\frac{L_A}{L_B} > 0|B)$, and $L_i$ the likelihood associated with model $i$. $p_i$ measures the probability that a likelihood ratio test selects model $j$, or equivalently $p_i$ is the probability of a model detection error.

**Proposition 2.2.6.** *Under assumptions 2.2.1 and 2.2.2, the detection error probabilities are given by:*

$$\begin{aligned}
RW: \quad p(\gamma) &= \Phi\left(-0.5\sqrt{T}\sigma_\epsilon\left|\gamma - 1\right|\right) \\
TS: \quad p(\gamma) &= \Phi\left(-0.5\sqrt{T}\sigma_\epsilon \frac{(1-\beta)\left|\gamma-1\right|}{1-\rho\beta}\right)
\end{aligned}$$

*Proof.* See A.4 □

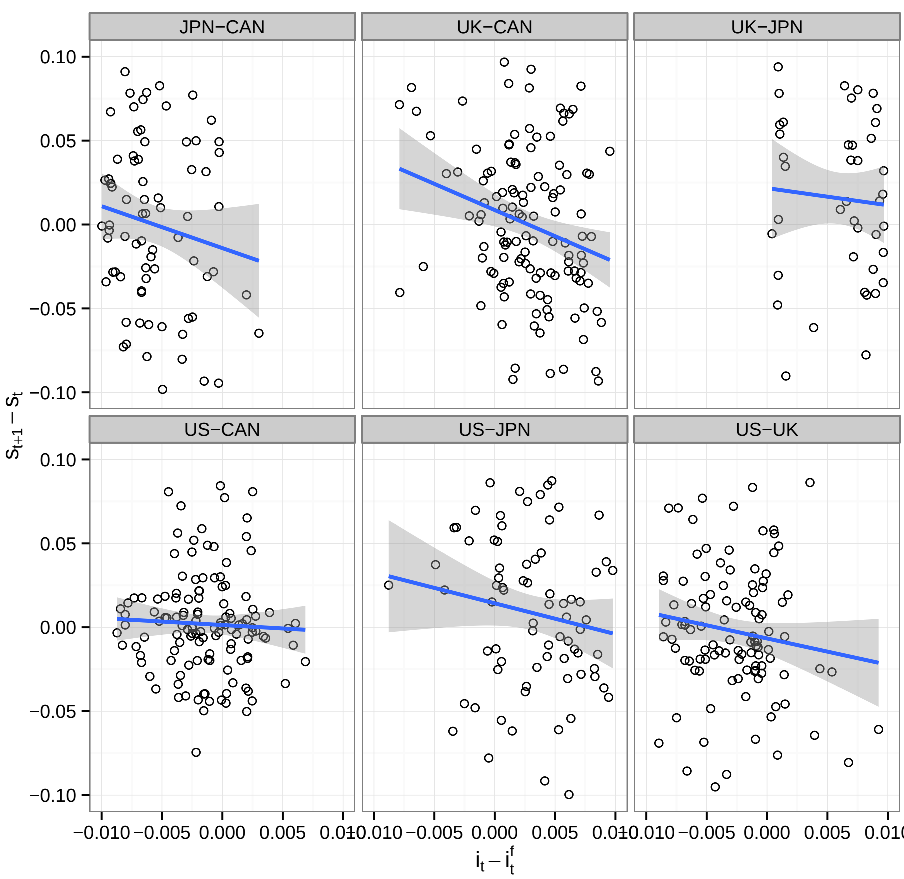
Proposition 2.2.6 implies that detection error probabilities are decreasing with the sample size ($T$), the variance of the temporary component ($\sigma_\epsilon$), and the first order autocorrelation $\rho$ and $\gamma$. Intuitively, as more data becomes available it becomes easier to discriminate between the approximating and worst case models. Thus, the detection error probability falls. Similarly, higher $\gamma$ translates into small $\theta$ which means that the evil agent can more freely perturb the actions of the agents. Thus, the worst case distribution will be easily distinguishable from the approximating model, entailing a smaller detection error probability. Note, the detection error probability is related to a significance level employed in traditional hypothesis testing. It varies from 0 to 0.5 and the larger it is, the more plausible is the model uncertainty explanation.

## 2.3 Applications

### 2.3.1 Data

The data consist of quarterly observations from 1979Q1 to 2012Q2 on spot exchange rates and nominal interest rates, all from Datastream. The consumer price indices are from the OECD database, while real consumption comes from the IFS database. I study four currencies: Canadian dollars, Japanese Yen, British pound and US dollars.

Figure 2.1: Excess return and relative depreciation




*Notes:* $i_t - i_t^f$ stands for $i_t - i_t^*$. The sample period is 1979Q1-2012Q2.

Figure 2.1 plots the return differential $(i_t - i_t^*)$ versus the realized depreciation $(s_{t+1} - s_t)$ for the six pairs of countries together with the fitted regression line and its confidence interval. This plot confirms that the slope coefficient is statistically far from being one, and is in fact negative for most of the pairs of countries considered. [13]

Table 2.1: MLE Estimate of consumption process

| Spec. | CAN | JPN | UK | US |
|---|---|---|---|---|
| **Random Walk** | | | | |
| $\mu$ | 0.01460 | 0.00625 | 0.01673 | 0.01491 |
| $\sigma_\varepsilon$ | 0.00832 | 0.01162 | 0.01179 | 0.00800 |
| **Trend Stationary** | | | | |
| $\mu$ | 0.01022 | -0.12680 | 0.00593 | 0.00470 |
| $\sigma_\varepsilon$ | 0.00648 | 0.01012 | 0.00895 | 0.00661 |
| $\rho$ | 0.96543 | 0.99889 | 0.98121 | 0.98855 |
| T | 134 | 134 | 134 | 134 |

Notes: T is the sample size 1979Q1-2012Q2.

A dynamic regression (see Figure A.1) using rolling and expanding window that the UIP coefficient is consistently negative and do not depend on the sample size. This result suggests that the investor earns excess return and that high interest currencies are more likely to experience an appreciation.

Furthermore, Table 2.1 reports estimates of the consumption growth process by the maximum likelihood method. Two features emerge from this table. First, for most countries, the two consumption specifications give almost identical estimates of the variance of the innovation $\sigma_\varepsilon$, but slightly different estimates of long run consumption growth, $\mu$. Second, error terms in the trend stationary specification are highly autocorrelated ($\rho$), indicating high persistence and weak mean reversion.

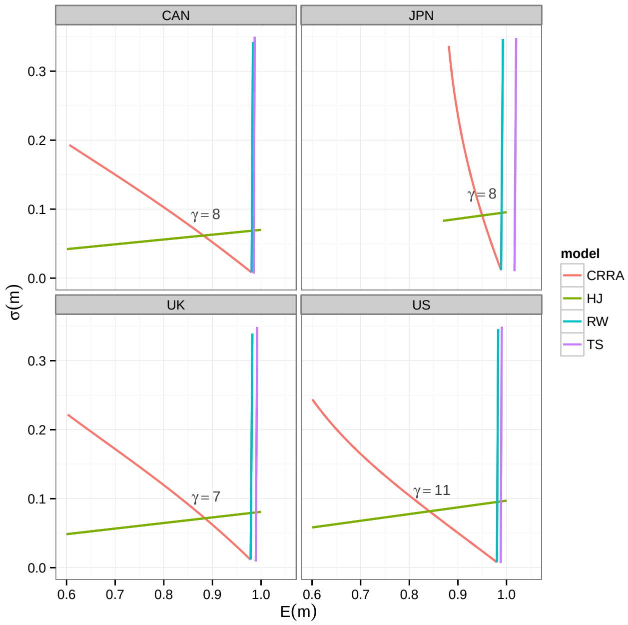
### 2.3.2 Robust Hansen-Jagannathan bound

This section presents the mean-variance frontier of the stochastic discount factor for the four countries. Figure 2.2 plots on the horizontal axis the first moment of the SDF and on the vertical axis its second moment, for different preferences specifications: (i) power utility (CRRA), and (ii) multiplier preferences with consumption process specified as either a random walk with drift(RW) or a Trend stationary(TS) process. The Hansen-Jagannathan volatility lower bound is also reported (HJ). Notice that with a conventional power utility function, implausibly high degree of risk aversion are needed to reach the HJ lower bound. These minimals degree of risk aversion are 8, 8, 7 and 11 respectively for Canada, Japan, UK and USA.

Interestingly, when I allow investors to have preference for objective probability and calibrate the taste for robustness using the detection error probability, the data provide

[13]Chinn (2006) shows that the intensity of the puzzle might depends of the sample period and the puzzle is more prominent in major currencies.

Figure 2.2: Robust Hansen Jagannathan Bound




*Notes:* This figure plots for each country, the Hansen and Jagannathan bound(HJ) and the loci $(E(m), \sigma(m))$ for the power utility(CRRA), multiplier preference with random walk (RW) and trend stationary (TS) consumption growth processes. Each point on the locus correspond to a particular value of $\gamma$. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$. RW: $E(m) = \beta \exp\left[-\mu + \frac{\sigma^2_\varepsilon}{2}(2\gamma - 1)\right]$, $\frac{\sigma(m)}{E(m)} = \sqrt{\exp\left[\sigma^2_\varepsilon \gamma^2\right] - 1}$. TS : $E(m) = \beta \exp\left[-\mu + \frac{\sigma^2_\varepsilon}{2}\left(1 - \frac{2(1-\beta)(1-\gamma)}{1-\beta\rho} + \frac{1-\rho}{1+\rho}\right)\right]$, $\frac{\sigma(m)}{E(m)} = \sqrt{\exp\left[\sigma^2_\varepsilon\left(\left\{\frac{(1-\beta)(1-\gamma)}{1-\beta\rho} - 1\right\}^2 + \frac{1-\rho}{1+\rho}\right)\right] - 1}$

Table 2.2: Distance to HJ bound and corresponding $\gamma$ and $p(\gamma)$

| $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^e}}$ | $\gamma$ | | | | $p(\gamma)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | CAN | JPN | UK | US | CAN | JPN | UK | US |
| | | | | | **Random Walk** | | | |
| 1.00 | 8 | 8 | 7 | 11 | 36.8 | 31.9 | 36.6 | 32.2 |
| 0.75 | 6 | 6 | 5 | 9 | 40.5 | 36.8 | 39.2 | 35.6 |
| 0.50 | 4 | 4 | 3 | 6 | 44.3 | 42.0 | 44.6 | 40.8 |
| 0.25 | 2 | 2 | 2 | 3 | 48.1 | 47.3 | 47.3 | 46.3 |
| | | | | | **Trend Stationary** | | | |
| 1.00 | 79 | 18 | 38 | 54 | 35.9 | 31.5 | 34.2 | 30.5 |
| 0.75 | 57 | 14 | 28 | 40 | 39.8 | 35.6 | 38.4 | 35.4 |
| 0.50 | 36 | 9 | 17 | 26 | 43.6 | 41.0 | 43.0 | 40.5 |
| 0.25 | 14 | 4 | 7 | 11 | 47.6 | 46.6 | 47.4 | 46.2 |

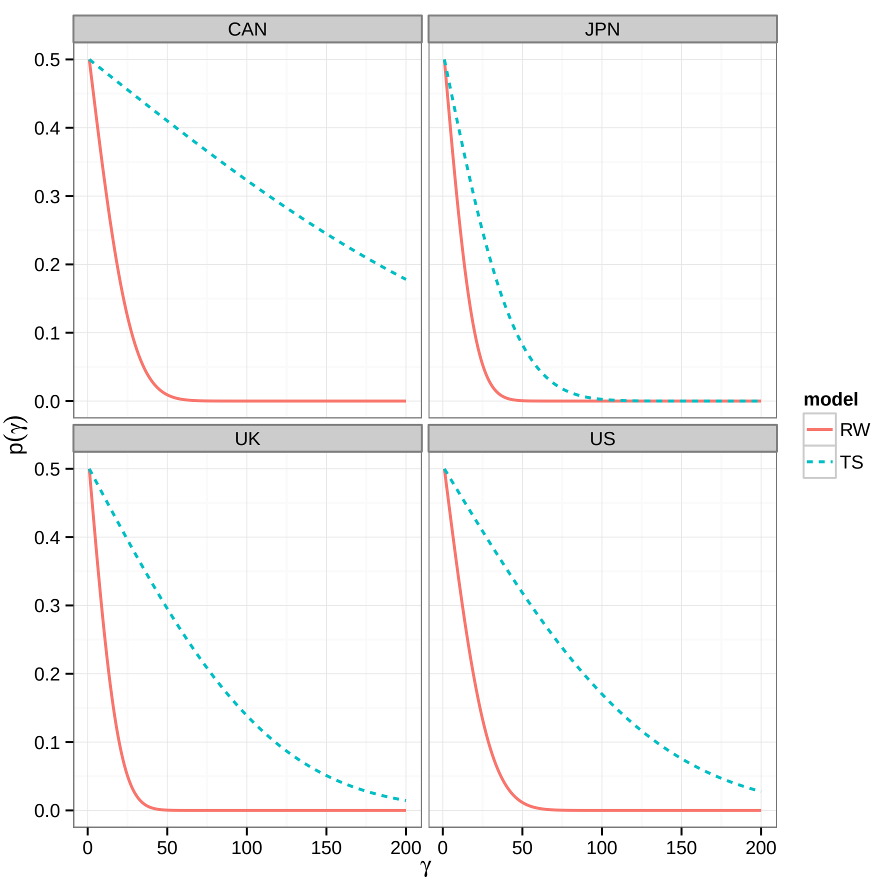
This table provides the minimal $\gamma$ and the corresponding detection error probability for different distance to the bound. The ratio $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^e}}$ equal to 1.00 means that the corresponding $\gamma$ puts the $(E(m), \sigma(m))$ on the HJ bound; while 0.50 put $(E(m), \sigma(m))$ half way to the bound. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

strong evidence for the ambiguity aversion explanation. In fact, the HJ lower bound is reached with detection error probabilities in the range 30.5% to 36.8 % using our two consumption growth specification(See fig. 2.2 and table 2.2) . How can we make sense of these probabilities? Taking for instance the case of USA and TS specification, there is 30.5% chance of making a mistake when the investor is trying to distinguish the worst case from the benchmark model. In contrast, usingTallarini (2000) approach and considering that consumption process is trend stationary, the minimal degree of risk aversion will be 79, 18, 38 and 54 for Canada, Japan, UK and US respectively.

### 2.3.3 Detection Error Probability

I use the estimates from table 2.1 together with the results of proposition 2.2.6 to calculate the detection error probability for each country and consumption specification. Figure 2.3 reports the results.

Figure 2.3: Detection Error Probablility




*Notes:*This figure plots the detection error probability by varying $\gamma$. RW: $p(\gamma) = \Phi\left(-0.5\sqrt{T}\sigma_\epsilon\,|\gamma-1|\right)$, TS : $p(\gamma) = \Phi\left(-0.5\sqrt{T}\sigma_\epsilon\frac{(1-\beta)|\gamma-1|}{1-\rho\beta}\right)$. The parameters estimates are from Table 2.1. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2.

From Figure 2.3 we observe that for a given value of $\gamma$, the corresponding detection error probability is higher with the TS specification than with the RW specification. This suggests that with TS the investor hardly discriminates between the approximating model and the worst case model, and thus there is a need to seek robustness. [14] In addition, the distinctive pattern for Japan in Figure2.3comes from the fact that consumption growth is highly persistent in Japan (see estimate Table 2.1) and the detection error probability (TS) is decreasing in $\rho$. Since, Japan has the highest estimate of the first order autocorrelation of consumption growth, the $p(\gamma)$ decreases rapidly compare to other countries.

### 2.3.4 Implications for UIP Regressions

Go back to the distorted Euler equation in 2.2.14. UIP regressions are based on the assumption that $m_{t+1}$ is constant, so that this Euler equation takes the simple form:

$$E_t R^e_{t+1} = 0$$

This is the restriction tested by UIP regressions, and the data suggest it is violated. If investors are risk averse, however, this regression is misspecified. In particular, if $m_{t+1}$ is stochastic we have

$$E_t R^e_{t+1} = -\frac{\mathrm{cov}_t(m_{t+1}, R^e_{t+1})}{E_t(m_{t+1})}$$

Writing this out and imposing Rational Expectations, we have

$$s_{t+1} - s_t = i_t - i^*_t - \frac{\mathrm{cov}_t(m_{t+1}, R^e_{t+1})}{E_t(m_{t+1})} + \varepsilon_{t+1}$$

Evidently, if $\frac{\mathrm{cov}_t(m_{t+1}, R^e_{t+1})}{E_t(m_{t+1})}$ is positively correlated with $(i_t - i^*_t)$, then OLS regressions that omit this variable will produce estimates that are biased downward, as seen in the data. One way to interpret the failure of conventional risk premium models is that attempts to produce an observable counterpart to $m_{t+1}$ produce an insufficiently strong positive correlation between $\frac{\mathrm{cov}_t(m_{t+1}, R^e_{t+1})}{E_t(m_{t+1})}$ and $(i_t - i^*_t)$. Ambiguity accentuates the volatility of $m_{t+1}$ by introducing the multiplicative model uncertainty premium, $g_{t+1}$, which can then accentuate the omitted variable bias. Unfortunately, showing this explicity would require us to explicitly solve the model, so we leave this to future research.

### 2.3.5 Sensitivity analysis

Over the past forty years, the global economy has experienced tremendous changes, e.g., rapid globabilization, integration of financial markets, the Asian financial crisis, and the 2008 subprime crisis. One might expect the dynamics of exchange rates and interest rates

[14] A low detection error probability means that the investor can easily discriminate, while a high detection error probability means he often makes mistakes.

to be influenced by these events, so that the severity of the UIP puzzle might vary over the sample. I argue in this paper that the deviation from UIP is attributed to underlying Knightian Uncertainty of consumption growth. So, if all these events contributed to a more turbulent or *uncertain* environment, one would expect the puzzle to still remain.

Table 2.3: Average distance to HJ bound using expanding window

| $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ | $\gamma$ | | | | $p(\gamma)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | CAN | JPN | UK | US | CAN | JPN | UK | US |
| | | | | | **Random Walk** | | | |
| 1.00 | 10 | 8 | 6 | 12 | 35.7 | 34.1 | 39.1 | 34.1 |
| 0.75 | 7 | 6 | 4 | 9 | 39.5 | 38.4 | 42.2 | 38.2 |
| 0.50 | 5 | 4 | 3 | 6 | 43.5 | 42.9 | 45.4 | 42.5 |
| 0.25 | 2 | 2 | 2 | 3 | 47.6 | 47.2 | 48.2 | 46.9 |
| | | | | | **Trend Stationary** | | | |
| 1.00 | 101 | 23 | 73 | 134 | 33.0 | 28.3 | 35.3 | 31.8 |
| 0.75 | 78 | 19 | 57 | 103 | 36.7 | 32.7 | 38.3 | 35.7 |
| 0.50 | 55 | 14 | 42 | 72 | 40.6 | 37.5 | 41.5 | 39.9 |
| 0.25 | 31 | 9 | 26 | 41 | 44.6 | 42.6 | 44.8 | 44.3 |

This table provides the minimal $\gamma$ and the corresponding detection error probability for different distance to the bound. These statistique are average over 67 samples constructed using expanding window. The first sample is 1979Q1-1995Q3 the second one is 1979Q1-1995Q4 etc. The ratio $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ equal to 1.00 means that the corresponding $\gamma$ puts the $(E(m), \sigma(m))$ on the HJ bound; while 0.50 put $(E(m), \sigma(m))$ half way to the bound. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

In order to assess the sensitivity[15] of our results with respect to sample selection, I selected 67 different samples using an expanding window technique. The first sample runs from 1979Q1-1995Q3, the second one is from 1979Q1-1995Q4, and so on. For each subsample I calculate the minimal multiplier parameter ($\gamma$) needed to reach the HJ bound and the corresponding detection error probability. Table 2.3 reports the average over the 67 samples. From this table, we see the results are pretty robust to sample selection.[16] The average detection error probability and $\gamma$ are in the same range as those obtain with the full sample.

## 2.4 Conclusion

This paper has proposed a solution to the forward premium puzzle in the foreign exchange market. This solution relaxes simultaneously the two underlying assumptions of the UIP hypothesis: risk neutrality and Rational Expectation Hypothesis. I consider a framework where risk averse investors fear model misspecification, and derive the implications for the stochastic discount factor in the spirit of Hansen and Jagannathan (1991). I showed that an empirically plausible preference for robustness can help attain the Hansen-Jagannathan

[15]Thanks to an anomous refree for suggesting the sensitivity analysis

[16]Tables A.2, A.3 and A.4 in the appendices contain respectively the standard deviations, minimun and maximum over the subsamples.

volatility bound. This result echoes the previous findings of Barillas, Hansen, and Sargent (2009), which showed that model uncertainty premia go a long way towards explaining the equity premium puzzle.

One potential extension of the present work would be to refine the econometric analysis by taking into account the effect of sampling variability on the minimal $\gamma$ needed to reach the HJ bound. For that purpose, Cecchetti, Lam, and Mark (1994) and Burnside (1994) would be useful. Another avenue would be to fully specify and solve a general equilibrium model that account for all puzzles surrounding the Forward Premium Puzzle (delayed overshooting, excess volatility andEngel (2012) puzzle). Instead of testing the model indirectly, it could be estimated using recent advances in particle filtering. In any case, robust filtering could be to be a promising and unified mechanism for understanding decision-making in financial markets.

# Chapter 3

# Robust Learning in the Foreign Exchange Market

This chapter studies risk premia in the foreign exchange market when investors entertain multiple models for consumption growth. Investors confront two sources of uncertainty: (1) individual models might be misspecified, and (2) it is not known which of these potentially misspecified models is the best approximation to the actual data-generating process. Following Hansen and Sargent (2010), agents formulate 'robust' portfolio policies. These policies are implemented by applying two risk-sensitivity operators. One is forward-looking, and pessimistically distorts the state dynamics of each individual model. The other is backward-looking, and pessimistically distorts the probability weights assigned to each model. A robust learner assigns higher weights to worst-case models that yield lower continuation values. The magnitude of this distortion evolves over time in response to realized consumption growth. It is shown that robust learning not only explains unconditional risk premia in the foreign exchange market, it can also explain the dynamics of risk premia. In particular, an empirically plausible concern for model misspecification and model uncertainty generates a stochastic discount factor that uniformly satisfies the spectral Hansen-Jagannathan bound of Otrok et. al. (2007).

## 3.1 Introduction

Economists have been struggling to understand exchange rate dynamics for more than forty years now, ever since the breakdown of the Bretton Woods system. Early attempts focused on linear present value models, which linked exchange rates to expectations of future fundamentals, such as money supplies and income levels. After these models failed, attention shifted to models with time-varying risk premia. However, these models have also failed.

A puzzling feature of all this research is that while economists struggle with one model after another, their models all assume that the agents *within* the model somehow know the true model generating exchange rates. Why the asymmetry? If economists don't know the model, but market participants do, why don't economists just ask them what's going on? This tension between the knowledge of econometricians and the presumed knowledge of agents within econometric models was recently highlighted by Hansen (2014, p.947) in his Nobel address:

> *Why is it fruitful to consider model misspecification? ... Part of a meaningful quantitative analysis is to look at models and try to figure out their deficiencies and the ways in which they can be improved. A more subtle challenge for statistical methods is to explore systematically potential modeling errors in order to assess the quality of the model predictions. This kind of uncertainty about the adequacy of a model or model family is not only relevant for econometricians outside the model, but potentially also for agents inside the models.*

This paper follows up on Hansen's conjecture in the context of the foreign exchange market, by assuming that agents within the model confront the same sort of fears of model misspecification that plague outside econometricians.

Given my focus on risk premia in the foreign exchange market, the model that matters to the representative agent here concerns the (equilibrium) consumption growth process.[1] As a practicing econometrician, the agent does not attempt to construct a single, all-encompassing, model. Instead, he considers several smaller, more parsimonious specifications, which differ along one or more key dimensions. In this paper, the key dimension along which models differ is their implied *persistence* of consumption growth.

Before explaining why allowing agents to consider multiple models is important, it is worth pausing to remember that according to traditional Bayesian decision theory, there is no meaningful distinction between models and parameters. A Bayesian forms a single nesting 'hypermodel', by assigning probability weights to each individual model. This Bayesian Model Averaging (BMA) procedure eliminates the distinction between models and parameters. Although many consider the Savage axioms to be normatively compelling, it is also worth remembering that Bayesian decision theory is just a theory, and the ultimate arbiter of any theory should be its ability to explain the data. As detailed by Hansen and Sargent (2008, 2011), and many others, there has been mounting empirical evidence against the Savage axioms. At the same time bayesian econometric methods have become popular among outside macroeconometricians, macroeconomic theorists are increasingly populating their models with inside agents that violate the Savage axioms. In this paper, the Savage axioms are violated in two ways. First, agents are unable to formulate a unique

[1] Of course, agents *choose* consumption, so there is never any uncertainty about it at the time. What really concerns agents is the future realization of the exogenous state, and how this will then determine what the ex post optimal level of consumption is. See Lucas (1978) for further discussion.

finite-dimensional prior distribution over model parameters. In response, they formulate an infinite-dimensional 'cloud' of unstructured alternatives, and then optimize against the worst-case model.[2] Second, agents violate the reduction of compound lotteries axiom. In particular, they have a preference for the early resolution of uncertainty. If models differ along this dimension then the distinction between models and parameters becomes important.

So we can now see why multiple consumption growth models might be important to the pricing of risk in the foreign exchange market. A currency is risky if its value moves procyclically (ie, if it appreciates when consumption growth is high). If investors are uncertain about consumption growth dynamics, this is reflected in currency risk premia. Given their assumed preference for the early resolution of uncertainty, a consumption growth process that is *persistent* is undesirable, since it confronts agents with a relatively late resolution of uncertainty. All else equal, a pessimistic, ambiguity averse, investor will therefore slant model probability estimates towards models that embody relatively more consumption growth persistence. At the same time, however, the value implications of persistence are also *state dependent*. During good times, persistent consumption growth is obviously desirable. To paraphrase Hansen and Sargent (2010), a pessimist is someone who thinks bad times are persistent and good times are transitory. Therefore, although on *average* robust model probabilities are biased toward persistent growth specifications, this bias becomes especially large during bad times. This implies robust learning can generate state dependence in the price of currency risk.

The main result of this paper is to show that the combination of an empirically plausible concern for individual model misspecification along with robust learning can explain both the average level of the risk premium in the foreign exchange market, and its state dependent dynamics. These dynamics are assessed by comparing the spectral density of the robust stochastic discount factor process with the frequency specific bound derived by Otrok, Ravikumar, and Whiteman (2007). This permits investigation of the properties of risk premia at different horizons, and therefore addresses the concern that the relationship between exchange rates and interest rates could be horizon-dependent (Chinn 2006, Chaboud and Wright 2005, Alexius 2001), or even non-monotonic (Hnatkovska, Lahiri, and Vegh 2013). I find that robust learning easily satisfies the bound for *all* frequencies, while conventional Bayesian learning only satisfies the bound for low frequencies. Following Hansen and Sargent (2008), I calibrate the degree of empirically plausible robustness to detection error probabilities. I find that risk premia can be explained with a 20% detection error probability.

This paper is closely related to prior work by Bansal and Yaron (2004), Hansen and Sargent (2010), and Djeutem (2014). Bansal and Yaron (2004) focus on the equity market,

[2] As discussed by Hansen and Sargent (2008), this can be interpreted as a weakening of Savage's 'Sure Thing Principle'.

and adopt the perspective of an outside econometrician. They observe that a consumption growth process with a small degree of high persistence is difficult to distinguish empirically from an i.i.d process. They show that if investors have recursive Epstein-Zin preferences that value early resolution of uncertainty, then the persistent growth specification can explain many observed asset pricing puzzles. However, they must resort to exogenous stochastic volatility to account for observed dynamics in the price of risk. Hansen and Sargent (2010) revisit the analysis of Bansal and Yaron (2004), but assume that the same model specification uncertainty that confronted Bansal and Yaron's outside econometrician, also confronts the agents in the model. They show that countercyclical slanting of model probabilities toward the persistent growth specification can generate countercyclical movements in the price of risk, without having to resort to exogenous stochastic volatility in consumption growth. In Djeutem (2014), I show that an empirically plausible model uncertainty premium can account for the forward premium puzzle in the foreign exchange market. However, in that paper there is no learning. Agents face unstructured uncertainty about a *given* benchmark model, but do not consider the possibility that some other benchmark model might be better. I show that robust learning can account not only for the average risk premium in the foreign exchange market, but also its state dependent dynamics.

Bansal and Shaliastovich (2013) also study risk premia in the foreign exchange market using a setting with long-run risk and Epstein-Zin investors. Their framework incorporates uncertain consumption growth along with uncertain inflation dynamics, specified as a vector autoregressive process with stochastic volatility. They show that a preference for the early resolution of uncertainty combined with stochastic volatility and non-neutral effects of inflation on growth can account for many observed features of excess returns in the foreign exchange market. Three features distinguish their model from mine. First, the two sources of uncertainty in their model (inflation and consumption growth) are modelled explicitly, using a so-called *structured uncertainty* approach. In contrast, I assume agents employ a robust, *unstructured uncertainty* approach. Second, I show that robust learning dynamics can explain observed risk premium dynamics *without* resort to exogenously specified stochastic volatility. Third, Bansal and Shaliastovich (2013) interpret their results in terms of risk aversion. They find that a risk aversion coefficient of 21 can explain observed risk premia. Although this is an improvement relative to many previous studies, one could argue that it is still implausibly high. In contrast, following Barillas, Hansen, and Sargent (2009), I interpret this parameter as a *model uncertainty* premium, and link it to detection error probabilities. Fourth, my model evaluation strategy, based on the spectral Hansen-Jagannathan bound, is simpler and less computationally intensive.

First generation risk premium explanations of observed deviations from Uncovered Interest Parity employed a risk averse representative agent with rational expectations (Fama 1984, Bekaert, Hodrick, and Marshall 1997). These models cannot fully account for the dynamics of excess returns, and they rely on implausibly high degrees of risk aversion.

This framework has been extended in many directions by using external habit persistence preferences (Wachter 2006, Verdelhan 2010, Moore and Roche 2002), recursive preferences (Colacito and Croce 2011b, Benigno, Benigno, and Nisticò 2011, Backus, Gavazzoni, Telmer, and Zin 2010), limited market participation (Alvarez, Atkeson, and Kehoe 2009), liquidity constraints (Rabitsch 2014), learning (Chakraborty and Evans 2008, Piazzesi and Schneider 2006, Lewis 1989), rare disaster(Lu and Siemer 2014, Farhi and Gabaix 2014, Gourio, Siemer, and Verdelhan 2013) or non rational expectation(Ilut 2012, Burnside, Han, Hirshleifer, and Wang 2011, Li and Tornell 2008, Gourinchas and Tornell 2004).[3] However, to the best of my knowledge, this is the first paper to explore the consequences of robust learning.

The remainder of the paper is organized as follows. The next section outlines a standard consumption-based model of risk in the foreign exchange market, and then shows how model uncertainty and robust learning can be introduced. Section 3.3 discusses the data and presents the empirical results. I consider quarterly data from 1979:1 to 2012:2 on three US dollar exchange rates: the Canadian dollar, the Japanese yen, and the British pound. In the main text I present results from the perspective of a US investor, who forms portfolios of these three currencies. I first construct two models that are difficult to distinguish empirically, but which embody different assumptions about the persistence of consumption growth. I then assume the investor confronts this uncertainty by constructing a robust, recursively updated, weighted average of these two models. I show that conventional Bayesian model averaging generates little variation in the stochastic discount factor, and can only satisfy the spectral Hansen-Jagannathan bound of Otrok, Ravikumar, and Whiteman (2007) at very low frequencies. In contrast, the robust learning stochastic discount factor is much more volatile, and satisfies the bound at *all* frequencies. Finally, Section 3.4 offers a few concluding remarks, while Appendix A shows that the results are similar when viewed from the perspective of foreign investors.

## 3.2 Framework

### 3.2.1 Basic Setup

[4] The setup is a standard consumption-based asset pricing model due to Lucas (1978, 1982). It is a two-country flexible exchange rate model with an exogenous nonstorable consumption good. There are two assets: a one period nominal bond in domestic currency, and a one period nominal bond in foreign currency, with interest rates $i_t$ and $i_t^*$ respectively. The domestic bond is a (nominally) risk free asset, while the foreign bond is a risky asset due to exchange rate risk. Given initial wealth, a representative consumer wants to maximize his

[3]For detailed literature review on exchange rates and interest parity see *inter alia*(Engel 2015, Sarno 2005, Engel 1996) and Lewis (2011) for international asset pricing.

[4]This set up is identical to the one in Djeutem (2014).

lifetime expected utility by choosing how much to consume each period $C_t$, and how much to save. The Euler equations can be combined and written compactly as follows:

$$E_t(m_{t+1}R^e_{t+1}) = 0 \tag{3.2.1}$$

where $R^e_{t+1} = \left((1+i_t) - (1+i^*_t)\frac{S_{t+1}}{S_t}\right)\frac{P_t}{P_{t+1}}$ is the real return differential, and $m_{t+1} = \beta\frac{U'(C_{t+1})}{U'(C_t)}$ is the agent's stochastic discount factor. Equation 3.2.1 is central to this paper, as it is in all consumption-based asset pricing models. This equation is mostly tested using the Fama reduced form regression or Generalized Method of Moments. In the following sections, I introduce more formally the framework by specifying the consumption growth dynamics and the agent's valuation.

### 3.2.2 Canonical consumption growth model

The consumption growth process is modelled in a state space framework. Let $y_t$ be the observable (or signal) time series of consumption growth, with conditional density function $p(y_{t+1}|x_t)$, where $x_t$ denotes an underlying latent process, with conditional transition density given by $p(x_{t+1}|x_t)$. These two densities fully characterize our state space model. I adopt throughout the convention that $p(.|.)$ denotes a density function specified by its arguments.

I further restrict attention to linear state spate models. In this case the model is :

$$\begin{aligned} x_{t+1} &= \quad Ax_t + C\epsilon_{t+1} \\ y_t &= \quad Dx_t + G\epsilon_t \end{aligned} \tag{3.2.2}$$

or alternatively

$$\begin{aligned} p(x_{t+1}|x_t) &= \quad \mathcal{N}(Ax_t, CC') \\ p(y_t|x_t) &= \quad \mathcal{N}(Dx_t, GG') \end{aligned} \tag{3.2.3}$$

where $y_t = \log C_t - \log C_{t-1}$, $x_t$ is a random vector, $\{\epsilon_t\}$ is a sequence of iid normally distributed random vectors with mean zero and covariance matrix $I$, and ($A$,$C$,$D$ $G$) are matrices of conforming dimensions. The representative investor is assumed to have unitary elasticity of intertemporal substitution. This implies the (equilibrium) value function for expected discounted utility is given by the following Bellman equation:

$$V(x, c) = (1-\beta)c + \beta E\left[V(x', c')|x, c\right] \tag{3.2.4}$$

where $c = \log C$. The log-linearity of instantaneous utility and consumption growth implies the consumption/wealth ratio is constant, which then delivers the following closed-form expression for the value function[5].

$$V(x,c) = c + \beta D(I - \beta A)^{-1}x \tag{3.2.5}$$

The agent does not observe $x_t$ and therefore has to make inferences about it using the signal history, $Y^{t-1} = \{y_1, y_2, \cdots, y_{t-1}\}$ and $y_t$. This signal extraction problem is solved via the Kalman filter, described recursively by:

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + K(\Sigma_t)\eta_{t+1} \\ y_t &= D\hat{x}_t + \eta_t \\ \Sigma_{t+1} &= A\Sigma_t A' + CC' - K(\Sigma_t)(A\Sigma_t D' + CG')' \\ K(\Sigma_t) &= (A\Sigma_t D' + CG')(D\Sigma_t D' + GG')^{-1} \end{aligned} \tag{3.2.6}$$

where $\hat{x}_t = E[x_t|Y^{t-1}]$, $\eta_t = y_t - E[y_t|Y^{t-1}]$, $\Sigma_t = E(x_t - \hat{x}_t)(x_t - \hat{x}_t)'$.

Given the forecast of the underling latent process, the conditional distribution of consumption growth is then given by:

$$p(y_t|\hat{x}_t) = \mathcal{N}\Big(D\hat{x}_t; GG^{'} + D\Sigma_t D^{'}\Big). \tag{3.2.7}$$

### 3.2.3 Multiple models and Bayesian learning

Now imagine a situation where the representative investor has in mind multiple models indexed by $M_1, M_2, \cdots, M_K$.[6] He does not know which model generates the data. Throughout this paper, a model is a probability distribution over the consumption growth process. Thus a model is specified by the state space formulation in 3.2.2, where the matrices specific to model $M_k$ are indexed by $k$. Let $L_t$ be a random variable indexing the model that generates the data at date $t$. The state space model class is then:

$$\begin{aligned} x_{t+1}^{(k)} = &\quad A^{(k)}x_t^{(k)} + C^{(k)}\epsilon_{t+1} \\ y_t = &\quad D^{(k)}x_t^{(k)} + G^{(k)}\epsilon_t \end{aligned} \tag{3.2.8}$$

Since $L_t$ is unknown to the investor it therefore becomes a state variable. Hence, the agent must infer both $x_t^{(k)}$ and $L_t$. As before, these latent processes are obtained recursively by Kalman filtering. Let $\pi_{0,k}$ be the initial prior distribution over models. Then:

$$\pi_{t,k} = \left[\frac{p(y_t|x_t^{(k)}, L_t = k)}{\sum_{l=1}^{K} \pi_{t-1,l}\, p(y_t|x_t^{(l)}, L_t = l)}\right]\pi_{t-1,k}. \tag{3.2.9}$$

[5] Letting $v = V(x,c) - c$, and rewrite the Bellman equation as $v = \beta E[v' + c' - c] = \beta E v' + \beta D x$. Iterate forward this expression using the 3.2.2 and obtain $v = \beta D(I - \beta A)^{-1}x$

[6] The notation is closely related to Raftery, Kárný, and Ettler (2010).

Notice the predicted signal is a weighted average of model-specific predictions, where the weights are the posterior probabilities, $\pi_{t,k}$. Specifically, let $\hat{y}_t^{BMA} = \sum_{k=1}^{K} \pi_{t,k} D^{(k)} \hat{x}_t^{(k)}$ be the Bayesian model average predicted signal.

### 3.2.4 Multiple models and Robust learning

With robust learning the agent still possesses multiple models, but now fears misspecification. The source of ambiguity is threefold. The first source of uncertainty is the distribution of the consumption growth process conditional on knowing the model and the latent process. That is, the agent mistrusts $p(y_t|x_t^{(k)}, L_t = k)$. The second source of doubt is the underlying conditional transition density of the latent process $p(x_{t+1}^{(k)}|x_t^{(k)}, L_t = k)$. Lastly, the agent also has doubts about his own prior distribution over models.

The multiple sources of uncertainty creates an ambiguous environment, where decision making is particularly challenging. To formalize this ambiguity, I adopt the multiplier preferences of Hansen and Sargent. In this case, our representative agent distorts probability distributions by trading off the plausibility of each alternative distribution and the loss it induces.

The previous Bayesian problem is now modified by introducing two risk sensitivity operators designed to cope with different aspects of this ambiguity. The $T^1$ operator is forward-looking, and guards against misspecification of each model's conditional signal dynamics $p(y_t|x_t^{(k)}, L_t = k)$. The $T^2$ operator is backward-looking, and guards against misspecification of the agent's prior $\pi_{t,k}$ across models, as well as the unobserved state transition dynamics $p(x_{t+1}^{(k)}|x_t^{(k)}, L_t = k)$ [7].

**Definition 3.2.1.** *The two risk sensitivity operators are given by:*

$$\mathbf{T^1}(W(x', c'))(x, c, ; \theta_1) = -\theta_1 \log E\Bigg( \exp\left(-\frac{W(x', c')}{\theta_1}\right) |x, c\Bigg)$$

$$\mathbf{T^2}(W(x, c))(\hat{x}, c, \Sigma; \theta_2) = -\theta_2 \log E\Bigg( \exp\left(-\frac{W(x, c)}{\theta_2}\right) |\hat{x}, c, \Sigma)\Bigg)$$

*where $\theta_1$ and $\theta_2$ are penalty parameters controlling the amplitudes of the two distortions.*

Given this definition, the new risk-sensitive Bellman equation is:

$$W(x, c) = (1-\beta)c + \beta \mathbf{T^2}\Big[\mathbf{T^1}\Big(W(x', c')(x, c, ; \theta_1)\Big)\Big](\hat{x}, c, \Sigma; \theta_2) \tag{3.2.10}$$

Solving this Bellman recursion taking into consideration the definition of the two risk sensitivity operators and the Kalman filtering problem produces a worst case density for each

[7] See Hansen and Sargent (2011). These risk sensitivity operators are mathematically equivalent to finding worst case distributions (see Proposition 1.4.2, Dupuis and Ellis (1997)).

of the probability distributions. The worst case distributions are obtained by distorting the approximating model inversely to their continuation value. This distortion mechanism is also known as *exponential tilting* in the robust control and large deviations literature.

**Proposition 3.2.2.** *The worst case probability distribution of consumption growth taking into consideration the first two sources of uncertainty is:*

$$\tilde{p}(y_t|\hat{x}^{(k)}, L_t = k) = \mathcal{N}\left( D^{(k)}\hat{x}_t^{(k)} - \frac{\beta}{\theta_1} G^{(k)} \left[ C^{(k)\prime}\lambda^{(k)} + G^{(k)\prime} \right] - \frac{1}{\theta_2} D^{(k)}\Sigma_t\lambda^{(k)}; G^{(k)}G^{(k)\prime} + D^{(k)}\Sigma_t D^{(k)\prime} \right)$$

*The worst case distribution of model averaging weights is:*

$$\tilde{\pi}_{t,l} \propto \pi_{t,l} \exp\left( -\frac{U(\hat{x}_t^{(k)}, \Sigma_t^{(k)}, k)}{\theta_2} \right)$$

*Where the continuation value function* $U(\hat{x}_t^{(k)}, \Sigma_t^{(k)}, k) = \lambda^{(k)}\hat{x}_t^{(k)} + \kappa^{(k)} - \frac{1}{2\theta_2}\lambda^{(k)\prime}\Sigma_t^{(k)}\lambda^{(k)}$*,* $\lambda^{(k)} = D^{(k)}\left(I - \beta A^{(k)}\right)^{-1}$*,* $\kappa^{(k)} = -\frac{\beta^2}{2(1-\beta)\theta_1}\left|\lambda^{(k)}C^{(k)} + G^{(k)}\right|^2$

*Proof.* See Hansen and Sargent (2010). □

Few remarks emerge from this proposition:

- the Mean Squared Error(MSE) forecast of consumption growth is unaffacted by the concern for model misspecification. This is mainly due to the fact that the investor has log-utility preferences over instantaneous consumption;

- the mean of conditional distribution of consumption growth is distorted with two terms: a constant forward looking adjustment term and a time varying term coming from the learning process of the latent process. The time varying term is dictated by the MSE of the best forecast of the latent process $\Sigma_t^{(k)}$;

- the priors over models are adjusted inversly with respect to the robust continuation value function. A model with large forecast error of the latent process induces a lower continuation value function and thus the prior models are shifted toward this model.

This last proposition completes the description of our framework. The main results are summarized in Table 3.1. This table provides the predictive model averaging signal and also the stochastic discount factor used to price risk in the context of the foreign exchange market. The two predictive model averaging conditional probabilities of consumption growth emerge from the optimization of our ambiguity averse investor. However how do we calibrate the desire for robustness practically? I address this question in the next section.

Table 3.1: Summary

| | CRRA with IES=1 | Bayesian (BMA) | Robustness(RMA) |
|---|---|---|---|
| Signal and state uncertainty | $p(y_t\|x_t)$ | $p(y_t\|x_t^{(k)}, L_t = k)$ | $\tilde{p}(y_t\|x_t^{(k)}, L_t = k)$ |
| Prior Uncertainty | NA | $\pi_{t,k}$ | $\tilde{\pi}_{t,k}$ |
| Model averaging signal | NA | $\hat{y}_t^{BMA} = \sum_{k=1}^{K} \pi_{t,k} D^{(k)} \hat{x}_t^{(k)}$ | $\tilde{y}_t^{RMA} = \sum_{k=1}^{K} \tilde{\pi}_{t,k} D^{(k)} \hat{x}_t^{(k)}$ |
| Mixture conditional density | NA | $f(y_t\|Y^{t-1}) = \sum_{k=1}^{K} \pi_{t,k} p(y_t\|x_t^{(k)}, L_t = k)$ | $\tilde{f}(y_t\|Y^{t-1}) = \sum_{k=1}^{K} \tilde{\pi}_{t,k} \tilde{p}(y_t\|x_t^{(k)}, L_t = k)$ |
| Stochastic Discount Factor | $m_{t+1} = \beta \frac{c_{t+1}}{c_t}$ | $m_{t+1} = \beta \exp(-\hat{x}_t^{BMA})$ | $\hat{m}_{t+1} = \beta \frac{c_{t+1}}{c_t} \frac{\tilde{f}(y_t\|Y^{t-1})}{f(y_t\|Y^{t-1})}$ |

### 3.2.5 Detection error probability

The robustness parameters $\theta_1$ and $\theta_2$ are calibrated to match the probability of incorrectly choosing the probability distribution that generates the data. In the statistics and information processing literatures this probability is usually called a detection error probability, the model discrimination rate, or the misclassification probability. An application of this concept in economics can be found in Anderson, Hansen, and Sargent (2003) or Hansen and Sargent (2008, chap 9.); for general textbook treatments see *inter alia* Gallager (2014, chap. 3) and Taniguchi and Kakizawa (2000, chap. 7). Let $Pr(p_i|p_j)$ be the probability of selecting model $i$ while the data are generated by $j$. And let $\alpha$ be the prior probability assigned to model $j$. The Bayesian detection error probability is given by $p = \alpha Pr(p_1|p_2) + (1-\alpha) Pr(p_2|p_1)$. In our case, the agent considers two alternatives: (i) the approximating Bayesian model averaging signal $\hat{y}_t^{BMA}$ and (ii) the worst case Robust model averaging version $\hat{y}_t^{RMA}$. Let $f^{BMA}(\omega)$ and $f^{RMA}(\omega)$ be the spectral density functions of these two processes. Then, treating symmetrically the two types of errors (related to the so called type I and type II errors) implies that:

$$p(\theta_1, \theta_2) = 0.5 Pr\Big(f^{BMA}|f^{RMA}\Big) + 0.5 Pr\Big(f^{RMA}|f^{BMA}\Big) \tag{3.2.11}$$

Applying the frequency domain approximation in Taniguchi and Kakizawa (2000) in theorem 7.3.1 gives[8]

$$p(\theta_1, \theta_2) = 0.5\Phi\left(-\sqrt{T}\frac{I(f^{BMA}, f^{RMA})}{V(f^{BMA}, f^{RMA})}\right) + 0.5\Phi\left(-\sqrt{T}\frac{I(f^{RMA}, f^{BMA})}{V(f^{RMA}, f^{BMA})}\right) \tag{3.2.12}$$

[8]See Djeutem and Kasa (2013) for an application of this formula to the calibration of robustness in a (risk-neutral) monetary model of exchange rates. Djeutem and Kasa (2013) show that a preference for robustness can account for observed violations of Shiller bounds.

where $T$ is the sample size, $I$ is the Kullback-Leibler distance between $f^{BMA}$ and $f^{RMA}$, and $V$ is an asymptotic standard error defined by:[9]

$$I(f^i, f^j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[ -\log \frac{|f^i(\omega)|}{|f^j(\omega)|} + \frac{f^i(\omega)}{f^j(\omega)} - 1 \right] d\omega \tag{3.2.13}$$

$$V^2(f^i, f^j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} f^i(\omega) \left[ \frac{1}{f^i(\omega)} - \frac{1}{f^j(\omega)} \right]^2 d\omega \tag{3.2.14}$$

with $i, j \in \{BMA, RMA\}$.

## 3.3 Empirical Analysis

### 3.3.1 Data

This paper studies three US dollar exchange rates: the Canadian dollar, the Japanese Yen, and the British pound. It uses quarterly data covering the period 1979:1 to 2012:2. The dataset is described in more detail in Djeutem (2014). The linear state space models considered here have the same formulation as in Hansen and Sargent (2010) and are given by :

$$\begin{aligned} x_{t+1} &= \begin{pmatrix} \rho & 0 \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} \sigma_x & 0 \\ 0 & 0 \end{pmatrix} \epsilon_{t+1} \\ y_t &= (1,1) x_t + (0, \sigma_y) \epsilon_t \end{aligned} \tag{3.3.15}$$

To capture the specification problem faced by our representative agent, I construct a set of two models that are hard to distinguish in the likelihood sense. I follow a two step procedure. Denote by $x_{2t} = \mu_y$ the deterministic latent process and define the vector $\psi = (\rho, \sigma_x, \sigma_y, \mu_y)$. In the first step, I find the maximum likelihood estimates $\psi_{MLE}$. Then in a second step, I fix the value of $\sigma_y$ to $\sigma_y^{MLE}$ and use profile maximum likelihood over $\rho$ to construct a set of alternative models. Profile maximum likelihood consists of defining a grid of $\rho$ values, and for each value of $\rho$ in the grid, finding the corresponding maximum likelihood estimates of $(\sigma_x, \mu_y)$. A model with log likelihood $\mathcal{L}$ is hard to distinguish from the MLE if $2|\mathcal{L}_{MLE} - \mathcal{L}| < \chi^2_{2,0.95}$. For simplicity, I picked from this set two models : models with the lowest and the highest persistence parameter $\rho$.[10] The model with highest persistence parameter will be labelled the long run risk model, and the other the non long run risk model. Using USA data (see table B.1 for other countries), the calibrated models are :

[9] The assumption here that signal distributions are known to be Gaussian implies that the asymptotic standard errors do not depend on fourth order cumulants.

[10] Profile maximun likelihood potentially provides a systematic approach to constructing the set of models that hard to distinguish using the available data. See Hansen and Sargent (2015) for an alternative approach.

1. Long run risk model, $k = 1$

$$
\begin{aligned}
x_{t+1} = &\quad \begin{pmatrix} 0.995 & 0 \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} 0.00179 & 0 \\ 0 & 0 \end{pmatrix} \epsilon_{t+1} \\
y_t = &\quad (1,1)x_t + (0, 0.00564)\epsilon_t
\end{aligned}
\tag{3.3.16}
$$

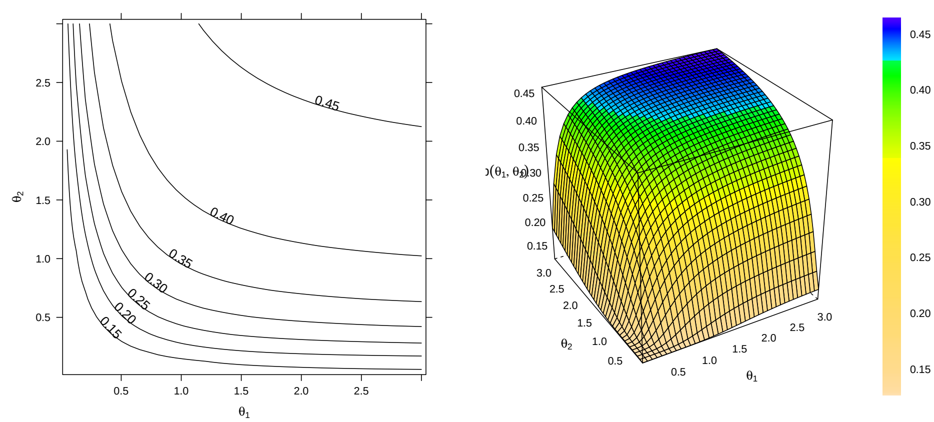
2. Non Long run risk model, $k = 2$

$$
\begin{aligned}
x_{t+1} = &\quad \begin{pmatrix} 0.746 & 0 \\ 0 & 1 \end{pmatrix} x_t + \begin{pmatrix} 0.00366 & 0 \\ 0 & 0 \end{pmatrix} \epsilon_{t+1} \\
y_t = &\quad (1,1)x_t + (0, 0.00564)\epsilon_t
\end{aligned}
\tag{3.3.17}
$$

With these two models in hand, I follow the procedure outlined in Section 3.2.5 to compute the locus of $(\theta_1, \theta_2)$ values that deliver the same detection error probability. I first apply the Kalman filter to each model, and then recursively compute the Bayesian and robust prior over models and the model averaging signals. Finally, Equation 3.2.12 is used to find the misclassification probability.

Figure 3.1: Detection error probabilities

(a) (b)




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2.

Throughout the rest of the paper, I present the results from a perspective of a US investor who constructs a portfolio of three return differentials with respect to the Canadian dollar, the Japanese Yen and the British pound. Then I use US consumption growth to study risk premia[11].

Figure 3.1 plots the detection error probability for each pair of values $(\theta_1, \theta_2)$. This plot reveals two features. First, the detection error probability increases with both $\theta_1$ and

[11]Figure B.3, B.4,B.5 andB.6 in the appendices contains the result from the perspective of a foreign investor pricing risk using foreign consumption. The results a similar to the case of a US investor

$\theta_2$. This is not surprising since an increase in the robustness parameter means that it is more costly to the evil agent to distort the approximating model. As a result, the size of the distortion is smaller and the approximating and worst case models will therefore be more similar. Thus, the detection error probability is higher. Second, the locus of values delivering the same detection error probability is downward sloping. Using this locus of values, I calibrate $(\theta_1, \theta_2)$ to yield a detection error probability of 20%. These calibrated values are used in the next two sections to derive the posterior model probabilities and the Hansen Jagannatan bound.

Table 3.2: Calibration

| | Parameter | Value |
|---|---|---|
| Subjective Discount factor | $\beta$ | 0.995 |
| $\mathbf{T^1}$ Parameter | $\theta_1$ | 0.770 |
| $\mathbf{T^2}$ Parameter | $\theta_2$ | 1.880 |
| Detection error probability | $p(\theta_1, \theta_2)$ | 0.200 |

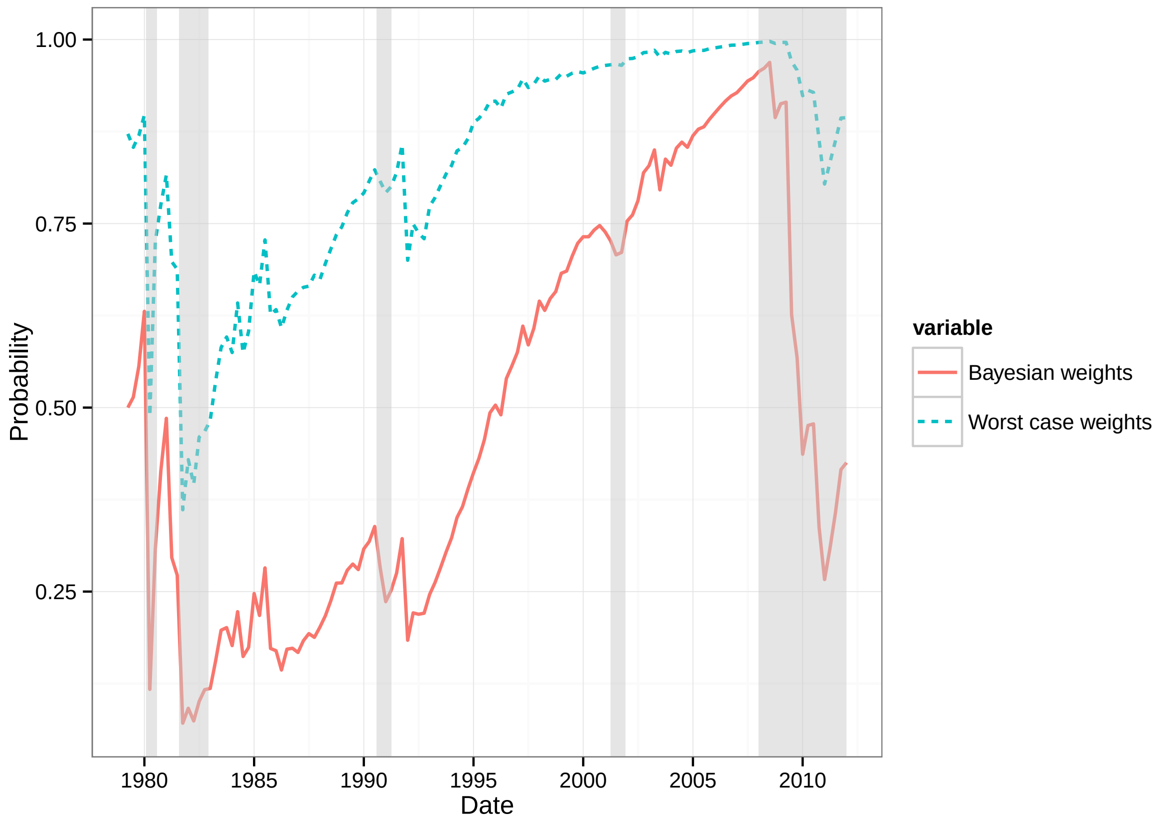
### 3.3.2 Posterior model probabilities

This section analyzes the posterior model probabilities contained in line 2 of Table 3.1 and depicted in Figure 3.2. The benchmark Bayesian probabilities can be subdivided roughly into 3 main sub-periods. In the 1980s, a Bayesian learner puts more weight on the non-long risk model characterized by low persistence, ie., the temporary component of the state space model plays a bigger role in the dynamics of consumption growth. This makes sense since this period was characterized by relatively volatile consumption growth. The next sub-period goes from the 1990s until 2007. During this period the long run risk model is the main driver of consumption growth which is less volatile. Therefore the process is mostly driven by the low frequency component. The last sub-period begins with the recent financial crisis and is again marked by an increase in the probability attached to the non long risk model.

This plot also shows that when the agent is averse to both misspecification of the state and to model priors, this leads to a uniform increase in the posterior probability assigned to the long run risk model. Notice the Robust Bayesian probabilities are consistently greater than 50%. More importantly, notice that although recessions trigger a reduction in the probability of the long-run risk model for both the Bayesian and the robust learner, the reduction is much greater for the Bayesian. That is, the bias of the robust learner is *state dependent*.

The next section investigates the properties of the stochastic discount factor generated by our robust model averaging investor. More specifically, we are interested in seeing whether it satisfies the Hansen and Jagannathan (1991) bound, since satisfying this bound has proven to be a challenge for conventional risk premium models.

Figure 3.2: Posterior model probabilities



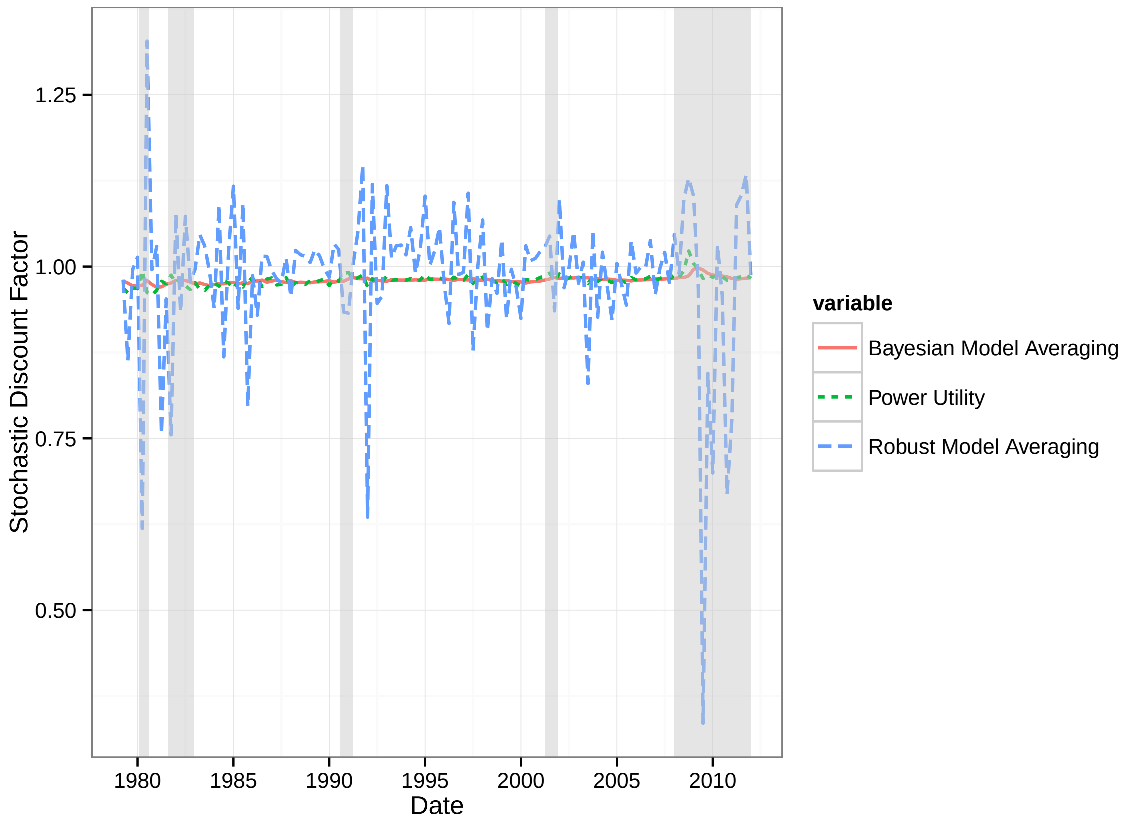

*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. Robustness parameters $(\theta_1, \theta_2) = (0.77, 1.88)$. Detection error probability $p(\theta_1, \theta_2) = 20\%$. The sample period is 1979Q1-2012Q2.

### 3.3.3 Hansen Jagannathan Bound

This subsection asks whether the robust stochastic discount factor is volatile enough to satisfy the Euler equation and the implied Hansen Jagannathan bound. To answer this question, I start with a visual inspection, and plot the stochastic discount factor implied by our model. This corresponds to line 5 of Table 3.1. Figure 3.3 clearly shows that relative to CRRA and Bayesian stochastic discount factors, the robust model averaging version is very volatile. Next, I plot the Hansen Jagannathan bound. This bound places restrictions on the first two moments of any valid stochastic discount factor, and thus provides a simple test of whether a given model is consistent with the data. In this paper I use the generalized version of this bound due to Otrok, Ravikumar, and Whiteman (2007). Their bound takes into account not only current returns, but also past and future returns. As a consequence, it provides a test for all horizons (e.g., short, medium and long). The bound is given by:

$$f_m(\omega) \geq \overline{f_{Rg}(\omega)}' f_R(\omega)^{-1} f_{Rg}(\omega) \ \forall \omega \tag{3.3.18}$$

Figure 3.3: Stochastic Discount Factors



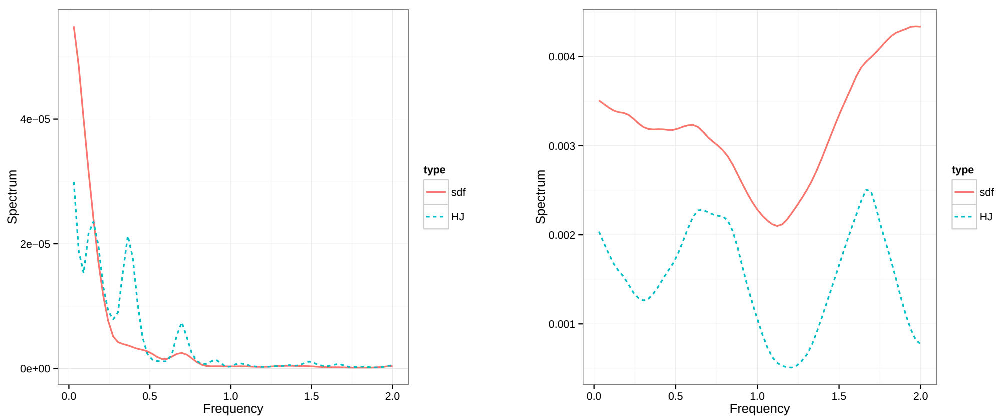

*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. Robustness parameters $(\theta_1, \theta_2) = (0.77, 1.88)$. Detection error probability $p(\theta_1, \theta_2) = 20\%$. The sample period is 1979Q1-2012Q2.

where $f_m$ is the spectral density of the stochastic discount factor, $f_{Rg}$ is the cross-spectrum between the assets' returns and the stochastic discount factor, $f_R$ is the spectral density matrix of assets returns. Denote the right hand side of Equation 3.3.18 as $f_g$ and refer to it as the *spectral bound.*

Figure 3.4 illustrates this spectral bound (dotted line) and the the spectrum of the stochastic discount factors(solid line). Each panel corresponds to a specific model and plots on the horizontal axis the frequency and on the vertical the corresponding spectrum. The spectral bound is almost always violated at all frequencies for the log utility model and the Bayesian model averaging version. In contrast, our robust learning version satisfies the bound at all frequencies. Moreover, this is achieved with an empirically plausible detection error probability of 20%. A formal model comparison procedure is conducted in the next section to more formally assess which of the three stochastic discount factors is more consistent with observed excess returns data in the foreign exchange market.

Figure 3.4: Hansen Jagannathan bound by frequency

(a) BMA (b) RMA




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. Robustness parameters $(\theta_1, \theta_2) = (0.77, 1.88)$. Detection error probability $p(\theta_1, \theta_2) = 20\%$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$.

### 3.3.4 Model Comparison

One candidate metric for model comparison would be to use the quadratic form introduced by Hansen and Jagannathan (1997). This distance depends on the implied pricing error of the stochastic discount factor and is given by:

$$d_{HJ} = \sqrt{m'_T \Sigma^{-1}_{R,T} m_T} \tag{3.3.19}$$

where $m_T$ is the vector of the sample average of pricing errors with $m_{i,T} = \frac{1}{T}\sum_{t=1}^{T} m_t R^e_{i,t}$, $\Sigma_{R,T} = \frac{1}{T} R^e R^{e'}$ is the sample second moment matrix of excess returns. Alternatively, I compute the symmetric Kullback-Leiber distance between the spectral density functions of the stochastic discount factor $f_m(\omega)$ and the spectral density function of the generalized Hansen Jagannathan bound $f_g(\omega)$.

$$d_{Spec} = I(f_m, f_g) + I(f_g, f_m) \tag{3.3.20}$$

I then find the pair of robustness parameters that minimizes this distance. The spectral distance for the robust model averaging is 1.153, about 23% lower than the benchmark Bayesian model averaging version and 26% lower than the time separable power utility with unit elasticity counterpart. It therefore appears that according to this spectral metric the proposed robust learning model is more consistent with the data. All these distances are reported in Table 3.3 together with the HJ distance. The HJ distance gives the same ranking of models.

Table 3.3: Pricing specification errors by models

| Models | HJ distance | | | | Spectral distance | | | |
|---|---|---|---|---|---|---|---|---|
| | $d_{HJ}$ | $\theta_1$ | $\theta_2$ | $p(\theta_1, \theta_2)$ | $d_{spec}$ | $\theta_1$ | $\theta_2$ | $p(\theta_1, \theta_2)$ |
| Robust Model Averaging | 0.040 | 15.261 | 0.317 | 0.160 | 1.153 | 3.979 | 0.843 | 0.268 |
| Bayesian Model Averaging | 0.096 | | | | 1.415 | | | |
| CRRA(IES=1) | 0.096 | | | | 1.561 | | | |

This table provides the HJ and spectral distance. The value of the robustness parameter are obtained by minimizing the corresponding distance. The subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

These results offer a reinterpretation of Bansal and Shaliastovich (2013). One could argue that their estimated coefficient of risk aversion is implausibly high. My results show that this parameter need not be interpreted in terms of risk aversion. Instead, a concern for model misspecification combined with learning provides an alternative interpretation in which the uncertainty is calibrated to detection error probabilities rather than risk. I find reasonable detection error probabilities (in the range 15-30%) can explain observed risk premium dynamics.

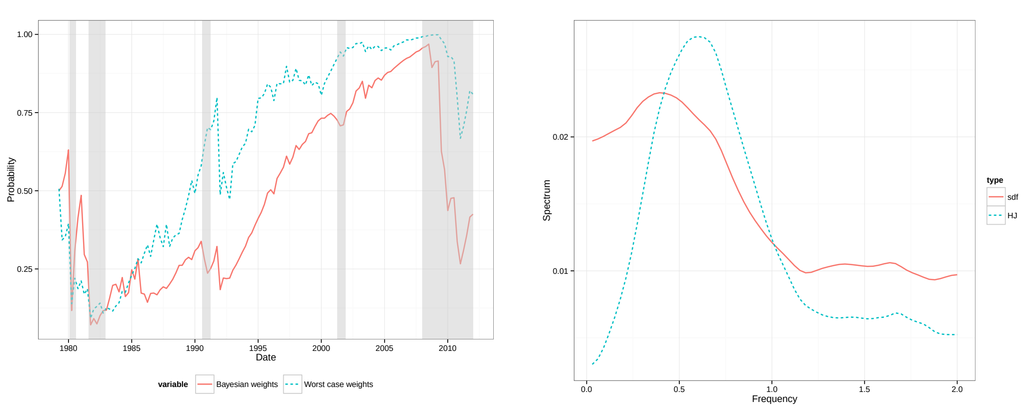
### 3.3.5 The Importance of Learning

A natural question at this point is to ask what is the relative contribution of the two risk-sensitivity operators. To address this question, I shut down the forward-looking learning operator $\mathbf{T^1}$ by letting $\theta_1 = +\infty$, and see how this affects the HJ bound and the probability weights assigned to each model. The robustness parameter associated with $\mathbf{T^2}$ is set to $\theta_2 = 0.74$ yielding a detection error probability of 30%.

Figure 3.5: Model probabilities and Hansen Jagannathan bound with $\mathbf{T^2}$ only

(a) Posterior models probabilities

(b) HJ bound




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. Robustness parameters $(\theta_1, \theta_2) = (+\infty, 0.74)$. Detection error probability $p(\theta_1, \theta_2) = 30\%$. The sample period is 1979Q1-2012Q2.

Two main features emerge from this experiment (see 3.5). First, the worst case probability weight assigned to the long run risk consumption growth is smaller. Second, the Hansen Jagannathan bound is now violated at high frequencies.

## 3.4 Conclusion

This paper has studied risk premia in the foreign exchange market using a consumption-based asset pricing framework with model uncertainty. I showed that fears of model misspecification combined with *robust learning* accounts for both the average level of the risk premium and its state dependent dynamics. The key mechanism is that our robust agent tilts asymmetrically model probabilities toward the consumption growth model with the highest degree of persistence, and this bias increases during bad times. This mechanism generates a stochastic discount factor that satisfies the Hansen-Jagannathan volatility bound at all frequencies, with a modest detection error probability of 20 %.

To preserve tractability, this paper used a linear state space model together with a unit elasticity of intertemporal substitution. One avenue for future research would be to relax this assumption, and use for instance a model with stochastic volatility. In this case nonlinear filtering techniques (eg, the particle filter) will be needed. Alternatively one could also use Ju and Miao (2012) preferences to model a concern for model misspecification, but at the price of losing the backward and forward interpretations of Hansen and Sargent's operators. Perhaps the most important direction will be to solve a fully specified general equilibrium model incorporating model uncertainty and learning, and study its implications for exchange rate dynamics and UIP regressions. In this regard Colacito and Croce (2012) would be a good starting point.

# Bibliography

Alexius, A. (2001): "Uncovered Interest Parity Revisited," *Review of International Economics*, 9(3), 505–517.

Alvarez, F., A. Atkeson, and P. J. Kehoe (2009): "Time-Varying Risk, Interest Rates, and Exchange Rates in General Equilibrium," *The Review of Economic Studies*, 76(3), 851–878.

Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003): "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1(1), 68–123.

Bacchetta, P., and E. van Wincoop (2010): "Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle," *American Economic Review*, 100, 870–904.

Backus, D. K., F. Gavazzoni, C. Telmer, and S. E. Zin (2010): "Monetary Policy and the Uncovered Interest Parity Puzzle," NBER Working Paper 16218, National Bureau of Economic Research, Inc.

Bansal, R., and M. Dahlquist (2000): "The forward premium puzzle: different tales from developed and emerging economies," *Journal of International Economics*, 51(1), 115–144.

Bansal, R., and I. Shaliastovich (2013): "A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets," *Review of Financial Studies*, 26(1), 1–33.

Bansal, R., and A. Yaron (2004): "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009): "Doubts or variability?," *Journal of Economic Theory*, 144(6), 2388–2418.

Bekaert, G., R. J. Hodrick, and D. A. Marshall (1997): "The implications of first-order risk aversion for asset market risk premiums," *Journal of Monetary Economics*, 40(1), 3–39.

Benigno, G., P. Benigno, and S. Nisticò (2011): "Risk, Monetary Policy and the Exchange Rate," Working Paper 17133, National Bureau of Economic Research.

Bidder, R., and M. Smith (2011): "Doubts and variability," Unpublished paper.

Burnside, C. (1994): "Hansen-Jagannathan Bounds as Classical Tests of Asset-Pricing Models," *Journal of Business & Economic Statistics*, 12(1), 57–79.

Burnside, C., B. Han, D. Hirshleifer, and T. Y. Wang (2011): "Investor Overconfidence and the Forward Premium Puzzle," *Review of Economic Studies*, 78, 523–58.

Campbell, J. Y., and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

Cecchetti, S. G., P.-S. Lam, and N. C. Mark (1994): "Testing Volatility Restrictions on Intertemporal Marginal Rates of Substitution Implied by Euler Equations and Asset Returns," *The Journal of Finance*, 49(1), 123–152.

Chaboud, A. P., and J. H. Wright (2005): "Uncovered interest parity: it works, but not for long," *Journal of International Economics*, 66(2), 349–362.

Chakraborty, A., and G. W. Evans (2008): "Can perpetual learning explain the forward-premium puzzle?," *Journal of Monetary Economics*, 55(3), 477–490.

Chinn, M. D. (2006): "The (partial) rehabilitation of interest rate parity in the floating rate era: Longer horizons, alternative expectations, and emerging markets," *Journal of International Money and Finance*, 25(1), 7–21.

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

Colacito, R., and M. M. Croce (2011a): "International Asset Pricing with Risk-Sensitive Agents," mimeo, University of North Carolina.

——— (2011b): "Risks for the Long Run and the Real Exchange Rate," *Journal of Political Economy*, 119(1), 153–181.

——— (2012): "International Robust Disagreement," *American Economic Review*, 102(3), 152–55.

Diba, B. T. (1987): "A Critique of Variance Bounds Tests for Monetary Exchange Rate Models," *Journal of Money, Credit and Banking*, 19, 104–111.

Djeutem, E. (2012): "Model Uncertainty and the Forward Premium Puzzle," Working Paper, Simon Fraser University.

——— (2014): "Model uncertainty and the Forward Premium Puzzle," *Journal of International Money and Finance*, 46, 16–40.

Djeutem, E., and K. Kasa (2013): "Robustness and Exchange Rate Volatility," *Journal of International Economics*, 91, 27–39.

Dupuis, P., and R. S. Ellis (1997): *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley-Interscience, New York.

Eichenbaum, M., and C. L. Evans (1995): "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *Quarterly Journal of Economics*, 110, 975–1009.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics*, 75(4), 643–669.

Engel, C. (1996): "The forward discount anomaly and the risk premium: A survey of recent evidence," *Journal of Empirical Finance*, 3(2), 123–192.

——— (2005): "Some New Variance Bounds for Asset Prices," *Journal of Money, Credit and Banking*, 37, 949–955.

——— (2012): "The Real Exchange Rate, Real Interest Rates, and the Risk Premium," Working Paper, University Of Wisconsin.

——— (2015): "Chapter 8 - Exchange Rates and Interest Parity," in *Handbook of International Economics*, ed. by K. R. a. G. G. Elhanan Helpman, vol. 4 of *Handbook of International Economics*, pp. 453–522. Elsevier.

Engel, C., N. Mark, and K. D. West (2007): "Exchange Rate Models Are Not As Bad As You Think," *NBER Macroeconomics Annual*, 22, 381–441.

Engel, C., and K. D. West (2004): "Accounting for Exchange Rate Variability in Present-Value Models When the Discount Factor is Near 1," *American Economic Review*, 94, 119–125.

——— (2005): "Exchange Rates and Fundamentals," *Journal of Political Economy*, 113(3), 485–517.

——— (2006): "Taylor Rules and the Deutschmark-Dollar Real Exchange Rate," *Journal of Money, Credit, and Banking*, 38, 1175–94.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Evans, G. W. (1986): "A Test for Speculative Bubbles in the Sterling-Dollar Exchange Rate: 1981-84," *American Economic Review*, 76, 621–36.

Fama, E. F. (1984): "Forward and spot exchange rates," *Journal of Monetary Economics*, 14(3), 319–338.

Farhi, E., and X. Gabaix (2011): "Rare Disasters and Exchange Rates," mimeo, Harvard University.

——— (2014): "Rare Disasters and Exchange Rates," Working Paper 71001, Harvard University OpenScholar.

Gallager, R. G. (2014): *Stochastic Processes: Theory for Applications*. Cambridge University Press, Cambridge, United Kingdom ; New York.

Gilboa, I., and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18(2), 141–153.

Gilles, C., and S. F. LeRoy (1991): "Econometric Aspects of the Variance-Bounds Tests: A Survey," *The Review of Financial Studies*, 4, 753–791.

Gourinchas, P. O., and A. Tornell (2004): "Exchange Rate Puzzles and Distorted Beliefs," *Journal of International Economics*, 64, 303–333.

Gourio, F., M. Siemer, and A. Verdelhan (2013): "International risk cycles," *Journal of International Economics*, 89(2), 471–484.

Hansen, L., and T. Sargent (1995): "Discounted linear exponential quadratic Gaussian control," *IEEE Transactions on Automatic Control*, 40(5), 968–971.

Hansen, L. P. (2014): "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122(5), 945–987.

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov (2007): "Chapter 61 Intertemporal Substitution and Risk Aversion," in *Handbook of Econometrics*, ed. by James J. Heckman and Edward E. Leamer, vol. Volume 6, Part A, pp. 3967–4056. Elsevier.

Hansen, L. P., and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88(5), 829–853.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), 225–262.

——— (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52(2), 557?590.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7–46.

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

——— (2010): "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1(1), 129?162.

——— (2011): "Chapter 20 - Wanting Robustness in Macroeconomics," in *Handbook of Monetary Economics*, ed. by Benjamin M. Friedman and Michael Woodford, vol. Volume 3, pp. 1097–1157. Elsevier.

——— (2015): "Sets of models and prices of uncertainty," Working paper, University of Chicago.

Hansen, L. P., T. J. Sargent, and T. D. Tallarini (1999): "Robust Permanent Income and Pricing," *The Review of Economic Studies*, 66(4), 873–907.

Hnatkovska, V., A. Lahiri, and C. A. Vegh (2013): "Interest rate and the exchange rate: A non-monotonic tale," *European Economic Review*, 63, 68–93.

Hong, H., and J. C. Stein (1999): "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *Journal of Finance*, 54, 2143–84.

Huang, R. D. (1981): "The Monetary Approach to Exchange Rate in an Efficient Foreign Exchange Market: Tests Based on Volatility," *The Journal of Finance*, 36, 31–41.

Huber, P. J. (1981): *Robust Statistics*. Wiley, New York.

Ilut, C. (2012): "Ambiguity Aversion: Implications for the Uncovered Interest Rate Parity Puzzle," *American Economic Journal: Macroeconomics*, 4(3), 33–65.

Jeanne, O., and A. Rose (2002): "Noise Trading and Exchange Rate Regimes," *Quarterly Journal of Economics*, 117, 537–69.

Ju, N., and J. Miao (2012): "Ambiguity, Learning, and Asset Returns," *Econometrica*, 80(2), 559?591.

Kasa, K. (2001): "A Robust Hansen-Sargent Prediction Formula," *Economics Letters*, 71, 43–48.

Kasa, K., T. B. Walker, and C. H. Whiteman (2010): "Heterogeneous Beliefs and Tests of Present Value Models," mimeo, University of Iowa.

Kim, Y. S. (2009): "Exchange Rates and Fundamentals Under Adaptive Learning," *Journal of Economic Dynamics and Control*, 33, 843–63.

King, R., W. Weber, and N. Wallace (1992): "Nonfundamental Uncertainty and Exchange Rates," *Journal of International Economics*, 32, 83–108.

Kleidon, A. W. (1986): "Variance Bounds Tests and Stock Price Valuation," *Journal of Political Economy*, 94, 953–1001.

Knight, F. H. (1921): *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin Company.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Lewis, K. F., and C. H. Whiteman (2008): "Robustifying Shiller: Do Stock Prices Move Enough to be Justified by Subsequent Changes in Dividends?," mimeo, University of Iowa.

Lewis, K. K. (1989): "Can learning affect exchange-rate behavior?: The case of the dollar in the early 1980's," *Journal of Monetary Economics*, 23(1), 79–100.

——— (2011): "Global Asset Pricing," *Annual Review of Financial Economics*, 3(1), 435–466.

Li, M., and A. Tornell (2008): "Exchange Rates Under Robustness: An Account of the Forward Premium Puzzle," mimeo, UCLA.

Ljungqvist, L., and T. J. Sargent (2012): *Recursive Macroeconomic Theory*. The MIT Press, third edition edn.

Lu, Y. K., and M. Siemer (2014): "Learning, Rare Disasters, and Asset Prices," SSRN Scholarly Paper ID 1977867, Social Science Research Network, Rochester, NY.

Lucas, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

——— (1982): "Interest rates and currency prices in a two-country world," *Journal of Monetary Economics*, 10(3), 335–359.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447ï£¡1498.

Mankiw, N. G., D. Romer, and M. D. Shapiro (1985): "An Unbiased Reexamination of Stock Market Volatility," *Journal of Finance*, 40, 677–87.

Manuelli, R. E., and J. Peck (1990): "Exchange Rate Volatility in an Equilibrium Asset Pricing Model," *International Economic Review*, 31, 559–74.

Mark, N. C. (1985): "On time varying risk premia in the foreign exchange market: An econometric analysis," *Journal of Monetary Economics*, 16(1), 3–18.

——— (2001): *International Macroeconomics and Finance: Theory and Econometric Methods*. Blackwell, Oxford.

Markiewicz, A. (2012): "Model Uncertainty and Exchange Rate Volatility," *International Economic Review*, 53, 815–44.

Meese, R. A. (1986): "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?," *Journal of Political Economy*, 94, 345–73.

Mehra, R., and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15(2), 145–161.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012): "Carry Trades and Global Foreign Exchange Volatility," *The Journal of Finance*, 67(2), 681–718.

Modjtahedi, B. (1991): "Multiple maturities and time-varying risk premia in forward exchange markets: An econometric analysis," *Journal of International Economics*, 30(1ï£¡2), 69–86.

Moore, M. J., and M. J. Roche (2002): "Less of a puzzle: a new look at the forward forex market," *Journal of International Economics*, 58(2), 387–411.

Otrok, C., B. Ravikumar, and C. H. Whiteman (2007): "A generalized volatility bound for dynamic economies," *Journal of Monetary Economics*, 54(8), 2269–2290.

Piazzesi, M., and M. Schneider (2006): "Equilibrium Yield Curves," Working Paper 12609, National Bureau of Economic Research.

Rabitsch, K. (2014): "An Incomplete Markets Explanation to the UIP puzzle," Working paper, Vienna University of Economics and Business,Department of Economics.

Raftery, A. E., M. Kárný, and P. Ettler (2010): "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52(1), 52–66, PMID: 20607102.

Sarno, L. (2005): "Viewpoint: Towards a solution to the puzzles in exchange rate economics: where do we stand?," *Canadian Journal of Economics/Revue canadienne d'Ãľconomique*, 38(3), 673âĂŞ708.

Savage, L. J. (1954): *The Foundations of Statistics*. New-York: John Wiley.

Shiller, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71(3), 421–436.

——— (1989): *Market Volatility*. MIT Press.

Strzalecki, T. (2011): "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 79, 47–73.

Tallarini, J. T. D. (2000): "Risk-sensitive real business cycles," *Journal of Monetary Economics*, 45(3), 507–532.

Taniguchi, M., and Y. Kakizawa (2000): *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Verdelhan, A. (2010): "A Habit-Based Explanation of the Exchange Rate Risk Premium," *The Journal of Finance*, 65(1), 123ï£¡146.

Wachter, J. A. (2006): "A consumption-based model of the term structure of interest rates," *Journal of Financial Economics*, 79(2), 365–399.

Weil, P. (1990): "Nonexpected Utility in Macroeconomics," *The Quarterly Journal of Economics*, 105(1), 29–42.

West, K. D. (1987): "A Standard Monetary Model and the Variability of the Deutschemark-Dollar Exchange Rate," *Journal of International Economics*, 23, 57–76.

——— (1988): "Dividend Innovations and Stock Price Volatility," *Econometrica*, 56, 37–61.

Woodford, M. (1990): "Learning to Believe in Sunspots," *Econometrica*, 58, 277–307.

Young, N. (1988): *An Introduction to Hilbert Space*. Cambridge University Press, Cambridge.

Zames, G. (1981): "Feedback and Optimal Sensitivity: Model Reference Transformation, Multiplicative Seminorms, and Approximate Inverses," *IEEE Transactions on Automatic Control*, 26, 301–20.

# Appendix A

# Model uncertainty and the forward premium puzzle

## A.1 Additional tables

Table A.1: Fama regression : $s_{t+1} - s_t = \alpha_0 + \alpha_1(i_t - i_t^*) + \varepsilon_{t+1}$

| | $\alpha_0$ | $\alpha_1$ | $R^2$ | F | $Pr(> F)$ |
|---|---|---|---|---|---|
| US-CAN | 0.000 | -0.347 | 0.002 | 2.015 | 0.137 |
| | (0.003) | (0.745) | | | |
| US-JPN | 0.024 | -2.377 | 0.058 | 8.188 | 0.000 |
| | (0.008) | (0.834) | | | |
| US-UK | -0.008 | -1.527 | 0.024 | 4.457 | 0.013 |
| | (0.006) | (0.856) | | | |
| UK-CAN | 0.012 | -3.261 | 0.075 | 9.025 | 0.000 |
| | (0.005) | (1.003) | | | |
| UK-JPN | 0.029 | -1.700 | 0.016 | 2.735 | 0.069 |
| | (0.014) | (1.167) | | | |
| JPN-CAN | -0.028 | -2.492 | 0.039 | 5.318 | 0.006 |
| | (0.011) | (1.079) | | | |

Notes: (1) F is the F-statistics for the joint test of $\alpha_0 = 0$ and $\alpha_1 = 1$.
(2) $Pr(> F)$ is the p-value for the joint test.
(3) Number in parentheses are standard errors.
(4) Sample 1979Q1-2012Q2

Table A.2: Distance to HJ bound and corresponding $\gamma$ and $p(\gamma)$ (Standard deviations)

| $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ | $\gamma$ | | | | $p(\gamma)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | CAN | JPN | UK | US | CAN | JPN | UK | US |
| | | | | | **Random Walk** | | | |
| 1.00 | 2.8 | 3.0 | 2.5 | 4.1 | 4.5 | 7.0 | 5.7 | 5.7 |
| 0.75 | 2.1 | 2.3 | 1.8 | 3.1 | 3.5 | 5.5 | 4.3 | 4.5 |
| 0.50 | 1.4 | 1.5 | 1.2 | 2.1 | 2.4 | 3.8 | 2.9 | 3.1 |
| 0.25 | 0.7 | 0.7 | 0.5 | 1.0 | 1.2 | 1.6 | 1.1 | 1.5 |
| | | | | | **Trend Stationary** | | | |
| 1.00 | 41.8 | 21.5 | 28.5 | 64.8 | 4.4 | 7.8 | 5.7 | 5.6 |
| 0.75 | 31.7 | 17.3 | 21.7 | 49.1 | 3.5 | 6.3 | 4.5 | 4.5 |
| 0.50 | 21.6 | 13.0 | 14.8 | 33.4 | 2.4 | 4.4 | 3.1 | 3.2 |
| 0.25 | 11.5 | 8.7 | 8.1 | 17.6 | 1.3 | 2.3 | 1.6 | 1.7 |

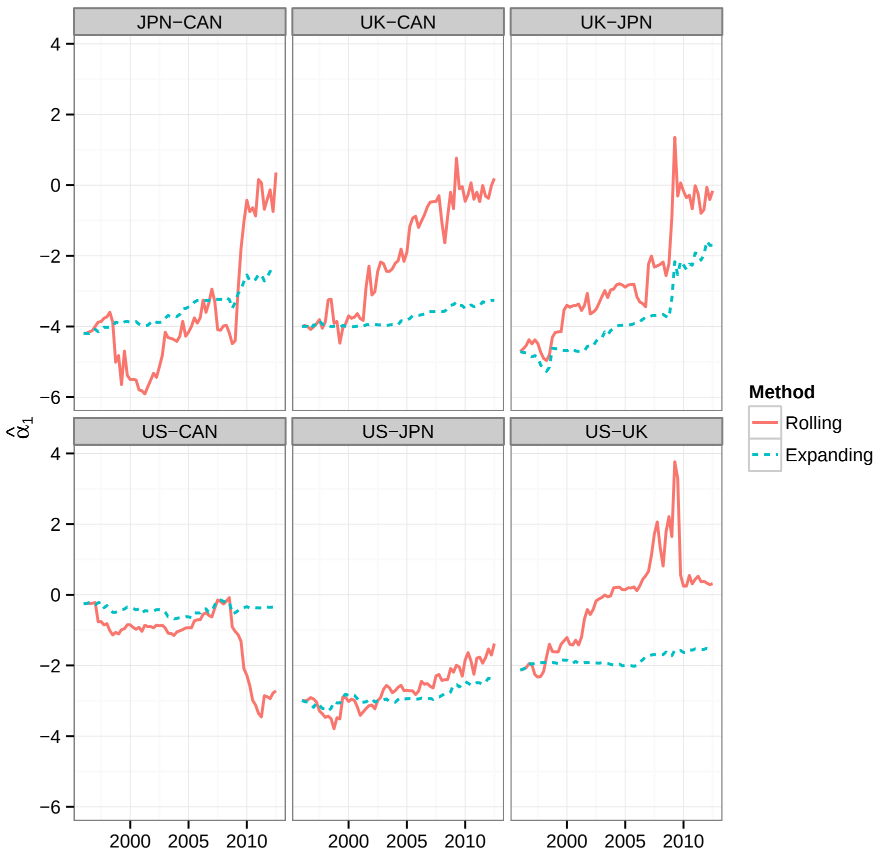
This table provides the minimal $\gamma$ and the corresponding detection error probability for different distance to the bound. These statistique are standard deviations over 67 samples constructed using expanding window. The first sample is 1979Q1-1995Q3 the second one is 1979Q1-1995Q4 etc. The ratio $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ equal to 1.00 means that the corresponding $\gamma$ puts the $(E(m), \sigma(m))$ on the HJ bound; while 0.50 put $(E(m), \sigma(m))$ half way to the bound. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

Table A.3: Distance to HJ bound and corresponding $\gamma$ and $p(\gamma)$ (Minimum)

| $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ | $\gamma$ | | | | $p(\gamma)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | CAN | JPN | UK | US | CAN | JPN | UK | US |
| | | | | | **Random Walk** | | | |
| 1.00 | 4 | 3 | 1 | 6 | 44.5 | 46.7 | 49.6 | 40.7 |
| 0.75 | 3 | 2 | 1 | 5 | 46.3 | 48.0 | 50.0 | 43.4 |
| 0.50 | 2 | 1 | 1 | 3 | 48.2 | 49.3 | 49.8 | 46.2 |
| 0.25 | 1 | 1 | 1 | 2 | 50.0 | 49.9 | 49.9 | 48.9 |
| | | | | | **Trend Stationary** | | | |
| 1.00 | 24 | 3 | 26 | 44 | 33.7 | 28.8 | 41.8 | 30.3 |
| 0.75 | 19 | 3 | 21 | 34 | 37.4 | 33.2 | 43.4 | 34.6 |
| 0.50 | 13 | 2 | 16 | 24 | 41.1 | 37.9 | 44.9 | 39.1 |
| 0.25 | 8 | 2 | 12 | 14 | 45.0 | 42.8 | 46.5 | 43.7 |

This table provides the minimal $\gamma$ and the corresponding detection error probability for different distance to the bound. These statistique are miniumun over 67 samples constructed using expanding window. The first sample is 1979Q1-1995Q3 the second one is 1979Q1-1995Q4 etc. The ratio $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ equal to 1.00 means that the corresponding $\gamma$ puts the $(E(m), \sigma(m))$ on the HJ bound; while 0.50 put $(E(m), \sigma(m))$ half way to the bound. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

Figure A.1: Dynamic UIP Regression using rolling and expanding window




*Notes:* This plot gives the UIP slope coefficient. The rolling window line corresponds to the regression using half sample and progressively add one observation at the top while removing one at the queue. This is done till the end of the full sample 2012Q2. The expanding window only adds new observation. Sample period: 1979Q1-2012Q2.

Table A.4: Distance to HJ bound and corresponding $\gamma$ and $p(\gamma)$ (Maximun)

| $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ | $\gamma$ CAN | JPN | UK | US | $p(\gamma)$ CAN | JPN | UK | US |
|---|---|---|---|---|---|---|---|---|
| | | | | | **Random Walk** | | | |
| 1.00 | 17 | 15 | 12 | 23 | 24.2 | 19.9 | 26.1 | 21.4 |
| 0.75 | 13 | 11 | 9 | 18 | 30.3 | 26.7 | 32.0 | 27.8 |
| 0.50 | 9 | 8 | 6 | 12 | 37.1 | 34.7 | 38.5 | 35.2 |
| 0.25 | 4 | 4 | 3 | 6 | 44.3 | 43.4 | 45.4 | 43.1 |
| | | | | | **Trend Stationary** | | | |
| 1.00 | 205 | 113 | 145 | 289 | 21.9 | 39.1 | 32.5 | 19.5 |
| 0.75 | 157 | 90 | 112 | 220 | 27.7 | 41.3 | 36.2 | 25.7 |
| 0.50 | 108 | 66 | 80 | 150 | 34.2 | 43.6 | 40.2 | 32.8 |
| 0.25 | 59 | 43 | 47 | 81 | 41.3 | 45.9 | 44.2 | 40.6 |

This table provides the minimal $\gamma$ and the corresponding detection error probability for different distance to the bound. These statistique are maximum over 67 samples constructed using expanding window. The first sample is 1979Q1-1995Q3 the second one is 1979Q1-1995Q4 etc. The ratio $\frac{\sigma(m)}{E(m)\sqrt{\bar{R}^{e\prime}\Sigma^{-1}\bar{R}^{e}}}$ equal to 1.00 means that the corresponding $\gamma$ puts the $(E(m), \sigma(m))$ on the HJ bound; while 0.50 put $(E(m), \sigma(m))$ half way to the bound. Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$

## A.2 Proof proposition 2.2.4

*Proof.* **RW case:**
Value Function:
The recursive formulation is fully characterized by the following system of equations:

$$
\begin{aligned}
V_t &= c_t - \beta\theta \, log \, E_t \left[ exp(-\frac{V_{t+1}}{\theta}) \right] \\
c_{t+1} &= \mu + c_t + \sigma_\epsilon \varepsilon_{t+1} \\
V_t &= A + B \, c_t
\end{aligned}
$$

Using the third equation into the first yields:

$$
\begin{aligned}
V_t &= c_t - \beta\theta \log \, E_t \left[ \exp(-\frac{A + Bc_{t+1}}{\theta}) \right] \\
&= c_t - \beta\theta \, \log \, E_t \left[ \exp(-\frac{A + B(\mu + c_t + \sigma_\epsilon \varepsilon_{t+1})}{\theta}) \right] \\
&= c_t - \beta\theta \, \log \, E_t \left[ exp(-\frac{A + B(\mu + c_t)}{\theta}) \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta}) \right] \\
&= c_t + \beta(A + B(\mu + c_t)) - \beta\theta \, \log \, E_t \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta}) \\
&= c_t + \beta(A + B(\mu + c_t)) - \beta\theta(\frac{B^2 \sigma_\epsilon^2}{2\theta^2}) \\
&= c_t + \beta(A + B(\mu + c_t)) - \beta(\frac{B^2 \sigma_\epsilon^2}{2\theta})
\end{aligned}
$$

The equality between line 4 and line 5 uses:

$$
\begin{aligned}
\varepsilon_{t+1} \sim \mathcal{N}(0,1) \Rightarrow \quad & -\frac{B}{\theta} \sigma_\epsilon \varepsilon_{t+1} \sim \mathcal{N}(0, \frac{B^2}{\theta^2} \sigma_\epsilon^2) \\
\Rightarrow \quad & \exp(-\frac{B}{\theta} \sigma_\epsilon \varepsilon_{t+1}) \sim log\mathcal{N}(0, \frac{B^2}{\theta^2} \sigma_\epsilon^2) \\
\Rightarrow \quad & E_t(\exp(-\frac{B}{\theta} \sigma_\epsilon \varepsilon_{t+1})) = \exp(0 + \frac{B^2 \sigma_\epsilon^2}{2\theta^2})
\end{aligned}
$$

Mapping the coefficient in the LHS to those in the RHS of the value function gives, the following system of equations:

$$
\begin{cases} B = & 1 + \beta B \\ A = & \beta A - \beta \frac{B^2 \sigma_\epsilon^2}{2\theta} + \beta B \mu \end{cases} \Rightarrow \begin{cases} B = & \frac{1}{1-\beta} \\ A = & \frac{\beta}{(1-\beta)^2} \left( \mu - \frac{\sigma_\epsilon^2}{2(1-\beta)\theta} \right) \end{cases}
$$

Using the fact that $\theta = -\frac{1}{(1-\beta)(1-\gamma)}$:

$$
V_t = \frac{\beta}{(1-\beta)^2} \left( \mu - \frac{\sigma_\epsilon^2 (\gamma - 1)}{2} \right) + \frac{1}{1-\beta} c_t
$$

Martingale increment:
The martingale increment is given by

$$
\begin{aligned}
g_{t+1} = &\quad \frac{\exp(-\frac{V_{t+1}}{\theta})}{E_t(\exp(-\frac{V_{t+1}}{\theta}))} \\
= &\quad \frac{\exp(-\frac{A+Bc_{t+1}}{\theta})}{E_t \exp(-\frac{A+Bc_{t+1}}{\theta})} \\
= &\quad \frac{\exp(-\frac{A+B(\mu+c_t)}{\theta}) \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta})}{E_t(\exp(-\frac{A+B(\mu+c_t)}{\theta}) \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta}))} \\
= &\quad \frac{\exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta})}{E_t(\exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta}))} \\
= &\quad \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\varepsilon^2}{2\theta^2}) \\
= &\quad \exp(-\sigma_\epsilon(\gamma-1)\varepsilon_{t+1} - \frac{(\gamma-1)^2\sigma_\varepsilon^2}{2})
\end{aligned}
$$

Consumption mean distortion:
Using the likelihood ratio interpretation of $g_{t+1}$ it follows that:

$$
\begin{aligned}
\hat{p}_\epsilon(\varepsilon_{t+1}) &= g_{t+1}(\varepsilon_{t+1}) p_\epsilon(\varepsilon_{t+1}) \\
&\propto \exp(-\frac{B\sigma_\epsilon \varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\varepsilon^2}{2\theta^2}) \exp(-\frac{\varepsilon_{t+1}^2}{2}) \\
&\propto \exp\left(-\frac{1}{2}(\varepsilon_{t+1} + \frac{B}{\theta}\sigma_\varepsilon)^2\right)
\end{aligned}
$$

And the mean distortion is : $w_{t+1} = -\frac{B}{\theta}\sigma_\varepsilon$

**TS case:**
First, the consumption process should be rewritten in a single equation taking into account the autocorrelated errors as follow:

$$
\begin{cases} c_{t+1} = \lambda + \mu t + z_{t+1} \\ z_t = \rho z_{t-1} + \sigma_\varepsilon \varepsilon_t \end{cases} \Rightarrow \begin{cases} c_{t+1} = ((1-\rho)\lambda + \rho\mu) + \mu(1-\rho)t + \rho c_t + \sigma_\varepsilon \varepsilon_{t+1} \\ c_{t+1} = \tilde{\lambda} + \tilde{\mu} t + \rho c_t + \sigma_\varepsilon \varepsilon_{t+1} \end{cases}
$$

Value function:
And the recursive formulation will be :

$$
\begin{aligned}
V_t = &\quad c_t - \beta\theta \log\, E_t \left[\exp(-\frac{V_{t+1}}{\theta})\right] \\
c_{t+1} = &\quad \tilde{\lambda} + \tilde{\mu} t + \rho c_t + \sigma_\epsilon \varepsilon_{t+1} \\
V_t = &\quad D + E\, t + F\, c_t
\end{aligned}
$$

$$
\begin{aligned}
V_t = \quad & c_t - \beta\theta \log\, E_t \left( \exp(-\frac{D + E(t+1) + F\, c_{t+1}}{\theta}) \right) \\
= \quad & c_t - \beta\theta \log\, E_t \left( \exp(-\frac{D + E(t+1) + F\,(\tilde{\lambda} + \tilde{\mu} t + \rho c_t + \sigma_\epsilon \varepsilon_{t+1})}{\theta}) \right) \\
= \quad & c_t - \beta\theta \log\, E_t \left( \exp(-\frac{D + E + Et + F(\tilde{\lambda} + \tilde{\mu} t + \rho c_t)}{\theta}) \exp(-\frac{F\sigma_\epsilon \varepsilon_{t+1}}{\theta}) \right) \\
= \quad & c_t + \beta(D + E + Et + F(\tilde{\lambda} + \tilde{\mu} t + \rho c_t)) - \beta\theta \log E_t(\exp(-\frac{F\sigma_\epsilon \varepsilon_{t+1}}{\theta})) \\
= \quad & c_t + \beta(D + E + Et + F(\tilde{\lambda} + \tilde{\mu} t + \rho c_t)) - \beta\theta(\frac{F^2\sigma_\epsilon^2}{2\theta^2}) \\
= \quad & c_t + \beta(D + E + Et + F(\tilde{\lambda} + \tilde{\mu} t + \rho c_t)) - \beta(\frac{F^2\sigma_\epsilon^2}{2\theta})
\end{aligned}
$$

Again, mapping the coefficient of the LHS and RHS of the last equality gives:

$$
\begin{cases} D = \beta(D + E + F\tilde{\lambda} - \frac{F^2\sigma_\varepsilon^2}{2\theta}) \\ E = \beta(E + F\tilde{\mu}) \\ F = 1 + \beta F \rho \end{cases} \Rightarrow \begin{cases} D = \frac{\beta}{(1-\beta)} \left( \frac{\beta\mu(1-\rho)}{(1-\beta)(1-\beta\rho)} + \frac{(1-\rho)\lambda + \rho\mu}{1-\beta\rho} - \frac{\sigma_\epsilon^2(1-\beta)(\gamma-1)}{2(1-\beta\rho)^2} \right) \\ E = \frac{\beta\mu(1-\rho)}{(1-\beta)(1-\beta\rho)} \\ F = \frac{1}{1-\beta\rho} \end{cases}
$$

And Finally:

$$
V_t = \frac{\beta}{(1-\beta)} \left( \frac{\beta\mu(1-\rho)}{(1-\beta)(1-\beta\rho)} + \frac{(1-\rho)\lambda + \rho\mu}{1-\beta\rho} - \frac{\sigma_\epsilon^2(1-\beta)(\gamma-1)}{2(1-\beta\rho)^2} \right) + \frac{\beta\mu(1-\rho)}{(1-\beta)(1-\beta\rho)} t + \frac{1}{1-\beta\rho} c_t
$$

Martingale Increment:
Using the same procedure as in the RW case, we have :

$$
\begin{aligned}
g_{t+1} = \quad & \exp\left( -\frac{F\sigma_\epsilon \varepsilon_{t+1}}{\theta} - \frac{F^2\sigma_\varepsilon^2}{2\theta^2} \right) \\
= \quad & \exp\left( -\frac{(1-\beta)(\gamma-1)}{1-\beta\rho} \sigma_\epsilon \varepsilon_{t+1} - \frac{(1-\beta)^2(\gamma-1)^2\sigma_\varepsilon^2}{2(1-\beta\rho)^2} \right)
\end{aligned}
$$

□

Consumption mean distortion:
Using the likelihood ratio interpretation of $g_{t+1}$ it follows that:

$$\begin{aligned}
\hat{p}_\epsilon(\varepsilon_{t+1}) = & \quad g_{t+1}(\varepsilon_{t+1})p_\epsilon(\varepsilon_{t+1}) \\
\propto & \quad \exp(-\frac{F\sigma_\epsilon\varepsilon_{t+1}}{\theta} - \frac{F^2\sigma_\varepsilon^2}{2\theta^2})\exp(-\frac{\varepsilon_{t+1}^2}{2}) \\
\propto & \quad \exp\left(-\frac{1}{2}(\varepsilon_{t+1} + \frac{F}{\theta}\sigma_\varepsilon)^2\right)
\end{aligned}$$

And the mean distortion is : $w_{t+1} = -\frac{F}{\theta}\sigma_\varepsilon$

## A.3 Proof proposition 2.2.5

**RW case:**

$$\begin{aligned}
m_{t+1} = & \quad \left(\beta\frac{C_t}{C_{t+1}}\right)\left(\frac{e^{-\frac{V_{t+1}}{\theta}}}{E(e^{-\frac{V_{t+1}}{\theta}}|\mathcal{F}_t)}\right) \\
= & \quad \left(\beta\frac{C_t}{C_{t+1}}\right)g_{t+1} \\
= & \quad \beta\exp\left(-\log(\frac{C_{t+1}}{C_t})\right)\exp\left(-\frac{B\sigma_\epsilon\varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\epsilon^2}{2\theta^2}\right) \\
= & \quad \beta\exp\left(-\Delta c_{t+1} - \frac{B\sigma_\epsilon\varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\epsilon^2}{2\theta^2}\right) \\
= & \quad \beta\exp\left(-\mu - \sigma_\epsilon\varepsilon_{t+1} - \frac{B\sigma_\epsilon\varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\epsilon^2}{2\theta^2}\right)
\end{aligned}$$

Taking the logarithm of this last expression gives:

$$\log(m_{t+1}) \quad = \log(\beta) - \mu - \sigma_\epsilon\varepsilon_{t+1} - \frac{B\sigma_\epsilon\varepsilon_{t+1}}{\theta} - \frac{B^2\sigma_\epsilon^2}{2\theta^2}$$

$$\Rightarrow \quad \left\{ \begin{aligned}
E(\log(m_{t+1})) = & \quad \log(\beta) - \mu - \frac{B^2\sigma_\epsilon^2}{2\theta^2} \\
var(\log(m_{t=1})) = & \quad \sigma_\epsilon^2\, var(\varepsilon_{t+1}) + \frac{B^2\sigma_\epsilon^2}{\theta^2}\, var(\varepsilon_{t+1}) + 2\frac{B\sigma_\epsilon^2}{\theta}\, cov(\varepsilon_{t+1}, \varepsilon_{t+1}) \\
= & \quad \sigma_\epsilon^2 + \frac{B^2\sigma_\epsilon^2}{\theta^2} + 2\frac{B\sigma_\epsilon^2}{\theta} \\
= & \quad \sigma_\epsilon^2\left(1 + \frac{B}{\theta}\right)^2
\end{aligned}\right.$$

Since $\log(m_{t+1})$ is a linear combination of Gaussian distributions, it is also Gaussian distribution fully characterized but its first two moments. Thus:

$$\log m_{t+1} \sim \mathcal{N}(\log\beta - \mu - \frac{B^2}{2\theta^2}\sigma_\varepsilon^2, \sigma_\varepsilon^2(1+\frac{B}{\theta})^2)$$

**TS case:**

$$\begin{aligned}
c_{t+1} = \tilde{\lambda} + \tilde{\mu}t + \rho c_t + \sigma_\varepsilon \varepsilon_{t+1} \Rightarrow \quad & \Delta c_{t+1} = & \tilde{\mu} + \rho\Delta c_t + \sigma_\varepsilon \Delta\varepsilon_{t+1} \\
\Rightarrow \quad & (1-\rho L)\Delta c_{t+1} = & \tilde{\mu} + \sigma_\varepsilon \Delta\varepsilon_{t+1} \\
\Rightarrow \quad & \Delta c_{t+1} = & \frac{\tilde{\mu}}{1-\rho L} + \frac{\sigma_\epsilon \Delta\varepsilon_{t+1}}{1-\rho L} \\
\Rightarrow \quad & \Delta c_{t+1} = & \frac{\tilde{\mu}}{1-\rho} + \sigma_\epsilon \sum_{j=0}^{\infty} \rho^j \Delta\varepsilon_{t+1-j} \\
\Rightarrow \quad & \Delta c_{t+1} = & \mu + \sigma_\epsilon \sum_{j=0}^{\infty} \rho^j \Delta\varepsilon_{t+1-j}
\end{aligned}$$

Using the same procedure as in the RW case we have:

$$\log(m_{t+1}) = \log(\beta) - \Delta c_{t+1} - \frac{F\sigma_\epsilon \varepsilon_{t+1}}{\theta} - \frac{F^2\sigma_\epsilon^2}{2\theta^2}$$

Plugging the $MA(\infty)$ representation of the TS process gives:

$$\log(m_{t+1}) = \log(\beta) - \mu - \sigma_\epsilon \sum_{j=0}^{\infty} \rho^j \Delta\varepsilon_{t+1-j} - \frac{F\sigma_\epsilon \varepsilon_{t+1}}{\theta} - \frac{F^2\sigma_\epsilon^2}{2\theta^2}$$

And

$$
\begin{aligned}
E(\log(m_{t+1})) &= \log(\beta) - \mu - \frac{F^2\sigma_\epsilon^2}{2\theta^2} \\
Var(\log(m_{t+1})) &= \sigma_\epsilon^2 Var\left[\sum_{j=0}^{\infty}\rho^j\Delta\varepsilon_{t+1-j} + \frac{F\varepsilon_{t+1}}{\theta}\right] \\
&= \sigma_\epsilon^2\left[Var(\sum_{j=0}^{\infty}\rho^j\Delta\varepsilon_{t+1-j}) + \frac{F^2}{\theta^2}Var(\varepsilon_{t+1}) + 2\frac{F}{\theta}Cov(\sum_{j=0}^{\infty}\rho^j\Delta\varepsilon_{t+1-j},\varepsilon_{t+1})\right] \\
&= \sigma_\epsilon^2\left[\sum_{j=0}^{\infty}\rho^{2j}Var(\Delta\varepsilon_{t+1-j}) + 2\sum_{j=0}^{\infty}Cov(\rho^j\Delta\varepsilon_{t+1-j},\rho^{j+1}\Delta\varepsilon_{t-j}) + \frac{F^2}{\theta^2} + 2\frac{F}{\theta}Cov(\varepsilon_{t+1}-\varepsilon_t,\varepsilon_{t+1}\right. \\
&= \sigma_\epsilon^2\left[2\sum_{j=0}^{\infty}\rho^{2j} - 2\rho\sum_{j=0}^{\infty}\rho^{2j} + \frac{F^2}{\theta^2} + 2\frac{F}{\theta}\right] \\
&= \sigma_\epsilon^2\left[\frac{2}{1-\rho^2} - \frac{2\rho}{1-\rho^2} + \frac{F^2}{\theta^2} + 2\frac{F}{\theta}\right] \\
&= \sigma_\epsilon^2\left[\frac{2}{1+\rho} + \frac{F^2}{\theta^2} + 2\frac{F}{\theta}\right]
\end{aligned}
$$

Hence :

$$\log m_{t+1} \sim \mathcal{N}\left(\log\beta - \mu - \frac{F^2}{2\theta^2}, \sigma_\varepsilon^2(\frac{2}{1+\rho} + 2\frac{F}{\theta} + \frac{F^2}{\theta^2})\right)$$

Finally the mean and the variance are computed using the properties of a log-normal distribution. If $\log X \sim \mathcal{N}(\mu,\sigma^2)$ then

$$
\begin{aligned}
E(X) &= \exp(\mu + \frac{\sigma^2)}{2} \\
Var(X) &= (\exp(\sigma^2) - 1)E(X)^2
\end{aligned}
$$

## A.4 Proof proposition 2.2.6

*Proof.* **RW case:**
Consider a sample $\{c_t\}_0^T$ of consumption following the approximating and worst case model outline previously. The conditional likelihood are respectively given by :

$$
\begin{aligned}
lnL_A &= -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}[\Delta c_{t+1} - \mu_A]^2 \\
lnL_B &= -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}[\Delta c_{t+1} - \mu_B]^2
\end{aligned}
$$

The log likelihood ratio is then:

$$
\begin{aligned}
lnL_A - lnL_B &= -\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\left[\Delta c_{t+1}-\mu_A\right]^2+\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\left[\Delta c_{t+1}-\mu_B\right]^2 \\
&= \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\left[(\Delta c_{t+1}-\mu_B)^2-(\Delta c_{t+1}-\mu_A)^2\right] \\
&= \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\Big[(\mu_A-\mu_B)\left(2\Delta c_{t+1}-\mu_A-\mu_B\right)\Big] \\
&= \frac{1}{\sigma_\epsilon^2}\left[\sum_{t=2}^{T}\Delta c_{t+1}-\frac{T-1}{2}\left(\mu_A+\mu_B\right)\right] \\
&= \frac{(\mu_A-\mu_B)}{\sigma_\epsilon^2}\left[(T-1)\hat{\mu}-(\frac{T-1}{2})(\mu_A+\mu_B)\right] \\
&= \frac{(T-1)(\mu_A-\mu_B)}{\sigma_\epsilon^2}\left[\hat{\mu}-\frac{(\mu_A+\mu_B)}{2}\right]
\end{aligned}
$$

Where $\hat{\mu}=\frac{1}{T-1}\sum\limits_{t=2}^{T}\Delta c_{t+1}$.

Since$\Delta c_{t+1}=\mu+\varepsilon_{t+1}$ with $\varepsilon_t\sim iid\mathcal{N}(0,\sigma_\epsilon^2)$, the Central Limit Theorem implies that: $\frac{\sqrt{T-1}(\hat{\mu}-\mu)}{\sigma_\epsilon}\sim\mathcal{N}(0,1)$ and :

$$
\begin{aligned}
p_A=\mathbb{P}\left(lnL_A-lnL_B<0|A\right)&=\mathbb{P}\left\{\frac{(T-1)(\mu_A-\mu_B)}{\sigma_\epsilon^2}\left[\hat{\mu}-\frac{(\mu_A+\mu_B)}{2}\right]<0|A\right\} \\
&=\mathbb{P}\left\{\frac{(T-1)(\mu_A-\mu_B)}{\sigma_\epsilon^2}\left[\hat{\mu}-\mu_A+\frac{(\mu_A-\mu_B)}{2}\right]<0|A\right\} \\
&=\mathbb{P}\left\{\frac{\sqrt{T-1}(\hat{\mu}-\mu_A)}{\sigma_\varepsilon}<\frac{\sqrt{T-1}(\mu_B-\mu_A)}{2\sigma_\varepsilon}|A\right\}\text{if } \mu_A>\mu_B \\
&=\Phi\left\{\frac{\sqrt{T-1}}{2\sigma_\varepsilon}(\mu_B-\mu_A)\right\}\text{if } \mu_A>\mu_B
\end{aligned}
$$

In the other case we have:

$$
\begin{aligned}
p_A=&\mathbb{P}\left\{\frac{\sqrt{T-1}(\hat{\mu}-\mu_A)}{\sigma_\varepsilon}\geq\frac{\sqrt{T-1}(\mu_B-\mu_A)}{2\sigma_\varepsilon}|A\right\}\text{ if } \mu_A\leq\mu_B \\
=&1-\Phi\left\{\frac{\sqrt{T-1}}{2\sigma_\varepsilon}(\mu_B-\mu_A)\right\}\text{ if } \mu_A\leq\mu_B
\end{aligned}
$$

In sum:

$$
p_A=\Phi\left\{-\frac{\sqrt{T-1}}{2\sigma_\varepsilon}\left|\mu_B-\mu_A\right|\right\}
$$

Similarly,

$$p_B = \Phi\left\{-\frac{\sqrt{T-1}}{2\sigma_\varepsilon}\left|\mu_B-\mu_A\right|\right\}$$

And thus :

$$p(\gamma) = \Phi\left\{-\frac{\sqrt{T-1}}{2\sigma_\varepsilon}\left|\mu_B-\mu_A\right|\right\} = \Phi\left(-0.5\sqrt{T}\sigma_\epsilon\left|\gamma-1\right|\right)$$

The equality comes comes from the fact that the mean distortion in random walk case is $|\mu_B-\mu_A| = \sigma_\epsilon^2\left|\gamma-1\right|$.

**TS case:**

The averaged conditional log-likelihood functions are :

$$lnL_A = -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2(T-1)}\sum_{t=2}^{T}\left[c_t-(1-\rho)\lambda_A-\mu[t-\rho(t-1)]-\rho c_{t-1}\right]^2$$

$$lnL_B = -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2(T-1)}\sum_{t=2}^{T}\left[c_t-(1-\rho)\lambda_B-\mu[t-\rho(t-1)]-\rho c_{t-1}\right]^2$$

And the log likelihood ratio is :

$$\begin{aligned}
lnL_A - lnL_B =& \frac{(1-\rho)^2(\lambda_A-\lambda_B)}{2\sigma_\epsilon^2(T-1)}\sum_{t=2}^{T}\left[\frac{2}{1-\rho}\left(c_t-\rho c_{t-1}-\mu(t-\rho(t-1))\right)-(\lambda_A+\lambda_B)\right]\\
=& \frac{(1-\rho)^2(\lambda_A-\lambda_B)}{2\sigma_\epsilon^2}\left\{\frac{1}{(T-1)(1-\rho)}\sum_{t=2}^{T}\left[c_t-\rho c_{t-1}-\mu(t-\rho(t-1))\right]-\frac{\lambda_A+\lambda_B}{2}\right\}\\
=& \frac{(1-\rho)^2(\lambda_A-\lambda_B)}{2\sigma_\epsilon^2}\left\{\hat{\lambda}-\frac{\lambda_A+\lambda_B}{2}\right\}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(lnL_A - lnL_B < 0|A) =& \mathbb{P}\left(\frac{(1-\rho)^2(\lambda_A-\lambda_B)}{2\sigma_\epsilon^2}\left\{\hat{\lambda}-\frac{\lambda_A+\lambda_B}{2}\right\}<0|A\right)\\
=& \begin{cases}\mathbb{P}\left(\hat{\lambda}-\frac{\lambda_A+\lambda_B}{2}<0|A\right) \text{ if } \lambda_A>\lambda_B\\ \mathbb{P}\left(\hat{\lambda}-\frac{\lambda_A+\lambda_B}{2}>\geq|A\right) \text{ if } \lambda_A\leq\lambda_B\end{cases}\\
=& \begin{cases}\mathbb{P}\left(\frac{\sqrt{T}(\hat{\lambda}-\lambda_A)}{\sigma_\epsilon}<\frac{\sqrt{T}(\lambda_B-\lambda_A)}{2\sigma_\epsilon}\right) \text{ if } \lambda_A>\lambda_B\\ \mathbb{P}\left(\frac{\sqrt{T}(\hat{\lambda}-\lambda_A)}{\sigma_\epsilon}\geq\frac{\sqrt{T}(\lambda_B-\lambda_A)}{2\sigma_\epsilon}\right) \text{ if } \lambda_A\leq\lambda_B\end{cases}\\
=& \Phi\left(-\frac{\sqrt{T}\left|\lambda_A-\lambda_B\right|}{2\sigma_\epsilon}\right)
\end{aligned}$$

Similarly,

$$\mathbb{P}(lnL_A - lnL_B > 0|B) = \Phi\left(-\frac{\sqrt{T}\,|\lambda_A - \lambda_B|}{2\sigma_\epsilon}\right)$$

and the detection probability

$$\begin{aligned} p(\gamma) = & \quad \frac{\mathbb{P}(lnL_A - lnL_B < 0|A) + \mathbb{P}(lnL_A - lnL_B > 0|B)}{2} \\ = & \quad \Phi\left(-\frac{\sqrt{T}\,|\lambda_A - \lambda_B|}{2\sigma_\epsilon}\right) \end{aligned}$$

Since the mean distortion is given by :

$$|\lambda_A - \lambda_B| = \frac{\sigma_\epsilon^2(1-\beta)\,|\gamma - 1|}{1-\rho\beta}$$

Plugging the mean distortion back into the expression of detection probability gives:

$$p(\gamma) = \Phi\left(-0.5\sqrt{T}\sigma_\epsilon\frac{(1-\beta)\,|\gamma-1|}{1-\rho\beta}\right)$$

$\square$

# Appendix B

# Robust learning in the foreign exchange market

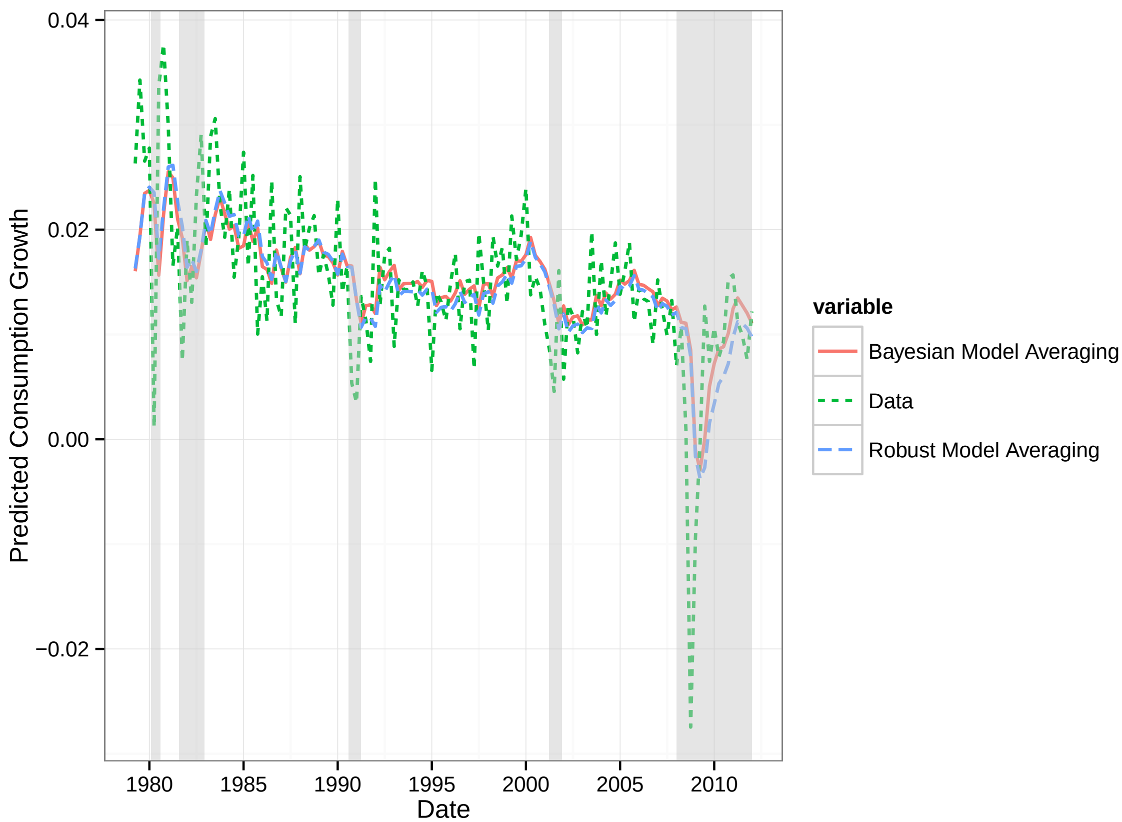
## B.1 Additional tables

Table B.1: Models

| | CAN | | JPN | | UK | | US | |
|---|---|---|---|---|---|---|---|---|
| | est. | se | est. | se | est. | se | est. | se |
| | | | | **MLE** | | | | |
| $\rho$ | 0.99110 | 0.01471 | 0.99426 | 0.00946 | 0.98982 | 0.01382 | 0.93972 | 0.06993 |
| $\sigma_x$ | 0.00129 | 0.00046 | 0.00092 | 0.00033 | 0.00240 | 0.00073 | 0.00230 | 0.00093 |
| $\sigma_y$ | 0.00610 | 0.00044 | 0.01000 | 0.00063 | 0.00810 | 0.00063 | 0.00564 | 0.00057 |
| $\mu_y$ | 0.01641 | 0.01272 | 0.00660 | 0.01249 | 0.02215 | 0.01679 | 0.01544 | 0.01047 |
| $-\mathcal{L}_{MLE}$ | -474.93998 | | -417.06674 | | -431.70185 | | -474.70339 | |
| | | | | **Long Run Risk** | | | | |
| $\rho$ | 0.99500 | | 0.99500 | | 0.99500 | | 0.99500 | |
| $\sigma_x$ | 0.00122 | 0.00035 | 0.00091 | 0.00028 | 0.00230 | 0.00060 | 0.00179 | 0.00051 |
| $\sigma_y$ | 0.00610 | 0.00044 | 0.01000 | 0.00063 | 0.00810 | 0.00063 | 0.00564 | 0.00057 |
| $\mu_y$ | 0.01680 | 0.01464 | 0.00661 | 0.01283 | 0.02339 | 0.02244 | 0.01705 | 0.01855 |
| $2\lvert\mathcal{L}_{MLE} - \mathcal{L}\rvert$ | 0.11134 | | 0.00669 | | 0.22305 | | 1.93452 | |
| | | | | **Non Long Run Risk** | | | | |
| $\rho$ | 0.88556 | | 0.93530 | | 0.90545 | | 0.74626 | |
| $\sigma_x$ | 0.00252 | 0.00045 | 0.00177 | 0.00044 | 0.00368 | 0.00064 | 0.00366 | 0.00051 |
| $\sigma_y$ | 0.00610 | 0.00044 | 0.01000 | 0.00063 | 0.00810 | 0.00063 | 0.00564 | 0.00057 |
| $\mu_y$ | 0.01477 | 0.01018 | 0.00640 | 0.01028 | 0.01806 | 0.01052 | 0.01501 | 0.01009 |
| $2\lvert\mathcal{L}_{MLE} - \mathcal{L}\rvert$ | 5.44966 | | 5.20789 | | 5.95974 | | 5.58613 | |

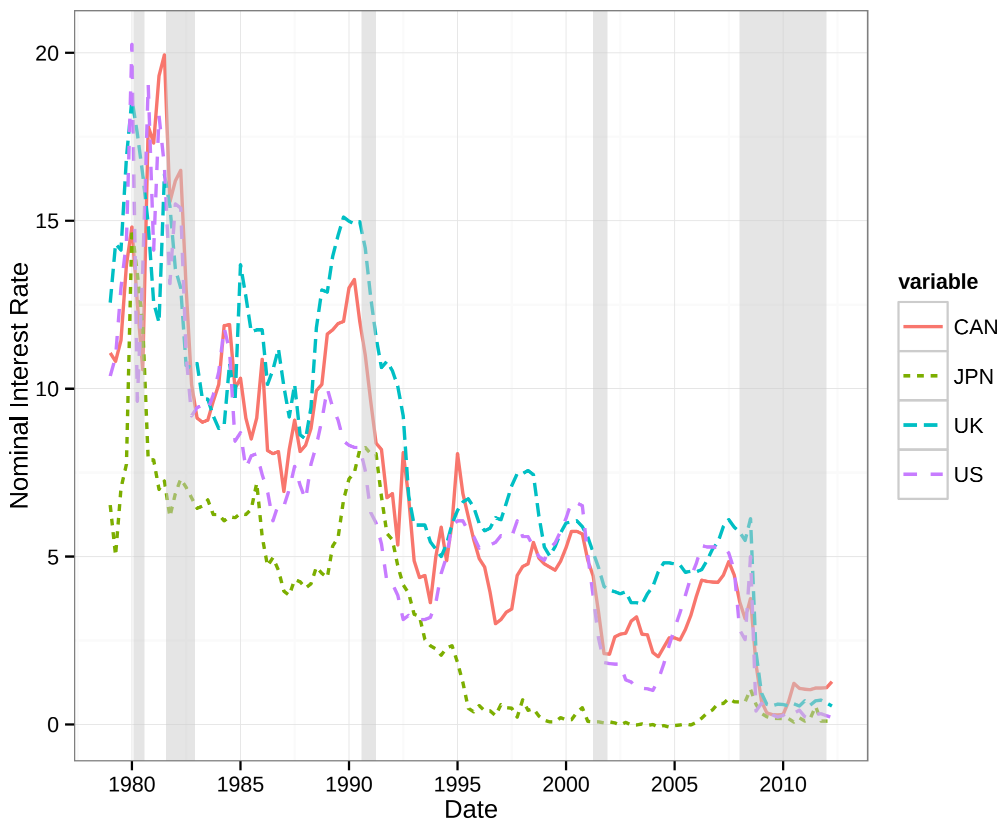
The sample period is 1979Q1-2012Q2.

Figure B.1: Predicted Model Averaging Consumption Growth - USA



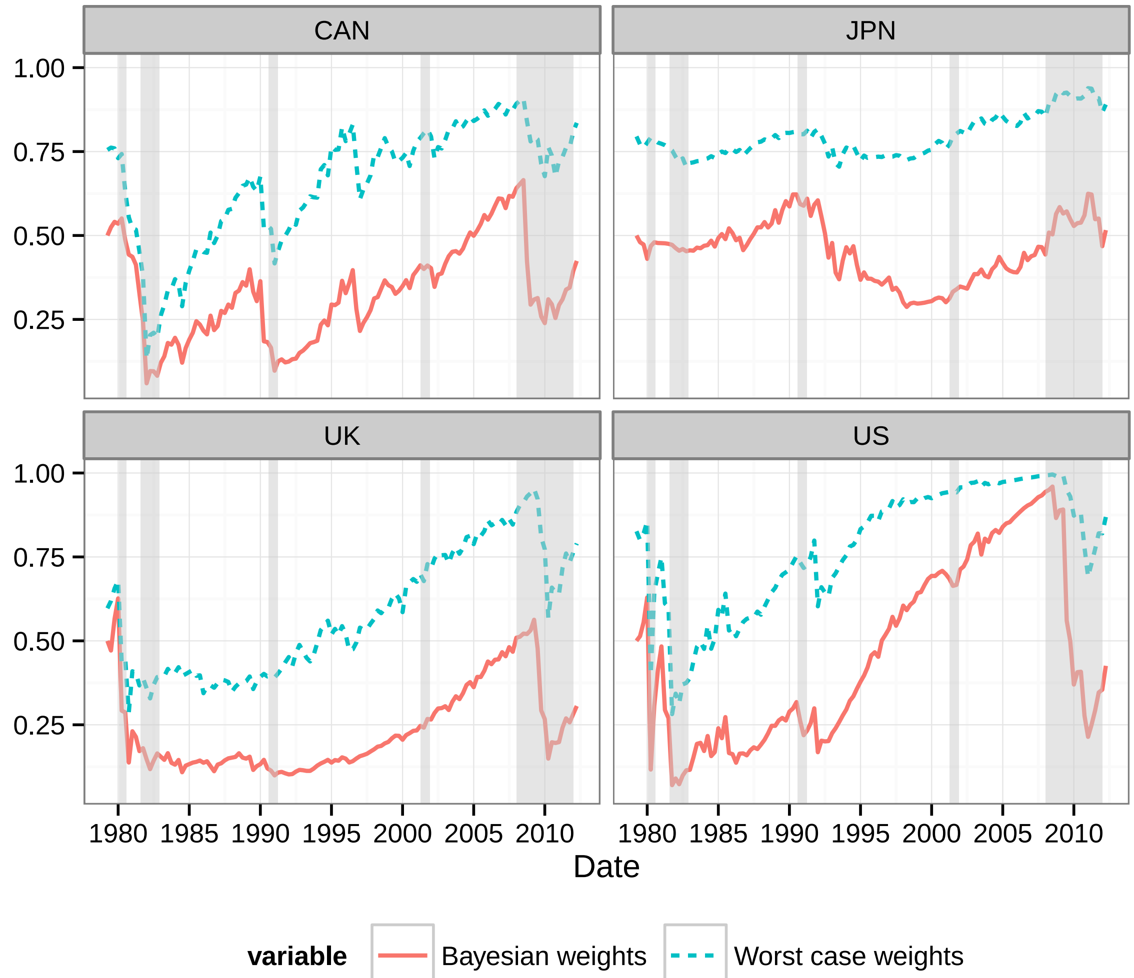

*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2.

Figure B.2: Nominal Interest Rate



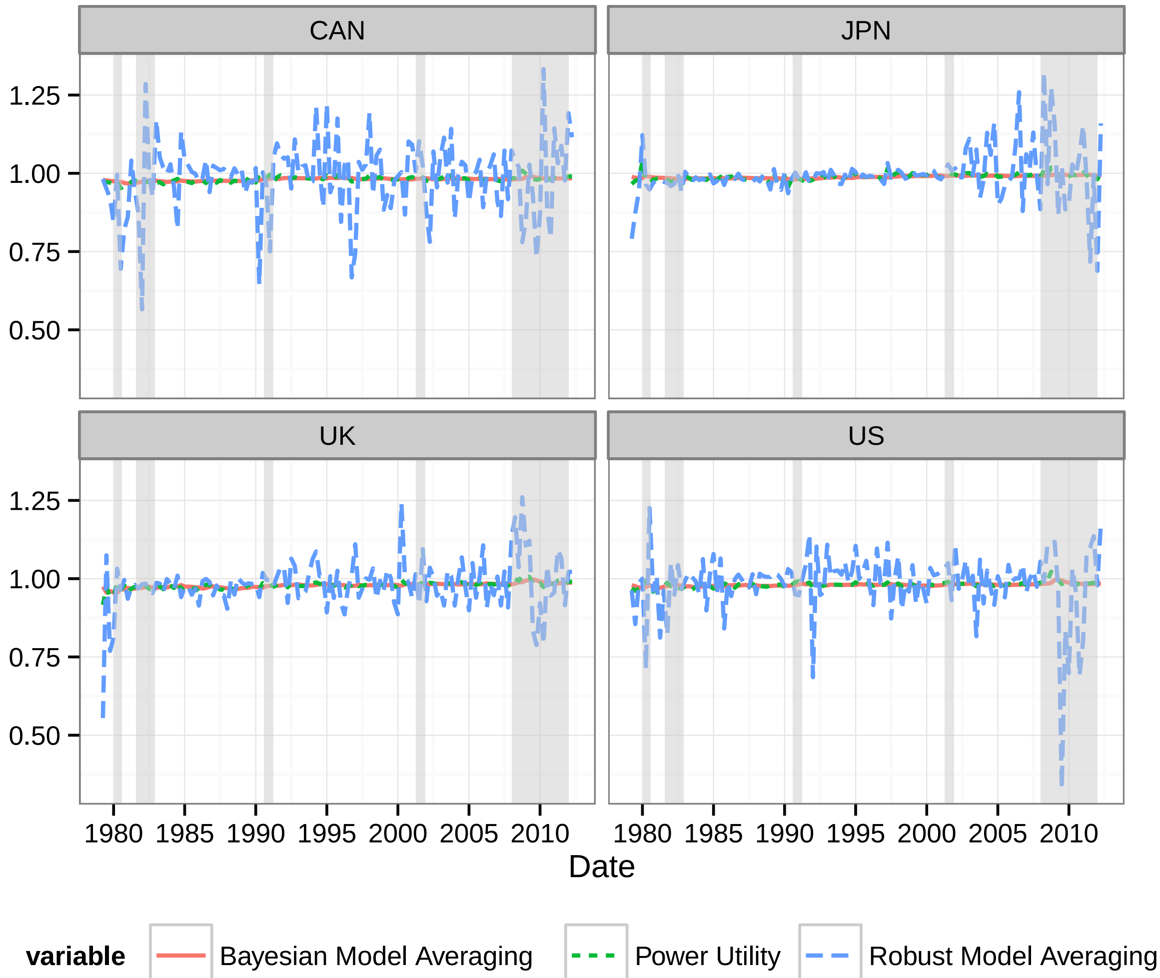

*Notes:* The sample period is 1979Q1-2012Q2.

Figure B.3: Posterior model probability by country




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2. Shaded area corresponds to NBER recession for in USA.

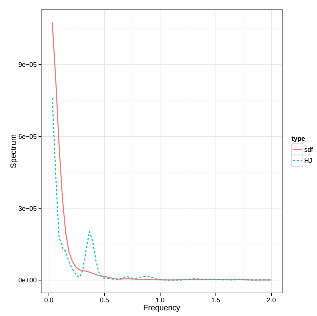
Figure B.4: Stochastic discount factor by countries



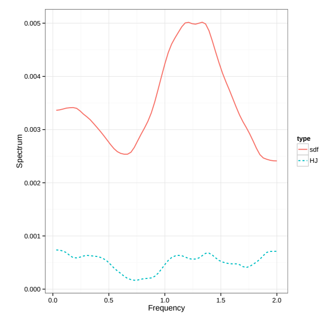

*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2. Shaded area corresponds to NBER recession for in USA.

Figure B.5: Hansen Jagannathan bound by frequency and by country

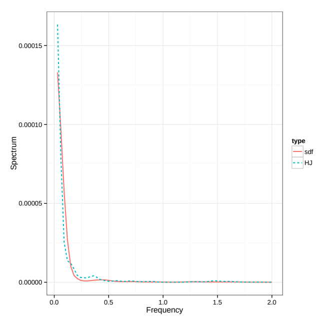
(a) CAN - BMA



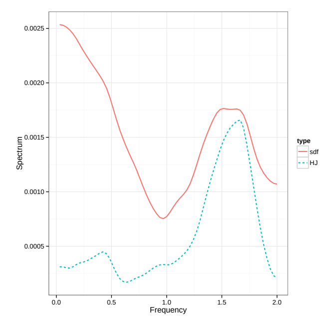

(b) CAN - RMA



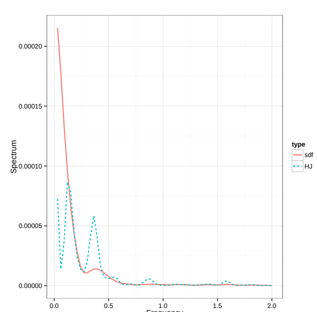

(c) JPN - BMA



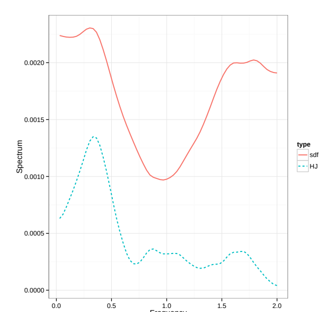

(d) JPN - RMA




(e) UK - BMA

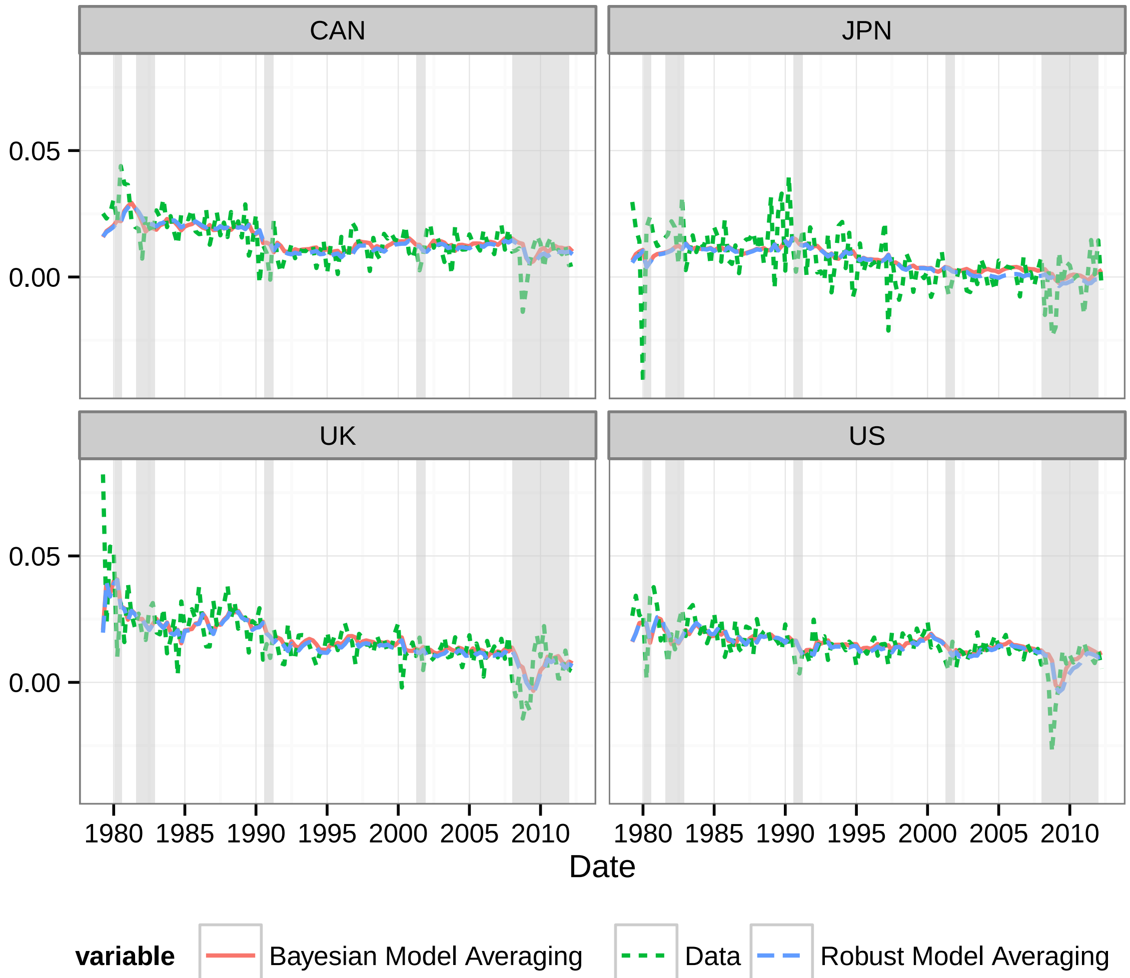



(f) UK - RMA




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. Robustness parameters are calibrated to give a detection error probability $p(\theta_1, \theta_2) = 20\%$. The sample period is 1979Q1-2012Q2 and the real interest rate differential is $R^e_{t+1} = \left((1+i^*_t)\frac{S_{t+1}}{S_t} - (1+i_t)\right)\frac{P_t}{P_{t+1}}$.

Figure B.6: Predicted consumption growth per country




*Notes:* Subjective discount factor is assumed to be $\beta = 0.995$. The sample period is 1979Q1-2012Q2. Shaded area corresponds to NBER recession for in USA.