Estimation of Heterogeneous Investor Beliefs from American Institutional Stock Holdings

Jason Yang

May 19, 2021

Abstract

We estimate a flexible time-varying distribution of investor beliefs regarding future returns of the Market Factor, Size Factor, and Value Factor using micro-data on institutional stock holdings in the United States. In the economic model, a distribution of mean-variance optimizing investors choose their portfolio risk exposures to two different factors based on their beliefs. We find that investor beliefs become substantially more pessimistic during an economic crisis and that disagreement among investors increases slightly. Impulse response analysis reveals that shocks to beliefs are fairly persistent, taking well over 3 quarters to fully dissipate. Finally, we incorporate these consensus belief estimates to test a Heterogeneous Belief Capital Asset Pricing model (H-CAPM). The H-CAPM fails to fit the data because returns are substantially more volatile than portfolios and because the average investor cannot predict returns.[1]

[1] I would like to thank the following people for helping me on this research journey. First and foremost, thanks to Prof. Ralph Koijen for meeting with me throughout the school year and providing the critical data to conduct research in this area of Asset Pricing and Finance. The meetings helped put my vague idea into a well-motivated and implementable research question closely related to the literature. Second, thanks to Prof. Kotaro Yoshida Prof. Yoshida for meeting with me several throughout second and third year to help plan my undergrad research plan and research-related courses and teaching an excellent macroeconomics class that really piqued my interest in financial economics and asset pricing; I found myself using all of the plans and coursework one way or another in this paper. Third, thank you to Prof. Victor Lima for really starting my "Chicago Economics Major" and making the course material really relevant to everyday life. Finally, thank you to Julie Wong for organizing the Economics thesis program, the coursework, and for meeting with me multiple times to go over available classes.

1 Introduction

Investor demand as explained by their beliefs are fundamental building blocks in financial economic models. These beliefs have very real economic consequences. For example, Sufi et al 2017 find that an increase in household debt to GDP predicts high short-run growth and a medium-term recession because of over-extrapolation in economic expectations [21]. Bordalo et al 2020 find that flawed expectations exacerbates business cycles [3].

However, there has been relatively little attention to research on uncovering and understanding the dynamics behind those beliefs. Earlier research focuses primarily on survey data. For example, Frankel and Froot 1990 analyze surveys of professional forecasters for depreciation of USD using a model with heterogeneous investment beliefs (extrapolative and fundamental investors) [12]. Nagel and Malmendier 2011 analyze surveys of financial risk taking in the United States and find that individuals experiencing low stock market returns during their early lives are less likely to participate in liquid financial markets and have pessimistic beliefs about future returns [19]. Shleifer and Greenwood analyze investor expectations of stock returns from several different surveys and find they are not consistent with classical models in financial economics; in particular, they show that investors become more pessimistic during economic crises while standard models predict that asset returns should be high [13]. Only recently has research begun structural estimation of investor beliefs. Egan et al nonparametrically estimate a distribution of investor beliefs about the market factor using data on S&P500 ETF holdings at different leverage points (-3x to +3x) [9]. Our paper seeks to build upon this new literature in investor beliefs and investor demand by providing and conducting an estimation procedure to identify investor beliefs about multiple asset pricing factors.

We model investors as mean variance optimizers with heterogeneous beliefs about future expected returns between two asset pricing factors and fixed beliefs about the covariance matrix. Using a revealed-preference approach similar to the one conducted in Egan et al, we identify a normal distribution of investor beliefs about future expected returns. Heterogeneity in beliefs then drives heterogeneity in investment decisions.

Our estimation results contributes to two areas in financial economics. First, it contributes to a growing literature of applying IO methods to investor demand by generalizing an earlier paper's approach to multiple factors. Second, we revisit a classic asset pricing model: the Capital Asset

Pricing Model with heterogeneous beliefs (H-CAPM). We test this model by using our estimation results as the consensus belief and find evidence rejecting this model.

The beta exposures are estimated by regressing quarterly investor portfolio returns on quarterly factor returns. The investor portfolio data contains the holdings of every asset manager that files a Form 13F with the United States Securities and Exchange Commission, required of funds managing over $100 million since 1980. The firms are divided into 6 types: banks, insurance companies, investment advisors, mutual funds, pension funds, and other institutions (endowments, nonprofit managers, etc.). The funds filing represent 68% of the stock market, while the residual 32% are held by households.

For each investor type and quarter between 1981 and 2017, we estimate a normal distribution of investor beliefs in expected returns. We find that the average investor believed that annual market excess returns were about 3%, size excess returns 0.7%, and value excess returns were -0.5%. These beliefs became more pessimistic during financial crises. Different investor types did not have substantially different consensus beliefs, explaining partially why the consensus factor returns were close to zero[2]. Furthermore, the average investor predicts actual returns very poorly. Regressing average beliefs on actual returns produces an R^2 value of less than 0.005.

We find a high degree of dispersion in investor beliefs, with an average standard deviation of about 0.7%, increasing during economic crises. The substantial standard deviation estimates shows that belief heterogeneity exists and is important.

Finally, using an approach inspired by Fama and Macbeth in 1973 we empirically investigate the predictions of a Heterogeneous Belief Capital Asset Pricing Model (H-CAPM)[11]. Multiple papers on this asset pricing model derive results similar to the standard CAPM but with consensus returns instead of actual market returns. Williams in 1977 show that security market line converges slowly towards the standard CAPM with some additional terms representing the risk that investors revise their beliefs[22]. Levy et al in 2006 show that convergence towards the true CAPM depends on the number of assets and number of investors and that large markets ought to be predicted well by the standard CAPM[18]. Chiarella in 2010 take preferences as given and derive a security market line with consensus belief returns instead of actual market returns[7]. We ultimately find that these

[2]It is not possible for the whole market to be net tilted towards Size or Value

models are not supported by the data, as H-CAPM predicted returns explain substantially less than 0.3% of actual returns primarily because investor beliefs explain les than 0.5% of actual returns.

The paper proceeds as follows: section 2 introduces and solves a Mean Variance optimization model with heterogeneous beliefs; section 3 explains the data used in estimation and simulation; section 4 discusses our identifying assumptions and identification procedure; section 5 presents the empirical findings; and section 6 concludes.

2 Investor Demand

Suppose there are N different investors choosing to expose themselves to one of two factors. Investor $i \in \{1, \ldots, N\}$ will choose their factor exposures $(\beta_{1,i}, \beta_{2,i})$ in a way that maximizes their portfolio return and minimizes their portfolio volatility. The risk exposures $\beta_{1,i}$ and $\beta_{2,i}$ can be interpreted as linear regression coefficients: if factor 1 goes up by 1%, then investor i's portfolio will go up by $\beta_{1,i}$. The investor has beliefs about the returns of the factors $(\mu_{1,i}, \mu_{2,i})$, variance of the factors $\left(\sigma^2_{1,i}, \sigma^2_{2,i}\right)$, and covariance between the factors (c_i).

The investor solves the following unconstrained optimization problem:

$$\max_{\{\beta_1, \beta_2\}} u_i\left(\beta_1, \beta_2\right) = \beta_1 \mu_{i,1} + \beta_2 \mu_{i,2} - \frac{1}{2}\beta_1^2 \sigma_1^2 - \frac{1}{2}\beta_2^2 \sigma_2^2 + \beta_1 \beta_2 c_i \tag{1}$$

This model follows the second order taylor expansion used in in Barseghyan et al 2013 and Egan et al 2020[2, 9].

In the appendix, solve the model by taking first order conditions, and demonstrate its mathematical relationship to the Markowitz mean-variance portfolio choice model[20].

The solution to this optimization problem and baseline for our empirical model is as follows:

$$\beta_{1,i} = \frac{\mu_{1,i}\sigma^2_{2,i} - c_i \mu_{2,i}}{\sigma^2_{1,i}\sigma^2_{2.i} - c_i^2}, \beta_{2,i} = \frac{\mu_{2,i}\sigma^2_{1,i} - c_i \mu_{1,i}}{\sigma^2_{1,i}\sigma^2_{2,i} - c_i^2} \tag{2}$$

Equation 2 can be interpreted most easily by first considering the case where $c_i = 0$. If $c_i = 0$, then the investor holds β_1 and β_2 in proportion to μ_1/σ_1^2 and μ_2/σ_2^2 respectively. In other words, the investor is most concerned with the Sharpe ratio (amount of return per unit variance, or risk). If $c_i > 0$, then the proportion adjusts to account for the fact that increasing the weight of β_1 affects the portfolio variance through both its own variance and the correlation to the other assets already held. In essence, it forms de-correlated versions of the variables before weighting them by variance. Indeed, rewritten in matrix form we can see that the mapping from expected returns to investment holdings simply divides by the covariance matrix.

$$\begin{bmatrix} \beta_{1,i} \\ \beta_{2,i} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & c_i \\ c_i & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_{1,i} \\ \mu_{2,i} \end{bmatrix}$$

When the market clears, the sum of betas demanded equals the sum of betas supplied. Mathematically, this means that the AUM-weighted average of betas demanded by the investment managers equals the Market-Cap weighted average of betas supplied by the stocks.

$$\frac{1}{\text{TOTAUM}} \sum_{i=1}^{N} \beta_{1,i} \cdot \text{AUM}_i = \frac{1}{\text{TOTMC}} \sum_{e=1}^{E} \beta_{1,e} \cdot \text{MC}_e$$

$$\frac{1}{\text{TOTAUM}} \sum_{i=1}^{N} \beta_{2,i} \cdot \text{AUM}_i = \frac{1}{\text{TOTMC}} \sum_{e=1}^{E} \beta_{2,e} \cdot \text{MC}_e$$

As an example, suppose factor 1 is the market factor and factor 2 is size (our baseline model specification). Let the RHS (supply) of market clearing equation 1 be $\mathcal{S}$ and the RHS of market clearing equation 2 be 0.[3]

Then plugging this into the demand equation we get

$$\sum_{i=1}^{N} \frac{\text{AUM}_i}{\text{TOTAUM}} \begin{bmatrix} \sigma_1^2 & c_i \\ c_i & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_{1,i} \\ \mu_{2,i} \end{bmatrix} = \begin{bmatrix} \mathcal{S} \\ 0 \end{bmatrix}$$

[3] Because by construction, SMB is long half the stocks and short half the stocks. In other words, not everyone can be tilted towards small stocks

$$\implies \sum_{i=1}^{N} \frac{\text{AUM}_i}{\text{TOTAUM}} \begin{bmatrix} \mu_{1,i} \\ \mu_{2,i} \end{bmatrix} = \begin{bmatrix} \mathcal{S} \cdot \sigma_1^2 \\ \mathcal{S} \cdot c_i \end{bmatrix}$$

This result is intuitive, as the belief view inducing the average investor to only expose themselves to the market factor is the one where the return on the first factor and second factor are proportional to the linear relationship between the first and second factor. In this case, the benefit from holding factor 2 is directly offset from the trade off of holding factor 1. Hence, even if investors hold net zero of the second asset, the average investor can still be "bullish" or "bearish" on the asset depending on the total asset supply (exogeneous) and the covariance matrix (endogeneous).[4]

3 Data

3.1 Stock Characteristics

We download the data on stock returns, prices, and shares outstanding from the Center for Research in Security Prices (CRSP) at a monthly frequency. We require that the stocks have non-missing prices and shares outstanding. We download data on the returns the Market, Value (HML), and Size (SMB) factors from the Kenneth-French data library.

3.2 Investor Holdings

We use the same data on institutional stock holdings as in Koijen et al 2019[17].[5] The complete description can be accessed in their paper, but we will describe briefly the data construction here.

Koijen et al retrieve the data on institutional common stock holdings from the Thomson Reuters Institutional Holdings Database, which are compiled from the SEC form 13F. These documents include stocks held by all institutional investment managers exercising investment discretion with assets under management exceeding $100 million. The holdings information do not include holdings of cash or bonds. There are six types of institutions: banks, insurance companies, investment advisers, mutual funds, pension funds, and other institutions. Data on institutional stock holdings cover about 68% of all assets held in the investment universe.

[4] In some time periods, the net demand of factor 2 is nonzero. The remaining demand may come from household demand, shorts, or measurement error.

[5] I am very thankful for Ralph Koijen, the adviser for this paper, for providing the data.

They then merge the holdings data onto the CRSP data and drop non-matching holdings by CUSIP. Dollar holding equals price times shares held. Assets under management equals sum of dollar holdings. Portfolio weight equals dollar holdings divided by assets under management.

3.3 Calculation of β_1, β_2

Using the investor holdings, we define portfolio return as the average of holding return weighted by portfolio weight. To compute the factor exposure, we estimate a rolling regression of portfolio returns on factor returns with an 8 quarter rolling window.[6] Let $f_{1,t}$ and $f_{2,t}$ be the returns of factor 1 and 2 respectively at time t. To estimate $\beta_{1,i,s}$ and $\beta_{2,i,s}$, we estimate

$$R_{i,t} = \beta_{1,i,s} f_{1,t} + \beta_{2,i,s} f_{2,t} + \beta_0$$

using data from $t \in [s-8, s]$. If we view f_1 and f_2 as systematic sources of risk, then $\beta_{1,i,t}$ and $\beta_{2,i,t}$ are historical measures of risk exposure to factors 1 and 2. β_0 captures portfolio returns unexplained by the risk factors.

We drop extremely high beta values above 3.75 (less than 0.5% of the observations) to avoid skewing the results with outliers.

At the end of this data pre-processing step, we then have a data-set of asset pricing factor exposures for every investor and every quarter. We use this data to structurally estimate the theoretical model of investor demand discussed in section 2.

3.4 Descriptive Statistics for β_1 and β_2

Tables 1 and 2 show mean and standard deviation of beta holdings for each investor type and each decade for the Market & Size model and the Market and Value model. The average market exposure is close to 1 almost all the time, while the average Size and Value exposures are close to 0 almost all the time. The standard deviations are much less consistent, which we analyze more closely using our empirical model.

[6] We find that changing the rolling window to 12 or 16 does not qualitatively change the results. We feel that 8 quarter rolling window balances the variation in estimation with the variation in new information or holdings.

[Table 1 about here.]

[Table 2 about here.]

4 Empirical Analysis

4.1 Empirical Model and Identifying Assumptions

For the empirical model, we make two notation modifications. First, we categorize investors into I different types with J_i investors in each type. So the factor exposure to asset f chosen by investor j of type i at time t is given by $\beta_{f,i,j,t}$. We add these additional subscripts to line up with the fact that the institutional holdings data is divided into multiple agent types.

Using these indices in the economic model, we then get the following solutions for investor demand:

$$\beta_{1,i,j,t} = \frac{\mu_{1,i,j,t}\sigma^2_{2,i,j,t} - c_{i,j,t}\mu_{2,i,j,t}}{\sigma^2_{1,i,j,t}\sigma^2_{2,i,j,t} - c^2_{i,j,t}}, \beta_{2,i,j,t} = \frac{\mu_{2,i,j,t}\sigma^2_{1,i,j,t} - c_{i,j,t}\mu_{1,i,j,t}}{\sigma^2_{1,i,j,t}\sigma^2_{2,i,j,t} - c^2_{i,j,t}}$$

As is, the model is not identified. For starters, $\beta_{1,i,j,t}$ and $\beta_{2,i,j,t}$ are not directly observed in the data. Furthermore, if we allow $\mu_1, \mu_2, c, \sigma_1^2, \sigma_2^2$ to be different for every investor for every time period, we will have 5 parameters to estimate for every 2 data points.

These identification problems motivate the following identification assumptions:

1. All investors agree on and calculate β_1, β_2 as described in section 3.3

2. All investors agree on a constant factor variance: $\sigma^2_{1,i,j,t} = \sigma^2_{1,k,\ell,s} = \sigma^2_1$ and $\sigma^2_{2,i,j,t} = \sigma^2_{2,k,\ell,s} = \sigma^2_2$

3. All investors agree on a constant factor covariance: $c_{i,j,t} = c \forall i, j, t$

4. Investors of the same type have beliefs about expected return drawn from the same (possibly time-varying) normal: $\mu_{1,i,j,t} \sim N(m_{1,t}, s_{1,t})$ and $\mu_{2,i,j,t} \sim N(m_{2,t}, s_{2,t})$, with $\mu_{1,i,j,t} \perp \mu_{2,i,j,t}$

Assumption 1 fixes the fact that $\beta_{1,i,j,t}$ and $\beta_{2,i,j,t}$ are not observed while assumption 2, 3, and 4 reduce the number of parameters that must be estimated. Let $\theta = (m_1, s_1, m_2, s_2)$ be the vector of parameters we are estimating. Under assumptions 1 through 4, we can rewrite the empirical model for each investor-type and time combination (dropping the time and individual subscript for ease of notation) accordingly:

$$\beta_1(\theta) \sim \frac{N\left(\sigma_2^2 m_1 - cm_2, \sigma_2^2 s_1 - cs_2\right)}{\sigma_1^2\sigma_2^2 - c^2} \tag{3}$$

$$\beta_2(\theta) \sim \frac{N\left(\sigma_1^2 m_2 - cm_1, \sigma_1^2 s_2 - cs_1\right)}{\sigma_1^2\sigma_2^2 - c^2} \tag{4}$$

We leave the derivation and proof of identification to the appendix. For now, we include the mathematical statement of identification as a proposition:

Proposition 1. *Suppose we have two sets of parameters* $\theta = (m_1, s_1, m_2, s_2)$ *and* $\tilde{\theta} = (\tilde{m_1}, \tilde{m_2}, \tilde{s_1}, \tilde{s_2})$ *for which the empirical model's distribution is the same:*

$$\beta_1(\theta) = \beta_1\left(\tilde{\theta}\right) \text{ and } \beta_2(\theta) = \beta_2(\theta)$$

Under assumptions 1 through 4, $\theta = \tilde{\theta}$

4.2 Discussion of Assumptions

The assumptions we adopt for identification largely follow those of Egan et al 2020, with two main changes.

Our first deviation is inferring values of beta (factor risk choices) from portfolio returns and stock holdings. This method of risk choice and risk evaluation is very standard, dating back to a classic paper by Jensen 1968 with the CAPM Market Factor and Daniel et al 1997 with characteristics based factors [16, 8]. In the asset management industry, this is a common metric to evaluate and understand the risks taken by different investment managers, as discussed in Brinson et al 1986 and Ibbotson & Kaplan 2000[4, 15].

Our second deviation in assumptions is fixing the covariance matrix. This is not a major deviation from the baseline specification in Egan et al, as their base specification fixes the market factor variance and estimate risk aversion once over the whole period. They find that this assumption is relatively robust, as the second specification (allowing for heterogeneous risk aversion) shows a relatively narrow distribution for beliefs in factor variance. We choose to interpret the second moment as beliefs in factor variance instead of risk aversion to be consistent with the mean-variance model. This is not a misappropriation of their model, as we demonstrate in the appendix that their model is equivalent to the tangency portfolio in Mean-Variance.

4.3 Estimation Procedure

Let $\theta_{i,t} = (m_{1,i,t}, m_{2,i,t}, s_{1,i,t}, s_{2,i,t})$ be the vector of parameters for investor type i at time t. Recall that we are estimating equations 3 and 4, which is simply a sum of 2 normal distributions.

Let $f(x; m, s)$ be the normal density with mean m standard deviation s evaluated at x. From these normal densities, we can define the log likelihood of a data point $\ell_d(\beta_{1,i,j,t}, \beta_{2,i,j,t}; \Theta_{i,t})$ as follows:

$$\ell_d(\beta_{1,i,j,t}, \beta_{2,i,j,t}; \theta_{i,t}) = \frac{\log\left[f\left(\beta_{1,i,j,t}; \sigma_2^2 m_{1,i,t} - c_{i,t}, \sigma_2^2 s_1 - c_{i,t} s_2\right)\right]}{\sigma_1^2\sigma_2^2 - c_{i,t}^2} +$$

$$\frac{\log\left[f\left(\beta_{2,i,j,t}; \sigma_2^2 m_{2,i,t} - c_{i,t} m_1, \sigma_1^2 s_2 - c_{i,t} s_1\right)\right]}{\sigma_1^2\sigma_2^2 - c_{i,t}^2}$$

We sum this over all data points for investor i at time t to define the log likelihood of the whole data set $\mathbb{L}(\Theta_{i,t})$

$$\mathbb{L}(\theta_{i,t}) = \sum_{j=1}^{J} \ell_d(\beta_{1,i,j,t}, \beta_{2,i,j,t})$$

Our optimization routine maximizes $\mathbb{L}$ subject to $m_1, m_2 \in [-0.3, 0.3]$ and $s_1, s_2 \in [0, 0.5]$. These constraints technically are necessary for the search space to be compact and the optimization to converge. However, they are not at all restrictive. Intuitively, these constraints require that the average investor does not believe that the factors will return more than 30% or less than -30% quarterly and that the standard deviation of such beliefs is investors is less than 50%.

4.4 Calibration Procedure

To calibrate the covariance matrix, we simply compute the sample covariance matrix for Market & Size and Market & Value from the years 1980 to 2017. This yields the results in table 3 and 4.

[Table 3 about here.]

[Table 4 about here.]

5 Results

5.1 Parameter Estimates

Interpretation

In this section, we will analyze the behavior of the series during economic crises, discuss the average levels of the series, and finally compare investor beliefs against actual returns. We present our estimates for investor beliefs weighted by assets under management in figure 1 and figure 2. The estimates for investor beliefs weighted equally are in figure 3 and 4. AUM weighted estimation captures the true market consensus, while equal weight estimation captures the diversity in asset manager beliefs. Differences come from the disproportionate representation of large players in the AUM weighted estimation.

In Figure 1 Panel A, we plot the time series of what the average investor believes about market expected returns in red and size expected returns in green. Panel B shows the disagreement (standard deviation) in investor beliefs about the market factor in red and the size factor in green. Figure 2 is similar but replacing Size with Value. Since we do pairwise estimation, we actually have two estimates for the market factor. However, the estimates for the market factor are qualitatively similar in both pairs.

Our estimates for the market factor suggests that investors of all types revised their beliefs downwards to varying degrees after financial crises.[7] The series for Mkt M after the black dotted lines (highlighting the years 2000 and 2008) dropped in both specifications for most agent types. The average investor decreased their beliefs about market expected returns 1.3% during the Dot Com bubble and 0.9% during the 2008 Financial Crisis. The average recovery period, or time it took for return beliefs to return to pre-crisis levels, was about 1 year for both crises. When looking at specific agent types, we find that banks (-1.8%) and investment advisers (-1.5%) changed their beliefs the most in response to the Dot Com bubble while all agent types responded roughly similarly during the 2008 financial crisis.

Dispersion in investor beliefs also increases during an economic crisis, as Mkt SD increases in both figure 3 and figure 4.[8] We find that disagreement increased by 1% in the Dot Com bubble and 1.5% in the financial crisis in the equal weighted specification. Disagreement increased the most (least) for investment advisers and mutual funds (insurance companies and pension funds) in both economic crises, which is intuitive because these investor types exhibit the greatest (least) degree of investment discretion.

Looking at the size factor, we see a similar story. The average investor decreased their beliefs about returns to size about 0.74% during the Dot Com bubble and 0.6% during the 2008 Financial Crisis. For this factor, there no investor types that were consistently revising their beliefs the most. Interestingly, the belief revisions for the value factor were either very small or slightly positive. The average investor increased their beliefs about returns to value about 0.5% during the dot com bubble and 1% during the 2008 financial crisis. This reflects the fact that investors view value as a 'safe' factor (since the calibrated covariance between market and value is slightly negative).

The estimated levels of the consensus beliefs are a bit lower than the actual value. We estimate 3% consensus market risk premium, which is slightly lower than the estimated value in Egan et al (4.5%) but about the same as their true estimated market risk premium (3.5%). We believe this difference primarily comes from our definitions of 'average investors'. In Egan et al, they exclude

[7] In our paper, we use the St. Louis Federal Reserve definition of a recession and add 1 quarter before and after the recession starts[14]. We define a belief revision during a crisis to be the max belief value during the crisis minus the min value during the crisis if the max value happened before the min value. We use max and min because different investors may have different timelines for updating their beliefs during a crisis.

[8] We choose to use the equal weighted estimation to quantify disagreement because it better captures the heterogeneity in the beliefs. The AUM weighted estimation does not capture disagreement because it would not capture variation in small managers that tend to be more active and exhibit contrarian opinions[17, 9].

a large amount of 1-beta holdings by large institutional managers. Furthermore, their sample does not include sophisticated investors that expose themselves to market risk without ETFs. Since our data is more complete in this regard, we believe our estimates better capture the notion of market consensus belief.[9]

We also estimate a 0.7% consensus size risk premium and -0.5% consensus value risk premium (with each average of each type holding close to the economy-wide average). These values are quite low and close to zero, making these non-market factors quite undesirable assets. However, this comes from the fact that the average beta to size and value is very close to zero as dicsussed in section 3. The only way to justify holding very little of an asset in the mean-variance paradigm requires that it is either very correlated to the primary asset or has very low mean return (or both). Size has a positive calibrated covariance, and so results in a low but positive estimated consensus belief return. On the other hand, value has a negative calibrated covariance and so results in a negative estimated consensus belief return.[10]

Although the average level of the consensus belief approximately matches the average market risk premium, the average investor predicts actual returns quite poorly. In figure 8a and 8b we plot the values of the market consensus belief against the actual market returns in the Market & Size model and the Market & Value model respectively. The red and green dots show the data points from economic recessions, while the gray dots show all other data points. The dashed black line shows the 45 degree line where average beliefs equals actual returns. The figures clearly show that the variation in the actual market risk premium are substantially higher than the variation in investor beliefs. Subsequently, regressing actual returns against return beliefs returns an R^2 value of less than 0.005. Furthermore, the actual return values that are the furthest below the 45 degree line are primarily colored (observations from financial crises).

[Figure 1 about here.]

[Figure 2 about here.]

[9] Of course, their approach has many other merits. For example, they achieve nonparametric identification because they are able to isolate several sources of independent variation in the ETF markets (such as fees). This would be very difficult to do in holdings data.

[10] At a first look, it may appear that the market doesn't clear because all investor types are bullish on size and bearish on value. However, a zero holding does not imply a zero return belief. We showed in the model section that in mean-variance, zero holding of one asset and positive holding of another asset corresponds to nonzero expected return beliefs if the variance and covariance beliefs are nonzero.

[Figure 3 about here.]

[Figure 4 about here.]

Biases in Interpretation

In assumption A2 and A3 we fix the covariance matrix. This implicitly fixes the investors' risk aversion coefficient, which could bias our results. This is most clear in the Market & Value specification, as the reallocation from the Market factor to the Value factor during financial crises may be explained by an increase in risk aversion instead of actually becoming more bullish on Value. Although we have no direct way to account for this in our analysis, the earlier estimation conducted in Egan et al 2020 finds that the qualitative results about the Market factor do not change when they allow for time-varying risk aversion. Since we find similar results as their paper in our Market factor estimates, the robustness of our results to time-varying risk aversion may apply to our paper as well.

5.2 Time Series Analysis

In this section, we conduct a time series analysis to assist describe the path of parameter estimates. First, we rescale the belief data so the units are in yearly percents. Next, we specify that the time series of $\theta = (m_1, s_1, m_2, s_2)$ (the vector of 4 parameters) follows a VAR(P) process with a constant mean θ_m.

$$\theta_t = \theta_m + \sum_{j=1}^{P} A_{t-j}\theta_t + \epsilon_{t-j}$$

We choose $P = 3$ using the Akaike Information Criterion (AIC) and estimate the series of A_{t-j} using multivariate least squares (MLS). Then, we convert the estimated VAR(3) model into an MA(∞) model.

$$\theta_t = \theta_m + \sum_{\ell=0}^{\infty} \Phi_{t-\ell} e_{t-\ell}$$

The ℓ period in the future response of variable j to a unit impulse in variable i is then in the (i, j) element of $\Phi_{t-\ell}$. Since we rescaled the data so that it is in yearly percents, the IRF values are also in yearly percents. To get an economy-wide impulse response function, we weight the impulse response functions for each investor type by the total assets owned by each investor type. We plot the impulse response functions in figures 5 and **??**. In these plots, the solid lines show the economy-wide average impulse response functions (along with 95% confidence bands) while the colored bars show the relative contribution of each agent type to the economy-wide impulse response function. We interpret these shocks as new information about the market, size, or value factors that update investor beliefs.

Along the diagonals of both specifications, we find that a positive shock to the average investor's beliefs dissipates 3 quarters after the shock. Average disagreement however, is more persistent. It takes a full year for such disagreement to finally dissipate. These findings suggest that investors take quite a bit of time to incorporate new information, which relates to confirms earlier models of of momentum investing and information arrival[6, 1, 5].

Now looking at the off diagonals, we see that a positive shock to beliefs about market returns makes investors more optimistic about size. This is expected because the investor knows that the covariance between size and the market is positive, and so a positive shock to the market factor provides a positive signal to the size factor. For a similar reason, a shock to beliefs about market returns makes investors pessimistic about Value.

Interestingly, a positive shock to beliefs about size returns does not make investors more optimistic about the market factor (nor does a positive shock to beliefs about value returns make investors more pessimistic about the market factor).

[Figure 5 about here.]

[Figure 6 about here.]

5.3 Heterogeneous CAPM

5.3.1 Discussion of Theory

We first turn to theory. Many different formulations of the CAPM with heterogeneous agent exist. Some formulations do not make significant deviations in the predictions of the classical CAPM. For example, Levy et al 2006 specifies that all investors have heterogeneous amd unbiased beliefs about expected returns and the covariance matrix[18]. When the number of assets and number of investors are large, the empirical predictions of CAPM hold because trading quickly results in a convergence to the average of those unbiased beliefs. Other papers specify a process for heterogeneous belief formation. Williams 1977 takes a similar approach to Levy but assumes that individuals estimate expected returns and the covariance matrix[22]. The mean vector is estimated with irreducible error, while the covariance matrix may be estimated to arbitrary accuracy. This specification results in slow convergence to the true CAPM; during the convergence phase individuals hold heterogeneous portfolios to hedge both asset risk (based on the consensus covariance matrix) the risk that they significantly revise their beliefs. Still, others take investor beliefs as given and derive a very similar identical CAPM but with the market portfolio replaced with the consensus market portfolio, such as Chiarella et al 2010 [7]. A key similarity of these three perspectives is that the market consensus returns, rather than actual market returns, drive pricing.

For the remainder of this section, we consider the following heterogeneous agent CAPM equation:

$$R_{i,t+1} = \beta_i m_t + r_{f,t} \tag{5}$$

Where $R_{i,t}$ is the actual return of stock i, β_i is the investor-perceived beta to the market portfolio, and m_t is the investor-perceived risk premium of the market portfolio. To test this prediction, it is not sufficient to use actual returns absent strong beliefs about investor belief formation.[11] This is especially true because our estimated series for the perceived market risk premium predicts the actual market risk premium very poorly.

In order to test this theoretical prediction, we take advantage of the fact that we already estimated

[11] For example, Williams suggests that one simple way to test their model is to assume that investors update their beliefs about the mean vector and covariance matrix by estimating off of historical data. Then, the model is identified and slowly converges to the true CAPM.

a time series for m_t and assume that investors perceive β_i as described in section 3. So, we simply estimate the following empirical model and test whether $\gamma = 1$:

$$R_{i,t+1} - r_{f,t} = \gamma [m_t \beta_i] + e_{t+1} \tag{6}$$

Practically, this requires sorting the individual stocks into portfolios due to the instability of individual stock betas[10]. We follow the portfolio sorting procedure in Fama and Macbeth 1973 to attain portfolios with stable estimates of β_i and $R_{i,t+1}$[11]. Essentially, the key difference between our test and the standard Fama & Macbeth regression is the fact that we estimate m_t separately from the holdings data while Fama & Macbeth estimate a second regression to find m_t from the returns data. This reflects the core difference in assumptions: both tests assume that investors percieve the 'correct' betas, but our test does not assume investors percieve the 'correct' risk premia.

5.3.2 Methodological Details

To estimate the empirical model, we must first form and estimate the β_i of the sorted portfolios. Fix a start quarter denoted D. We conduct the regression regression for the test date D in three main steps. For notation, let $R_{m,t}$ denote the actual market risk premium at time t and let $R_{j,t}$ denote the actual return of stock j at time t.

First, we use quarters D to $D + 16$ (the portfolio sorting period) to sort every stock in the universe into 40 different portfolios. The first substep regresses actual returns of each stock against the actual returns of the market factor. Specifically, we estimate the following regression for stock j using data from $t \in [D, D + 4)$

$$R_{j,t} = \beta_j R_{m,t} + \alpha$$

The second substep takes these estimated β_j and sorts them into 40 equal-sized levels in order. Now we have map from stock j to portfolio p. Let S_p be the set of stocks in portfolio p and let $|S_p|$ be the cardinality of S_p.

Second, we use years $D + 16$ to $D + 48$ to estimate $\beta_{p,t}$ and $R_{p,t}$ of each sorted portfolio using a rolling regression with an 8 quarter window. We estimate the beta values of every stock using the following rolling regression:

$$R_{j,t} = \beta_{j,s} R_{m,t} + \alpha$$

Where $\beta_{j,s}$ uses data from time periods $t \in [s-8, s]$.

We then define the beta of the portfolio as the average beta of all its components:

$$\beta_{p,s} = \frac{1}{|S_p|} \sum_{j \in S_p} \beta_{j,s}$$

Third, we define the return of the portfolio as the average return of all its components:

$$R_{p,s} = \frac{1}{|S_p|} \sum_{j \in S_p} R_{j,t}$$

After these three steps, we now have a data set containing $40 \times 24 = 960$ observations of $\beta_{p,s}$ and $40 \times 24 = 960$ observations of $R_{p,s}$. Using the estimation results from section 4, we also have a data set of m_t (the percieved market risk premium). So, we have all the data to estimate the empirical model in equation 6.

Since this empirical test uses 12 years of data and we have 26 years of data in the sample, we can repeat this procedure for every year D in the first 14 years of our data.

5.3.3 Discussion

We show the results of our estimation in figure 7a and 7b. The plot shows the estimated value of γ on the y-axis and the test period ($D+4$ in the methodological details section).[12] The blue and red lines show the upper and lower confidence intervals at the 95% level, while the green line shows the point estimate. For reference, we also have a dotted line at $\gamma = 1$ because the hypothesis of interest is $\gamma = 1$ (i.e. that the H-CAPM predicts stock returns).

[12] We choose to use the start of the testing period because it shows the start of the data used in the regression. The sorting period only forms the portfolios, the data is not directly used in estimating equation 6

Throughout the series, the confidence intervals for the gamma coefficient are very wide. Even in the few years when the point estimate is close to 1, it is not significantly different from 0. However, most of the time we see that the estimated coefficient is very far from 1, particularly in the earliest and most recent testing periods. Furthermore, the R^2 value of these regressions are all very low and for the most part less than 0.003. If we accept the modelling assumptions, we can interpret this as strong evidence that assets are not priced according to the heterogeneous belief CAPM.

There are two features of the data that explain why the model does not fit. First and foremost is the fact that investors do a bad job of predicting actual returns. Then, $\beta_e m_t$ cannot predict returns well because it is simply a scaled version of m_t. This highlights an issue with classes of models that draw a simple (linear) mapping from average beliefs to returns. Since average-joe investors fail to predict market returns, it would be unreasonable for their predictions to explain equity returns. We can see the phenomenon most obviously in figures 8a and 8b, where the economic crises (when the average investors significantly underpredicted returns) correspond to very negative leverage points. This mechanically increases the γ estimate in 6 for the 2000 and onwards test periods.

Second, we can see that the variance in $\beta_e m_t$ is much lower than actual returns. Again we can see this in figures 8a and 8b where the variation in the actual market factor an order of magnitude larger than the variation in beliefs about the market factor. This happens because beliefs are estimated directly off of portfolios, which tables 3 and 4 show do not change frequently.

[Figure 7 about here.]

[Figure 8 about here.]

6 Conclusion

Our paper utilizes revealed-preference approach to estimate investor beliefs about market, size, and value expected returns. Variation in risk exposure choices allows us to identify the mean (consensus belief in expected returns) and standard deviation (disagreement among investors) of the belief distribution.

The estimation procedure allows us to identify a normal distribution of investor beliefs every quarter from 1981 to 2017. The time series of belief distribution shows that pessimism and disagreement increase during recessions, and the impulse response function of distribution parameters shows that shocks to beliefs are persistent. This agrees with prior research on investor expectations from survey data and structural estimations that also find evidence of extrapolative beliefs and increased pessimism during financial crises.

We then use our estimated time series of consensus beliefs to empirically test a Heterogeneous Belief Capital Asset Pricing Model (H-CAPM) under the assumption that investors perceive the correct betas but form their own beliefs about expected returns. Our test rejects the model because investor beliefs predict actual returns very poorly.

References

[1] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.

[2] Levon Barseghyan, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, 103(6):2499–2529, October 2013.

[3] Pedro Bordalo, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–82, September 2020.

[4] Gary P. Brinson, L. Randolph Hood, and Gilbert L. Beebower. Determinants of portfolio performance. *Financial Analysts Journal*, 42(4):39–44, 1986.

[5] Christopher D. Carroll. Macroeconomic Expectations of Households and Professional Forecasters*. *The Quarterly Journal of Economics*, 118(1):269–298, 02 2003.

[6] LOUIS K. C. CHAN, NARASIMHAN JEGADEESH, and JOSEF LAKONISHOK. Momentum strategies. *The Journal of Finance*, 51(5):1681–1713, 1996.

[7] Carl Chiarella, Roberto Dieci, and Xue-Zhong He. *A Framework for CAPM with Heterogeneous Beliefs*, pages 353–369. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[8] KENT DANIEL, MARK GRINBLATT, SHERIDAN TITMAN, and RUSS WERMERS. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3):1035–1058, 1997.

[9] Mark L Egan, Alexander MacKay, and Hanbin Yang. Recovering investor expectations from demand for index funds. Working Paper 26608, National Bureau of Economic Research, January 2020.

[10] Eugene F. Fama and Kenneth R. French. The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, 18(3):25–46, 2004.

[11] Eugene F. Fama and James D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

[12] Jeffrey A. Frankel and Kenneth A. Froot. Chartists, fundamentalists, and trading in the foreign exchange market. *The American Economic Review*, 80(2):181–185, 1990.

[13] Robin Greenwood and Andrei Shleifer. Expectations of Returns and Expected Returns. *The Review of Financial Studies*, 27(3):714–746, 01 2014.

[14] James Hamilton. Dates of u.s. recessions as inferred by gdp-based recession indicator [jhdusrgdpbr]. Data series, Federal Reserve Bank of St. Louis.

[15] Roger G. Ibbotson and Paul D. Kaplan. Does asset allocation policy explain 40, 90, or 100 percent of performance? *Financial Analysts Journal*, 56(1):26–33, 2000.

[16] Michael C. Jensen. The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2):389–416, 1968.

[17] Ralph S. J. Koijen and Motohiro Yogo. A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515, 2019.

[18] Haim Levy, Moshe Levy, and Golan Benita. Capital asset prices with heterogeneous beliefs. *The Journal of Business*, 79(3):1317–1353, 2006.

[19] Ulrike Malmendier and Stefan Nagel. Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*. *The Quarterly Journal of Economics*, 126(1):373–416, 02 2011.

[20] Harry Markowitz. Portfolio selection*. *The Journal of Finance*, 7(1):77–91, 1952.

[21] Atif Mian, Amir Sufi, and Emil Verner. Household Debt and Business Cycles Worldwide*. *The Quarterly Journal of Economics*, 132(4):1755–1817, 05 2017.

[22] Joseph T. Williams. Capital asset prices with heterogeneous beliefs. *Journal of Financial Economics*, 5(2):219–239, 1977.

7 Appendix

7.1 Proof of Identification

The resulting distribution for investor beliefs is quite flexible. The beliefs are constructed by a sequence of I 2-parameter distributions.

Let $\theta = (m_1, s_1, m_2, s_2)$ be a vector of belief parameters. Then we can rewrite the empirical model for each investor-type and time combination (dropping the time and individual subscript for ease of notation) accordingly:

$$\beta_1 \sim \frac{\sigma_2^2 N(m_1, s_1) - cN(m_1, s_1)}{\sigma_1^2\sigma_2^2 - c^2}$$

$$\beta_2 \sim \frac{\sigma_1^2 N(m_2, s_2) - cN(m_1, s_1)}{\sigma_1^2\sigma_2^2 - c^2}$$

Using the fact that the normal distribution is closed under convolution, we can rewrite the model as:

$$\beta_1(\theta) \sim \frac{N\left(\sigma_2^2 m_1 - cm_2, \sigma_2^2 s_1 - cs_2\right)}{\sigma_1^2\sigma_2^2 - c^2}$$

$$\beta_2(\theta) \sim \frac{N\left(\sigma_1^2 m_2 - cm_1 + c, \sigma_1^2 s_2 - cs_1\right)}{\sigma_1^2\sigma_2^2 - c^2}$$

Proposition 2. *Suppose we have two sets of parameters $\theta = (m_1, s_1, m_2, s_2, c)$ and $\tilde{\theta} = (\tilde{m_1}, \tilde{m_2}, \tilde{s_1}, \tilde{s_2}, c)$ for which the empirical model's distribution is the same for at least two different values of r_f, i.e. that*

$$\beta_1(\theta) = \beta_1\left(\tilde{\theta}\right) \text{ and } \beta_2(\theta) = \beta_2(\theta)$$

Under assumptions 1 through 4, $\theta = \tilde{\theta}$

Proof. Since the distributions on RHS and LHS are equivalent. Then, we have the following system of 4 equations (setting the mean and standard deviations equal)

$$\sigma_2^2 m_1 - c m_2 - \sigma_1^2 \tilde{m_1} + c \tilde{m_2} = 0$$

$$\sigma_1^2 m_2 - c m_1 - \sigma_1^2 \tilde{m_2} + c \tilde{m_1} = 0$$

$$\sigma_2^2 s_1 - c s_2 - \sigma_1^2 \tilde{s_2} - c \tilde{s_1} = 0$$

$$\sigma_1^2 s_2 - c s_1 - \sigma_1^2 \tilde{s_2} + c \tilde{s_1} = 0$$

Let $d_{m1} = m_1 - \tilde{m_1}$, $d_{m2} = m_2 - \tilde{m_2}$, $d_c = c - \tilde{c}, d_{s1} = s_1 - \tilde{s_1}$ and $d_{s2} = s_2 - \tilde{s_2}$. It is sufficient to show that all of these new variables are zero.

We can rewrite the above system in terms of these variables as follows:

$$\sigma_2^2 d_{m1} - c d_{m2} = 0$$

$$\sigma_1^2 d_{m2} - c d_{m1} = 0$$

$$\sigma_2^2 d_{s1} - c d_{s2} = 0$$

$$\sigma_1^2 d_{s2} - c d_{s2} = 0$$

Rewrite as two matrix systems:

$$\begin{bmatrix} \sigma_2^2 & -c \\ -c & \sigma_1^2 \end{bmatrix} \begin{bmatrix} d_{m1} \\ d_{m2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \sigma_2^2 & -c \\ -c & \sigma_1^2 \end{bmatrix} \begin{bmatrix} d_{s1} \\ d_{s2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Invoking our identification assumption, $\sigma_1^2\sigma_2^2 - c^2 \neq 0$ implies that the matrix is nonsingular and only has $\vec{0}$ in the kernel. So, $\begin{bmatrix} d_{m1} \\ d_{m2} \end{bmatrix} = \begin{bmatrix} d_{s1} \\ d_{s2} \end{bmatrix} = \vec{0}$.

We have then proven that $0 = d_{m1} = d_{m2} = d_{s1} = d_{s2}$ as required.

$\square$

7.2 Estimation of Covariance Beliefs: Proof of Concept

We can relax the identification procedure to identify a covariance constant per investor type if we have another source of identification. For example, suppose there were some measure of aggregate and time-varying transaction costs τ_t.

Explicitly, we adopt the following set of assumptions:

1. All investors agree on and calculate β_1, β_2 as described in section 2.3
2. All investors agree on a constant factor variance: $\sigma^2_{1,i,j,t} = \sigma^2_{1,k,\ell,s} = \sigma_1^2$ and $\sigma^2_{2,i,j,t} = \sigma^2_{2,k,\ell,s} = \sigma_2^2$
3. Investors of the same type agree on the factor covariance each time period: $c_{i,j,t} = c_{i,k,t}\ \forall j, k$

4. Investors of the same type have beliefs about expected return drawn from the same (possibly time-varying) normal: $\mu_{1,i,j,t} \sim N(m_{1,t}, s_{1,t})$ and $\mu_{2,i,j,t} \sim N(m_{2,t}, s_{2,t})$

This differs from the baseline specification only in assumption number 3. We proceed in the same way as the earlier section, but condition the observed variables (β_1, β_2) on the risk free rate.

Let $\theta = (m_1, s_1, m_2, s_2, c)$ be a vector of belief parameters. Write the empirical model for each investor-type and time combination conditioned on the risk free rate accordingly:

$$\beta_1 \mid \tau \sim \frac{\sigma_2^2 N(m_1, s_1) - \sigma_2^2 \tau + cN(m_1, s_1) - c\tau}{\sigma_1^2 \sigma_2^2 - c^2}$$

$$\beta_2 \mid \tau \sim \frac{\sigma_1^2 N(m_2, s_2) - \sigma_1^2 \tau + cN(m_1, s_1) - c\tau}{\sigma_1^2 \sigma_2^2 - c^2}$$

Using the fact that the normal distribution is closed under convolution, we can rewrite the model as:

$$\beta_1(\theta) \mid \tau \sim \frac{N\left(\sigma_2^2 m_1 - \sigma_2^2 \tau + cm_2 + c\tau, \sigma_2^2 s_1 + c\tau\right)}{\sigma_1^2 \sigma_2^2 - c^2}$$

$$\beta_2(\theta) \mid \tau \sim \frac{N\left(\sigma_1^2 m_2 - \sigma_1^2 \tau + cm_1 + cr_f, \sigma_1^2 s_2 + c\tau\right)}{\sigma_1^2 \sigma_2^2 - c^2}$$

We now prove identification using the following proposition:

Proposition 3. *Suppose we have two sets of parameters $\theta = (m_1, s_1, m_2, s_2, c)$ and $\tilde{\theta} = (\tilde{m_1}, \tilde{m_2}, \tilde{s_1}, \tilde{s_2}, c)$ for which the empirical model's distribution is the same for at least two different values of τ, i.e. that*

$$\beta_1(\theta) \mid \tau_a = \beta_1\left(\tilde{\theta}\right) \mid \tau_a \text{ and } \beta_2(\theta) \mid \tau_b = \beta_2(\theta) \mid \tau_a$$

$$\beta_1(\theta) \mid \tau_b = \beta_1\left(\tilde{\theta}\right) \mid \tau_b \textit{ and } \beta_2(\theta) \mid \tau_b = \beta_2(\theta) \mid \tau_b$$

If we make assumptions 1 through 5, then $\theta = \tilde{\theta}$

Proof. Since the distributions on RHS and LHS are equivalent. Then, we have the following system of 6 equations (setting the mean and standard deviations equal)

$$\sigma_2^2(m_1 - \tau_a) - c(m_2 - \tau_a) - \sigma_1^2(\tilde{m}_1 - \tau_a) + \tilde{c}(\tilde{m}_2 - \tau_a) = 0$$

$$\sigma_1^2(m_2 - \tau_a) - c(m_1 - \tau_a) - \sigma_1^2(\tilde{m}_2 - \tau_a) + \tilde{c}(\tilde{m}_1 - \tau_a) = 0$$

$$\sigma_2^2(m_1 - \tau_b) - c(m_2 - \tau_b) - \sigma_1^2(\tilde{m}_1 - \tau_b) + \tilde{c}(\tilde{m}_2 - \tau_b) = 0$$

$$\sigma_1^2(m_2 - \tau_b) - c(m_1 - \tau_b) - \sigma_1^2(\tilde{m}_2 - \tau_b) + \tilde{c}(\tilde{m}_1 - \tau_b) = 0$$

$$\sigma_2^2 s_1 - c s_2 - \sigma_1^2 \tilde{s}_2 - \tilde{c}\tilde{s}_1 = 0$$

$$\sigma_1^2 s_2 - c s_1 - \sigma_1^2 \tilde{s}_2 + \tilde{c}\tilde{s}_1 = 0$$

Let $d_{m1} = m_1 - \tilde{m}_1$, $d_{m2} = m_2 - \tilde{m}_2$, $d_\tau = \tau_a - \tau_b$, $d_c = c - \tilde{c}, d_{s1} = s_1 - \tilde{s_1}$ and $d_{s2} = s_2 - \tilde{s}_2$. It is sufficient to show that all of these new variables are zero.

We can rewrite the above system in terms of these variables as follows:

$$\sigma_2^2 d_{m1} - c d_{m2} = d_c\left(\tilde{m}_2 - \tau_a\right) = d_c\left(\tilde{m}_2 - \tau_b\right)$$

$$\sigma_1^2 d_{m2} - c d_{m1} = d_c\left(\tilde{m}_1 - \tau_a\right) = d_c\left(\tilde{m}_2 - \tau_b\right)$$

$$\sigma_2^2 d_{s1} - c d_{s2} = d_c \tilde{s_2}$$

$$\sigma_1^2 d_{s2} - c d_{s2} = d_c \tilde{s_1}$$

Rewrite as two matrix systems:

$$\begin{bmatrix} \sigma_2^2 & -c \\ -c & \sigma_1^2 \end{bmatrix} \begin{bmatrix} d_{m1} \\ d_{m2} \end{bmatrix} = \begin{bmatrix} d_c\left(\tilde{m}_2 - \tau_a\right) \\ d_c\left(\tilde{m}_1 - \tau_a\right) \end{bmatrix} = \begin{bmatrix} d_c\left(\tilde{m}_2 - \tau_b\right) \\ d_c\left(\tilde{m}_1 - \tau_b\right) \end{bmatrix}$$

$$\begin{bmatrix} \sigma_2^2 & -c \\ -c & \sigma_1^2 \end{bmatrix} \begin{bmatrix} d_{s1} \\ d_{s2} \end{bmatrix} = \begin{bmatrix} d_c \tilde{s_2} \\ d_c \tilde{s_1} \end{bmatrix}$$

Since d_c is a constant, and $r_{f1} \neq r_{f2}$, it must be that $d_c\left(\tilde{m}_2 - \tau_a\right) = d_c\left(\tilde{m}_2 - \tau_b\right) \implies d_c = 0$. Therefore, $\begin{bmatrix} d_c\left(\tilde{m}_2 - \tau_a\right) \\ d_c\left(\tilde{m}_1 - \tau_a\right) \end{bmatrix} = \begin{bmatrix} d_c\left(\tilde{m}_2 - \tau_b\right) \\ d_c\left(\tilde{m}_1 - \tau_b\right) \end{bmatrix} = \vec{0}$.

Since $\sigma_1^2, \sigma_2^2 - c^2 \neq 0$, the matrix is non-singular and only has the zero vector in the kernel. So, $\begin{bmatrix} d_{m1} \\ d_{m2} \end{bmatrix} = \begin{bmatrix} d_{s1} \\ d_{s2} \end{bmatrix} = \vec{0}$.

We have then proven that $0 = d_{m1} = d_{m2} = d_{s1} = d_{s2} = d_c$ as required.

$\square$

7.3 Model Solution

7.3.1 Quadratic Model

Here, the investor solves the following problem:

$$\max_{\{\beta_1,\beta_2\}} u_i\left(\beta_1,\beta_2\right) = \beta_1\mu_{i,1} + \beta_2\mu_{i,2} - \frac{1}{2}\beta_1^2\sigma_1^2 - \frac{1}{2}\beta_2^2\sigma_2^2 + \beta_1\beta_2 c_i$$

The first order conditions are

$$[\beta_1] : \mu_{i,1} - \beta_1\sigma_1^2 - \beta_2 c_i = 0$$

$$[\beta_2] : \mu_{i,2} - \beta_2\sigma_2^2 - \beta_1 c_i = 0$$

Rewritten in matrix form

$$\implies \begin{bmatrix} \sigma_1^2 & c_i \\ c_i & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \mu_{i,1} \\ \mu_{i,2} \end{bmatrix}$$

Invert the covariance matrix to get

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \frac{1}{\sigma_1^2\sigma_2^2 - c_i^2} \begin{bmatrix} \sigma_2^2 & -c_i \\ -c_i & \sigma_1^2 \end{bmatrix} \begin{bmatrix} \mu_{i,1} \\ \mu_{i,2} \end{bmatrix}$$

$$\implies \beta_1 = \frac{\sigma_2^2\mu_{i,1} - c_i\mu_{i,2}}{\sigma_1^2\sigma_2^2 - c_i^2}, \beta_2 = \frac{\sigma_1^2\mu_{i,2} - c_i\mu_{i,1}}{\sigma_1^2\sigma_2^2 - c_i^2}$$

7.3.2 Mean Variance

Here the investor solves the following optimization problem (in vector notation)

$$\min_{\vec{\beta}} \frac{1}{2}\vec{\beta}^T C\vec{\beta}$$

$$\text{s.t. } \vec{\beta}^T\vec{\mu} \geq \mu_0$$

$$\vec{\beta}^T\vec{1} = 1$$

Following the approach in Markowitz, we first take the Lagrangian.

$$\mathcal{L} = \frac{1}{2}\vec{\beta}^T C\vec{\beta} + \lambda\left(\vec{\beta}^T\vec{\mu} - \mu_0\right) + \gamma\left(\vec{\beta}^T\vec{1} - 1\right)$$

Using the fact that C is positive definite and symmetric, we can take the FOC for $\vec{\beta}$ as:

$$\left[\vec{\beta}\right] : \vec{0} = C\vec{\beta} - \lambda\vec{\mu} - \gamma\vec{1}$$

$$[\lambda] : \mu_0 = \vec{\beta}^T\vec{\mu}$$

$$[\gamma] : \vec{\beta}^T\vec{1} = 1$$

$$[\text{Slackness}] : \lambda^T\left(\vec{\beta}^T\vec{\mu} - \mu_0\right) = 0$$

From here Markowitz demonstrates the “Two Fund Theorem” for investors facing only risky assets. Markowitz defines the following 2 portfolios: first is the minimum variance portfolio, and the second is the fronteir portfolio which we denote $\vec{\beta}_{\text{min}}$ and $\vec{\beta}_f$.

$$\vec{\beta}_{\min} = \frac{\vec{\mu}^T C^{-1} \vec{1}}{\vec{1}^T C^{-1} \vec{1}}$$

$$\vec{\beta}_f = \frac{C^{-1} \vec{\mu}}{\vec{1}^T C^{-1} \vec{\mu}}$$

There are 2 cases: a) $\mu_0 \leq \vec{\beta}_{\min}^T \vec{\mu}$. Then, the solution is $\vec{\beta}_{\min}$. By definition, $\vec{\beta}_{\min}$ is the minimum variance portfolio (subject to no constraints). So it solves this problem and meets the minimum return constraint.

b) If $\mu_0 \geq \vec{\beta}_{\min}^T \vec{\mu}$, then take

$$\vec{\beta} = \vec{\beta}_{\min} + c \left(\vec{\beta}_f - \vec{\beta}_{\min} \right)$$

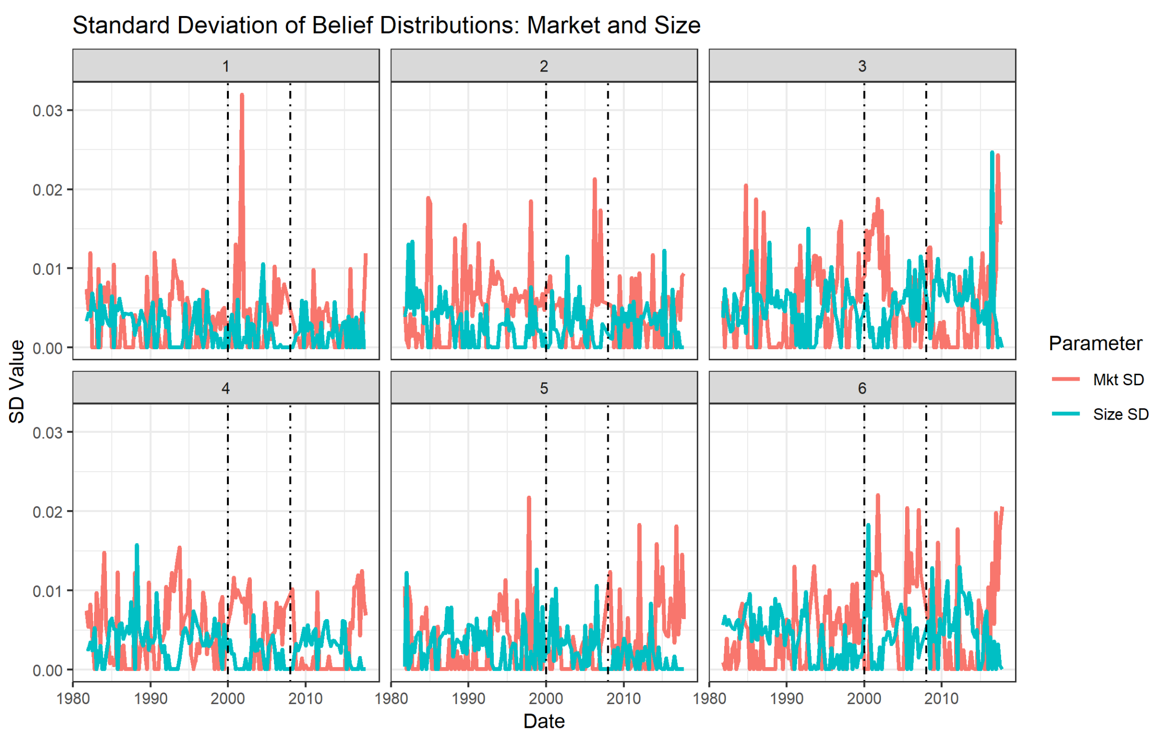
Where

$$c = \frac{\mu_0 - \vec{\mu}^T \vec{\beta}_{\min}}{\vec{\mu}^T \left(\vec{\beta}_f - \vec{\beta}_{\min} \right)}$$

Under the two-fund theorem, $\vec{\beta}_f$ (the frontier portfolio which maximizes the sharpe ratio) equals the solution to the quadratic model in the earlier subsection. Furthermore, if μ_0 is significantly larger than $\vec{\mu}^T \vec{\beta}_{\min}$ then the mean-variance optimizing solution will be significantly weighted towards $\vec{\beta}_f$. This is reasonable in this context because the market return is much higher than the return on size or value. So, we would expect the investor behavior under mean-variance to be well approximated by $\vec{\beta}_f$ and therefore the quadratic model.

List of Figures

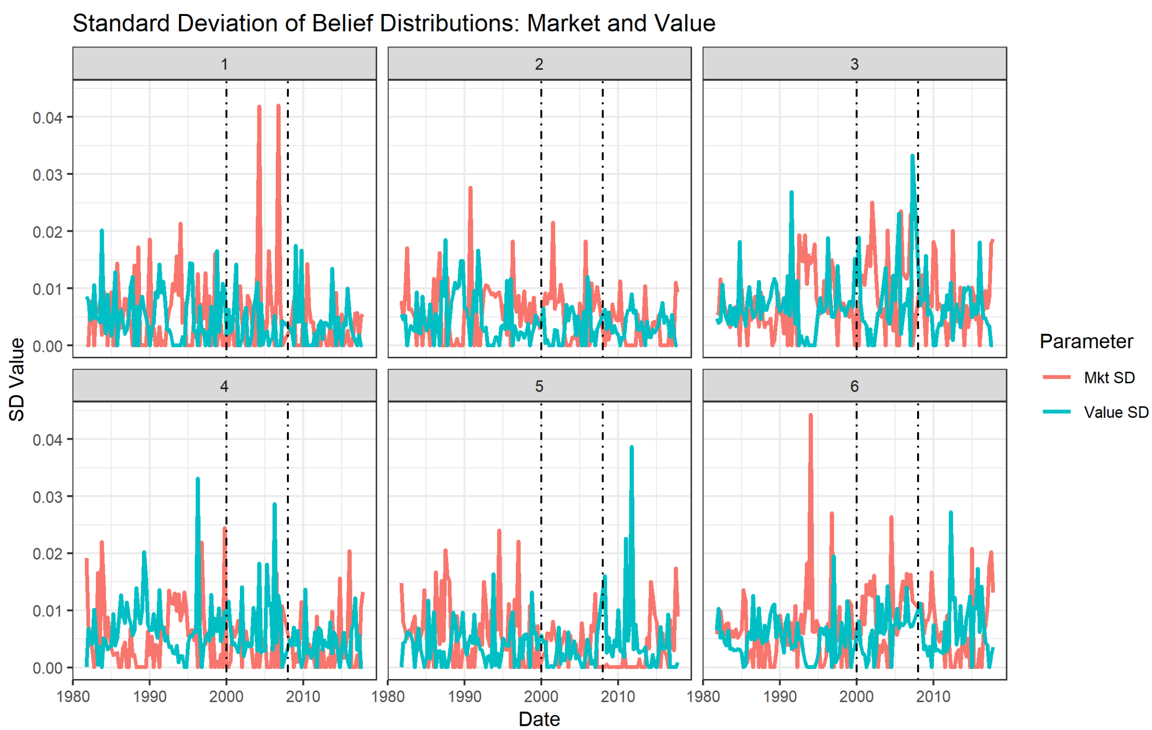
(a) Mean Parameters



(b) Standard Deviation Parameters

Figure 1: Parameter Estimates, Market and Size (Weighted by AUM)

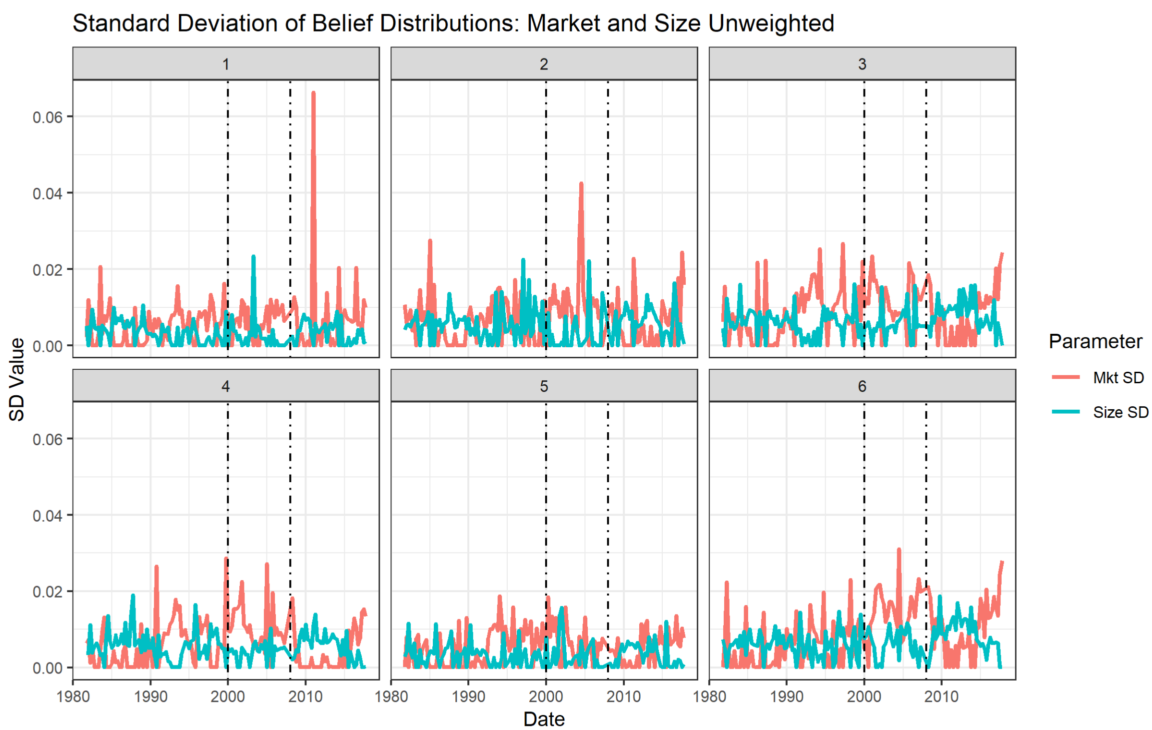
(a) Mean Parameters



(b) Standard Deviation Parameters

Figure 2: Parameter Estimates, Market and Value (Weighted by AUM)

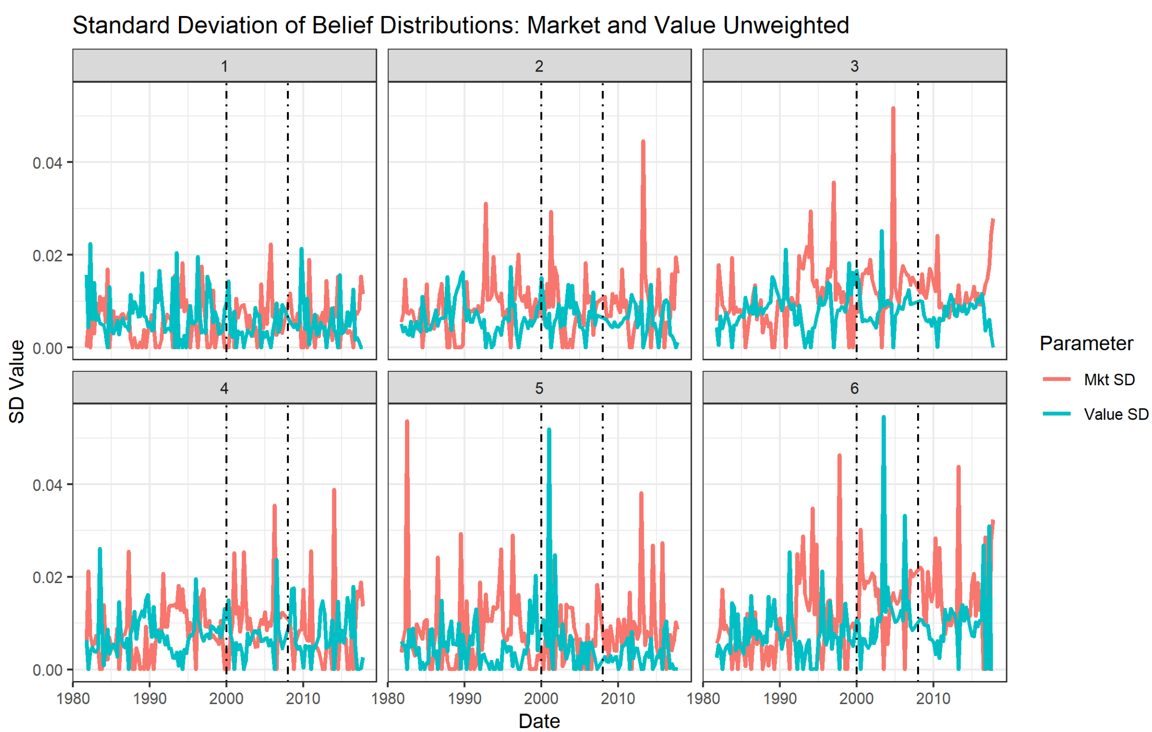
(a) Mean Parameters



(b) Standard Deviation Parameters

Figure 3: Parameter Estimates, Market and Size (Equal Weight)

(a) Mean Parameters



(b) Standard Deviation Parameters

Figure 4: Parameter Estimates, Market and Value (Equal Weight)

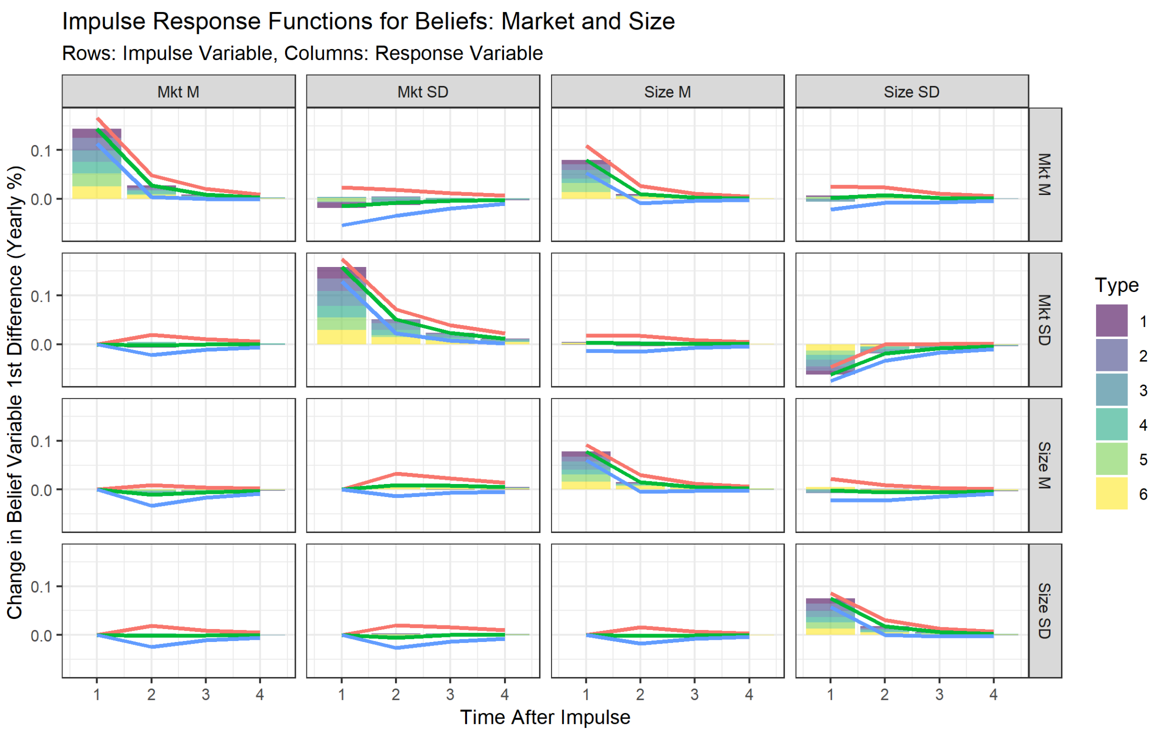


Figure 5: Belief Impulse Response Function, Market and Size Factors

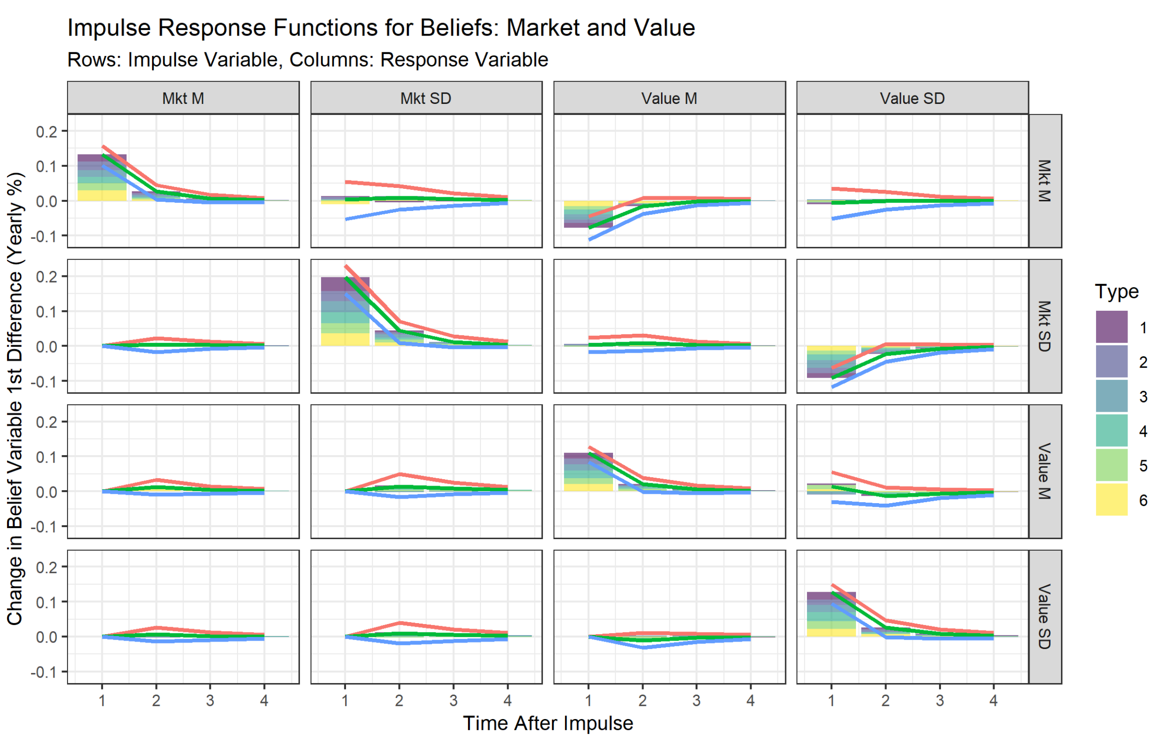
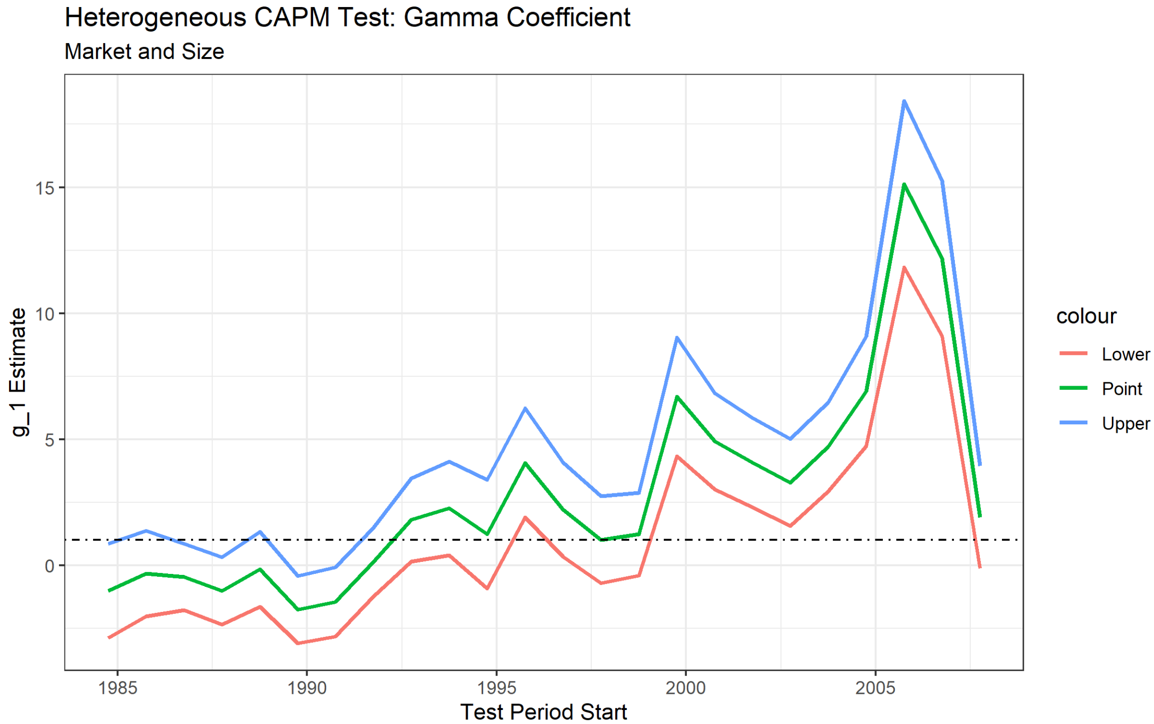
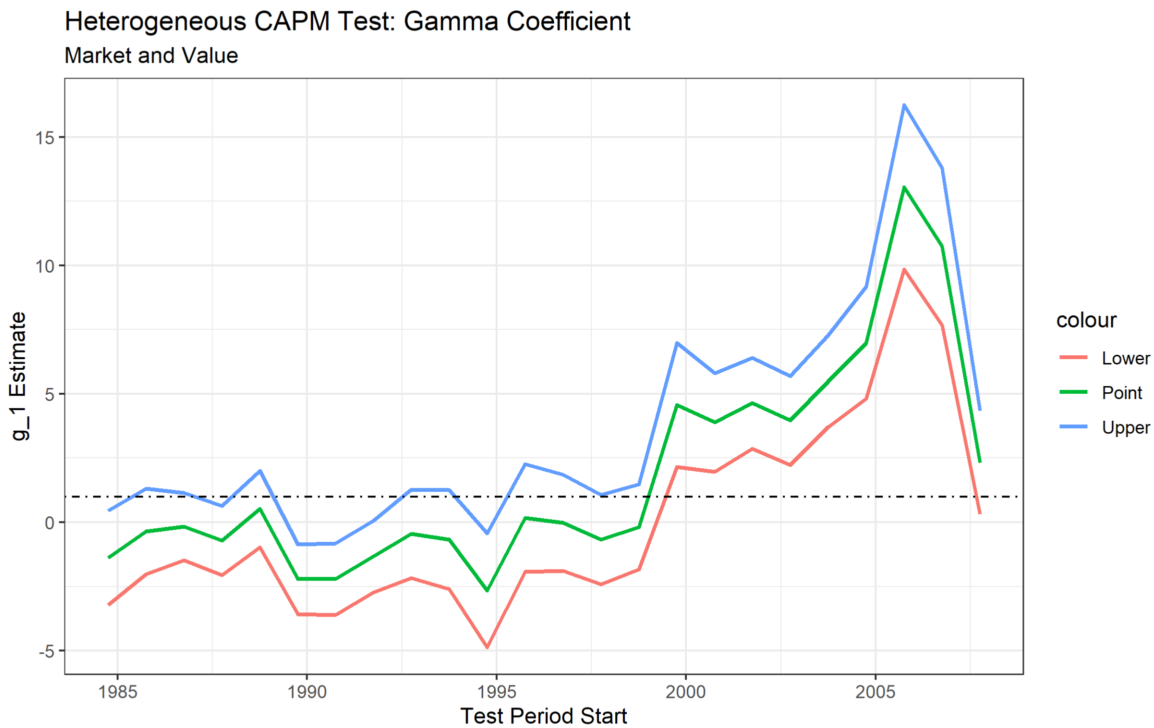


Figure 6: Belief Impulse Response Function, Market and Value Factors

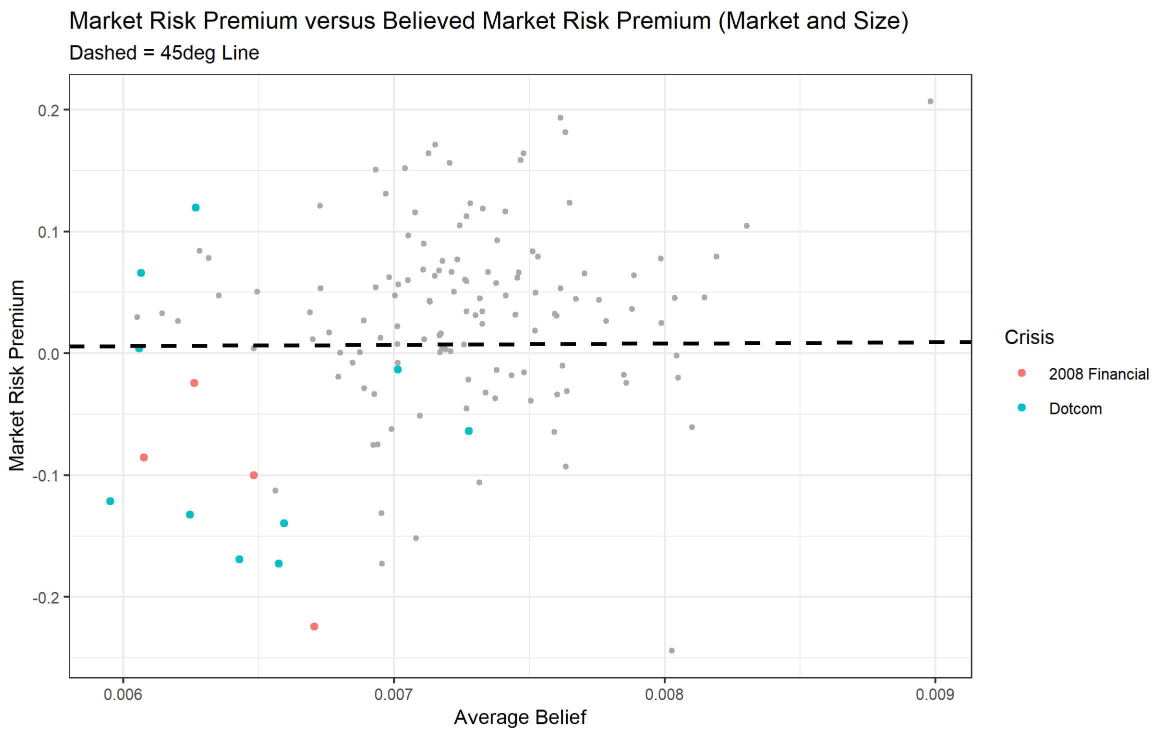


(a) Market and Size

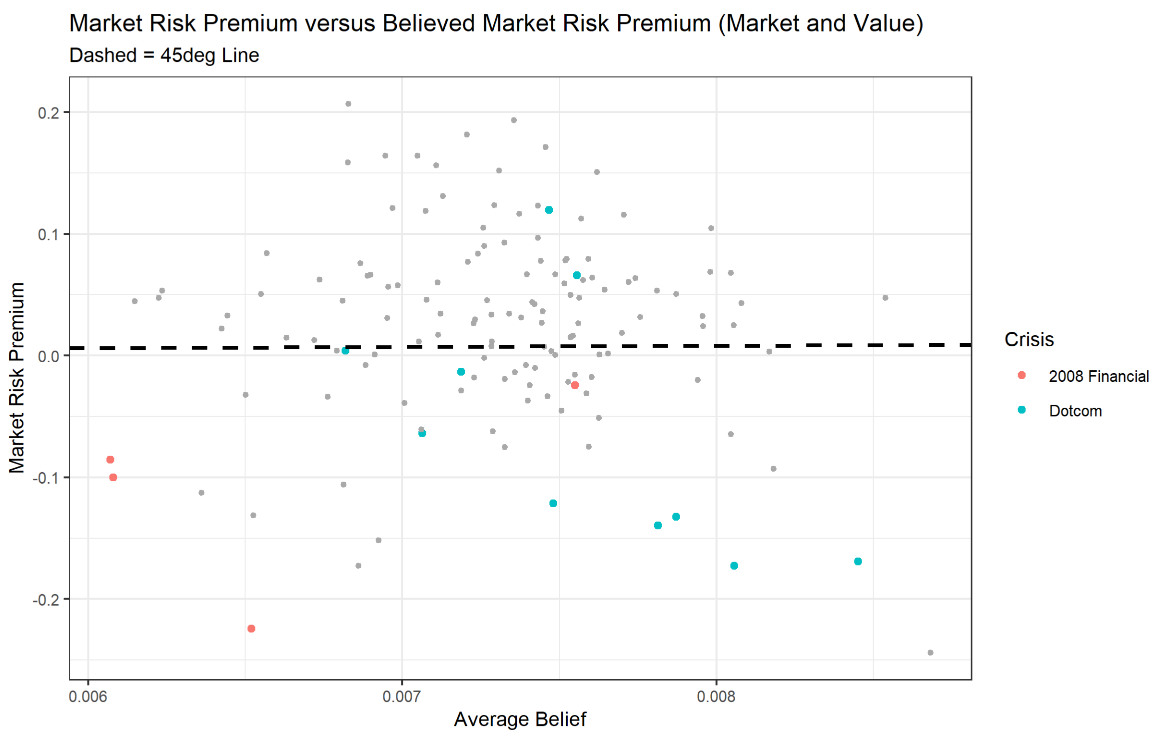


(b) Market and Value

Figure 7: Heterogeneous CAPM Regression



(a) Market and Size



(b) Market and Value

Figure 8: Consensus versus Actual Returns

List of Tables

Type	Decade	Market Mean	SMB Mean	Market SD	SMB SD
1	1980	0.97	-0.12	0.21	0.44
1	1990	0.95	-0.19	0.22	0.27
1	2000	0.80	-0.06	0.25	0.37
1	2010	0.93	-0.08	0.22	0.38
2	1980	1.02	-0.12	0.25	0.53
2	1990	1.00	-0.05	0.34	0.41
2	2000	0.88	0.03	0.33	0.45
2	2010	0.94	0.05	0.33	0.62
3	1980	1.07	0.07	0.26	0.62
3	1990	1.06	0.14	0.44	0.61
3	2000	1.00	0.18	0.50	0.67
3	2010	0.99	0.15	0.45	0.73
4	1980	1.08	-0.00	0.22	0.53
4	1990	1.07	0.14	0.36	0.51
4	2000	0.98	0.15	0.40	0.54
4	2010	1.01	0.16	0.25	0.50
5	1980	1.03	-0.12	0.14	0.33
5	1990	0.99	-0.06	0.26	0.31
5	2000	0.95	0.00	0.27	0.37
5	2010	1.00	-0.08	0.26	0.41
6	1980	1.02	-0.12	0.26	0.58
6	1990	0.97	-0.04	0.33	0.53
6	2000	1.01	0.19	0.61	0.82
6	2010	1.02	0.15	0.59	0.90

Table 1: Beta Choices Summary Statistics (Market and Size)

Type	Decade	Market Mean	SMB Mean	Market SD	SMB SD
1	1980	0.93	-0.05	0.21	0.38
1	1990	0.89	0.02	0.23	0.33
1	2000	0.85	0.18	0.20	0.30
1	2010	0.90	0.04	0.20	0.27
2	1980	1.01	0.02	0.24	0.47
2	1990	1.00	0.03	0.35	0.45
2	2000	0.94	0.19	0.31	0.48
2	2010	0.94	0.08	0.31	0.42
3	1980	1.08	-0.01	0.27	0.63
3	1990	1.09	-0.05	0.45	0.64
3	2000	1.11	0.09	0.48	0.63
3	2010	1.04	0.03	0.45	0.58
4	1980	1.07	-0.04	0.21	0.53
4	1990	1.10	-0.05	0.37	0.54
4	2000	1.10	0.11	0.35	0.52
4	2010	1.05	-0.01	0.27	0.39
5	1980	1.00	-0.04	0.17	0.35
5	1990	0.98	0.02	0.27	0.33
5	2000	0.99	0.09	0.26	0.35
5	2010	0.98	0.03	0.23	0.27
6	1980	0.99	-0.06	0.25	0.50
6	1990	0.95	-0.03	0.39	0.51
6	2000	1.13	0.07	0.59	0.80
6	2010	1.06	0.07	0.60	0.70

Table 2: Beta Choices Summary Statistics (Market and Value)

Symbol	Meaning	Value
σ_1^2	Var (Market)	0.00735
σ_2^2	Var(Size)	0.00308
c	Cov(Market, Size)	0.00201

$$\Sigma = \begin{bmatrix} 0.00735 & 0.00201 \\ 0.00201 & 0.00308 \end{bmatrix}$$

Table 3: Covariance Matrix (Market and Size)

Symbol	Meaning	Value
σ_1^2	Var (Market)	0.00735
σ_2^2	Var(Value)	0.00358
c	Cov(Market, Value)	-0.00134

$$\Sigma = \begin{bmatrix} 0.00735 & -0.00134 \\ -0.00134 & 0.00358 \end{bmatrix}$$

Table 4: Covariance Matrix (Market and Value)