Essays on Time-Varying Discount Rates

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:9306478">http://nrs.harvard.edu/urn-3:HUL.InstRepos:9306478</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
ABSTRACT

This dissertation consists of three essays that explore the interaction between various discount rates and the macroeconomy.

The first essay studies the cross-section of discount rates, specifically, the term structure of interest rates. When physical capital is discounted like a bond with a similar duration, a high term spread is associated with low average duration for investment. I document a strong negative correlation between the term spread and the duration of investment, implying an important role for the cost of capital in determining the composition of aggregate investment. The results are robust to including a variety of controls. Consumer durable goods purchases display similar behavior.

The second essay develops a new utility specification that incorporates Campbell–Cochrane–type habits into the Epstein–Zin class of preferences. It is a model in which risk premia change over time. In a simple calibration of a real business cycle model with EZ-habit preferences, the model generates a strongly countercyclical equity premium, substantial equity return predictability, and a stable riskless interest rate, as in the data. Moreover, conditional on the average level of risk aversion, time-variation in risk aversion increases the volatility and mean return of equities. On the real side, the model matches the short and long-term variances of output, consumption, and investment growth. As an additional empirical test, I measure implied risk aversion and find that it has an R² of over 50 percent for 5-year stock returns in post-war data.

The third essay develops a New-Keynesian model in which households have Epstein–Zin preferences with time-varying risk aversion and the central bank has a time-varying inflation target. The model matches the dynamics of nominal bond prices in the US economy well: the fitting errors for individual bond yields are roughly as large as those ob-
tained from a non-structural three-factor model, and two thirds smaller than in models with constant risk aversion or a constant inflation target.
<table>
<thead>
<tr>
<th>CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>1. Investment and the Cost of Capital in the Cross-Section: The Term Spread Predicts the Duration of Investment</td>
</tr>
<tr>
<td>1.1 Introduction</td>
</tr>
<tr>
<td>1.2 Data</td>
</tr>
<tr>
<td>1.3 Results</td>
</tr>
<tr>
<td>1.4 Model</td>
</tr>
<tr>
<td>1.5 Alternative explanations</td>
</tr>
<tr>
<td>1.6 Consumer durables</td>
</tr>
<tr>
<td>1.7 The firm-level mechanism</td>
</tr>
<tr>
<td>1.8 Conclusion</td>
</tr>
<tr>
<td>2. A model of time-varying risk premia with habits and production</td>
</tr>
<tr>
<td>2.1 Introduction</td>
</tr>
<tr>
<td>2.2 The model</td>
</tr>
<tr>
<td>2.3 Calibration and simulation</td>
</tr>
<tr>
<td>2.4 Empirical return forecasting</td>
</tr>
<tr>
<td>2.5 Extensions</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
</tr>
<tr>
<td>3. Bond pricing with a time-varying price of risk in an estimated medium-scale Bayesian DSGE model</td>
</tr>
</tbody>
</table>
### 3.1 Introduction ........................................... 93
### 3.2 Household preferences .............................. 97
### 3.3 Aggregate supply ...................................... 101
### 3.4 Model solution ........................................ 108
### 3.5 Empirics .................................................. 112
### 3.6 Asset pricing .............................................. 118
### 3.7 The real economy ....................................... 137
### 3.8 Conclusion ................................................. 143

**Appendix** 146

**A. Appendix to Chapter 1** ................................. 147
   A.1 The approximation for average duration ............... 147
   A.2 Further robustness tests ............................... 148

**B. Appendix to Chapter 2** ................................. 150
   B.1 The certainty equivalent ............................... 150
   B.2 Derivation of the SDF ................................. 152
   B.3 The log-linear model with production .................. 153
   B.4 Details of return forecasting ......................... 165

**C. Appendix to Chapter 3** ................................. 175
   C.1 Results from the text ................................. 175
   C.2 Approximation method ................................. 177
   C.3 Estimation .............................................. 182
ACKNOWLEDGMENTS

I completed this dissertation with the support of many people. John Campbell was endlessly patient, and amazingly diligent about reading and commenting on any work I sent him. He went above and beyond any reasonable expectation. I also benefited from numerous comments and conversations with my other advisors, Emmanuel Farhi, Effi Benmelech, and also David Laibson, Robin Greenwood, and Jim Stock, among others. Countless conversations with fellow students made this possible, including in particular Jason Beeler, Stefano Giglio, Kelly Shue, and Eric Zwick. Economists too numerous to name also generously gave me constant feedback on my work which is reflected here. But none of the above should be implicated in any errors or omissions.
1. INVESTMENT AND THE COST OF CAPITAL IN THE CROSS-SECTION: THE TERM SPREAD PREDICTS THE DURATION OF INVESTMENT

1.1 Introduction

This paper studies the cross-section of investment. While there is an enormous amount of work studying the aggregate level of investment and the determinants of firm-level investment, there is essentially no analysis of the determinants of investment in different types of assets. This paper begins that task by analyzing the distribution of investment across assets according to their depreciation rates. I show that when interest rates for long-duration assets are higher than those for short-duration assets, aggregate investment shifts relatively towards high-depreciation assets.

The response of investment to the cost of capital is a key mechanism in macroeconomics and finance. It is central to production-based asset pricing theories (e.g. Cochrane 1991, 1996); a primary feedback mechanism in standard general-equilibrium models; one of the key drivers for the response of the economy to monetary policy shocks; the source of the classical crowding-out effect of government spending; and an important determinant of the size of distortions from taxes. This paper considers a novel method for uncovering an empirical relationship between investment and the cost of capital.

There is a long literature that studies the effect of the cost of capital on investment. Simple methods have, in general, failed to find important effects.\(^1\) Bernanke and Gertler (1995) find that nonresidential investment seems to respond only weakly to shocks to the Federal funds rate.\(^2\) The discount rate is also a determinant of Tobin’s Q, but estimates of the impact of Q on investment tend to be small (Summers, 1981; Eberly, Rebelo, and Vincent, 2002).

---

1 See Chirinko, 1993, for an extensive review.

2 However, in recent unpublished papers, Gilchrist and Zakrajsek (2008) and Guiso et al. (2002) find a relationship in micro data between investment and interest rates.
2009, give a recent review). The primary contribution of this paper is to show that interest rates affect what assets firms invest in at the aggregate level. Furthermore, the effect is relevant at the cyclical frequency: it is neither centered around discrete (and somewhat rare) policy changes, such as tax changes, or dependent on very long-term effects.³

The basic idea here is to forecast the cross-section of investment using the cross-section of interest rates, instead of forecasting the level of investment with the level of interest rates. Long-term assets are discounted with long-term interest rates, and short-term assets with short rates. When long rates are higher than short rates—the term spread is high—a cost-of-capital effect implies investment should shift towards short-duration assets. The negative relationship holds strongly in the data: it explains roughly one third of the cross-sectional variation in investment by duration, and the effect holds both within and across industries.

Standard regressions of aggregate investment on the level of interest rates have the fundamental identification problem that periods of high interest rates may also be periods when investment demand is high, so the correlation between investment and interest rates could be zero or even positive. By studying the cross-section, I abstract from aggregate shocks, hopefully reducing this endogeneity problem. The strong empirical results suggest that in fact endogeneity is less of an issue in the cross-section.

The data is simple to construct. I obtain nominal investment by asset and year from the Bureau of Economic Analysis. I study an index of average duration defined as the average of the assets’ economic life-spans, weighted by their share in aggregate investment in each year.⁴ Figure 1.1 shows that this index of average duration is highly negatively correlated with the spread between interest rates on ten and one-year nominal Treasury bonds (note that the axis for average duration is reversed for the sake of clarity). When interest rates are relatively high for long-duration assets, investment shifts towards short-duration assets, creating a strong negative correlation between average duration and the

³ Caballero (1994) and Schaller (2006) use cointegration methods to show that in the long run the cost of capital is meaningfully related to the size of the capital stock (and hence the level of investment). A number of researchers have also focused on high-frequency changes in taxes which produce large movements in the cost of capital and investment (Hassett and Hubbard, 2002, provide an extensive review).

⁴ Specifically, “lifespan” is measured as a Macaulay duration using data on economic depreciation rates.
A negative raw correlation between investment and interest rates suggests that the cost of capital has an important role in determining the cross-sectional distribution of investment, but there are alternative mechanisms that could produce this result. I therefore build a simple Q-theory model to help elucidate the possible sources of bias in the basic result in figure 1.1 and try to account for them in subsequent regressions. I control for the level of productivity and expected productivity growth in a variety of ways and find that they do not eliminate the basic effect. More importantly, I find that the term spread–average duration relationship is not driven by changes in demand across industries. When the term spread is high, investment shifts towards low-duration assets within individual industries, in addition to shifting from industries that use more long-duration assets to ones that use more short-duration assets.

In addition to contributing to the literature on the determinants of the level of investment, this paper is related to the recent literature on production-based asset pricing with projects that have differing characteristics (e.g. Berk, Green, and Naik, 1999, and Gomes, Kogan, and Yogo, 2009). While those papers show that variation in the types of capital owned by firms can lead to differences in their stock prices, I find that variation in the cross-section of asset prices can affect the types of investment that firms undertake.

The findings here are also relevant to understanding the relationship between interest rates and debt issues. The final section of the paper provides novel evidence that firms match the maturity of their debt issues to their physical investment, consistent previous evidence in the finance literature (e.g. Stohs and Mauer, 1996). My findings suggest that the timing of debt issues to the term spread documented by Baker, Greenwood, and Wurgler (2003) could be explained by the dynamics of physical investment and the fact that firms match the maturity of their debt to their assets.

The remainder of the paper is organized as follows. Section 1.2 describes the data and section 1.3 reports the main result. Next, I outline in section 1.4 a simple model that justifies the regression of the average duration of investment on the term spread. Section 1.5 controls for a number of possible biases suggested by the investment model and shows that the term spread is the single most powerful predictor of the average duration of in-
Figure 1.1: The average duration of investment versus the term spread

Note: The term spread is the gap between the 10 and 1-year treasury yields averaged over the previous year. Both variables are HP-detrended. The axis for average duration is reversed. Grey bars indicate NBER-dated recessions.
vestment. In section 1.6 I show that the relationship between duration and the term spread also appears in purchases of consumer durable goods. Section 1.7 examines the relationship between the type of debt that firms sell and the duration of their assets. I find a positive relationship (consistent with maturity-matching theories), which gives added support for the idea that long-term interest rates are the relevant cost of capital for long-duration assets and short rates for short-term assets. Finally, section 1.8 concludes.

1.2 Data

To study the relationship between investment and the cost of capital in the cross-section, we need a relevant measure of the cost of capital that differs across assets. The duration of assets is a natural source of variation because it is easy to quantify for both physical assets (through their depreciation rates) and bonds (through maturities). Of course, the cost of capital depends on more than simply the level of interest rates. The equity premium is large and variable (e.g. Lettau and Ludvigson, 2001). The advantage of focusing on interest rates here is that we can directly observe the cost of capital for assets of different durations. While there have been studies of the term structure of equity (Lettau and Wachter, 2007), there is no simple way to actually measure the term structure of expected returns on equity, let alone the variation in the slope of that term structure.

I obtain data on Treasury yields measured at year-end from the Federal Reserve. Treasury data has the advantage of including bonds with a large variety of maturities over a long period of time. However, firms do not in general borrow at the Treasury yield. I therefore also study the spread between yields on 3-month commercial paper and the Moody’s seasoned Baa corporate bond yield (from Global Financial Data and the Federal Reserve, respectively). The Moody’s index is meant to measure bonds with remaining maturities near 30 years. The main results focus on the Treasury yield spread.\footnote{Ideally, we would measure the true cost of capital for each asset, including the cost of capital for equity, in particular. While there is research on the term structure for equity (Lettau and Wachter, 2007), it is not obvious how to construct an equity cost of capital for each asset simply by looking at its depreciation rate.}

A potential concern is that the relevant discount rate for investment is the real interest rate, not the nominal rate. One method for obtaining the real interest rate would be...
to subtract an inflation forecast from the nominal rate. In general, random-walk inflation forecasts are competitive with more sophisticated methods (Atkeson and Ohanian, 2001). With a random-walk forecast, the nominal term spread and the spread obtained after subtracting expected inflation will be identical, which suggests that there is little to be gained by forecasting inflation here.\(^6\)

Another option is to look at yields on inflation-protected bonds. The time series of inflation-protected bonds in the United States is relatively short, but inflation protected bonds have been sold in the United Kingdom since the 1980’s. Figure 1.2 plots the 10/5 year term spread in the UK for both nominal and inflation-protected bonds since 1985. Over the sample, the two series move together closely, even through the financial crisis. Their variances differ, but they are over 70 percent correlated. This result suggests that by studying the nominal term spread, we will obtain results that are similar to what we would obtain with the unobservable real term spread.

Data on capital stocks and investment come from the Bureau of Economic Analysis’s (BEA) fixed asset tables. The main results focus on aggregate investment by asset, but the BEA also reports data at the asset \(\times\) industry level. Data on depreciation rates is from Fraumeni (1997), the source for current depreciation rates used by the BEA.\(^7\) The BEA uses geometric (declining balance) depreciation for nearly all assets.\(^8\) Depreciation rates are estimated primarily from data on service lives and sales of vintage assets. Given the resale value of an asset for each age along with a service life, one can estimate an approximate geometric depreciation rate.\(^9\) These depreciation rates are closer to economic depreciation than the straight-line method used for accounting purposes by many firms (Hulten and Wykoff, 1981).

I use 36 asset classes from the BEA tables, excluding household and government as-

---

\(^6\) Furthermore, we would need to estimate a 10-year inflation forecast, which would be difficult even if inflation were relatively easy to forecast at short horizons. Another option would be to use survey data on inflation forecasts, but this would substantially limit the available time series.

\(^7\) Her depreciation rates closely match depreciation obtained by simply dividing BEA reported depreciation by the capital stock.

\(^8\) Missiles and nuclear fuel rods, for example, are modeled with straight-line depreciation.

\(^9\) The BEA’s current estimates are a combination of data from a variety of studies on resale values reviewed in Fraumeni, 1997.
Figure 1.2: United Kingdom 10/5 year term spreads, 1985–2010

Note: gap between yields on 10 and 5-year nominal and inflation-protected bonds
sets and educational, health, and religion-related structures. The majority of the analysis focuses on equipment investment. The investment literature generally finds that models have substantial trouble explaining structures investment (Oliner, Rudebusch, and Sichel, 1995). This may be partly caused by the fact that nonresidential building projects take fourteen months to complete on average (Edge, 2000).10 The main results below go through when structures are included, but the relationships are far less clear. I therefore leave the analysis of structures to future work so as not to distract from a complete analysis of equipment investment.

For each asset class, the BEA reports total stocks (on a current-cost basis) and investment for the private nonresidential economy. The asset classes accounting for the most nominal investment in 2007 were software (16 percent of total investment), petroleum and natural gas exploration and wells (8 percent), communication equipment (7 percent), and computers and peripheral equipment (6 percent). Except for oil and gas, these assets all have high depreciation rates, and substantial investment is necessary just to keep the stocks at constant levels.

For an asset with geometric depreciation rate δ_i, if we assume that productivity is constant and there is a fixed discount rate r^*, Macaulay’s duration, D_i, will be

\[ D_i = \sum_{j=1}^{\infty} j \left(1 - \delta_i\right)^{j-1} \left(1 + r^*\right)^{-j} = \frac{1 + r^*}{r^* + \delta_i} \]  

(1.1)

When measuring durations I fix r^* = 0.03.11 Table 1.1 lists the assets used in this study along with their depreciation rates and durations. Software, computers, and office and accounting equipment have the highest depreciation rates, all above 20 percent per year. Types of heavy industrial machinery tend to have lower depreciation rates, as low as 5 percent.

Finally, for the purpose of summarizing the cross-section of investment, I define an

10 See Edge (2000) for an empirical model of residential and nonresidential structures investment that takes into account building lags.

11 Allowing for a constant rate of productivity growth would be the equivalent of choosing a lower value of r^*. The results are not sensitive to the choice of r^*.
Table 1.1: Assets, depreciation rates, and durations

<table>
<thead>
<tr>
<th>Depreciation rate (percent)</th>
<th>Duration (years)</th>
<th>Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>3.73</td>
<td>Information processing equipment and software</td>
</tr>
<tr>
<td>0.40</td>
<td>2.37</td>
<td>Computers and peripheral equipment</td>
</tr>
<tr>
<td>0.14</td>
<td>6.11</td>
<td>Communication equipment</td>
</tr>
<tr>
<td>0.14</td>
<td>6.24</td>
<td>Medical equipment and instruments</td>
</tr>
<tr>
<td>0.14</td>
<td>6.24</td>
<td>Nonmedical instruments</td>
</tr>
<tr>
<td>0.18</td>
<td>4.90</td>
<td>Photocopy and related equipment</td>
</tr>
<tr>
<td>0.31</td>
<td>3.01</td>
<td>Office and accounting equipment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial equipment</td>
</tr>
<tr>
<td>0.09</td>
<td>8.46</td>
<td>Fabricated metal products</td>
</tr>
<tr>
<td>0.05</td>
<td>12.62</td>
<td>Engines and turbines</td>
</tr>
<tr>
<td>0.12</td>
<td>6.75</td>
<td>Metalworking machinery</td>
</tr>
<tr>
<td>0.10</td>
<td>7.74</td>
<td>Special industry machinery, n.e.c.</td>
</tr>
<tr>
<td>0.11</td>
<td>7.51</td>
<td>General industrial, including materials handling, equipment</td>
</tr>
<tr>
<td>0.05</td>
<td>12.88</td>
<td>Electrical transmission, distribution, and industrial apparatus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transportation equipment</td>
</tr>
<tr>
<td>0.15</td>
<td>5.72</td>
<td>Trucks, buses, and truck trailers</td>
</tr>
<tr>
<td>0.22</td>
<td>4.12</td>
<td>Autos</td>
</tr>
<tr>
<td>0.12</td>
<td>7.10</td>
<td>Aircraft</td>
</tr>
<tr>
<td>0.06</td>
<td>11.31</td>
<td>Ships and boats</td>
</tr>
<tr>
<td>0.06</td>
<td>11.59</td>
<td>Railroad equipment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other equipment</td>
</tr>
<tr>
<td>0.14</td>
<td>6.15</td>
<td>Furniture and fixtures</td>
</tr>
<tr>
<td>0.12</td>
<td>6.87</td>
<td>Agricultural machinery</td>
</tr>
<tr>
<td>0.16</td>
<td>5.51</td>
<td>Construction machinery</td>
</tr>
<tr>
<td>0.15</td>
<td>5.72</td>
<td>Mining and oilfield machinery</td>
</tr>
<tr>
<td>0.16</td>
<td>5.42</td>
<td>Service industry machinery</td>
</tr>
<tr>
<td>0.18</td>
<td>4.83</td>
<td>Electrical equipment, n.e.c.</td>
</tr>
<tr>
<td>0.15</td>
<td>5.81</td>
<td>Other nonresidential equipment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Structures</td>
</tr>
<tr>
<td>0.02</td>
<td>18.83</td>
<td>Office, including medical buildings</td>
</tr>
<tr>
<td>0.02</td>
<td>19.73</td>
<td>Commercial</td>
</tr>
<tr>
<td>0.03</td>
<td>16.89</td>
<td>Manufacturing</td>
</tr>
<tr>
<td>0.02</td>
<td>19.18</td>
<td>Electric</td>
</tr>
<tr>
<td>0.02</td>
<td>19.18</td>
<td>Other power</td>
</tr>
<tr>
<td>0.02</td>
<td>19.18</td>
<td>Communication</td>
</tr>
<tr>
<td>0.08</td>
<td>9.80</td>
<td>Petroleum and natural gas</td>
</tr>
<tr>
<td>0.05</td>
<td>13.73</td>
<td>Mining</td>
</tr>
<tr>
<td>0.02</td>
<td>19.62</td>
<td>Other buildings</td>
</tr>
<tr>
<td>0.03</td>
<td>17.91</td>
<td>Railroads</td>
</tr>
<tr>
<td>0.02</td>
<td>19.11</td>
<td>Farm</td>
</tr>
</tbody>
</table>

Note: Depreciation rates are obtained from the BEA. Duration is measured as $1.03/(0.03+δ).
index measuring the average duration of investment,

\[ \bar{D}_t \equiv \frac{\sum_i I_{it} D_i}{\sum_i I_{it}} \]  

(1.2)

\( \bar{D}_t \) is simply a weighted average of the durations of the assets, where the weights are the assets’ shares in aggregate nominal investment. When investment shifts relatively towards short-duration assets, e.g. computers or software, \( \bar{D}_t \) falls. Furthermore, \( \bar{D}_t \) is constructed so that it is not mechanically related to the level of investment. There is no particular reason why there need be a positive or negative relationship between the level of investment (or the state of the business cycle) and \( \bar{D}_t \).\(^{12}\)

Figure 1.3 plots \( \bar{D}_t \) for 1948–2008. As might be expected, average duration has been falling over time. The fastest rate of decline appears in the late 1980’s, and the series flattens out after 1994, actually rising substantially between 2006 and 2008. We should not expect transitory changes in the term spread to explain the long-term changes in the duration of investment. Long run changes are driven by technological shifts, e.g. the introduction of computers, software, and other electronic equipment. Instead, the term spread will explain the year-to-year variation in \( \bar{D}_t \).\(^{13}\)

Table 1.1 shows that computers have a depreciation rate of 25 percent, and software 40 percent. Their combined share of nominal investment rises from 7.4 percent in 1978 to 32.7 percent in 2008. Figure 1.3 also includes a version of \( \bar{D}_t \) that excludes investment in computers and software, and we can see that if not for computers and software, there is no decline in the average duration of investment over time. Over the sample, though, the correlation of the first differences of the two versions of \( \bar{D}_t \) is over 90 percent. For the main regressions, I detrend all of the variables using the Hodrick-Prescott (HP) filter with a smoothing parameter of 25. I obtain similar results when I use a polynomial trend or take first differences (see table 1.2 for results in first differences).

\(^{12}\) Rather than using an index of average duration, which involves a discount rate, we could also simply use an index of the average depreciation rate of investment. All of the results below go through with this alternative measure.

\(^{13}\) Tevlin and Whelan (2003) give a more extensive discussion of the recent decrease in the duration of the capital stock.
Figure 1.3: Average Duration of Investment, 1948–2008

Note: average duration is duration summed across all assets, weighted by nominal investment shares. Investment is obtained from the BEA fixed asset tables.
1.3 Results

Figure 1.1 plots HP-detrended $\bar{D}_t$ and the 10/1 year term spread at the end of the previous year (with the axis for $\bar{D}_t$ reversed). The negative relationship is immediately apparent. The term spread and average duration have a correlation of -54 percent. Gray bars indicate NBER-dated recessions. In most recessions, the term spread rises due to the Fed cutting interest rates, and the duration of investment falls. Duration is often high just prior to recessions, e.g. 1970, 1990, 2001, and 2007, when the yield curve is inverted. Looking more closely, we can see that over time the term spread has become more volatile while $\bar{D}_t$ has become somewhat less volatile. This is a common finding: the real economy has become less volatile (the great moderation), while Federal Reserve policy has become more aggressive, causing higher volatility in interest rates.

Table 1.2 reports results of regressions of $\bar{D}_t$ on the first lag of the term spread. All of the variables in table 1.2 are standardized to have unit variance so that the regression coefficients indicate how a one standard deviation increase in the independent variables affects $\bar{D}_t$ in terms of its own standard deviation. The units of $\bar{D}_t$ have no deep economic meaning on their own.

As expected, in the first column we find a highly significant negative coefficient on the term spread and an $R^2$ of 0.30. This is a high value; Oliner, Rudebusch, and Sichel (1995), when forecasting the level of aggregate investment using models with as many as 11 lags of quarterly data, obtain at best an $R^2$ of 0.34. With a single variable, I am able to get an $R^2$ nearly as high for $\bar{D}_t$. Column two uses the term spread on corporate bonds instead of Treasuries and finds a nearly identical coefficient and $R^2$.

The third column of table 1.2 controls for the lagged level of $\bar{D}_t$. The coefficient is only marginally significant and the coefficient on the term spread is essentially unchanged. Column 4 shows that leading values of the term spread have no explanatory power for average duration, which is consistent with the theory that firms are responding to the cost of capital, rather than there being some underlying variable that causes the term spread and $\bar{D}$ to generally move together.

Finally, the fifth column runs the basic regression using investment in all assets instead
Table 1.2: Regressions of the average duration of investment

<table>
<thead>
<tr>
<th>Assets:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equip.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term spread(t-1)</td>
<td>-0.56 ***</td>
<td>-0.49 ***</td>
<td>-0.58 ***</td>
<td>-0.31 ***</td>
<td>-0.41 ***</td>
<td>-0.28 ***</td>
<td>-0.14 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.11]</td>
<td>[0.12]</td>
<td>[0.14]</td>
<td>[0.12]</td>
<td>[0.11]</td>
<td>[0.06]</td>
<td>[0.06]</td>
<td></td>
</tr>
<tr>
<td>Corporate TS(t-1)</td>
<td></td>
<td>-0.52 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration(t-1)</td>
<td></td>
<td></td>
<td>0.20 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.11]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term spread (t)</td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Spread(t+1)</td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.08]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>59</td>
<td>59</td>
<td>47</td>
<td>58</td>
<td>59</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>R2</td>
<td>0.30</td>
<td>0.25</td>
<td>0.37</td>
<td>0.31</td>
<td>0.10</td>
<td>0.22</td>
<td>0.21</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Note: * indicates significance at the 10 percent level, ** 5 percent level, *** 1 percent level. Annual data, 1950–2008, where available. The dependent variable is the average duration of investment. Investment and depreciation rates are obtained from BEA. The term spread is the 10-year minus the 1-year treasury yield at the end of the calendar year. The corporate term spread is the spread between the Moody’s AAA corporate 30 year index and the St. Louis Fed’s 3 month commercial paper yield. Columns 7 through 9 give results from first differenced regressions. Column 8 uses the effect of within-industry reallocation on average duration as the dependent variable. Column 9 is defined analogously using cross-industry reallocation. All variables are detrended with the HP filter with a smoothing parameter of 25 and standardized to have unit variance. Newey-West standard errors with a 3-year window are reported in brackets.
of equipment alone. The results still go through. The symmetrical regression using only structures investment is unenlightening because there is not enough variation in duration within structures to provide reasonable statistical power.

To test for a break in the relationship between $\bar{D}_t$ and the term spread, I use the sup-$F$ test (also known as the Quandt likelihood ratio test). We might expect that the break in this relationship would have appeared following the great moderation, when monetary policy became more aggressive and the economy less volatile. The $F$-test for a break, though, is maximized in 1958. Looking at figure 1.1, it is clear that after 1958 the volatility of $\bar{D}_t$ fell and the volatility of the term spread rose. The $F$-statistic for a break is never above the critical value reported in Andrews (1993) except for in 1958 and 1959. The highest value outside those two years is 4.56 in 1992, well below the 10 percent critical value of 5.00. There is thus evidence for a structural break, but not where we might have thought. For the period since 1960, we cannot reject the hypothesis that the relationship between $\bar{D}_t$ and the term spread has been stable.

1.4 Model

With the basic result in hand, it is useful to build a simple and stylized model to help understand where this correlation might come from. It is tempting to immediately jump to the conclusion that there is variation in the cost of capital (i.e. shocks to the supply of investment goods), which drives the result in figure 1.1. The model helps identify what other factors might induce a similar correlation.

I consider a standard infinite-horizon setup with a few simplifications for analytic tractability. Firms face a linear production function in each type of capital, where the current level of productivity for asset $i$ is $B_{it}$. That is, revenue is equal to

$$\sum_i B_{it} K_{it}$$

where $K_{it}$ is the stock of asset $i$ at date $t$. Note that this revenue function ignores complementarities between types of assets. In general, if a decline in the term spread is expected to shift investment towards long-duration assets, complementarity across assets will at-
tenuate this effect (in the limit of a Leontief production function, firms would never vary the composition of the capital stock).

I follow Baxter and Crucini (1993) and Jermann (1998) in specifying the update process for capital as

\[ K_{it+1} = (1 - \delta_i) \phi_i (I_{it-1}) + \phi_i (I_{it}) \]  

(1.4)

The update process for capital assumes that capital only operates for two periods for the sake of analytic simplicity. Each asset depreciates by the factor \((1 - \delta_i)\) between its first and second period of operation, and is subsequently obsolete. \(\phi_i\) incorporates adjustment costs in investment so that a unit of investment may create less than one unit of capital. \(\phi_i\) takes the form

\[ \phi_i (I_{it}) = \frac{\eta_{1i}}{1 - 1/\gamma} t_{it}^{1-1/\gamma} + \eta_{2i} \]  

(1.5)

\(\phi_i\) has the useful property that the elasticity of investment with respect to Tobin’s Q will equal the constant \(\gamma\).\(^{14}\) The parameters \(\eta_{1i}\) and \(\eta_{2i}\) determine the level of investment and the size of the adjustment costs paid and are allowed to vary across assets.\(^{15}\)

Denoting the discount rate between dates \(t\) and \(t + 1\) as \(r_{t+1}\), the firm maximizes the discounted value of its revenue net of investment costs,

\[ \Pi_t = \max_{I_{it}} \sum_{j=0}^{\infty} \sum_{l} \left[ \exp \left( -\sum_{k=1}^{l} r_{t+k} \right) E_t B_{it+j} K_{it+j} - I_{it+j} \right] \]  

(1.6)

where \(E_t\) denotes the expectation operator conditional on information available at date \(t\). All profits are discounted at the riskless rate \(r_t\).

For productivity growth, I assume that different assets may have different current levels of productivity, but expected productivity growth in the future is the same for all assets,\(^{14}\) This functional form has the drawback that it is not necessarily consistent with negative investment. However, asset-level investment is always positive in the data, so this is not a practical concern here.\(^{15}\) These two parameters allow us to choose a steady state level of investment \(x_i\) where \(Q_i = 1\) and \(\phi_i (x_i) = x_i\) and \(\phi_i' (x_i) = 1\). That is, they allow us to choose a point where the firm pays no adjustment costs overall and on the margin.
\[ E_t \log \left( \frac{B_{it+j}}{B_{it+j-1}} \right) = \mu_{t+j} \]. Taking the first-order condition for \( I_{it} \) gives

\[
\frac{1}{\phi'_i(I_{it})} \exp (\mu_{t+1} - r_{t+1}) B_{it} + \exp (\mu_{t+1} + \mu_{t+2} - r_{t+1} - r_{t+2}) B_{it} (1 - \delta_i) \quad (1.7)
\]

The appendix shows that, using a first-order approximation, we can derive an approximate expression for the index of average duration, \( \bar{D}_t \),

\[
\bar{D}_t \approx d_0 + \gamma N^{-1} \sum_i \left( \log \left( \frac{B_{it}}{\bar{B}_{it}} \right) \right) D_i + F \left( \mu_{t+2} - r_{t+2} \right) \quad (1.8)
\]

where \( d_0 \) is a constant, \( N \) is the number of assets, \( \log \left( \frac{B_{it}}{\bar{B}_{it}} \right) \equiv \log \left( \frac{B_{it}}{N^{-1} \sum_i \log B_{it}} \right) \) is the deviation of the productivity of asset \( i \) from the period average, and \( F \) is a strictly increasing function.\(^{16}\)

We can rewrite the term \( \mu_{t+2} - r_{t+2} \) as

\[
\mu_{t+2} - r_{t+2} = \left( \frac{\mu_{t+1} + \mu_{t+2}}{\text{Total prod. growth}} \right) - \left( \frac{r_{t+1} + r_{t+2}}{\text{Total discount rate}} \right) + \left( \frac{\mu_{t+2} - \mu_{t+1}}{\text{Productivity spread}} \right) - \left( \frac{r_{t+2} - r_{t+1}}{\text{Term spread}} \right) \quad (1.9)
\]

The first line is total productivity growth and the total discount rate between periods \( t \) and \( t + 2 \). The term \( (r_{t+2} - r_{t+1}) \) is the relevant concept of the term spread, and I refer to \( (\mu_{t+2} - \mu_{t+1}) \) as the productivity growth spread.

The previous section considered a simple regression of \( \bar{D}_t \) on the term spread, \( (r_{t+2} - r_{t+1}) \). Holding all else equal (including total expected productivity growth and the total discount rate), equation (1.8) confirms the simple intuition that this relationship should be negative. Equation (1.8) shows, however, that there are at least four potential omitted variables in this regression: total expected productivity growth and discount rates between \( t \) and \( t + 2 \) (\( \mu_{t+1} + \mu_{t+2} \) and \( r_{t+1} + r_{t+2} \)); the productivity growth spread, \( (\mu_{t+2} - \mu_{t+1}) \), and the levels of idiosyncratic productivity, \( N^{-1} \sum_i \left( \log \left( \frac{B_{it}}{\bar{B}_{it}} \right) \right) D_i \).

\(^{16}\) Specifically, \( F(x) = -\frac{\exp(x)}{1 + \exp(x)(1 - \delta)} N^{-1} \sum_i \hat{\delta}_i D_i \) where \( \hat{\delta} \equiv N^{-1} \sum_i \delta_i \) and \( \hat{\delta}_i \equiv \delta_i - \hat{\delta} \). Note that since \( D_i \) is decreasing in \( \delta_i \), \( F \) is strictly increasing.
First, holding the term spread and the productivity spread fixed, an increase in productivity growth \((\mu_{t+1} + \mu_{t+2})\) or a decrease in discount rates \((r_{t+1} + r_{t+2})\) will tilt the distribution of investment towards long-duration assets. This effect is the primary feature of duration: long-duration assets gain more value from a decline in interest rates or an increase in expected productivity growth than do short-duration assets. To the extent that the term spread is correlated with long-term average productivity growth and interest rates, then, a regression of the average duration of investment on the term spread will be biased.

Specifically, we could spuriously find a negative relationship between the term spread and \(\bar{D}_t\) if expected long-term productivity growth is low in periods when the term spread is high. The term spread is countercyclical, so this would correspond to a situation in which expected long-term productivity growth \((\mu_{t+1} + \mu_{t+2})\) is low during recessions. I will try to control for these effects by controlling for the level of aggregate investment and various other indicators of the state of the business cycle.

The second source of bias is that the productivity spread \((\mu_{t+2} - \mu_{t+1})\) could be correlated with the term spread. In particular, if productivity growth is expected to slow down in the same periods that the term spread is high, we would find a spurious negative relationship between average duration and the term spread. In this case, recessions would have to be periods in which productivity growth is expected to decelerate in the future, which seems unlikely given that recessions are periods when growth is already slow in the first place (by definition).

Finally, the levels of productivity across assets could be related to duration, affecting \(\bar{D}_t\) through the \(N^{-1} \sum_i \left( \log \left( B_{it} \right) \right) D_i\) term, which can be thought of as the covariance between duration and productivity across assets. If this covariance changes over time and is systematically related to the level of the term spread, then omitting it from the regression would bias the coefficient on the term spread.

Over long horizons, investment and productivity shift substantially across different assets. The most notable of these changes is the long-run decline in prices and increase
in investment in computers and software (Tevlin and Whelan, 2003). The model would interpret this phenomenon as an increase in $B_{it}$ for low-duration assets, which drives $\bar{D}_t$ downward. A simple way to control for those movements is to detrend $\bar{D}_t$.

Short-run movements in idiosyncratic productivity are more difficult to account for, though. If changes in the term spread are correlated with shifts in productivity that favor certain assets, then the regression of $\bar{D}_t$ on the term spread will be biased. In the empirical analysis below, I discuss and control for some specific mechanisms, most importantly industry demand shifts, that could drive high-frequency movements in $N^{-1} \sum \left( \log (B_{it}) \right) D_i$.

Instead of running a regression of average duration on investment, it would be nice to estimate a more fundamental parameter, such as the coefficient on marginal Q, which tells us about the size of adjustment costs in investment. One way to do that would be to calculate Tobin’s Q for each asset individually, as in Abel and Blanchard (1986), using the full term structure of interest rates. The problem is that we do not actually directly measure the marginal product of any individual asset at any point in time. Moreover, we do not measure anything like the true discount rate for each asset. Rather, the term spread in this paper is measured using Treasury yields and is taken as an indicator of differences in discount rates across assets. A deeper problem is that Abel and Blanchard’s method would also require forecasting inflation at very long horizons, when the literature generally finds that inflation is difficult to forecast even at quarterly and annual horizons (e.g. Atkeson and Ohanian, 2001).

---

17 See also Caballero, 1994, and Schaller, 2006, for studies of the relationship between investment and the cost of capital in the long-run.

18 Euler equation estimation is also an option. In a pair of papers, Oliner, Rudebusch, and Sichel (1995, 1996) study the effectiveness and internal consistency of Euler equation models for investment. They obtain parameter estimates that are somewhat difficult to reconcile with economic theory, find that supposedly “structural” parameters are unstable over time, and that the models have little forecasting power. There are also legitimate concerns about the validity and relevance of the instruments used in these models (especially when extended to asset-level data).

I attempted to estimate an Euler equation using the panel of data on asset-level investment. Between two-stage least squares, LIML, and GMM methods, there were substantial differences in results indicating that the model is misspecified or there are problems with the instruments. I also replicated some of the troubling results found by Oliner, Rudebusch, and Sichel. Furthermore, Euler equations are clearly difficult to estimate even with quarterly data, and I only have annual data on asset-level investment.

The Euler-equation method is also more restrictive than the methods used in this paper because it is difficult or impossible to incorporate all of the controls that I consider. Euler equations are useful for estimating specific parameters in tightly theorized models. The regressions used here are meant to test a broader range of possible explanations for the correlation between average duration and the term spread and to measure the explanatory...
What the regression of average duration on the term spread is useful for is testing whether the term spread drives investment in the direction that we would expect and how much explanatory power the term spread has for the cross-section of investment. A high $R^2$ in a regression of average duration on the term spread is evidence that the cross-section of interest rates is an important determinant of the cross-sectional distribution of investment.

1.5 Alternative explanations

The working hypothesis is that the negative relationship between average duration and the term spread is a simple cost-of-capital effect. The model in the previous section shows that there are a number of other factors that could cause us to find the correlation we observe in figure 1.1. This section considers a range of possible alternative explanations. I find that the correlation is driven to some extent by these other factors, but that the cost of capital retains a substantial amount of explanatory power and is generally the most powerful variable for explaining average duration.

1.5.1 Correlations by asset and industry

One possible explanation for the correlation between the term spread and $\bar{D}$ is that demand for the products of different industries depends on the term spread. For example, suppose when the term spread is high consumers demand fewer durable goods (the term spread tends to be countercyclical, as are durables purchases; Yogo, 2006). If durable goods industries tend to use relatively more long-duration capital than services providers (for example, a car manufacturer may use more heavy machinery than a barber shop), then we would see investment shift towards low-duration assets. In the terms of the model, this is a story about the covariance term $\sum_i D (\delta_i) \left[ \log (B_{i0}) - N^{-1} \sum_i \log (B_{i0}) \right]$. The correlation between $\bar{D}$ and the term spread then would be driven by consumer demand (and hence the variation in the marginal product across assets) instead of the cost of capital. We can power of the term spread. I therefore leave the Euler equation analysis of this panel dataset for future work.
test this hypothesis by decomposing $\bar{D}$ into components driven by within-industry reallocation and changes in the composition of investment across industries.

As noted above, the BEA not only reports data on aggregate investment; it also gives levels of investment at the asset×industry level. Denoting the first difference of $\bar{D}_t$ as $\Delta \bar{D}_t$, we can decompose $\Delta \bar{D}_t$ following van Ark and Inklaar (2006) using the industry-level data as

$$\Delta \bar{D}_t = \sum_j \left[ \frac{1}{2} \left( \frac{I_{jt}}{I_t} - \frac{I_{jt-1}}{I_{t-1}} \right) (\bar{D}_{jt} + \bar{D}_{jt-1}) \right] + \sum_j \left[ \frac{1}{2} \left( \frac{I_{jt}}{I_t} + \frac{I_{jt-1}}{I_{t-1}} \right) (\bar{D}_{jt} - \bar{D}_{jt-1}) \right]$$

(1.10)

where $\bar{D}_{jt} = \sum_i \frac{I_{ijt}}{I_{jt}} D_i$ is the average duration of industry $j$ at time $t$. The first part of equation (1.10) can be thought of as a cross-industry reallocation effect. It sums the changes in the industry investment shares weighting by their average depreciation rates at dates $t$ and $t - 1$. The second term is the within-industry reallocation term. It represents the effects of industries changing their mix of investment among different assets. I refer to the two effects as the between and within-industry effects, respectively.

The final three columns of table 1.2 report results from first-differenced regressions of $\Delta \bar{D}$ and its decomposition (1.10) on the change in the term spread. $\Delta \bar{D}$ and $\Delta TS$ are standardized to have unit variance as in the remainder of the table. The three columns report results from regressions with different dependent variables. The first column uses $\Delta \bar{D}$. The coefficient on the term spread is similar to though somewhat smaller than the coefficient in column 1. In other words, the relationship between $\bar{D}$ and the term spread is somewhat weaker in high frequency data, which is perhaps not surprising considering the effects of planning, ordering, and building lags. The coefficients in columns 7 and 8 by definition sum to the coefficient in column 6. The within-industry coefficient is twice the size of the between-industry coefficient; in other words, two thirds of the aggregate effect comes from reallocation within industries. The hypothesis that industry demand is correlated with the term spread seems to be true, but it explains only a minority of the variation in average duration over time.
To analyze how the relationship in figure 1.1 and table 1.2 differs across assets, I run a regression of each asset’s share of aggregate investment on the term spread.\textsuperscript{19} Specifically, for each asset we run the regression

$$\frac{I_{it}}{\sum_i I_{it}} = \alpha_i + \beta_i T S_t + \epsilon_{it} \quad (1.11)$$

It is straightforward to show that if $\beta_i$ is negatively related to each asset’s duration, then there will be a negative relationship between the term spread and $\bar{D}_t$. This is a way of asking whether the relationship we observe at the aggregate level is pervasive across assets, or is driven by a few outlier assets.

Figure 1.4 plots the coefficients $\beta_i$ against duration. The black boxes are for equipment, grey diamonds structures. Regression lines are included for the sample of all assets and for equipment only. The correlations between $\beta_i$ and $D_i$ are -0.42 and -0.31 for equipment only and all assets, respectively. Looking across equipment, the relationship between the composition of investment and the term spread is broadly based not driven by a few outliers.

The plot includes labels for the assets that make up the largest part of investment over the last 15 years. Numbers in parentheses represent their percentage shares over that period. Within equipment, auto purchases as a share of total investment are far more positively correlated with the term spread than any other asset, though they represent a relatively small part of aggregate investment. Software is the single largest component of investment and it is well above the best fit line. Communication equipment and computers are next in the rankings and are somewhat closer to the regression line.

Structures do not match the results for equipment very well. While the shares of structures are generally negatively related to the term spread, they are not as negative as we would think from just looking at equipment. Electric-power plants, in particular, are a large positive outlier. As noted above, the fact that building lags average over a year (a time that does not take into account the time required for planning) is likely to distort the

\textsuperscript{19} To control for long-term changes in the composition of investment I first detrend the dependent variable and the term spread using the HP filter with a smoothing parameter of 25 as above.
Figure 1.4: Coefficients from regressions of investment shares on the term spread

Note: Coefficients from regressions of each asset's share of aggregate investment on the term spread. Both variables are detrended with the HP filter with a smoothing parameter of 25. Numbers in parentheses are investment shares over 1993–2008.
1.5.2 The business cycle, volatility, and other explanations

Table 1.3 explores a number of other mechanisms that could cause the observed correlation beyond changes in demand across industries. Columns 1 and 2 control for the business cycle with the lagged detrended unemployment rate and level of output. In both cases the coefficient on the term spread is smaller but still statistically and economically significant. This is perhaps not surprising: even if the term spread does represent a true cost-of-capital effect, it is also a proxy for the business cycle. Controlling for other business cycle indicators will probably lower its coefficient. Including the current value and longer lags of unemployment and output do not change the results of the regressions.

Another obvious question is whether there is a mechanical relationship between average duration and the level of investment. Suppose a firm has equal stocks of two assets, one with a depreciation rate of 1 percent, the other 10 percent. In a maintenance phase with no net capital growth, there will be 10 times as much investment in the high depreciation as the low depreciation asset. However, in an expansion phase, assuming both assets are expanded equally, investment will shift towards being equally balanced between the two assets. If the term spread is correlated with the level of investment, it might also then be correlated with average duration. Column 3 tests that hypothesis by including detrended aggregate equipment investment. Puzzlingly, unlike the example just given, when investment is high, duration actually tends to be low. However, the coefficient on the term spread is still large and significant. The term spread thus has explanatory power beyond its indication of either the business cycle of overall level of investment. Column 4 shows that if we include all three aggregate indicators, unemployment, GDP, and investment, the coefficient on the term spread is the highest, and has the highest t-statistic, of any of the variables (implying that the marginal $R^2$ of the term spread is higher than any of the business-cycle indicators).

Abel et al. (1996), among many others, study the effects of irreversibility on investment. With irreversibility, when idiosyncratic uncertainty is high, firms may be less willing to
Table 1.3: Robustness tests

<table>
<thead>
<tr>
<th>Term</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term Spread(t-1)</td>
<td>-0.27 **</td>
<td>-0.33 ***</td>
<td>-0.65 ***</td>
<td>-0.44 ***</td>
<td>-0.41 ***</td>
<td>-0.36 ***</td>
<td>-0.37 ***</td>
</tr>
<tr>
<td></td>
<td>[0.12]</td>
<td>[0.11]</td>
<td>[0.11]</td>
<td>[0.08]</td>
<td>[0.09]</td>
<td>[0.10]</td>
<td>[0.10]</td>
</tr>
<tr>
<td>Unemployment(t-1)</td>
<td>-0.45 ***</td>
<td></td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.14]</td>
<td></td>
<td>[0.19]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP(t-1)</td>
<td>0.38 ***</td>
<td>0.40 **</td>
<td>0.37 ***</td>
<td>0.31 **</td>
<td>0.38 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.10]</td>
<td>[0.17]</td>
<td>[0.12]</td>
<td>[0.14]</td>
<td>[0.09]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investment(t)</td>
<td>-0.23 **</td>
<td>-0.29 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.11]</td>
<td>[0.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD_profits(t+1)</td>
<td></td>
<td></td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD_returns(t+1)</td>
<td></td>
<td></td>
<td>-0.31 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.08]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank tightness(t)</td>
<td></td>
<td></td>
<td></td>
<td>-0.24 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value spread(t)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.22 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.08]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>58</td>
<td>59</td>
<td>59</td>
<td>58</td>
<td>44</td>
<td>37</td>
<td>59</td>
</tr>
<tr>
<td>R2</td>
<td>0.45</td>
<td>0.45</td>
<td>0.35</td>
<td>0.52</td>
<td>0.65</td>
<td>0.62</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Note: See table 2. The dependent variable is the detrended average duration of equipment investment. The value spread is the gap between log book/market (B/M) for the top and bottom 30 percent of firms ranked by B/M, among the smaller 50 percent of firms, measured at the beginning of the year. SD_profits and SD_returns are the cross-sectional standard deviations of quarterly firm profit growth and stock returns, controlling for a time trend and 3-digit industry dummies. The unemployment rate is the national rate obtained from the BLS. GDP is real GDP from the BEA. Bank tightness is the Fed's Survey of Senior Loan Officers index (from Morgan and Low, 2006). Investment is aggregate real nonresidential equipment investment. All variables are detrended with the HP filter with a smoothing parameter of 25 (except for the value spread, for which it is 100), and standardized to have unit variance.
invest in long-duration assets. Intuitively, if it is more difficult to sell a long-duration asset (e.g., a large wind turbine) because it is more costly to disassemble than a short-duration investment, then there is option value to delaying investment which is increasing in uncertainty.\(^{20}\) Campbell et al. (2001) and Bloom (2009) find that when the volatility of returns on the aggregate stock market is high, so is idiosyncratic firm volatility. If the term spread is partially driven by aggregate volatility (a finding of Bloom, 2009, and implied by many term structure models, e.g., Longstaff and Schwartz, 1992), and volatility drives \(\bar{D}\), then we would find a spurious correlation between the term spread and \(\bar{D}\).

I use two measures of cross-sectional volatility that are also used in Bloom (2009): the period-by-period cross-sectional standard deviations of firm quarterly profit growth and stock returns, including controls for 3-digit SIC industries.\(^{21}\) Column 5 of table 1.3 reports results of a regression of \(\bar{D}\) on the volatility indexes. Both measures of volatility are positively correlated with the next year’s term spread, which is consistent with Bloom’s (2009) results. He finds that volatility shocks lead to economic contractions and reductions in the short rate. Table 1.3 shows that conditional on the term spread and the state of the business cycle, high stock return volatility (though not profit growth volatility) in the following year is associated with low duration investment. This is consistent with the hypothesis that long-duration investment involves a bigger commitment for firms than short-duration investment. That is, the hypothesis that high volatility interacts with fixed costs of adjustment to decrease investment seems to apply more strongly to long than short-term assets. Note, though, that even when controlling for volatility, the term spread remains significant and has a large coefficient.

Another alternative hypothesis is that the term spread does not reflect the cost of capital but is simply an indicator of the stance of monetary policy. When the Federal Reserve contracts the money supply, this may inhibit bank lending, as in Kashyap and Stein (2000). If banks are more likely to finance projects of a certain duration (either high or low), then the term spread might simply be correlated with movements in \(\bar{D}\) because it is correlated

\(^{20}\) House and Shapiro, 2008, discuss the relationship between real option-type effects and asset duration.

\(^{21}\) The original data was retrieved from Compustat and CRSP. I obtained the data used here from Nick Bloom’s website.
with bank lending standards. One way to test this hypothesis is to try to directly measure bank lending standards. The Federal Reserve has administered a Survey of Senior Loan Officers since 1967 (with a gap between 1983 and 1989) that asks banks about the level of their lending standards. Column 6 includes the tightness index from this survey in the regression. The coefficient on the term spread remains significant. When bank lending standards are relatively tight (a high value of the index), average duration is low. This is perhaps surprising, since banks are usually thought of as financing short-duration projects, while firms go to credit markets for longer-term financing. One possible explanation is that lending standards tend to be high when other factors are driving firms towards short-duration investment. In particular, standards might be high in times of high uncertainty.

The appendix includes further robustness tests. When all of the controls are included simultaneously, the term spread is the only significant variable and it has more explanatory power than any of the other variables individually.

Lettau and Wachter (2007) argue that the differences in returns between high and low book/market (B/M) stocks can be explained by differences in the duration of their cash flows (see also Hansen, Heaton, and Li, 2008). A high value spread is associated with a high valuation for growth stocks, or long-duration assets, which implies investment in long-duration assets should be high. Since stock prices represent claims on capital, whereas Treasury bonds are claims on currency, we might expect that the value spread would have more predictive power than the term spread. Column 7 of table 1.3 reports the results of a regression including the value spread. I measure the value spread here as the ratio of the book to market ratios for the top and bottom third of stocks sorted by book to market (as reported on Kenneth French’s website). The coefficient is significantly negative: the opposite of what the duration theory of the value spread would predict. One possible explanation for this result is that firms with growth stocks tend to have lower-durational assets—e.g., technology firms—so when their values are high average duration

---

22 I obtain data from Lown and Morgan, 2006.

23 Specifically, French reports value spreads for small and large stocks, split at the median of market capitalization. I average these two value spreads. Furthermore, I detrend the value spread using the HP filter with a smoothing parameter of 100.
falls. To many readers, that may have been the obvious result all along. Nevertheless, it runs against Lettau and Wachter’s theory.

1.6 Consumer durables

If the term spread truly represents a cost of capital effect then we would expect household purchases of durable goods to respond to it in a manner similar to nonresidential investment. Households face some of the same choices as firms when deciding what types of durable goods to purchase. In particular, long-lasting durable goods may have financing arrangements with longer terms than those of shorter duration assets.24

Denoting the duration of durable good of type \(i\) as \(C_i\) and purchases as \(P_{it}\), I define the average duration of consumer durables purchases as

\[
\bar{C}_t \equiv \frac{\sum_i C_i P_{it}}{\sum_i P_{it}}
\]  

(1.12)

Table 1.4 lists the assets available from the BEA, along with their depreciation rates and durations. The two assets with the lowest depreciation rates are luggage and furniture at 13 percent. Computer software and motor vehicle parts have the highest rates at 76 and 90 percent, respectively. The assets are mostly clustered in a small range of depreciation rates, though: three fourths have depreciation rates between 16 and 25 percent.

Figure 1.5 plots HP-detrended \(\bar{C}_t\) against the detrended term spread. As in figure 1.1, the axis for \(\bar{C}_t\) is reversed so that a negative correlation in the data is an easier-to-read positive correlation in the figure. For most of the sample, there is a strong negative correlation, just as we observe for nonresidential investment. In a regression similar to those in table 1.2, consumer durables on the lagged term spread, the coefficient is -0.31 with a p-value of 0.008. There is thus a significant relationship over the full sample, though the correlation is somewhat weaker than what we observe for nonresidential investment. The correlation is clearest between 1965 and 1991. For nonresidential investment the correlation is more consistent over time, which explains why the QLR test in section 1.3 indicated a break point

---

24 Attanasio, Goldberg, and Kyriazidou (2008) show that auto loan terms tend to be between three and five years, while home loans may be as long as 30 years.
Table 1.4: Consumer durables, depreciation rates, and durations

<table>
<thead>
<tr>
<th>Depreciation rate (percent)</th>
<th>Duration (years)</th>
<th>Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>3.27</td>
<td>Autos</td>
</tr>
<tr>
<td>0.25</td>
<td>3.70</td>
<td>Light trucks</td>
</tr>
<tr>
<td>0.90</td>
<td>1.11</td>
<td>Motor vehicle parts &amp; accessories</td>
</tr>
<tr>
<td>0.13</td>
<td>6.63</td>
<td>Furniture</td>
</tr>
<tr>
<td>0.18</td>
<td>4.90</td>
<td>Clocks, lamps, lighting fix &amp; other</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Carpets and other floor coverings</td>
</tr>
<tr>
<td>0.18</td>
<td>4.89</td>
<td>Window coverings</td>
</tr>
<tr>
<td>0.16</td>
<td>5.37</td>
<td>Household appliances</td>
</tr>
<tr>
<td>0.18</td>
<td>4.90</td>
<td>Glassware, tableware, &amp; household uten</td>
</tr>
<tr>
<td>0.18</td>
<td>4.90</td>
<td>Tools &amp; equipment for house &amp; garden</td>
</tr>
<tr>
<td>0.20</td>
<td>4.45</td>
<td>Video &amp; audio equipment</td>
</tr>
<tr>
<td>0.18</td>
<td>4.92</td>
<td>Photographic equipment</td>
</tr>
<tr>
<td>0.44</td>
<td>2.21</td>
<td>Personal computers and peripheral equip</td>
</tr>
<tr>
<td>0.76</td>
<td>1.31</td>
<td>Computer software &amp; accessories</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Cales, typewrtrs, &amp; oth info proc equip</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Sporting equip, supplies, guns, &amp; ammo</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Motorcycles</td>
</tr>
<tr>
<td>0.18</td>
<td>4.92</td>
<td>Bicycles &amp; accessories</td>
</tr>
<tr>
<td>0.18</td>
<td>4.90</td>
<td>Pleasure boats</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Pleasure aircraft</td>
</tr>
<tr>
<td>0.26</td>
<td>3.52</td>
<td>Other recreational vehicles</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Recreational books</td>
</tr>
<tr>
<td>0.20</td>
<td>4.46</td>
<td>Musical instruments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other durable goods</td>
</tr>
<tr>
<td>0.16</td>
<td>5.36</td>
<td>Jewelry &amp; watches</td>
</tr>
<tr>
<td>0.32</td>
<td>2.95</td>
<td>Therapeutic appliances &amp; equip</td>
</tr>
<tr>
<td>0.18</td>
<td>4.91</td>
<td>Educational books</td>
</tr>
<tr>
<td>0.13</td>
<td>6.63</td>
<td>Luggage &amp; similar personal items</td>
</tr>
<tr>
<td>0.18</td>
<td>4.88</td>
<td>Telephone &amp; facsimile equipment</td>
</tr>
</tbody>
</table>

Note: Depreciation rates are obtained from the BEA. Duration is measured as $1.03/(0.03+\delta)$. 

28
only in the very beginning of the sample.

The relationship between $\bar{C}_t$ and the term spread seems to abruptly break down after 1991. If we run a QLR test as before, we can reject the hypothesis of no break at the 1 percent level. The F-statistic is maximized in 1991, only one year different from the local maximum that is obtained in the F-statistic for nonresidential investment. The fact that these two break tests are maximized around the same time suggests that the breakdown in the consumer durables plot is not due to a factor that is specific to consumers.

One possible consumer-specific explanation is that there was some sort of change in consumer credit markets around 1991. Perhaps easier access to credit cards made consumers less dependent on long-term financing for some durables purchases, which made them less sensitive to long-term credit conditions. The Flow of Funds accounts measure total credit card balances and household net worth. The ratio of consumer credit debt to net worth rises from 1.0 to 3.8 percent between 1945 and 1965, but then stays flat subsequently. While there were certainly changes in consumer credit markets following 1965, the total quantity of credit has remained in this sense stable.

1.7 The firm-level mechanism

To augment the analysis above, this section studies two aspects of firm-level investment. I begin by asking whether firms that invest in long-duration assets also tend to sell long-term debt. Next, I look at whether industries with larger cash holdings are more sensitive to the term spread. The data answer both these questions in the affirmative, but when we include a full set of industry and year dummies the results go away, possibly because of insufficient statistical power. The first result indicates that when firms go to debt markets, the interest rate that they face depends on the duration of the investment they plan on undertaking. The second result shows that firms that are more likely to have to go to debt markets seem to vary the composition of their investment more strongly in response to interest rates.

25 Note, again, that the local maximum for nonresidential investment is not statistically significant.
Figure 1.5: Average duration of consumer durable purchases versus the lagged term spread

Note: Average duration of consumer durables is defined analogously to that for durable equipment. Both lines represent HP-detrended values. The axis for average duration is reversed. Grey bars indicate NBER-dated recessions.
The link between the term spread and the cost of capital will be most clear to managers if investment in long-duration assets is financed with long-duration debt. If, for example, firms always borrow at the same maturity and simply roll over their debt, then they might only pay attention to the interest rate for the maturity at which they borrow, instead of the full term structure.

Baker, Greenwood, and Wurgler (BGW, 2003) find that firms time the debt market when they sell bonds. In particular, when the term spread is high firms sell short-term debt. BGW argue that firms do this because when the term spread is high, the prices of short-term bonds are expected to fall in the future. Firms are selling expensive or overpriced debt, which BGW claim represents arbitrage.

But if it is true that firms try to match the maturity of their debt to the maturity of their investments, then the results in the previous sections could explain the BGW result. Matching the maturity of debt to assets reduces potential deadweight losses from bankruptcy (see, e.g., Stohs and Mauer, 1996). Graham and Harvey (2002) report evidence from surveys that maturity matching is the single most important determinant of debt maturity choice.

Section 1.3 showed that when short-term yields are low, firms invest in short-duration assets. If the maturity-matching hypothesis is correct then those firms should also sell short-duration debt. That matches the Baker et al. result: low short yields are associated with short-duration investment, which is associated with sales of short-duration debt. BGW claim that firms are arbitraging debt markets; I claim they are managing risk through maturity-matching. The key to completing the argument is showing that firms actually do try to match the duration of their debt to that of their assets. In this section I provide evidence in support of this proposition.

---

26 See also Barclay and Smith, 1995, and Guedes and Opler, 1996, among many others.

27 Baker at al. tried measuring the duration of assets with a similar strategy to mine. However, rather than using industry depreciation reported by the BEA, they used the amount of depreciation reported to the IRS by individual firms. Presumably this data was substantially more noisy than the BEA data, which caused them to find inconclusive results. Moreover, accounting depreciation is in general not the same as economic depreciation. The majority of firms use straight line depreciation, rather than the declining balance method found to better match the resale value of assets (Hulten and Wykoff, 1981).
I obtain data from two sources. Data on capital stocks come from the BEA’s detailed fixed asset tables as before.\textsuperscript{28} I continue to measure average duration within industry \(j\) as 

\[
\bar{D}_{jt} = \frac{\sum_i D_{ijt}}{\sum_i I_{ijt}} 
\]  

(1.13)

where \(i\) indexes assets, \(j\) indexes industries, and \(I\) is investment.

I obtain data on corporate debt from Compustat. Following Baker et al. (2003) and Greenwood et al. (2009), the long-term share in a given industry and year is the sum of all outstanding long-term debt reported by firms in that industry divided by all long and short-term debt.\textsuperscript{29} I estimate issuance of long-term debt as the change in the level of long-term debt, and short-term issuance as simply the level of short-term debt (since short-term debt has, by definition, a maturity of less than one year). The long-term issuance share is then just the ratio of long-term issuance to total issuance.\textsuperscript{30}

An important issue here is that Compustat only covers publicly traded firms, whereas the BEA’s fixed-asset data covers all firms. To the extent that private firms have limited access to long-term credit markets, this will bias the level of the long-term share upwards.\textsuperscript{31} It is less clear, though, that selection should cause us to spuriously find that high-depreciation industries have a low long-term share. The selection would need to occur in such a way that firms in high-depreciation industries are more likely to go public but are no more likely to have access to long-term credit markets.

Table 1.5 reports regressions of the long-term level and issuance shares on industry average duration. The first two columns use the level share, the second two the issue share. Columns 2 and 4 include industry fixed effects. Each regression includes year dummies and the standard errors are corrected for clustering within industries. Columns 1 and 3

\textsuperscript{28} The BEA has its own industry classification which is slightly different from NAICS. I use industries that roughly correspond to a 2-digit NAICS classification, but I combine some industries to ensure that I have sufficient firm observations to get good financial data. I end up with 22 industries.

\textsuperscript{29} I measure long term debt as the sum of items 9 (long term borrowing) and 44 (long term debt about to retire), and short term debt as item 9 plus item 34 (current liabilities) minus long term debt.

\textsuperscript{30} I drop firm observations if the level of long term debt drops by more than one half (as this amount of retirement is implausible). Industry-year observations are dropped if they have a negative level of long term debt issuance.

\textsuperscript{31} For example, Titman and Wessels, 1988, find that small firms are less likely to use long-term debt financing.
show that there is a significant negative relationship between long-term debt levels and issuance and the depreciation rate of assets in an industry. However, columns 2 and 4 show that when we include industry dummies the effect goes away. That is, there is not evidence that when an industry shifts towards higher-depreciation assets, it also changes the composition of its debt.

One reason I do not find within industry effects in table 1.5 could be that the data is not sufficiently precise. The median number of firms that is used to create the industry×year observations is only 148, and the 25th percentile is 30. Moreover, the measure of the duration of debt is extremely rough. Firms could easily be changing the maturity of their long-term issues, rather than substituting between long and short-term issues.32

1.7.2 Cash reserves and investment

If firms match the maturities of assets and debt, as suggested by table 1.5, then when they borrow to finance investment, shifts in the term spread directly feed into their cost of capital, and presumably their investment decisions. However, when firms finance investment internally, we might think they simply use a rule-of-thumb method for the cost of capital, ignoring the term structure of interest rates (e.g. Graham and Harvey, 2002).

I study this question by looking at how investment differs across industries with different cash holdings. I study the following regression

\[
\bar{D}_{j,t} = \alpha_j + \beta_1 TS_{t-1} + \beta_2 CH_{j,t} + \beta_3 TS_{t-1} \times CH_{j,t} + \epsilon_{j,t} \tag{1.14}
\]

where, as before, \(TS_t\) is the term spread and \(CH_{j,t}\) is a measure of cash holdings in industry \(j\) at time \(t\). The coefficient \(\beta_3\) measures the effect of cash holdings on the response of an industry’s average duration of investment to the term spread. Under the hypothesis that firms that can finance investment internally respond less to the term spread, we should observe a positive value for \(\beta_3\).

I use two measures of an industry’s ability to finance investment internally: its total

32 While there is data with more detail on the duration of corporate debt, it does not have a long enough time series to be useful for finding the aggregate effects that I am looking for here.
### Table 1.5: Regressions of the long-term corporate level and issues shares

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Levels</td>
<td>Levels</td>
<td>Issues</td>
<td>Issues</td>
</tr>
<tr>
<td>Duration</td>
<td>0.013 ***</td>
<td>-0.008</td>
<td>0.021 **</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>[0.005]</td>
<td>[0.006]</td>
<td>[0.11]</td>
<td>[0.012]</td>
</tr>
<tr>
<td>Fixed Effects?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1,040</td>
<td>1,040</td>
<td>1,023</td>
<td>1,023</td>
</tr>
</tbody>
</table>

Note: The long-term level share is the share of total corporate debt accounted for by long term (>1 year maturity) debt. The issues share is the share of issues accounted for by long term debt. Duration is the average duration rate of the industry’s capital stock. All regressions include year dummies. Standard errors reported in brackets are corrected for clustering within industries. Annual data for 22 industries, 1950–2008.
current cash holdings and its cash flows (income before extraordinary items), both scaled by current property, plants, and equipment (PPE). One of the measures for \( CH_{jt} \) is a flow, while the other is a stock. If industries differ in the amount of cash that they prefer to hold at any given time, then cash flows might be more relevant for their ability to finance investment internally. On the other hand, cash reserves could be saved precisely for use in future investment, and hence represent a source of funds for investment. I test both possibilities.

As above, I obtain data from Compustat and detrend \( \bar{D}_{jt} \) and \( TS_t \) with the HP filter. I subtract industry means from the measure of cash holdings, \( CH_{jt} \), so that the coefficient on the interaction term, \( \beta_3 \), represents the change in the response of \( \bar{D}_{jt} \) to the term spread depending on the difference between the industry’s current cash holdings and the sample average for that industry.

Columns 1 and 2 of table 1.6 report estimates of equation (1.14). Column 1 shows that cash flows seem to have no significant relationship with the response of \( \bar{D}_{jt} \) to the term spread. In column 2 we see, though, that cash holdings have a strong effect. For an industry with its sample average of level of cash holdings, the response of \( \bar{D} \) falls by 0.15 standard deviations for every one standard deviation increase in the term spread. For an industry with cash holdings one standard deviation above their mean, this response falls to only 0.08 standard deviations. This result fits with the hypothesis that when firms are forced to finance investment in credit markets they are more sensitive to the cost of capital.

Columns 3 and 4 are the same as columns 1 and 2 except that I generate a variable \( \hat{CH}_{jt} \), by regressing \( CH_{jt} \) on a set of year and industry dummies. \( \hat{CH}_{jt} \) thus measures the deviation of an industry’s cash holdings relative to both its sample average and the average for the year. This controls for general trends in cash holdings over time. The coefficients on the interaction terms are no longer significant. As before, this may simply be a power issue. Note that the standard error on the interaction term is substantially larger in column

---

33 The results are unchanged if we use free cash flow instead of income.

34 See Opler et al., 1999, for an analysis of the determinants of cash holdings and their use for investment, and the large literature following Fazzari, Hubbard, and Petersen, 1988, on the relationship between investment and cash flows.
Table 1.6: Interaction of the term spread with industry cash holdings

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term spread (t-1)</td>
<td>-0.15 ***</td>
<td>-0.15 ***</td>
<td>-0.15 ***</td>
<td>-0.15 ***</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
<td>[0.02]</td>
<td>[0.03]</td>
<td>[0.02]</td>
</tr>
<tr>
<td>Cash flows (t-1)</td>
<td>0.03</td>
<td>0.08 **</td>
<td>0.03</td>
<td>0.08 **</td>
</tr>
<tr>
<td></td>
<td>[0.02]</td>
<td>[0.03]</td>
<td>[0.02]</td>
<td>[0.03]</td>
</tr>
<tr>
<td>Cash holdings (t-1)</td>
<td>0.03 *</td>
<td>0.07</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>[0.02]</td>
<td>[0.06]</td>
<td>[0.02]</td>
<td>[0.06]</td>
</tr>
<tr>
<td>TS(t-1)xCF(t-1)</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
<td>[0.02]</td>
<td>[0.03]</td>
<td>[0.02]</td>
</tr>
<tr>
<td>TS(t-1)xCH(t-1)</td>
<td>0.07 **</td>
<td>-0.03</td>
<td>0.07 **</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
<td>[0.06]</td>
<td>[0.03]</td>
<td>[0.06]</td>
</tr>
<tr>
<td>Industry dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year dummies</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R2</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>N</td>
<td>770</td>
<td>806</td>
<td>770</td>
<td>806</td>
</tr>
</tbody>
</table>

Note: Regressions of the average duration of investment by industry on the term spread interacted with measures of cash holdings. Cash data is obtained from Compustat. TS is the term spread. CF is cash flows. CH is cash holdings. All variables are measured at the end of the year. All regressions include industry dummies. Standard errors are clustered by industry.
4 than column 2. We in fact cannot reject that the two coefficients are equal.

1.8 Conclusion

This paper shows that there is a strong relationship in aggregate data between investment and the cost of capital. I find that the term spread can explain a third of the variation of the cross section of investment. While this relationship does not quantify the magnitude of internal adjustment costs facing firms, it does show that the cost of capital is a major factor driving the variation in the type of investment that firms do. The composition of investment changes meaningfully over the business cycle, and a substantial portion of these changes can be explained by the term spread alone.

The results are robust to including a variety of controls, including multiple indicators of the state of the business cycle. None of the controls eliminate the coefficient on the term spread. Moreover, when we include all of the controls at once, the term spread is the only variable that remains significant. Of all of the variables I study, the term spread is the most robust and powerful explanator of the distribution of investment. The dimension of investment studied here has not been examined before. The results extend also to consumer durables purchases: households tend to buy less-durable durables when the yield curve is steep.

Cochrane (2011) gives an extensive review of the literature on return predictability and variation in the price of risk, arguing that shifts in discount rates are part of "the central organizing question of asset pricing research." As Treasury bonds have (nominally) riskless payoffs, shifts in the term spread are purely driven by discount rates. The finding that the term spread determines the composition of investment is thus connected to Cochrane’s organizing question by showing its relevance for the aggregate economy, and not just financial markets.

There are many other cross-sectional sources of variation in the cost of capital beyond differences in asset lives. Tax policies, e.g. R&D tax credits and bonus depreciation, distort the cost of capital, as will changes in the price of risk. The finding here that shifts in the term structure of interest rates affect the composition suggests that tax policy can succeed
distorting investment choices. Similarly, to the extent that the price of risk varies over time, an interesting question is whether a high price of risk causes businesses to shift relatively towards low-risk/low-reward projects.
2. A MODEL OF TIME-VARYING RISK PREMIA WITH HABITS AND PRODUCTION

2.1 Introduction

Stock prices are more volatile than can be explained by movements in expected dividends. Moreover, excess returns on the aggregate stock market are predictable over time. The two phenomena are connected: changes in the discount rates applied to future dividends can induce excess volatility in asset prices. This paper develops a new preference specification with time-varying risk aversion that generates realistically predictable and volatile stock returns. When combined with a production framework, the model can match the short and long-run volatilities of output, consumption, and investment growth and at the same time generate a high and volatile price of risk. Simulated stock-return forecasting regressions are consistent with empirical results, and the model also delivers a new method for forecasting stock returns. The structural estimate of risk aversion has an $R^2$ for 5-year stock returns in the post-war period of over 50 percent.

The standard model of time-varying risk aversion is the habit specification of Campbell and Cochrane (1999). In their model, when an agent’s consumption falls close to her habit, her risk aversion rises. Using aggregate consumption data, they find that their implied risk aversion measure can explain a large proportion of the movements in the price-dividend ratio on the stock market. Campbell and Cochrane study an endowment economy, though, so they never test whether their utility function generates a realistic consumption process in equilibrium. In fact, Lettau and Uhlig (2000) and Rudebusch and Swanson (2008) find that Campbell–Cochrane preferences imply that consumers smooth

---

1 A partial selection of other early papers studying habit formation is Abel (1990), Constantinides (1990), Boldrin, Christiano, and Fisher (2001) and Jermann (1998).

For other papers that study return predictability in a production setting, see Gourio (2010), Campanale, Castro, and Clementi (2010), and Guvenen (2009), though note that the latter two papers do not match the degree of predictability observed in the data.
consumption growth extremely and implausibly strongly following technology shocks in
standard general-equilibrium models.

This paper embeds the intuition behind Campbell and Cochrane (1999)—that persis-
tent external habits can induce time-varying risk aversion—into the framework devel-
Zin specification allows us to model risk aversion and intertemporal substitution separ-
ately, while the Campbell–Cochrane intuition motivates time-variation in risk aversion.
In particular, consumers are modeled as having a time-varying external habit, which is
a benchmark to which they compare their own lifetime utility. When lifetime utility is
farther above the benchmark, risk aversion over proportional shocks to future welfare is
lower. By explicitly separating variation in risk aversion from intertemporal substitution,
the Epstein–Zin framework eliminates the problems that arise when standard Campbell–
Cochrane preferences are used in a production setting. I refer to the new preference speci-
fication as the EZ-habit model for its combination of these two frameworks.\(^2\)

The simple real business cycle (RBC) model with fixed labor supply provides a trans-
parent laboratory in which to study the effects of time-variation in risk aversion on the
macroeconomy in general equilibrium. I find that the dynamics of real variables and real
interest rates under the EZ-habit specification are highly similar to a model with Epstein–
Zin utility and constant relative risk aversion.\(^3\) The model can match both the short and
long-run variances of output, investment, and consumption.

Since consumption and wealth are cointegrated under balanced growth, their long-run
variances must be the same. But empirically, the short-run (quarterly) variance of con-
sumption growth is much smaller than the variance of changes in wealth. To match both
the long and short-run moments, a model must have either mean-reversion in wealth or
strong persistence in consumption growth. A number of recent asset-pricing papers (e.g.
Bansal and Yaron, 2004, Kaltenbrunner and Lochstoer, 2010) have gone the route of choos-

---

\(^2\) Melino and Yang (2003) study a utility specification that is highly similar to mine in reduced form. How-
ever, they do not discuss the inclusion of a habit, and they do not insert the preferences into a production
setting.

\(^3\) For other recent studies of asset pricing in production economies, see Danthine, Donaldson and Mehra
ing very strong persistence for consumption growth. In Kaltenbrunner and Lochstoer’s (2010) analysis of asset prices in the RBC model, for example, innovations to the permanent component of consumption have a standard deviation of 8 percent per year, which is at odds with the data. The EZ-habit model, on the other hand, implies that consumption is roughly a random walk—the short and long-run variances are nearly equal—but wealth is mean-reverting: declines in risk aversion raise current asset prices and lower expected returns. Whereas other papers in the production-based asset-pricing literature do not check the fit of their models to the long-run variance of consumption and output, I show that the EZ-habit model can match this moment along with the short-run variances.

In addition to matching macro moments, the EZ-habit model improves the fit of the RBC model to financial moments. Previous habit-based models designed to generate high or volatile risk premia tended to have implausibly volatile interest rates, a flaw not found here.\(^4\) The reasonable behavior of interest rates is an important innovation of this paper; the EZ-habit model is able to have stable interest rates but still generate substantial asset price volatility because it has variation in discount rates on risky assets that is driven by variation in risk aversion.\(^5\) Movements in discount rates imply that asset returns should be predictable, and extensive tests show that the degree of predictability in the model is similar to what is observed in the data. Variation in risk aversion not only raises the volatility of asset returns, I find that it also makes the equity premium roughly 1/3 larger on average than it would be otherwise. Countercyclical movements in risk aversion thus increase both the quantity and price of risk in financial markets: good times seem even better and bad times worse.

There are numerous empirical methods of forecasting stock returns, but the majority of them are not based on equilibrium theories. For example, regressions of stock returns on price-dividend ratios are motivated simply by an identity that links the price-dividend ratio to future returns and dividend growth. Under the EZ-habit model, though, it turns


out to be possible to directly measure risk aversion. As is standard in the habit literature, I assume that positive innovations to household welfare reduce risk aversion. So if we can measure welfare, we can also measure risk aversion. Under Epstein–Zin preferences with a constant elasticity of intertemporal substitution, welfare is a function of current household wealth and consumption. And this result holds generally; it is not dependent on the RBC model I analyze. Using data on consumption and wealth, I construct an empirical estimate of risk aversion and show that it is a strong forecaster of aggregate stock returns: it outperforms the price-dividend ratio, Lettau and Ludvigson’s (2001) measure of the consumption-wealth ratio, and Campbell and Cochrane’s (1999) excess consumption ratio. This result differentiates my paper from models of time-varying disaster risk because it does not rely on an unobservable latent process to drive risk premia.⁶

The model also can match forecasting results for consumption growth. Lettau and Ludvigson (2001) find little ability to forecast consumption growth using their measure of the consumption-wealth ratio. Campbell and Shiller (1988) obtain similar results for the stock market. As in the empirical data, it is essentially impossible to forecast consumption growth in the EZ-habit model using the consumption-wealth ratio, but forecasts of risk premia are highly effective.

An alternative way to forecast consumption growth is with interest rates. Hall (1988) and Campbell and Mankiw (1989), in trying to estimate the elasticity of intertemporal substitution (EIS), essentially ask whether consumption growth can be forecasted with interest rates. They find little forecasting power, suggesting the representative household has a small or even zero EIS. In this paper, the EIS is set to 1.5, but I still replicate the regression results from Hall (1988) and Campbell and Mankiw (1989). The EZ-habit model explains the failure of those regressions through a time-varying precautionary-saving effect. When risk aversion is high, households want to save more to protect themselves against future shocks, which drives interest rates downward. This effect biases standard Euler-equation estimation based on models with constant relative risk aversion.

After testing the model’s fit to macro and asset pricing moments and the predictions for

⁶ See Gourio (2010), and Wachter (2010), for recent models with time-varying disaster risk.
the EIS regressions and return forecasting, I consider two extensions to the model. First, I examine the effect of time-varying risk aversion on labor supply. Following positive technology shocks, risk aversion falls, raising consumption (through a decline in precautionary saving demand). This effect also lowers the response of labor supply to technology shocks. Intuitively, intratemporal optimization means that when households are willing to spend more money to raise consumption, they are also willing to sacrifice in terms of opportunity costs to raise leisure. Endogenous labor supply has little effect on risk premia in the economy, though. The reason is simply that under Epstein–Zin preferences with a high elasticity of intertemporal substitution, the volatility of the stochastic discount factor is driven mainly by the permanent component of consumption; so even if households smooth consumption growth by varying labor supply, the total amount of risk in the economy is essentially unchanged.

The second extension is a log-linearization of the model using methods similar to Campbell (1994) and Lettau (2003). Unlike standard perturbation methods, the log-linearization used here does not impose certainty equivalence, so we can obtain expressions that take into account potentially time-varying risk premia even in the first-order approximation. I am able to derive explicit expressions for the Sharpe ratio in the model as a function of current risk aversion and the underlying parameters of the model and find that the results are highly similar to those from accurate numerical solutions. Much of the previous production-based asset-pricing literature has focused on simulations to study the implications of various models, so this paper introduces an important methodological contribution in extending and simplifying the analytic results of Campbell (1994) and Lettau (2003). Further, in the case where risk aversion is constant, I give an analytic characterization of how endogenous consumption smoothing generates long-run risks in a production setting (Bansal and Yaron, 2004; Kaltenbrunner and Lochstoer, 2010). The log-linearization thus provides an analytic explanation for results that were previously supported only with simulation-based evidence.

The log-linear solution returns a stochastic discount factor (SDF) that takes on the essentially affine form that is widely used in the empirical asset-pricing literature. This is possibly the first paper to derive an essentially affine SDF with a time-varying price of risk.
from a production-based model. It thus connects the standard modeling framework in macroeconomics with one of the most widely used asset-pricing specifications in empirical finance.

The paper is organized as follows. Section 2.2 discusses the preference specification and lays out the economic environment. Section 2.3 calibrates a production economy and compares its behavior to the data. Section 4 tests the empirical implications of the model for return forecasting, and section 2.5 studies extensions to the basic framework. Section 2.6 concludes.

2.2 The model

2.2.1 Household preferences

For households with a constant elasticity of intertemporal substitution (EIS), Epstein–Zin (1989) utility can be expressed as

\[
V_t = \left\{ (1 - \exp(-\beta)) C_t^{1-\rho} + \exp(-\beta) \left[ G_t^{-1} (E_t [G_t (V_{t+1})]) \right]^{1-\rho} \right\}^{1/(1-\rho)}
\] (2.1)

for some function $G_t$, where $C_t$ is household consumption and $E_t$ is the expectation operator conditional on information available at date $t$. The term $G_t^{-1} (E_t [G_t (V_{t+1})])$ is a certainty equivalent. When there is no uncertainty about $V_{t+1}$, $G_t^{-1} (E_t [G_t (V_{t+1})]) = V_{t+1}$.

The usual choice for $G_t$ (going back to Weil, 1989, and Epstein and Zin, 1991) is power utility,

\[
G_t^{Power} (V_{t+1}) = V_{t+1}^{1-\alpha}
\] (2.2)

Epstein and Zin (1989) show that the coefficient of relative risk aversion for a household with preferences of the form (2.1) is equal to the coefficient of relative risk aversion for $G_t$, while the EIS is equal to $1/\rho$.

---

7 The preferences can be further generalized to study alternative time aggregators, instead of the constant elasticity of substitution form.
Now consider a habit-formation utility function for $G$,

$$G_{t}^{\text{Habit}}(V_{t+1}; H_t) = (V_{t+1} - H_t)^{1-\alpha} \quad (2.3)$$

Value functions involving $G_{t}^{\text{Habit}}$ are related to those using $G_{t}^{\text{Power}}$ in the same way that usual habit specifications, e.g. Constantinides (1991), are related to time-separable power utility. Rather than caring only about the absolute level of their continuation value, $G_{t}^{\text{Habit}}$ says that households care about the spread between tomorrow’s value and a benchmark $H_t$. Since the utility function adds a habit to Epstein–Zin, I refer to it as the EZ-habit specification.\(^8\) I refer to the version of $V_t$ using $G_{t}^{\text{Power}}$ for the certainty equivalent as canonical Epstein–Zin in deference to its popularity in the literature.

The coefficient of relative risk aversion for $G_{t}^{\text{Habit}}$ is equal to $\alpha \frac{V_{t+1}}{V_{t+1} - H_t}$. As the spread between value and habit rises, the coefficient of relative risk aversion falls. Intuitively, when the continuation value falls close to its benchmark, proportional shocks to $V_{t+1}$ loom much larger than when the household has a cushion between its continuation value and $H_t$.

In principle it is possible to analyze a model with $G_{t}^{\text{Habit}}$, but it has three important drawbacks. First, if the support of the shocks to $V_{t+1}$ is sufficiently wide, there is a non-zero probability that $V_{t+1}$ will fall below $H_t$, leaving the certainty equivalent undefined.\(^9\) Second because $G_{t}^{\text{Habit}}$ is not log-linear in $V_{t+1}$, obtaining simple analytic results with it is difficult or impossible. Third, because $G_{t}^{\text{Habit}}$ is not log-linear, standard arguments for the existence of a representative agent do not apply.\(^10\)

\(^8\) Other papers, for example Rudebusch and Swanson (2010) and Yang (2008), incorporate consumption habits into Epstein–Zin preferences. That is, the $C_{t-\rho}^{1-\rho}$ term is replaced by $(C_t - X_t)^{1-\rho}$ where $X_t$ is the habit. Rudebusch and Swanson (2008) show that in general equilibrium this does not lead to a time-varying Sharpe ratio because households endogenously smooth consumption to reduce their overall risk exposure. That said, the specification in Rudebusch and Swanson (2008) is meant to generate smooth consumption growth rather than a high risk premium. In principle, there is no reason that this type of habit formation could not be added to the EZ-habits model to help generate smoother consumption (e.g. to help explain the excess smoothness puzzle of Campbell and Deaton, 1989). Dew-Becker (2011) studies preferences with both time-varying risk aversion and consumption habits in a medium-scale DSGE model.

\(^9\) This issue also arises in other habit specifications. When models are solved with standard perturbation methods, the problem is simply ignored. I use a more precise global numerical solution technique that forces me to grapple with the problem.

\(^10\) A representative agent may exist, but their preferences need not actually look like the preferences of any
For the remainder of the paper I therefore replace $G_t^{\text{Habit}}$ with the alternative

$$G_t^{TV}(V_{t+1}) = V_{t+1}^{1-\alpha_t}$$

$$\alpha_t = \frac{V_t}{V_t - H_t},$$

(2.4)

$G_t^{TV(-1)}(E_t G_t^{TV}(V_{t+1}))$ (where TV stands for time-varying) is a second-order approximation to $G_t^{\text{Habit}(-1)}(E_t G_t^{\text{Habit}}(V_{t+1}))$ around the non-stochastic version of the model.\textsuperscript{11} Moreover, the appendix shows that in the continuous-time limit (i.e. under stochastic differential utility), preferences with $G_t^{TV}$ are exactly equivalent to preferences using $G_t^{\text{Habit}}$.\textsuperscript{12} $G_t^{TV}$ is locally equivalent to $G_t^{\text{Habit}}$ in terms of risk preferences, but it solves the problems of integrability inside the certainty equivalent and the existence of a representative agent.

As in Campbell and Cochrane (1999), I assume that households take the excess value ratio, $\frac{V_t}{V_t - H_t}$, and hence the coefficient of relative risk aversion, $\alpha_t$, as external to their own decisions. The final step, then, is to specify a dynamic process for risk aversion. I assume a simple log-linear process, which we will find to be highly tractable,

$$\alpha_{t+1} = \phi \alpha_t + (1 - \phi) \bar{\alpha} + \lambda \left( \Delta v_t^A - E_t \Delta v_{t+1}^A \right)$$

(2.5)

where $v_t^A$ is the log of $V_t$ for the representative agent. Intuitively, when value unexpectedly rises, it moves away from the habit and risk aversion falls, so $\lambda < 0$. Movements in the habit, and hence risk aversion, depend on aggregate value so that they are not affected by an individual household’s decisions. The AR(1) specification for risk aversion is approximately equivalent to a specification where $\log H_t$ is a geometrically weighted

\textsuperscript{11} More precisely, the second-order approximation also assumes no growth. Adding a constant growth rate $\mu$ to $V$ would change the result to $\alpha_t = \frac{(1+\mu)V_t}{(1+\mu)V_t - H_t}$. The remainder of the analysis is identical.

\textsuperscript{12} Melino and Yang (2003) study a utility function with the same form as $G_t^{TV}$, but they take $\alpha_t$ as a latent variable and give no theoretical motivation for its variation. This paper is original for proposing inserting habits into the certainty-equivalent part of Epstein–Zin preferences to motivate movements in $\alpha_t$. 

46
moving average of past values of $v_t^A$.\textsuperscript{13}

The appendix shows how to derive the marginal rate of intertemporal substitution (the stochastic discount factor, or SDF) for the general form of Epstein–Zin preferences in (2.1). In the case of $G^TV$, we end up with the expression,

$$M_{t+1} = \frac{\partial V_t}{\partial C_t} = \frac{\partial V_t}{\partial C_t} + \frac{\partial V_t}{\partial C_t} = \exp(-\beta) \frac{V_{t+1}^{\rho - \alpha_t}}{E_t V_{t+1}^{1-\alpha_t} C_t^{\rho - \alpha_t}}$$ \hspace{1cm} (2.6)

with the only difference from the SDF under canonical Epstein–Zin preferences being the subscript on $\alpha_t$. The SDF is a critical piece of the model since its volatility determines the price of risk in the economy.\textsuperscript{14} As usual, changes in expected consumption growth or volatility will affect the SDF through their effects on $V_{t+1}$. Changes in $\alpha_{t+1}$ (or $H_{t+1}$) will also affect the SDF in the same way. Specifically, when the habit rises and households are more risk averse, they penalize consumption uncertainty more, driving $V_{t+1}$ down. High risk-aversion states thus have high Arrow–Debreu prices.

It is also straightforward to derive the standard result that

$$W_t = V_t^{1-\rho} C_t^{\rho} / (1 - \exp(-\beta))$$ \hspace{1cm} (2.7)

where $W_t$ is the equilibrium price of a claim on the household’s consumption stream, which I refer to as the aggregate wealth portfolio. This formula holds regardless of whether risk aversion varies over time. Intuitively, the market price of the consumption stream is equal to the utility value that a household places on it, $V_t$, divided by the marginal utility of consumption, $V_t^{\rho} C_t^{-\rho} (1 - \exp(-\beta))$. This leads to the familiar result from Epstein and Zin (1991),

$$M_{t+1} = \exp(-\beta)^{1-\alpha_t} \left( \frac{C_{t+1}}{C_t} \right)^{-\rho^{1-\alpha_t}} R^{\rho^{1-\alpha_t}}_{\omega,t+1} \hspace{1cm} (2.8)$$

where $R^{\omega}_{\omega,t+1}$ is the return on the wealth portfolio.

\textsuperscript{13} It is straightforward to derive the actual process that $H_t$ must follow in order for risk aversion to follow the process in (2.5).

\textsuperscript{14} Hansen and Jagannathan (1991) show that the maximum Sharpe ratio (expected excess return divided by standard deviation) attained by any asset in the economy is equal to the standard deviation of the SDF divided by its mean.
2.2.2 Discussion

The model is motivated as an extension of habit-based preferences. Rather than consumers having a habit level of consumption that they target, I assume they have a habit level of value. Since equation (2.7) shows that there is a direct link between value and wealth, we could also think of the model as saying that households have a benchmark level of wealth. The house-money effect of Thaler and Johnson (1990) has a somewhat similar intuition. They find that when subjects in lab experiments have recently gained money in betting games, they play more aggressively.\(^{15}\)

Abel (1990) interprets habits in consumption as a "keeping up with the Joneses" effect. That intuition extends to the EZ-habit model. What households try to keep up with in this model, though, is fundamentally different. For example, consider a college senior who is trying to decide between following her friends into consulting or getting a law degree. With the J.D., she knows that in the short run her consumption will be lower than that of her friends, but in the long run she will likely be better off. In a model with an external consumption habit, three years of consumption below that of her friends looks painful. But when the habit appears as a function of value, the student is comfortable giving up consumption in the short run as long as she knows she will do well compared to her friends in the long run. Since the habit appears only in the risk aggregator, an agent with EZ-habit preferences is willing to substitute consumption over time in a way that an agent with standard habit-forming preferences is not. For the same reason, the EZ-habit model is not inconsistent with the mixed evidence on the effects of classic consumption habits at the micro level (e.g. Dynan, 2000, and Ravina, 2007).

There are a number of papers that use investment choices to measure variation in risk aversion. Carroll (2002) finds that households with higher wealth tend to tilt their investment portfolios towards more risky assets. Brunnermeier and Nagel (2008), though, argue that there is little evidence that changes in wealth affect portfolio choices in household data. Rather, they find that inertia is the dominant characteristic of household portfolio

\(^{15}\) Barberis, Huang, and Santos (2001) embed the house-money effect in a full asset-pricing model. See Gertner (1993) and Post et al. (2008) for evidence on the house-money effect from game shows.
Calvet, Campbell, and Sodini (2009), after controlling for the inertia studied by Brunnermeier and Nagel, find a strong and significant relationship between innovations to wealth and the riskiness of a household’s portfolio. Furthermore, they show that weakness in the instruments for wealth shocks can cause a researcher to erroneously find that wealth does not affect risk-taking. Calvet and Sodini (2010) show that higher past income, controlling for current wealth and genetic differences in risk attitudes, is also negatively related to the share of household portfolios invested in risky assets. On net, with the notable exception of Brunnermeier and Nagel (2008), the empirical literature supports the idea that increases in wealth reduce risk aversion.

2.2.3 Production

Aggregate output is a function of the capital stock, $K_t$, and productivity $A_t$

$$Y_t = A_t^{1-\gamma}K_t^\gamma$$

(2.9)

In section 2.5.3 I add endogenous labor supply and show that it does not substantially change the dynamics of the model. The production function (2.9) can be thought of Cobb–Douglas with labor supply held fixed at unity.

The aggregate resource constraint is

$$K_{t+1} = (1 - \delta)K_t + Y_t - C_t$$

where $\delta$ is the depreciation rate of capital.

---

16 See also Tanaka, Camerer, and Nguyen (2010), who find that income, both its raw level and instrumented for with exogenous shocks, has a negative impact on loss aversion, and Guiso, Sapienza, and Zingales (2011) who find that following the financial crisis of 2008, households both reduced the risky shares of their portfolios and became more averse to gambles in survey questions.
For the benchmark calibration, productivity follows a random walk in logs,$^{17}$

$$
\log A_{t+1} = \log A_t + \mu + \sigma_a \epsilon_{t+1} \\
\epsilon_{t+1} \sim N(0, 1)
$$

The drawback of using random-walk technology is that it is difficult to generate the degree of volatility for output and investment that is observed in the data.$^{18}$ I therefore also consider a dual-shock version of the model that can match both the short and long-run variances of output,

$$
A_t = \bar{A}_t X_t \\
\log \bar{A}_{t+1} = \log \bar{A}_t + \mu + \sigma_a \epsilon_{t+1} \\
\log X_{t+1} = \phi_x \log X_t + \sigma_x \epsilon_{x,t+1} \\
\epsilon_{t+1}, \epsilon_{x,t+1} \sim i.i.d. N(0, 1)
$$

$\bar{A}_t$ here is the permanent component of output, while $X_t$ can be interpreted as a simple method of trying to capture forces that drive short-run fluctuations in output and consumption, e.g. shocks to monetary policy or energy prices. I refer to the version of the model with random-walk technology as the benchmark model, while the model with permanent and temporary technology shocks is the dual-shock model.

---

$^{17}$ An alternative is a trend-stationary process for productivity. Alvarez and Jermann (2005) argue that permanent shocks to the level of productivity (more generally, to the level of state prices) are necessary to explain asset-pricing facts. Also, in models with Epstein–Zin preferences, because the SDF depends not only on current consumption but also on the level of the value function itself, an I(1) process for productivity tends to increase the volatility of the SDF compared to models with trend-stationary productivity, which helps explain the equity premium. Kaltenbrunner and Lochstoer (2010) find that in order to match the empirical equity premium in a model with trend-stationary productivity, their model needs an implausibly small EIS (0.05). With difference-stationary productivity they are able to choose a more reasonable value (1.5).

$^{18}$ In particular, without a mean-reverting component, it is impossible for the model to replicate the result from Cochrane (1994) that the long-run variance of output is smaller than the unconditional variance. In the RBC model, output does not overshoot its long-run trend following a permanent increase in technology: it does not have a mean-reverting component to its dynamics. Because they rely only on permanent shocks in the RBC model, Kaltenbrunner and Lochstoer (2010) have to set the annual standard deviation of technology shocks to an implausibly high 8.2 percent per year to match the unconditional standard deviation of output growth.
I solve the model with projection methods, which entails fitting a polynomial approximation to the decision rule and searching for coefficients so that the equilibrium conditions hold exactly at certain specified points in the state space. The Euler equation errors in the simulations imply households misprice a claim on capital by uniformly less than 1/100th of 1 basis point (i.e. one part in one million) over the range of the state space that the simulations visit, and the median simulated error is an order of magnitude smaller.

The model is parameterized to match quarterly data. Table 2.1 lists the parameter values and the target moments. Many of the parameters, e.g. the exponent on capital in the production function, take standard values. I discuss here the parameters that are unique to this paper or do not have standard and agreed-upon values.

I set $\rho = 2/3$ as in Bansal and Yaron (2004), for an EIS of 1.5. Bansal and Yaron note that an EIS greater than 1 is necessary for increases in volatility to lower asset prices (specifically, the wealth-consumption ratio) in an endowment economy. In a production economy this result does not hold exactly (because consumption is endogenous), but it is approximately true. Similarly, an EIS greater than 1 ensures that increases in risk aversion increase the expected return on the wealth portfolio and lower its current price. Many studies attempting to estimate the EIS have obtained values much smaller than 1 (Hall, 1988; Campbell and Mankiw, 1989). An important test of the model will be whether it can match that result even though the calibrated EIS is larger than 1.

I choose the variance of permanent innovations to technology to match the long-run variance of consumption growth in the data. Since technology and consumption are cointegrated in the model, the long-run variance of consumption growth is equal to the variance of the permanent technology shocks. I estimate the empirical long-run variance (i.e. the spectral density at frequency zero) of consumption growth with a third-order univari-

---

19 See Caldara et al. (2009) for a good description of the method as applied to models with recursive utility. When solving the RBC model with Epstein–Zin preferences, they find that projection methods are orders of magnitude more accurate than the perturbation methods used in the majority of the macro literature.

20 Intuitively, an increase in risk aversion or volatility has two effects – it lowers the risk-free rate and raises the excess return on the wealth portfolio. Which of these effects dominates depends on the EIS.
Table 2.1: Calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>0.33</td>
<td>Capital income share</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9975</td>
<td>2% annual real risk-free rate</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.02</td>
<td>8% annual depreciation (BEA data)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.005</td>
<td>2% annual output growth</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.94</td>
<td>Persistence of price/dividend ratio</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.67</td>
<td>A priori (see text)</td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>0.0088</td>
<td>Long-run standard deviation of consumption growth</td>
</tr>
<tr>
<td>mean($\alpha$)</td>
<td>14</td>
<td>Mean Sharpe ratio (0.32 annualized)</td>
</tr>
<tr>
<td>stdev($\alpha$)</td>
<td>6.2</td>
<td>Stock return predictability</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>0.012</td>
<td>Variance of output growth</td>
</tr>
<tr>
<td>$\phi_x$</td>
<td>0.9</td>
<td>Variance of output growth</td>
</tr>
</tbody>
</table>

Note: Parameters used for the structural models. In Table 2, the CRRA model uses with stdev($\alpha$)=0; the benchmark EZ-habit model (column 3) sets $\sigma_x=0$. 
ate AR model (where the lag length was selected with the Bayesian information criterion) and obtain a value of 0.0088. That is, the quarterly innovations to the permanent component of consumption have a standard deviation of 0.88 percent. For the dual-shock model, I select the parameters $\sigma_x$ and $\phi_x$ to match the short-run volatility of consumption and output growth. The parameters imply that the temporary component of technology has an unconditional standard deviation of 2.7 percent.

The persistence of risk aversion, $\phi$, is set to match the empirical persistence of the price-dividend ratio for the aggregate stock market, as in Campbell and Cochrane (1999). The mean and volatility of risk aversion ($\bar{\alpha}$ and, implicitly, $\lambda$) are chosen to match the average Sharpe ratio for the stock market in the post-war sample and the degree of predictability observed using the price-dividend ratio to forecast stock returns. Mean risk aversion is 14 and the standard deviation is set to 6.2.

2.3.1 Comparisons across models

Table 2.2 reports basic moments from the three models. The first column gives the moments from the data while the second column gives results from the canonical Epstein–Zin model with constant relative risk aversion (EZ-CRRA). Columns 3 and 4 give results for the EZ-habit model under the benchmark calibration and with temporary technology shocks added.

The first row simply shows that all three models are calibrated to match the long-run variance of consumption exactly, which, under balanced growth, means they also match the long-run variances of output and investment growth. Rows 2 through 4 give the standard deviations of quarterly output, consumption, and investment growth. Rows 2 through 4 give the standard deviations of quarterly output, consumption, and investment growth. Both

---

21 By choosing a smaller value for the long-run variance than the long-run risks literature, I only make the task of matching the equity premium harder. I also make the model consistent with the point estimate of the long-run variance of consumption, rather than choosing a value in the upper end of the confidence interval. Empirically, I measure consumption as real per-capita nondurable and service consumption from the BEA.

22 Smets and Wouters (2007) estimate that the 1-quarter autocorrelation of stationary technology shocks is 0.95. On the other hand, the 1-quarter autocorrelation of detrended real GDP is 0.85. I take $\phi_x = 0.90$ as the midpoint between these two values.

23 When $\alpha_t < 0$, I still use the standard Euler equation even though the household’s optimization problem is convex. In the simulations, $\alpha_t < 0$ only 1.5 percent of the time. Treating households as if they are risk-neutral in periods when $\alpha_t < 0$ (i.e. censoring $\alpha_t$ at zero) has no discernible effect on the results.
Table 2.2: Comparison of preference specifications

<table>
<thead>
<tr>
<th>Model:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>EZ-CRRA</td>
<td>EZ-habit</td>
<td>Dual-shock</td>
<td></td>
</tr>
<tr>
<td><strong>Real moments:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Long-run SD(dC, dY, dI) (%)</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>2 StdDev(dY) (%)</td>
<td>0.99</td>
<td>0.59</td>
<td>0.59</td>
<td>1.03</td>
</tr>
<tr>
<td>3 StdDev(dC) (%)</td>
<td>0.46</td>
<td>0.28</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>4 StdDev(dI) (%)</td>
<td>2.65</td>
<td>1.11</td>
<td>0.83</td>
<td>2.37</td>
</tr>
<tr>
<td>5 corr(dC(t), Rf(t-1))</td>
<td>-0.09</td>
<td>0.28</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Financial moments:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Mean SR (annualized)</td>
<td>0.32</td>
<td>0.22</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>7 Std. dev. SR</td>
<td>0.22</td>
<td>0.12</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>8 p-value SR std. dev.</td>
<td>N/A</td>
<td>0.16</td>
<td>0.50</td>
<td>0.42</td>
</tr>
<tr>
<td>9 Mean Rw (annualized %)</td>
<td>6.78</td>
<td>1.04</td>
<td>4.17</td>
<td>4.15</td>
</tr>
<tr>
<td>10 StdDev(Rw) (annualized %)</td>
<td>21.19</td>
<td>4.71</td>
<td>13.30</td>
<td>12.98</td>
</tr>
<tr>
<td>11 Mean Rf (annualized %)</td>
<td>0.91</td>
<td>2.20</td>
<td>2.04</td>
<td>1.94</td>
</tr>
<tr>
<td>12 StdDev(Rf) (annualized %)</td>
<td>1.16</td>
<td>0.21</td>
<td>0.25</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Note: Column 2 gives results under Epstein–Zin preferences with constant relative risk aversion, column 3 uses EZ-habit preferences and random-walk technology, and column 4 EZ-habit preferences with the dual-shock specification. All models are calibrated as in table 1. All variables are measured using quarterly values. dI is investment growth, dY output growth, and dC consumption growth. Rf is the risk-free rate (measured empirically as the nominal 3-month yield minus an inflation forecast), and Rw is the annualized return on a levered consumption claim (with a leverage ratio of 2.74). The long-run SD is the square root of the spectral density at frequency zero multiplied by 2π. SR is the annualized Sharpe ratio; the standard deviation of the Sharpe ratio is measured by the standard deviation of the fitted values in forecasts of one-quarter-ahead returns in 228-quarter simulated samples divided by the unconditional standard deviation of returns. Row 7 reports the median of the standard deviations of the Sharpe ratio in the simulated samples. Row 8 reports the fraction of simulated samples in which the standard deviation of the Sharpe ratio is as large as in the data.
the EZ-CRRA and single-shock EZ-habit models have volatilities for output and investment growth that are well below the empirical values. The dual-shock model rectifies this problem, matching both the short-run and long-run variances well. Both versions of the EZ-habit model match the empirical variance of consumption growth.

Row 5 reports the correlation between the risk-free rate and the next period’s consumption growth. Empirically, the real risk-free rate is measured as the 3-month nominal interest rate minus an inflation forecast.24 In the EZ-CRRA model, the risk-free rate has a substantial amount of forecasting power for consumption growth, while in the data interest rates and consumption growth seem essentially unrelated. The two EZ-habit calibrations come much closer to matching that fact.

Rows 1 through 5 show that the EZ-habit model can capture the basic unconditional moments of output, consumption, and investment. Rows 6 through 12 of table 2.2 summarize the financial side of the model. We can begin by looking at a measure of the price of risk. The Sharpe ratio on an asset is the ratio of its expected excess return over the risk-free rate divided by its standard deviation, so it measures the risk–return tradeoff. Hansen and Jagannathan (1991) show that the maximum Sharpe ratio obtained by any asset in the economy is equal to the standard deviation of the SDF divided by its mean. Recall that all three calibrations have the same average coefficient of relative risk aversion. The Hansen–Jagannathan bound and the mean Sharpe ratio for the consumption claim are roughly 1/3 higher in the two EZ-habit models than the EZ-CRRA case. The reason for this is that the household’s value, \( V_t \), a component of the SDF (equation 2.6), is more volatile in the EZ-habit models. In all the models, a technology shock permanently raises expected consumption and hence \( V_t \). In the EZ-habit case, the coefficient of relative risk aversion also falls. Households become less averse to future uncertainty, so \( V_t \) rises even more. Countercyclical variation in risk aversion thus makes good times even better and bad times even worse, raising the volatility of the SDF. This effect allows the model to explain the equity premium (or at least the Sharpe ratio on equities) with a lower coefficient of relative risk aversion than we would need in the EZ-CRRA model.

---

24 Expected inflation is measured as a forecast of quarterly inflation based on lagged levels of inflation and the nominal risk-free interest rate.
To test whether the models can match the degree of predictability for stock returns that is observed in the data, I regress simulated quarterly excess returns on the consumption claim on its lagged price-dividend ratio. I then estimate the standard deviation of the conditional Sharpe ratio as the standard deviation of the fitted returns divided by the unconditional standard deviation of returns (i.e., assuming a constant volatility). Row 7 reports the median standard deviation from 5,000 simulations of 228 quarters of data, while row 8 reports the proportion of the simulations that have a standard deviation as high as observed empirically (0.22).25

In column 2, we can see that there is actually a nontrivial amount of implied predictability on average in the EZ-CRRA model due to small-sample overfitting, but only 16 percent of the simulations match the variability observed in the data. For the EZ-habit model, the predictability observed in the data is calibrated to be exactly the median value in the simulations.

Rows 9 and 10 report the mean and standard deviation of the excess return on a levered consumption claim in the model. For comparability to past results, I follow Abel (1999) and Gourio (2010) in assuming a leverage ratio of 2.74 (i.e., the asset pays a dividend of $C_{t-74}^{2.74}$). The two EZ-habit models are able to generate means and volatilities for returns that are far closer to the equity return observed in the data than the EZ-CRRA model can. Part of the reason for this success is that consumption growth, and hence dividend growth, is more volatile in the EZ-habit models than in the EZ-CRRA case, and part of the reason is that discount rates are more volatile. Following a positive technology shock, not only do dividends rise, but discount rates fall, thus making the returns on the wealth portfolio and the levered consumption claim more volatile.26

Rows 11 and 12 show that the means and standard deviations of the real risk-free rate in the three models are all reasonably close to the data. The volatility of interest rates is similar across all three models, and somewhat lower than in the data. The real risk-free

---

25 The fitted Sharpe ratio is measured empirically by forecasting the CRSP value-weighted aggregate excess return with the aggregate price/dividend ratio.

rate is measured empirically as the nominal 3-month Treasury yield minus a forecast of inflation. Errors in the inflation forecast will make the estimated real risk-free rate more volatile than the true real risk-free rate, which explains some of the divergence between the empirical and simulated volatilities.

A common problem in early attempts to generate a high equity premium (e.g. Constantinides, 1990; Boldrin, Christiano, and Fisher, 2001, and Jermann, 1998), is a highly volatile risk-free rate. The EZ-habit specification replaces movements in discount rates coming from the risk-free rate with movements coming from risk premia.

To summarize, table 2.2 shows that the EZ-habit model can match a broad array of features of the economy—the short and long-run variances of output growth, the relative volatilities of investment and consumption growth, and the mean and standard deviation of the Sharpe ratio on equities. The model also helps generate a larger premium on a levered consumption claim, closing roughly half the gap in the equity premium between the EZ-CRRA model and the data. Finally, the behavior of the risk-free rate is reasonably similar to the data, unlike previous general-equilibrium attempts at generating a high and volatile Sharpe ratio.

2.3.2 Predictability in the simulated model

2.3.2.1 The magnitude of return predictability

Figure 2.1 plots $R^2$s from univariate regressions of excess aggregate stock returns over various horizons on the log price-dividend ratio on the CRSP value-weighted portfolio (e.g. Campbell and Shiller, 1988, among many others), Lettau and Ludvigson’s (2001) measure of the consumption-wealth ratio, $cay$, Campbell and Cochrane’s (1999) excess consumption ratio, and an estimate of risk aversion derived from the EZ-habit model in section 2.4. For the four different variables used in the empirical sample, the $R^2$s generally rise as the sample length grows, and estimated risk aversion outperforms $cay$, excess consumption, and the price-dividend ratio.

The gray line labeled "Simulated mean" gives the mean $R^2$ from 5000 regressions of excess returns on a consumption claim on the price-dividend ratio (equivalently, the wealth-
Figure 2.1: Simulated and empirical $R^2$'s

Note: $R^2$'s from univariate regressions of stock returns on various predictors. The forecast horizon is reported in quarters. Data for $cay$ is obtained from Sydney Ludvigson's website. Price/dividend data comes from CRSP; the Campbell-Cochrane excess consumption ratio is computed using their parameter values and consumption data from the BEA. The gray lines give the mean and 95th percentiles in the simulation of the EZ-habit model.
consumption ratio) over 228-quarter spans in the benchmark simulation of the single-shock model (the same length as the empirical sample). The upper gray line gives the 95th percentile of the simulations. As in the data, the simulated $R^2$s rise as the horizon lengthens. The model compares favorably with the price-dividend and excess-consumption ratios, with the simulated mean tracking the empirical values closely (the median follows almost the same path). The empirical $R^2$s for $cay$ are at or below the 95th percentile in the simulations. The only variable that the simulations cannot match is estimated risk aversion, but raising the volatility of risk aversion in the calibration would solve this problem.

The $R^2$s generated here are substantially higher than those obtained in production models such as Campanale, Castro, and Clementi’s (2010) model of time-varying first-order risk aversion and Guvenen (2009) and De Graeve et al.’s (2010) studies of limited participation. The population $R^2$s are also essentially identical to those found by Wachter (2010) and Gourio (2010) in endowment-economy and production-based models, respectively, with time-varying disaster risk.

The top panel of table 2.3 reports the percentage of simulated samples in which the simulated $R^2$ is as high as we observe in the data for $cay$ and the price-dividend ratio (results for excess consumption are similar to the price-dividend ratio), and where a high price-dividend ratio forecasts low returns. The table reports values for horizons of one quarter and one through five years. The EZ-CRRA model matches empirical $R^2$s for $cay$ less than 5 percent of the time at horizons shorter than 16 quarters, but can match the $R^2$s for the price-dividend ratio 15 to 25 percent of the time. The habit model substantially raises the likelihood of the simulations of matching the data, by a factor of three or more at every horizon, and it never matches less than 5 percent of the time except for $cay$ at the one-quarter horizon.

As an alternative to the $R^2$, I also consider the test statistic suggested by Kiefer, Vogelsang, and Bunzel (KVB, 2000) based on Newey–West standard errors with the lag window equal to the sample size. At various horizons, I calculate the t-statistic on the coefficient in a regression of stock returns on the price-dividend ratio in the simulated samples. The bottom panel of table 2.3 repeats the analysis from the top half, but with the KVB test statistics. In every case, the habit model matches the empirical t-statistics at least five per-
Table 2.3: Proportion of simulated samples that match empirical statistics

<table>
<thead>
<tr>
<th>Forecasting $R^2$</th>
<th>EZ-CRRA</th>
<th>EZ-habit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model:</td>
<td>cay</td>
<td>P/D</td>
</tr>
<tr>
<td>Predictor:</td>
<td>cay</td>
<td>P/D</td>
</tr>
<tr>
<td>Horizon 1 (quarters)</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>8</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>12</td>
<td>0.04</td>
<td>0.16</td>
</tr>
<tr>
<td>16</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>20</td>
<td>0.10</td>
<td>0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KVB t-statistics</th>
<th>EZ-CRRA</th>
<th>EZ-habit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model:</td>
<td>cay</td>
<td>P/D</td>
</tr>
<tr>
<td>Predictor:</td>
<td>cay</td>
<td>P/D</td>
</tr>
<tr>
<td>Horizon 1 (quarters)</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>8</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
<td>0.15</td>
</tr>
<tr>
<td>16</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>20</td>
<td>0.06</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Note: The top panel reports, for each predictor, the proportion of simulated 228-quarter samples in which the $R^2$ in a return-forecasting regression is at least as large as observed in the data and where the predictive relationship has the correct sign. The bottom panel reports the proportion of simulated samples that generate Kiefer-Vogelsang-Bunzel t-statistics for each variable that are as large as in the data and have the same sign. The EZ-habit model is the benchmark (single-shock) model. cay is the consumption/wealth ratio from Lettau and Ludvigson (2001). P/D is the price/dividend ratio from CRSP. All simulated regressions use as the predictor the wealth/consumption ratio (equivalently, the price/dividend ratio on a claim to aggregate consumption) and the dependent variable is the excess return on a consumption claim. The regressions are run at horizons listed in the left-hand column. Bold numbers are less than 0.05, bold italics less than 0.01.
cent of the time. The EZ-CRRA model again has trouble matching the results for \( cay \), and only replicates the statistics for the price-dividend ratio in 5 to 20 percent of the samples, compared to 20 to 50 percent of the samples for the habit model.

### 2.3.2.2 Other return predictors

Table 2.4 reports the simulated correlation between five-year excess returns on the aggregate wealth portfolio and a variety of return predictors. The first row gives the correlation for actual risk aversion, which we would expect would be highest of all of the variables. The second row shows that the predictive power of the price-dividend ratio is nearly as high as that of \( \alpha_t \) in the benchmark model, but somewhat lower in the dual-shock calibration (though still not as much lower as in the data).

Fama and Schwert (1977) and Campbell (1987) find that short term interest rates negatively predict future stock returns.\(^{27}\) In table 2.4, I do not replicate the result that the real interest rate negatively forecasts returns, but the risk-free rate minus its 4-quarter moving average (denoted RREL as in Campbell, 1987), does weakly negatively forecast returns. Table 2.4 shows that the correlation of the five-year excess stock return with the real risk-free rate and RREL is substantially negative and nearly as large as that of \( \hat{\alpha} \). In the model, two effects cause interest rates to forecast stock returns. First, positive technology shocks raise interest rates and lower risk aversion. Second, even if risk aversion were driven by shocks unrelated to technology, interest rates might still forecast stock returns since a decline in risk aversion lowers the precautionary saving effect, raising interest rates. Intuitively, there is a flight-to-quality effect in interest rates, linking them to expected stock returns.

Table 2.4 reports the mean and standard deviation of the real term spread for the EZ-habit model. For the sake of simplicity, I follow the literature in modeling long-term debt as an asset that has a constant probability of paying its principal of one unit of the consumption good and retiring.\(^{28}\) If the bond does not retire and pay out, the holder retains

\(^{27}\) Campbell (1991) subtracts the 12-month moving average of the nominal risk-free rate from itself as a way to detrend the short term interest rate, since the nominal rate may be nonstationary if there are changes in trend inflation. Detrending in that way should be unnecessary in the model since interest rates are stationary, but I still check this variable.

\(^{28}\) See, e.g. Rudebusch and Swanson (2008) and Miao and Wang (2010).
### Table 2.4: Various return predictors

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>EZ-CRRA</th>
<th>EZ-habit</th>
<th>Dual-shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five-year excess stock return correlations:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>0.69</td>
<td>0.05</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>0.09</td>
<td>-0.05</td>
<td>-0.30</td>
<td>-0.23</td>
</tr>
<tr>
<td>RREL</td>
<td>-0.09</td>
<td>-0.03</td>
<td>-0.22</td>
<td>-0.17</td>
</tr>
<tr>
<td>Term spread</td>
<td>0.20</td>
<td>0.05</td>
<td>0.29</td>
<td>0.10</td>
</tr>
<tr>
<td>P/D</td>
<td>-0.39</td>
<td>-0.05</td>
<td>-0.30</td>
<td>-0.23</td>
</tr>
<tr>
<td>Term spread summary statistics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.43</td>
<td>-0.14</td>
<td>-0.23</td>
<td>-0.26</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>1.17</td>
<td>0.11</td>
<td>0.07</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Note: The bottom two rows are the mean and standard deviation of the spread between the yield on a ten-year equivalent bond and a one-quarter riskless bond (measured in the data using nominal Treasuries). The remaining rows give population correlations between various variables and five-year stock returns. For the simulations, population values are reported. RREL is the gap between the risk-free rate and its four-quarter moving average; P/D is the price/dividend ratio. In the simulations, P/D is measured as the wealth/consumption ratio. $\bar{a}$ is the value of risk aversion in the model, and estimated in the data as in section 4.
the bond for another period. I assume that the quarterly probability of payout is $1 - 0.9^{1/4}$ so that the expected maturity of the bond is ten years. The term spread is the yield to maturity on this bond minus the one-quarter riskless yield. The term spread in the model is on average negative, whereas the nominal Treasury yield curve is almost always upward-sloping in the data. The reason we have a negative term spread in the model is that in good times the marginal product of capital, and hence the risk-free rate, is above average. So in good times, short-term bonds have low prices, and hence they are a hedge and have a negative risk premium.

Fama and French (1989) show that the term spread forecasts stock returns. Interestingly, table 2.4 shows that even though the term spread is negative on average in the model, it still positively predicts future stock returns as in the data. This essentially comes through an expectations-hypothesis effect. In periods when the risk premium is low, the risk-free rate is high and expected to fall. To the extent that long-term yields are just averages of expected future short yields, long yields will rise less than short yields. So in periods when risk aversion is low, the term spread falls, and the term spread thus positively predicts stock returns.

In the model, the equity premium is a nearly a constant multiple of $\alpha_t$. The variables that forecast returns in table 2.4 are all correlated with $\alpha_t$, but imperfectly. For example, the price-dividend ratio also depends on expected consumption growth and interest rates. The fourth column of table 2.4 shows that the dual-shock model can qualitatively, if not quantitatively, match the empirical result in column 1 that estimated risk aversion is a more powerful forecaster of excess stock returns than any of the other variables, since it is uncontaminated by factors like expected consumption growth.

2.3.2.3 Consumption growth predictability

The aggregate price-dividend and wealth-consumption ratios may be driven by either movements in expected dividend (consumption) growth or movements in discount rates. For the aggregate stock market, Campbell and Shiller (1988) and Cochrane (2008) find that

---

$^{29}$ This result is exact in the log-linear approximation.
the price-dividend ratio has at best weak forecasting power for dividend growth. Similarly, Lettau and Ludvigson (2001) find that the wealth-consumption ratio has little forecasting power for consumption growth.

Figure 2.2 shows that the EZ-habit model is consistent with those results. First, to get a general sense of the dynamic properties of consumption growth, the top panel of figure 2.2 plots the autocorrelations of consumption growth against their empirical counterparts. The shaded region is the 95-percent confidence interval for the empirical estimates using the Newey–West method with a lag window of 12 quarters. In the model, the autocorrelations are near zero at all horizons. The data suggests that the first three autocorrelations are positive, which the model does not match. At longer lag lengths, though, there is no evidence for persistence in consumption growth, consistent with the model.

The bottom panel of figure 2.2 simulates 228-quarter samples as in figure 1 and calculates correlations between consumption growth between dates $t$ and $t + k$ and the consumption-wealth ratio at date $t$. What we see is that while many of the simulated correlations are far from zero, the mean sample correlation between the wealth-consumption ratio and future consumption growth is nearly zero. The figure also plots empirical correlations between $cay$ and future consumption growth at various horizons, and they are similar to the simulated mean. Figure 2.2 thus shows that the EZ-habit model not only matches the short and long-run variances of consumption growth, but it also replicates relevant features of the dynamics of consumption.

### 2.3.3 Impulse response functions

Figure 2.3 plots impulse response functions (IRFs) in the EZ-CRRA and benchmark EZ-habit models for four variables: consumption, household value, the risk-free rate, and the Sharpe ratio on the consumption claim. The lines give log deviations from steady-state, except for the risk-free rate, for which I report the absolute change in annualized percentage points. The shock is a unit standard deviation (88 basis-point) permanent increase in the level of technology, which will lead to an identical long-run increase in consumption, capital, and output.
Figure 2.2: Consumption predictability

Note: The top panel reports the empirical autocorrelation function for consumption. The gray shaded region is the 95% confidence interval using Newey–West standard errors with a lag window of 12 quarters. The bottom panel reports the correlation of consumption growth between periods $t$ and $t + x$ with the wealth-consumption ratio ($c_{ay}$) at date $t$, where $x$ is reported in quarters on the horizontal axis. The solid lines give the mean and 5th and 95th percentiles of the same correlation in simulated 228-quarter samples.
Figure 2.3: Impulse response functions

Note: Impulse responses for the EZ-CRRA and EZ-habit models. The shock is a positive unit-standard-deviation increase in technology. The dotted lines are for EZ-CRRA, solid are for EZ-habit. All functions are reported as fractions of the variables' means except for the risk-free rate, for which the response is in annualized percentage points. Value is lifetime utility; the Sharpe ratio is for an asset that pays aggregate consumption as its dividend.
The top-left panel shows the response of household value. For the EZ-CRRA model, value immediately jumps to a point just below its new steady state, and then slowly rises as households accumulate capital. For the EZ-habit model, though, value actually overshoots its new steady state. The reason is that the positive shock to productivity drives risk aversion down. When households are less risk-averse, they place a higher value on their future consumption stream because they penalize uncertainty less strongly. This effect helps increase the volatility of the SDF (equation 2.6), raising the Hansen–Jagannathan bound. The top-right panel shows that on the impact of a shock, the Sharpe ratio in the EZ-habit model falls by 12.5 percent (as a fraction of its mean), and then gradually rises again, with a half-life of 12 quarters.

The bottom-left panel shows the dynamics of the risk-free rate. The initial response is essentially identical for the two models. The reason for this is that the risk premium on an unlevered claim on capital is very small in the model, so the return on capital is roughly equal to the risk-free rate. Since the size of the capital stock is essentially fixed in the short-run, an increase in productivity directly increases the return on capital and hence the risk-free rate.

The final panel of figure 2.3 shows the response of consumption in the two models. The EZ-habit model shows a larger initial response of consumption, with lower expected consumption growth going forward. To see why this is, we can write consumption growth in equilibrium as

\[ E_t \Delta c_{t+1} = \bar{c} + \rho^{-1} r_{f,t+1} + \alpha_t \times vol \]  (2.15)

where \( r_{f,t+1} \) is the risk-free interest rate between dates \( t \) and \( t + 1 \), \( vol \) represents a measure of the total volatility in the model, and \( \bar{c} \) is a constant. \( \alpha_t \times vol \) represents the precautionary saving effect and is a function of the current level of risk aversion and the variances of the shocks in the model. The standard interpretation in an endowment economy is that conditional on consumption growth, a strong precautionary saving motive leads to a low risk-free rate. In a production setting, though, it is the risk-free rate that is held roughly fixed since it is tied to the marginal product of capital, which is hard to change quickly through investment. Conditional on the risk-free rate, then, a small precautionary saving
motive leads to lower expected consumption growth (more consumption today, saving less for tomorrow). In the EZ-habit model, a positive technology shock lowers risk aversion, and hence consumption rises more than in the canonical EZ case. This effect also serves to increase the volatility of the SDF, just as the higher response of value does.

Given the results in figure 2.3, it is straightforward to see what would happen in this economy if there were a pure shock to the coefficient of relative risk aversion. Since the risk-free rate is tied to the marginal product of capital, it would not move on the impact of a shock. The only effect on real variables of a pure decline in risk aversion then is that households would want a smaller buffer stock of savings, so they would raise consumption and lower investment: shocks to risk aversion look like simple demand shocks.

### 2.3.4 Estimating the EIS from interest rate regressions

The value of the elasticity of intertemporal substitution is controversial. Regressions based on aggregate consumption and asset returns often find a very small EIS (Hall, 1988; Campbell and Mankiw, 1989). Campbell (2003) reviews the literature and estimates the EIS using a variety of specifications and data from a broad range of countries, finding values generally less than 0.5 and often less than 0.2.\(^{30}\) This result is in conflict with the calibration used here and in other recent production-based asset pricing studies (Kaltenbrunner and Lochstoer, 2010; Gourio, 2010), which assume that the EIS is greater than 1. The question is whether the EZ-habit model generates small EIS estimates in regressions similar to those estimated in Campbell (2003).

The standard aggregate EIS regressions start from a model in which the risk-free rate takes the form

\[
r_{f,t+1} = b_0 + \rho E_t \Delta c_{t+1} \tag{2.16}
\]

where \(r_{f,t+1}\) is the riskless interest rate between periods \(t\) and \(t + 1\). \(b_0\) is a parameter depending on the discount rate and underlying volatility in the model (which are taken

---

to be constant). This relationship is straightforward to derive in an endowment economy with homoskedastic consumption growth and where households have a constant EIS and coefficient of relative risk aversion. It is also obtained in a log-linearization of the standard RBC model with homoskedastic technology shocks.

In principle, the EIS can be estimated from a regression of interest rates on consumption growth or vice versa. However, since the reduced-form relationship between consumption growth and interest rates is nearly zero in the EZ-habit model, in some of the simulations regressing $E_t \Delta c_{t+1}$ on $r_{f,t+1}$ produces explosive estimates for $\rho^{-1}$ (since we have to invert the coefficient estimate). Moreover, consumption in the EZ-habit model nearly follows a random walk, so it is essentially unpredictable and there are serious weak-instruments problems in an IV regression of interest rates on consumption growth. I therefore focus on the regression of consumption growth on interest rates,

$$E_t \Delta c_{t+1} = b' + \rho^{-1} r_{f,t+1}$$  \hspace{1cm} (2.17)

In the simulations of the model in section 2.3, we have the ability to directly measure $E_t \Delta c_{t+1}$. The first row of table 2.5 reports the population estimate of $\rho^{-1}$ in regression (2.17) under the EZ-CRRA and EZ-habit models. In the EZ-CRRA case, the regression identifies $\rho^{-1}$ exactly. On the other hand, the estimate of $\rho^{-1}$ is biased substantially downwards in the EZ-habit specifications.

The bias comes from the fact that the time-varying precautionary saving effect (equation 2.15) is omitted from the regression. Since precautionary saving is correlated with both expected consumption growth and interest rates, omitting it biases the simple IS-curve regression usually used to identify the EIS. An alternative way to see the source of the bias is to go back to the IRFs in figure 2.3. In both models the risk-free rate rises by the same amount following a shock. In the EZ-habit specification, though, because of the decline in precautionary saving, expected consumption growth is lower following a shock than in the EZ-CRRA case. That means that the estimate of $\rho^{-1}$ will fall.\textsuperscript{31}

\textsuperscript{31} In Bansal and Yaron (2004), time-variation in the volatility of shocks in principle causes EIS regressions to be biased. However, Beeler and Campbell (2010) show that their calibration generates almost no actual bias—the median sample EIS estimates are well above 1. This paper thus represents an improvement in being able
Table 2.5: Regressions estimating the elasticity of intertemporal substitution

<table>
<thead>
<tr>
<th>Model:</th>
<th>Data</th>
<th>EZ-CRRA</th>
<th>EZ-habit</th>
<th>Dual-shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Population, infeasible ($E_t[\Delta c_{t+1}]$)</td>
<td>N/A</td>
<td>1.50</td>
<td>0.64</td>
<td>0.78</td>
</tr>
<tr>
<td>2 Population</td>
<td>N/A</td>
<td>1.50</td>
<td>0.56</td>
<td>0.71</td>
</tr>
<tr>
<td>3 Small sample</td>
<td>0.14</td>
<td>1.16</td>
<td>0.03</td>
<td>0.35</td>
</tr>
<tr>
<td>4 [2.5%, 97.5%]</td>
<td>N/A</td>
<td>[0.03, 1.79]</td>
<td>[-1.98, 1.02]</td>
<td>[-1.32, 1.34]</td>
</tr>
<tr>
<td>5 Population, RRA control</td>
<td>N/A</td>
<td>N/A</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>6 Small sample, RRA control</td>
<td>0.18</td>
<td>N/A</td>
<td>0.07</td>
<td>0.94</td>
</tr>
<tr>
<td>7 [2.5%, 97.5%]</td>
<td>N/A</td>
<td>N/A</td>
<td>[-3.08, 3.11]</td>
<td>[-1.07, 2.78]</td>
</tr>
</tbody>
</table>

Note: Values reported are the coefficient from regressions of consumption growth or expected consumption growth on the risk-free rate. The dependent variable in row 1 is expected consumption growth (computed numerically in the simulations); all other rows use actual consumption growth. The small-sample regressions are based on 228 quarters of data, and median coefficient estimates are reported; 2.5 and 97.5 percentiles are reported in brackets. The RRA control is actual risk aversion in the simulations and estimated risk aversion (section 4) in the empirical regressions.
The regression in the first row of table 2.5 is in some sense ideal, but it is not the regression that we are actually able to run in the data since $E_t \Delta c_{t+1}$ is unobservable. Rows 2 through 4 report results for estimates of $\rho^{-1}$ from regressions of actual consumption growth, $\Delta c_{t+1}$, on the risk-free rate, $r_{f,t+1}$. Row 2 gives the population estimates, while rows 3 and 4 give the median and 95-percent range of the estimates from 228-quarter simulations. With constant relative risk aversion, the population regression in row 2 estimates the EIS exactly. The median estimate from the small-sample regressions in row 3 is 1.16. The 95-percent range is wide, and it only just barely contains the estimate from the data. So it is in principle possible for the EZ-CRRA model to generate an estimate of the EIS as small as what we observe in the data, but the probability is small (less than 10 percent).

In the EZ-habit models, the bias is far larger. The population estimate in the single-shock case is 0.56, and the median sample estimate is 0.03. For the dual shock model, the estimates are only slightly better—0.71 in population and a median of 0.35 in small samples. The reason for this slight improvement is that the temporary shocks have little effect on risk aversion, so they induce variation in consumption and the risk-free rate that is closer to the usual EZ-CRRA case. The upper end of the 95 percent range for the simulated estimates is well below the true value of the EIS.

The first four rows of table 2.5 show that in general, regressions of interest rates on expected consumption growth are not a very good way to estimate the EIS, and in the EZ-habit model they are biased and inconsistent. It is worth noting, though, that if we could observe $\alpha_t$, we could completely eliminate the bias in the EIS regressions. Empirically, this suggests that regressions designed to estimate the EIS could be improved by including a control for risk aversion, such as the price-dividend ratio on the stock market. The final three rows of table 2.5 try to estimate the EIS including a control for risk aversion. In the data, I use a measure of risk aversion derived from the EZ-habit model below in section 32

In principle, the real risk-free rate, $r_{f,t+1}$, is also unobservable in the data. As above, I form $r_{f,t+1}$ as the difference between the nominal 3-month interest rate and a forecast of inflation based on lagged inflation and nominal interest rates. Errors in the estimate of the true real-risk-free rate would bias the estimate of $\rho^{-1}$ towards zero. Instrumental-variables methods can theoretically eliminate this bias.
2.4, denoted $\hat{\alpha}_t$. The empirical estimate of the EIS is essentially unchanged from when $\hat{\alpha}_t$ is not included. In population, when $\alpha_t$ is included in the simulated regressions, the EIS is estimated exactly. In small-sample regressions, though, the estimate of the EIS in the model is still biased downward. In the single-shock model, the median estimate is 0.07, and in the dual-shock model 0.94. Row 7 shows, though, that the 2.5 percentile of the small-sample estimates is -3.08 in the dual shock model, while the 97.5 percentile is 3.11. So even though the median estimate in the dual-shock model is not enormously biased, the empirical value of 0.18 is well within the simulated range. In the end, while controlling for risk aversion should, in principle, allow us to estimate the EIS consistently, in small samples the regressions still do not seem to provide useful estimates because of weak-identification problems.

As an alternative to these regressions, the EIS could be consistently estimated if we had an instrument for the risk-free rate that was uncorrelated with risk aversion. Standard aggregate instruments like lagged interest rates and consumption growth (e.g. Campbell, 2003) will certainly not be valid instruments under the EZ-habit model since the precautionary saving effect is persistent. Household-level instruments would work better if my model is correct in treating risk aversion as being driven by aggregate factors. The EZ-habit model thus has the ability to explain the divergence between micro and macro estimates of the EIS if the micro instruments are valid and the macro instruments invalid.

2.4 Empirical return forecasting

This section shows how the model suggests we can directly estimate the coefficient of relative risk aversion in the data, and then demonstrates that the estimate is a powerful predictor of stock returns. Second, I present novel evidence that technology growth forecasts stock returns, just as it does in the production model. The results differentiate the EZ-habit paper from models with time-varying disaster risk. Gourio (2010) predicts that when

33 Gruber (2005) discusses precisely these issues and tries to resolve them by using household-specific variation in tax rates as an instrument for consumption growth. Dynan (1993) explicitly controls for the precautionary savings effect at the household level by predicting the conditional volatility of consumption, but she does not deal with the possibility that risk aversion varies over time.
there are changes in the probability of a large disaster occurring, price-dividend ratios will forecast returns, which is also true in the EZ-habit model. The EZ-habit model also predicts, though, that technology and estimated risk aversion will forecast stock returns, and that estimated risk aversion will be the single most powerful forecaster of returns, which would not be true in the time-varying disaster model or models based on other forms of time-varying volatility (e.g. Bansal and Yaron, 2004, Bloom, 2009, or Fernandez-Villaverde et al., 2011).

2.4.1 Estimating risk aversion

If risk aversion follows the AR(1) process given in (2.5), then we can measure current risk aversion if we simply observe the history of aggregate value, $v_t^A$. For a given value of the EIS and observed data on wealth and consumption, it is possible to calculate $v_t^A$ by rearranging equation (2.7)

$$v_t^A = \frac{1}{1-\rho} w_t^A - \frac{\rho}{1-\rho} c_t^A + \frac{1}{1-\rho} \log (1 - \exp (-\beta)) \quad (2.18)$$

If we can measure household wealth and consumption, then we can measure value. We then simply plug the estimates of $v_t^A$ into equation (2.5) to obtain estimates of $\alpha_t$.

Lettau and Ludvigson (2001) study a cointegrating relationship between consumption and aggregate wealth. Their method is valid in my model since consumption and wealth are cointegrated under balanced growth. While their analysis was designed to estimate the consumption-wealth ratio, it also delivers, as a byproduct, a measure of aggregate wealth (since we can always add consumption to the wealth-consumption ratio to obtain wealth). The estimate of wealth derived using their method is a combination of asset wealth data obtained from the flow of funds accounts plus an estimate of human wealth. They treat labor income as the dividend from the stock of human wealth. Assuming the price-dividend ratio for human wealth is stationary, we can use labor income as a proxy for human wealth. Denoting asset wealth as $a_t$ and labor income as $y_t$, the appendix shows
that we then have a cointegrating relationship,

\[ c_t = \zeta_0 a_t + \zeta (1 - \omega) y_t + \zeta_t' \]  

(2.19)

where \( \zeta \) and \( \omega \) are parameters and \( \zeta_t' \) is a stationary error term. Lettau and Ludvigson (2001) refer to the residual \( \zeta_t' \) as \( cay \). This variable essentially represents an estimate of the consumption-wealth ratio. Since I want to estimate wealth, I define

\[ a_y t \equiv \omega a_t + (1 - \omega) y_t \]  

(2.20)

which, under the assumptions above, will be a statistically unbiased estimate of total wealth, but will include error due to the fact that we do not assume we directly measure human wealth.\(^{34}\)

With our measure of wealth \( a_y t \), we estimate \( \nu_t^A \) as

\[ \hat{\nu}_t^A = \frac{1}{1 - \rho} a y_t - \frac{\rho}{1 - \rho} c_t \]  

(2.21)

where we ignore constants, and a circumflex indicates an estimated variable. Note that since the parameters of the cointegrating relationship for \( c_t, a_t, \) and \( y_t \) are estimated super-consistently we do not have to modify any standard errors in the subsequent analysis to take into account the fact that \( a_y t \) is a generated regressor (which is why it does not receive a circumflex). That said, to the extent that there is measurement error in the consumption or wealth data, \( \hat{\nu}_t^A \) will inherit that same error. When we use \( \hat{\nu}_t^A \) to forecast market returns, this measurement error should only weaken the results. For measurement error to generate a spurious predictive relationship, it would have to be correlated with other predictors of returns.\(^{35}\)

---

\(^{34}\) In simulations with variable labor supply, The price/dividend ratio on human wealth does vary over time, but the variation is relatively small: risk aversion calculated using the method here (assuming a constant price-dividend ratio on human wealth) is over 95 percent correlated with actual risk aversion.

\(^{35}\) One obvious source of measurement error is that human capital is not a perfect estimator of the value of human wealth. Suppose risk aversion rises above average and lowers the price-dividend ratio on human wealth below average. Labor income will then be overestimating human wealth (compared to its average). High levels of wealth drive out measure of \( \hat{\alpha}_t \) downward, so this measurement error should bias the results against correctly forecasting returns (high risk aversion in the data leads to low risk aversion in our estimates).
This definition of \( \hat{\sigma}_t^A \) is similar to Lettau and Ludvigson’s \( cay_t \), except they have equal weights on \( c_t \) and \( ay_t \), whereas equation (2.21) uses a combination where the weights depend on the EIS. Also, \( cay_t \) is stationary by construction, whereas \( \hat{\sigma}_t^A \) is growing over time (it is cointegrated with consumption and wealth).

In equation (2.21) a high EIS (low \( \rho \)) raises the weight on consumption relative to asset wealth. If the EIS is less than 1 (\( \rho > 1 \)), the weight on wealth, \( ay_t \), is actually negative, and the weight on consumption greater than 1. Bansal and Yaron (2004) and Kaltenbrunner and Lochstoer (2010) both find that an EIS of 1.5 allows their models to fit asset pricing facts, so I use the same value. This value is also consistent with the micro evidence of Vissing-Jorgensen and Attanasio (2003). The results reported below are broadly similar as long as the EIS is greater than 1.1 (at that level and below, \( \hat{\sigma}_t^A \) becomes very volatile). The appendix reports a sensitivity analysis for various values of the EIS.

Figure 2.4 plots \( \hat{\sigma}_t^A \) both in its raw levels and with a linear trend taken out. As we would expect, \( \hat{\sigma}_t^A \) follows a strong upward trend. There seem to be both low and high frequency components to detrended \( \hat{\sigma}_t^A \). In particular, there are long-run swings with peaks in the early 1970’s and 2000’s and a trough around 1994, generally consistent with movements in the aggregate price-dividend ratio and variation in average output growth.\(^{36}\) At the same time, there are business-cycle frequency movements, e.g. the troughs in 1973, 2001, and 2008.

I construct an estimate of \( \alpha_t, \hat{\alpha}_t \), using the update process for risk aversion, equation (2.5), and the data on \( \hat{\sigma}_t^A \). In particular, we have

\[
\hat{\alpha}_{t+1} = \phi \hat{\alpha}_t + (1 - \phi) \bar{k} + \lambda \left( \Delta \hat{\sigma}^A_{t+1} - E_t \Delta \hat{\nu}^A_{t+1} \right) \tag{2.22}
\]

As above, I assume that \( \phi = 0.96 \). \( E_t \Delta \hat{\nu}^A_{t+1} \) is estimated simply as the sample average of \( \Delta \hat{\sigma}^A_t \).\(^{37}\)

---

\(^{36}\) Note, though, that linear detrending will tend to make the series look as if it is mean-reverting even if it follows a random walk. The linear trend is only used to make the graph legible; none of the results involve it.

\(^{37}\) In principle, it is possible to forecast \( \Delta \hat{\nu}^A_{t+1} \), but the amount of predictability in \( \Delta \hat{\nu}^A_{t+1} \) is sufficiently small that the results are nearly identical to assuming that \( \nu^A_{t+1} \) simply follows a random walk.
Table 2.6: Value, raw and linearly detrended

<table>
<thead>
<tr>
<th>Year</th>
<th>Value, raw</th>
<th>Value, linearly detrended</th>
</tr>
</thead>
<tbody>
<tr>
<td>1952</td>
<td>-0.2</td>
<td>11.5</td>
</tr>
<tr>
<td>1957</td>
<td>-0.15</td>
<td>11.3</td>
</tr>
<tr>
<td>1962</td>
<td>-0.1</td>
<td>11.1</td>
</tr>
<tr>
<td>1967</td>
<td>-0.05</td>
<td>11.0</td>
</tr>
<tr>
<td>1972</td>
<td>0.05</td>
<td>10.9</td>
</tr>
<tr>
<td>1977</td>
<td>0.1</td>
<td>10.8</td>
</tr>
<tr>
<td>1982</td>
<td>0.15</td>
<td>10.7</td>
</tr>
<tr>
<td>1987</td>
<td>0.2</td>
<td>10.6</td>
</tr>
<tr>
<td>1992</td>
<td>0.25</td>
<td>10.5</td>
</tr>
<tr>
<td>1997</td>
<td>9.5</td>
<td>10.4</td>
</tr>
<tr>
<td>2002</td>
<td>9.7</td>
<td>10.3</td>
</tr>
<tr>
<td>2007</td>
<td>9.9</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Note: Household value (lifetime utility) is measured using data on wealth and consumption from Sydney Ludvigson’s website. The thin line is the absolute level of value (left-hand axis); the thick line is value linearly detrended. Both variables are measured in logs. Grey bars are NBER-dated recessions.
The parameter \( \lambda \) governs the volatility of \( \alpha_t \), but it has only a multiplicative effect on \( \hat{\alpha}_t \). That is, two estimates of \( \hat{\alpha}_t \) will be perfectly correlated with each other, regardless of what values are chosen for \( \lambda \). The same argument applies for \( \bar{\alpha} \). As long as we are simply trying to forecast stock returns using a linear regression, we can ignore any additive or multiplicative shifts in \( \hat{\alpha}_t \). Therefore, I set \( \bar{\alpha} = 0 \) and choose \( \lambda \) so that \( \hat{\alpha}_t \) has unit variance, normalizations that will have no effect on the regression-based measures of forecasting power (and I choose a negative value of \( \lambda \) to match the habit-formation motivation of the model). In the first period of the sample I assume \( \hat{\alpha} = \bar{\alpha} \).

An important feature of this method of forecasting is that it is based only on the preference specification. None of the assumptions we made about the production side of the economy are required for this method to be valid. We simply take advantage of the relationship between household value and changes in risk aversion and the relationship under Epstein–Zin preferences between household value and wealth.

### 2.4.2 Forecasting market returns

The next question is to what extent the model-implied variation in expected returns is related to actual returns. Figure 2.5 plots \( \hat{\alpha}_t \) and 5-year excess returns on the stock market (the value-weighted excess return from Kenneth French). The strong correlation between the two series (0.68) is immediately apparent. There are both high and low-frequency movements in \( \hat{\alpha}_t \) associated with changes in growth in value. In the periods when value is growing quickly, e.g. the late 1990’s, risk aversion falls. At the same time, there are higher-frequency movements, such as the temporary increases in estimated risk aversion around the recessions in 1991 and 2001.

Figure 2.1 plots \( R^2 \)'s from regressions of future stock returns on \( \hat{\alpha}_t \), \( cay_t \), the price-dividend ratio (P/D), and the excess consumption ratio from Campbell and Cochrane (1999). Each line gives the \( R^2 \) from a univariate regression. The x-axis gives the horizon for the return in quarters. The \( n^{th} \) point is the \( R^2 \) from a regression of \( \sum_{j=1}^{n} r_{t+j} \) on the predictor at time \( t \). The regressions are all run on quarterly data from 1952 to 1999 (to

The appendix shows that the results are robust to different choices for \( \phi \).
Figure 2.4: Estimated risk aversion and 5-year excess stock returns

Note: Excess stock returns are for the CRSP value-weighted index minus the risk-free rate, from Ken French’s website. Returns are forward-looking five-year averages. Risk aversion is estimated from data on aggregate wealth and consumption and is normalized to have zero mean and unit variance.
ensure we have data for the 40-quarter regression). Each regression uses the same sample for the predictors.

At every horizon, $\hat{\alpha}$ is dominant. At the five-year horizon, the $R^2$ for estimated risk aversion peaks at more than twice that of the other variables. The $R^2$s are impressively high for just a single variable: at the 5-year horizon, $\hat{\alpha}$ explains 50 percent of the post-war variation in stock returns. Furthermore, in horse-race regressions (reported in the appendix), $\hat{\alpha}$ dominates $cay$ at all horizons.

An important consideration in long-horizon forecasting regressions is that the residuals are highly persistent. Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2005) show that by using Newey–West standard errors with a very long lag-window, we can obtain test statistics with better size properties than techniques that use a fixed (and usually short) lag window. I choose a lag window equal to half the sample size and use the critical values reported in Kiefer and Vogelsang (2005). For $cay$, every regression except for those with horizons greater than 30 quarters is significant at the 5 percent level. For $\hat{\alpha}$, the largest p-value is 0.0008. The price-dividend ratio is significant at the 5 percent level for forecasts of 14 quarters or longer. In other words, these regressions all imply that we have substantial ability to forecast stock returns in the post-war period, and $\hat{\alpha}$ is the strongest of the predictors. Out-of-sample tests with both asymptotic and bootstrapped critical values give similar results (appendix D.3). Appendix D examines the sensitivity of the results in this section to the various parameters we had to calibrate (e.g. the EIS and the persistence of habits). The basic results hold across a broad range of parameter sets.

2.4.3 Forecasts from estimates of technology

The method of estimating the level of risk aversion studied above does not rely on any assumptions about the structure of production in the economy, being derived purely from the preference specification. However, in the production model, changes in value are

---

38 Kiefer and Vogelsang (2002) note that there is a size-power tradeoff. When the lag window is increased, the size of the test statistics gets closer to their nominal size, but there is a loss of power. I choose a lag window of half the sample to balance these considerations. The results are basically identical when using a lag window equal to the sample size as in Kiefer, Vogelsang, and Bunzel (2000), though the price-earnings is significant in more of the regressions.
closely related to changes in productivity. If we can measure innovations to technology, then risk aversion should follow an AR(1) process where the innovations are equal to the shocks to the stochastic trend in technology.

There is a large literature that tries to estimate aggregate technology shocks. I consider two methods here. The first builds off of Solow (1957) and uses restrictions from a constant-returns production function:

$$a_t = y_t - \gamma k_t - (1 - \gamma) l_t$$  \hspace{1cm} (2.23)

$a_t$ measures technology if the economy has a Cobb–Douglas production function, with $l$ denoting log labor supply. I also consider a simpler metric, labor productivity, $lp_t = y_t - l_t$. Labor productivity does not take into account the effects of capital accumulation and simply models technology as the average product of labor. Capital can be difficult to measure, whereas the number of hours supplied in the economy is a fairly concrete quantity (though the quality of those hours is difficult to account for). 39

To extract the stochastic trend from the two productivity series, I estimate univariate ARMA models for each variable. The Bayesian information criterion implies that TFP growth is best fit with an MA(2), while labor productivity growth should be treated as i.i.d. $\epsilon_{TFP}^t$ is defined as the residual in the MA(2), while $\epsilon_{LP}^t$ is simply equal to labor productivity growth. That is, $\epsilon_{TFP}^t$ and $\epsilon_{LP}^t$ are innovations to the Beveridge–Nelson (1981) trends in productivity.

Section 2.5.1 shows that, at least in the case where log technology follows a random walk, risk aversion follows an AR(1) process of the form,

$$\alpha_t = (1 - \theta) \bar{\alpha} + \theta \alpha_{t-1} + \epsilon_{X}^t$$  \hspace{1cm} (2.24)

where $\epsilon_{X}$ denotes a measure of technology growth. We then have two measures of $\alpha_t$.

39 Furthermore, labor productivity determines the tradeoff that households face between consumption and leisure. If the capital stock rises because foreigners want to invest more in the US, household welfare will increase even if TFP does not. Similarly, a tax increase that reduced desired saving could lower welfare and labor productivity, without having any effect on TFP. And welfare is the relevant input in estimating $\bar{\alpha}$.
which I denote $\hat{\alpha}^{TFP}$, $\hat{\alpha}^{LP}$, using $\epsilon_i^{TFP}$ and $\epsilon_i^{LP}$, respectively.\footnote{Note that $\hat{\alpha}^{TFP}$ includes some forward-looking information since its construction requires the estimation of an MA(2) on the full sample. $\hat{\alpha}^{LP}$ does not suffer from this flaw. It is true that in both cases we have to estimate mean productivity growth, but shifts in the estimated mean simply correspond to shifts in the mean of $\hat{\alpha}_i$; they have no effect on its dynamics. In regressions of returns on $\hat{\alpha}_i$, the constant will thus always absorb shifts in $\hat{\alpha}$, so the estimation of the mean of productivity growth is irrelevant for forecasting returns.} The two measures turn out to be highly correlated (93 percent).

Figure 2.6 plots five-year excess returns against $\hat{\alpha}^{TFP}$ and $\hat{\alpha}^{LP}$. The two series are both clearly highly correlated with future excess returns. The p-values in regressions of quarterly excess returns on $\hat{\alpha}^{TFP}$ and $\hat{\alpha}^{LP}$ are 0.032 and 0.026, respectively (using Kiefer, Vogelsang, and Bunzel, 2000, t-type-statistics to account for autocorrelation). The relationship between the three series is most clear around the turning points. Productivity growth begins slowing down around 1970, driving risk aversion upwards. Forward-looking stock returns reach their trough at roughly the same point. Productivity growth rises again starting in the mid-1990’s, which is exactly when stock returns begin falling again.

The two measures of risk aversion in figure 2.6 clearly do not have the explanatory power of the variables studied above. The wealth and consumption data used above have a forward-looking component that is not present in instantaneous measures of technology. On the other hand, my measure of risk aversion is highly correlated with the consumption-wealth ratio. In almost any model with time-varying discount rates, the consumption-wealth ratio will forecast stock returns. It is not the case, though, that any model will predict that measures of technology should forecast stock returns. Figure 2.6 thus provides evidence in favor of the model presented here over other explanations of time-varying expected returns.

\section*{2.5 Extensions}

\subsection*{2.5.1 Log-linearization}

This section studies a log-linear version of the production economy from above. I use the solution to build a better understanding of the basic results reported in the previous sections. In particular, I derive analytic approximations for the consumption function, the risk-free rate and the conditional Sharpe ratio for the wealth portfolio. I also derive
Figure 2.5: Stock returns and estimates of risk aversion from productivity growth

Note: Total factor productivity is the quarterly Solow residual obtained from John Fernald’s website. Labor productivity is output per hour in the non-farm private business sector from the BLS. Risk aversion is an AR(1) with innovations equal to the (negative) innovations to the Beveridge-Nelson trend in productivity.
an essentially affine model of the term structure with a time-varying price of risk. This result connects the standard production theory in the macro literature to one of the most commonly used empirical asset-pricing frameworks.

2.5.1.1 Approximation method and solution

I derive a log-linear consumption function as the solution to a model that represents a log-linearization of the environment derived above with permanent technology shocks only. Specifically, if the capital accumulation equation, the return on capital, and the return on the wealth portfolio are log-linearized, then we are able to obtain an exact formula for the consumption function under EZ-habit utility (with canonical Epstein–Zin and power utility as special cases). The methods build on Campbell (1994) and Lettau (2003). Unlike the usual techniques in the macro literature, the solution is not based on certainty-equivalent or higher-order approximations to expectations. Rather, I take advantage of log-normality to calculate expectations exactly. That feature of the method is critical for accurately capturing risk premia and precautionary saving effects.

It is straightforward to show that the approximation technique delivers policy functions that are identical to those obtained from perturbation up to the first order (depending on where the derivatives are taken). The difference is that because I take advantage of formulas for log-normal expectations, a term involving risk aversion appears in the solution, meaning that the approximation captures the time-varying precautionary saving effect that is central to driving the difference in the consumption response between the EZ-habit model and the RBC model with canonical Epstein–Zin preferences.

The approximation method involves log-linear approximations to three components of the model: the budget constraint, the return on the wealth portfolio, and the return on

\footnote{The usual technique is the perturbation method of Judd (1999). See Woodford (2003) for a representative application and Rudebusch and Swanson (2009) for extensions to higher order approximations.}

\footnote{In perturbation, the equilibrium equations are not only approximated with respect to the endogenous and exogenous variables, but also the volatility of the technology shock, \( \sigma_a \). The first-order perturbation solution therefore does not include any terms involving interactions of state variables with \( \sigma_a \). The approximation used here includes a term for \( \alpha_t \sigma_a^2 \). The solutions are otherwise identical.}
capital,

\[ k_{t+1} \approx \lambda_0 + \lambda_k k_t + \lambda_a a_t + \lambda_c c_t \]  
(2.25)

\[ r_{w,t+1} \approx E_t r_{w,t+1} + \Delta E_{t+1} \sum_{j=0}^{\infty} \theta^j \Delta c_{t+j+1} - \Delta E_{t+1} \sum_{j=1}^{\infty} \theta^j r_{w,t+j+1} \]  
(2.26)

\[ r_{k,t+1} \approx r_0 + r_k (k_{t+1} - a_{t+1}) \]  
(2.27)

where \( \{\lambda_0, \lambda_k, \lambda_a, \lambda_c, \theta, r_0, r_k\} \) are linearization coefficients that I solve for in the appendix and lower-case letters denote logs.\(^{43}\)

Define \( \tilde{c}_t \equiv c_t - a_t \) and \( \tilde{k}_t \equiv k_t - a_t \). With the three log-linear approximations (2.25), (2.26), and (2.27), the appendix shows that we can obtain the following result:

**Proposition 2.1.** Given the log-linear budget constraint, the log-linear return on the consumption claim, and the production function, the optimal consumption plan takes the form

\[ \tilde{c}_t = \eta_{c0} + \eta_{ck} \tilde{k}_t + \eta_{ca} a_t \]  
(2.28)

and the return on the wealth portfolio can be written as

\[ r_{w,t+1} = \eta_{w0} + \eta_{wa} a_t + \rho E_t \Delta c_{t+1} + \kappa_r \varepsilon_{t+1} \]  
(2.29)

where the coefficients \( \{\eta_{c0}, \eta_{ck}, \eta_{ca}, \eta_{w0}, \eta_{wa}, \kappa_r\} \) are solved for in the appendix. Furthermore, the log price-dividend ratio of the consumption claim is linear in risk aversion and scaled capital.

The first important implication of this result is that even though risk aversion is time-varying, both consumption growth and returns on the wealth portfolio are homoskedastic. We did not assume the existence of a log-linear policy for consumption or homoskedastic wealth return. Variation in risk aversion induces variation in expected returns on risky assets, but not in their volatility. This same result is obtained in the numerical solution above.

\(^{43}\) (2.25) is a log-linearization of the resource constraint, \( K_{t+1} = (1 - \delta) K_t + A_t^{-1 - \gamma} K_t^\gamma - C_t \); (2.26) is the Campbell–Shiller (1988) approximation for the return on the wealth portfolio; (2.27) is a linearization of \( R_{k,t+1} = \gamma K_{t+1}^{\gamma - 1} A_t^{1 - \gamma} + 1 - \delta \).
Remark 2.1. $\eta_{ck}$ does not depend on the level or volatility of the coefficient of relative risk aversion (i.e. on $\bar{\alpha}$ or $\lambda$).

The finding that $\eta_{ck}$ is not affected by the time-variation in the coefficient of relative risk aversion helps us build intuition as to why the IRF for consumption changes when risk aversion varies. Consumption responds to a technology shock more strongly in the EZ-habit model than in the RBC model purely because the coefficient of relative risk aversion falls in response to positive technology shocks. If the economy experienced a hypothetical shock to the size of the capital stock holding the coefficient of relative risk aversion fixed, the behavior of consumption and saving would be identical under EZ-habit and EZ-CRRA preferences.

2.5.1.2 The risk-free rate and excess returns on the wealth portfolio

Denote the conditional standard deviation of a variable $x$ as $\sigma(x)$. We have the following formulas for the risk-free rate and Sharpe ratio,

Proposition 2.2. In the log-linearized model, the risk-free rate follows

$$r_{f,t+1} = \eta_{f0} + \rho E\Delta c_{t+1} - \eta_{fa}\alpha_t$$  \hspace{1cm} (2.30)

and the Sharpe ratio of the consumption claim is

$$E_t(r_{w,t+1} - r_{f,t+1}) + \frac{1}{2}\sigma^2(r_w) = \rho \sigma(\Delta c) + (\alpha_t - \rho) \sigma(\Delta \nu)$$  \hspace{1cm} (2.31)

As usual, expected consumption growth affects the risk-free rate in proportion to the inverse of the EIS. There is an additional term $\eta_{fa}\alpha_t$ reflecting the time-varying precautionary saving motive. When risk-aversion is high, precautionary saving demand is high, driving the risk-free rate downwards, all else equal. It is immediately clear that a simple regression of consumption growth on interest rates will not identify the EIS, $1/\rho$, unless the instruments used for the interest rate are uncorrelated with current risk aversion or risk aversion is controlled for.

Note also that the Sharpe ratio is strictly increasing in $\alpha_t$. The terms $\sigma(\Delta c)$ and $\sigma(\Delta \nu)$
are the standard deviations of growth in household consumption and value, respectively (both of which are constant in equilibrium). Tallarini (2000) and Lettau (2003) show that in an RBC model with power utility, an increase in risk aversion need not increase the Sharpe ratio because consumers can endogenously smooth consumption.\textsuperscript{44} In this setting, though, risk aversion unambiguously increases the Sharpe ratio. The reason is that the preference for consumption smoothing comes from the EIS. An increase in risk aversion does not cause consumers to smooth consumption endogenously, and so the only effect is to raise the Sharpe ratio on the wealth portfolio.

An obvious question is what the term $\sigma (\Delta v)$ actually is. The appendix derives the following results,

**Proposition 2.3.** In equilibrium, the coefficient of relative risk aversion $\alpha_t$ follows the process,

$$\alpha_t = \phi \alpha_{t-1} + (1 - \phi) \bar{\alpha} + \sigma_{aa} \epsilon_t$$  \hspace{1cm} (2.32)

where $\sigma_{aa}$ depends on the parameters $\lambda$, $\theta$, $\phi$, and $\sigma_a$.

The standard deviation of innovations to value is

$$\sigma (\Delta v) = \frac{-1 + \sqrt{1 + 2 \frac{\theta}{1 - \phi^2} \sigma_{aa} \sigma_a}}{1 - \phi \sigma_{aa}}$$  \hspace{1cm} (2.33)

$\sigma (\Delta v)$ is increasing in $\sigma_a$ for $\sigma_a > 0$ and $\sigma_{aa} < 0$.

Result 2.3 first shows that the coefficient of relative risk aversion follows an AR(1) process with innovations that are perfectly correlated with technology shocks. This result is a consequence of household value being a log-linear function of the level of technology, so that $v_{t+1} - E_t v_{t+1}$ is a linear function of the technology shock.

Result 2.3 shows that the standard deviation of innovations to household value is constant. Furthermore, in the benchmark case where technology shocks drive risk aversion down, an increase in their volatility raises the volatility of innovations to value, as we would expect.

\textsuperscript{44} Lettau and Uhlig (2004) and Rudebusch and Swanson (2008) obtain similar results for Campbell–Cochrane preferences.
In the case where $\sigma_{aa} = 0$, which corresponds to power utility, we obtain a surprisingly simple formula:

**Remark 2.2.** For the RBC model where technology follows a random walk and consumers have constant relative risk aversion, the Sharpe ratio on a consumption claim is approximately

\[
SR_t \approx \alpha \sigma_a (1 - \eta_{ck}) + (\alpha - \rho) \sigma_a \eta_{ck}
\]

(2.34)

\[
\approx \alpha \sigma_a - \rho \sigma_a \eta_{ck}
\]

(2.35)

This formula is similar to the formula obtained by Bansal and Yaron (2004) for the Sharpe ratio in the presence of long-run risks in an endowment economy. $\sigma_a$ represents long-run shocks to consumption growth, since consumption eventually catches up to a technology shock. Of that total response, $\sigma_a (1 - \eta_{ck})$ comes in the first period, with $\sigma_a \eta_{ck}$ in subsequent periods.\(^{45}\) We can thus think of the first component of the Sharpe ratio, $\alpha \sigma_a (1 - \eta_{ck})$, as Bansal and Yaron’s short-run risk term, and $(\alpha - \rho) \sigma_a \eta_{ck}$ as long-run risks. Kaltenbrunner and Lochstoer (2010) also show that production models generate long-run risk endogenously, but this simple formula for the Sharpe ratio has not been obtained elsewhere.

The second line shows that we can isolate the $\eta_{ck}$ term. If the EIS is large ($\rho$ is small) then the endogenous response of consumption in the model is unimportant and the Sharpe ratio is determined simply by the volatility of technology shocks and the coefficient of relative risk aversion.

2.5.2 Affine bond pricing

It turns out that the log-linear solution to the model allows us to connect the standard macro framework to the bond-pricing literature through the following result:

---

\(^{45}\) Recall that $\eta_{ck}$ is the coefficient on scaled capital in the consumption function. A unit increase in the technology shock $\varepsilon_{t+1}$ raises consumption by $\sigma_a$; the associated decline in scaled capital of $\sigma_a$ lowers consumption $\eta_{ck} \sigma_a$. 

87
Proposition 2.4. The log stochastic discount factor can be expressed as

\[ m_{t+1} = -r_{f,t+1} - \frac{1}{2} (\omega_0 + \omega_1 \alpha_t)^2 \sigma^2 + (\omega_0 + \omega_1 \alpha_t) \epsilon_{t+1} \]  

(2.36)

The SDF takes the tractable essentially affine form studied in much of the recent bond-pricing literature (see Duffee, 2002, and Piazzesi, 2010, for a recent review). We have a production-based general-equilibrium affine model of the term structure with an endogenously varying price of risk. Not only is the one-period risk-free rate affine in the state variables, but so are the prices and yields for all longer-term zero-coupon bonds. The fact that the SDF is affine is convenient because it means that the model could be estimated using the Kalman filter, either through Bayesian or frequentist methods. I am not aware of an affine model of the term structure with a time-varying price of risk being derived in a production setting previously.

2.5.3 Labor supply

I model labor supply as in Gourio (2010) and van Binsbergen et al. (2010). The household’s value function takes the form

\[ V_t = \left\{ (1 - \exp(-\beta)) \left( C_t^{1-\nu} (1 - N_t) \right)^{1-\rho} + \exp(-\beta) \left( E_t V_{t+1}^{1-\rho} \right)^{1-\rho} \right\}^{\frac{1}{1-\rho}} \]  

(2.37)

where \( N_t \) represents market labor and \( \alpha_t \) follows the same process as above. The household’s labor supply condition is

\[ \frac{1}{1 - N_t} = \frac{1 - \nu \omega_t}{\nu C_t} \]  

(2.38)

where \( \omega_t \) is the wage.

Note that risk aversion and habits only affect labor supply to the extent that they affect consumption. When consumption rises, labor falls, all else equal. Since positive permanent technology shocks drive consumption up farther in the EZ-habit model compared to the EZ-CRRA case, labor supply will rise by less in the EZ-habit model. Following a temporary technology shock, there is little change in risk aversion, so labor supply in that case
will look similar in the EZ-habit and EZ-CRRA models.

This result is not specific to the Cobb-Douglas utility specification studied here. In general, preferences consistent with balanced growth will specify labor supply as some function $H(\omega_t/C_t)$ (see King, Plosser, and Rebelo, 1988). Since $\omega_t/C_t$ is stationary with balanced growth, labor supply will be too. If $H$ is monotonically increasing, regardless of its functional form, the increase in consumption induced by a decline in risk aversion will also lower labor supply.

To see how habits affect labor supply here, figure 2.7 plots the response of employment to a shock to technology in the EZ-habit model versus a model with constant relative risk aversion. With EZ-CRRA we have the usual RBC result that the increase in technology increases $\omega_t/C_t$ thus raising employment. Employment then slowly falls back down to its steady state.

In the simple RBC model, it is possible to make labor supply fall following a shock by varying the parameters, but it always monotonically returns to steady state. In the EZ-habit specification, though, the response of labor supply has a hump shape. On the impact of the shock, employment barely increases at all, and it then rises slowly thereafter. This behavior actually matches the response of employment to technology shocks in the literature following Gali (1999). In particular, all of those papers, though they use different methods, and though they find different initial responses of employment to technology, find a pronounced hump shape. Basu, Fernald, and Kimball (2006) argue that this could be explained by a New Keynesian model. Figure 2.7 shows that variation in risk aversion could also explain that behavior.

Boldrin, Christiano, and Fisher (2001) and Jaccard (2011) note that in the RBC model with power utility and additive habits, variable labor supply undermines the ability of the RBC model to generate a volatile SDF. Intuitively, households can use labor supply to smooth consumption growth. Under power utility, the volatility of consumption growth

\[ 46 \text{ All of the parameters are identical to the main text, and } v \text{ is set to 0.33 as in Gourio (2010).} \]

\[ 47 \text{ See, e.g. Christiano, Eichenbaum, and Vigfusson (2004), Francis and Ramey (2005), and Basu, Fernald, and Kimball (2006).} \]

\[ 48 \text{ Note, though, that those papers also find hump-shaped responses for output and consumption, which I do not obtain.} \]
Figure 2.6: Response of employment to a technology shock

Note: Response of labor supply to a unit-standard-deviation permanent increase in the level of technology.
is what determines the volatility of the SDF. Under Epstein–Zin utility, though, the ability to smooth consumption shocks does not reduce the volatility of the SDF since the SDF loads almost purely on the permanent component of consumption, as we saw in the previous subsection.\(^{49}\) The EZ-habit model thus does not suffer from the drawback of previous habit-based models that freely variable labor supply could substantially reduce the Hansen–Jagannathan bound.

### 2.6 Conclusion

This paper presents a model of time-varying risk aversion. It simultaneously matches the basic behavior of macroeconomic and financial aggregates. The EZ-habit model gives a framework in which consumption, output, and investment growth are all realistically volatile in both the short and long-run, consumption growth is nearly a random walk, and risk premia are high and volatile.

More generally, this paper provides a general framework for modeling time-varying discount rates that can be used with other macro models. As pointed out by Cochrane (2011), asset-pricing research has recently focused on understanding variation in the price of risk over time. This paper gives a way of analyzing time-varying risk prices in the standard macro framework. I show that for the RBC model, the effect of an increase in risk aversion on consumption and investment looks similar to a decline in the household’s rate of time preference in the sense that it temporarily increases investment and reduces consumption.

An obvious next step is to study the EZ-habit preferences in a richer setting. Dew-Becker (2011) estimates a standard medium-scale DSGE model with sticky prices and wages, but with the added feature that risk aversion varies over time, as here. Complementing the results in this paper on equity pricing, Dew-Becker (2011) shows that the EZ-habit model, when augmented with a model of inflation, can match the behavior of the nominal term structure well, generating a strongly upward-sloping term structure of

---

\(^{49}\) I confirm this result numerically; the Hansen–Jagannathan bound is essentially identical with and without labor supply.
nominal interest rates and a volatile term premium.
3. BOND PRICING WITH A TIME-VARYING PRICE OF RISK IN AN ESTIMATED MEDIUM-SCALE BAYESIAN DSGE MODEL

3.1 Introduction

Non-structural models are widely used in both macroeconomics and the study of the term structure of interest rates. Recently, Smets and Wouters (2003) have shown that a structural New Keynesian model can match the dynamics of the macroeconomy as well as or better than a benchmark non-structural VAR. This paper extends that work by showing that a suitably augmented version of their model can also match the dynamics of the term structure of interest rates as well as a standard non-structural model. In addition, including information from the term structure has substantial effects on the estimated sources of variation in the real economy.

Bekaert, Cho, and Moreno (2010) show that a log-linearized macro model naturally also delivers closed-form expressions for bond prices. Their approximation method, however, is not able to describe risk premia, and even if it could, the model assumes that risk premia are constant. This paper builds on their work by using an approximation method that allows for positive and time-varying risk premia. I then estimate the model using Bayesian methods, and show that it fits interest rates with errors that are similar to those generated by a non-structural three-factor model. The errors in fitting annualized yields on bonds with maturities ranging from 1 quarter to 10 years have a standard deviation of 8 basis points.

For the production side of the economy, I take the model described in Justiniano, Primiceri, and Tambalotti (JPT; 2010) and combine it with a preference specification that endogenously generates the essentially affine stochastic discount factor of Duffee (2002). Households are assumed to have Epstein–Zin preferences with time-varying risk aversion as in
Melino and Yang (2003) and Dew-Becker (2011a), which induces a time-varying price of risk. I also allow the central bank to have a time-varying inflation target, movements in which shift the entire term structure, inducing a so-called level factor in interest rates.

The steady-state term spread in the model simply represents the average risk premium on long-term bonds. The steady-state term spread is estimated to be 152 basis points, similar in magnitude to the 207-basis-point average observed in the sample. To understand why that risk premium would be large, we first need to understand what drives the variance of the pricing kernel. When the representative household has Epstein–Zin preferences with a coefficient of relative risk aversion that is substantially larger than the inverse of its EIS (preferring an early resolution of uncertainty), state prices are almost entirely driven by innovations to the household’s lifetime utility, i.e. the value placed on its entire future stream of consumption and leisure. With a high EIS, transitory changes in consumption have a small effect on lifetime utility. Permanent technology shocks, though, will have large effects. Shifts in risk aversion also affect lifetime utility because they affect how much the household penalizes future uncertainty.

Even though there are nine shocks in the economy, only two of them turn out to be relevant for the pricing kernel—labor-neutral technology and risk aversion. Since all of the other shocks (e.g. monetary policy, markups, government spending) are purely transitory, they have little effect on permanent income or welfare (because the household is estimated to have a relatively high EIS of 1.33), and thus they do not have a strong effect on state prices.

Following a positive innovation to the level of technology, nominal interest rates are estimated to fall, making long-term bonds risky and inducing a positive slope in the term structure. This result is common to a variety of New-Keynesian models, e.g. JPT, Smets and Wouters (2004), and Christiano, Trabandt, and Walentin (2011). In this paper, the reason is that the central bank’s inflation target falls following positive technology shocks. Intuitively, a positive supply shock lowers inflationary pressure, which the central bank takes as an opportunity to drive inflation lower for an extended period. The fact that the negative correlation between technology shocks and interest rates is obtained in numerous other models that assume a constant inflation target suggests that this is in fact a well-
identified feature of the data.

Variation in risk aversion also makes an important contribution to the model’s ability to the term structure of interest rates, though. Standard statistical tests easily reject a model with constant risk aversion in favor of one with time-varying risk aversion. The pricing errors for bonds are smaller by a factor of three when risk aversion is allowed to vary over time. Movements in risk aversion account for a large fraction of the variance of the term spread, particularly outside of recessions.

While the variance decompositions imply that the pricing kernel is driven entirely by the labor-neutral technology and risk aversion shocks, I find that those two shocks have only minor effects on the dynamics of the real economy in the short-run. Risk aversion explains less than 5 percent of the variance of output, consumption, investment, and hours worked at business-cycle frequencies. The variable that is most responsive to the technology shock is hours worked, and the technology shock still explains only 25 percent of its variance. The variance decompositions also differ substantially from the results found by JPT. Whereas JPT find that investment technology shocks are an important driver of the business cycle, I find that they explain little other than investment, and monetary policy and markup shocks play much larger roles. This finding suggests that including information about bond prices in estimation has important effects on estimation results.

In addition to matching the behavior of the term structure, the estimated parameters imply reasonable behavior for equity prices. The steady-state annualized Hansen–Jagannathan bound is estimated to be 0.47, which is consistent with the observed Sharpe ratio for the stock market in the data sample, even though data on equity returns is not included in the estimation. Furthermore, the estimated degree of variation in risk aversion is similar to (though somewhat higher than) the value used in Dew-Becker (2011a), who calibrates a general-equilibrium model that can match the both the average Sharpe ratio on equities and also empirical stock return forecasting regressions. At business-cycle frequencies, estimated risk aversion displays similar behavior to Cochrane and Piazzesi’s (2005) tent-shaped bond return forecasting factor (and they are both strongly correlated with the term spread).

This paper is related to a small but growing literature on bond pricing in production
economies. Bekaert, Cho, and Moreno (2010) and Doh (2011) estimate New-Keynesian macro models, but they do not focus on the size and volatility of the term premium, whereas that is the feature of the term structure that this paper concentrates on. Rudebusch and Swanson (2011) generate a large and volatile term premium in a calibrated model. This paper moves beyond them by considering a substantially more complex model and showing that it can be dynamically estimated through standard Bayesian methods using the Kalman filter. Models of the business cycle have strong implications for the term structure of interest rates, so adding that information can have strong effects on estimation results. For example, I find that when the model is estimated without bond price information, the shock to investment technology is estimated to account for a large fraction of the variance of short-term interest rates and the term spread. But when the term spread is included as part of the information set, the effects of investment technology shocks are much smaller. The implications of the model for wage-setting also change when interest rates are added: I estimate a substantially larger Frisch elasticity than JPT do, coming in closer agreement with micro evidence.

The remainder of the paper is organized as follows. Section 3.2 describes household preferences and derives the pricing kernel. Section 3.3 describes the remainder of the economy including the production process, price setting, and monetary and fiscal policy.

Next, section 3.4 explains how the model is solved. If we used perturbation methods, a third-order approximation would be necessary to capture time-variation in risk premia. The estimation of the model turns out to be sufficiently difficult, however (due to numerous local extrema in the likelihood function, a common feature of models of the term structure), that the use of a nonlinear filter for calculating the model’s marginal likelihood is infeasible. I therefore use the essentially affine solution method described in Dew-Becker (2011b). The method approximates the pricing kernel separately from the remainder of the model, allowing it to take the essentially affine form with a time-varying price of risk described in Duffee (2002). The essentially affine method is equivalent to a first-order perturbation local to the non-stochastic steady-state, but it includes corrections for volatility that allow it to substantially outperform perturbation in stochastic simulations. The key feature of the essentially affine method is that risk premia may vary over time and affect
real variables, not just asset prices.

Section 3.5 describes the Bayesian methods used to estimate the model. Sections 3.6 and 3.7 examine the implications of the estimates for asset prices and the dynamics of the real economy, respectively. Finally, section 3.8 concludes.

3.2 Household preferences

3.2.1 Objective function and budget constraint

I assume the household has recursive preferences over consumption and leisure

\[
V_t = \left\{ (1 - \beta B_t) U (C_t, \bar{C}_{t-1}, N_t, Z_t) + \beta B_t \left( E_t V_{t+1}^{1-\alpha} \right)^{1-\rho} \right\}^{1-\rho} \tag{3.1}
\]

where \(C_t\) is consumption, \(\bar{C}_{t}\) is aggregate consumption, \(N_t\) is the number of hours worked outside the home, and \(E_t\) denotes the expectation operator conditional on information available at date \(t\). The term \(\bar{C}_{t-1}\) allows the period utility function to potentially include external habit formation. The level of technology, \(Z_t\), may also affect household utility in order to ensure balanced growth (as in Rudebusch and Swanson, 2010).

\(B_t\) is an exogenous shock to the household’s rate of time preference. The choice of exactly how to specify this preference shock is not trivial. The goal is to generate variation in consumption demand conditional on the level of interest rates. However, because \(B_t\) enters the value function, it may also affect the level of \(V_t\), and hence asset prices. The specification (3.1) has the feature that if period utility, \(U (C_t, \bar{C}_{t-1}, N_t, Z_t)\), is constant over time, then a change in \(B_t\) will have no effect on \(V_t\). So in some sense it purely affects the relative preference for consumption today versus in the future, as opposed to also affecting the household’s overall level of welfare.\(^1\) This specification thus imposes the restriction that intertemporal preference shocks are per se unpriced (in the sense that if they have no effect on consumption or leisure, they have no effect on the pricing kernel) since they have no direct effect on the level of welfare.

\(^1\) Variance decompositions for the estimated model reported below confirm that shocks to \(B_t\) have essentially no effect on the pricing kernel.
The household’s coefficient of relative risk aversion, $\alpha_t$, is allowed to vary over time. Dew-Becker (2011a) motivates variation in $\alpha_t$ by considering adding a time-varying benchmark to the standard Epstein–Zin certainty equivalent, $E_t (V_{t+1} - H_t)^{1-\alpha}$. When $V_{t+1}$ is close to $H_t$, the household’s effective risk aversion over shocks to $V_{t+1}$ rises. The formulation (3.1) has the advantage that it is log-linear and we do not have to worry about the possibility that $V_{t+1}$ falls below $H_t$. In Dew-Becker (2011a), movements in $\alpha_t$ are connected to movements in the household’s welfare. I loosen that constraint here and allow for independent shocks to risk aversion (equivalently, independent shocks to the habit). Melino and Yang (2003) study a similar specification, but without the emphasis on the habit. Unlike the intertemporal preference shock, since $\alpha_t$ directly affects the level of welfare, shocks to $\alpha_t$ will be per se priced – that is, even if they have no effect on consumption or leisure, they will still affect the pricing kernel through their impact on the level of welfare.

The household’s budget constraint is

$$P_t C_t + P_t I_t + H_t + D_t = (1 + i_t) H_{t-1} + W_t N_t + \Pi_t + R_{k,t} u_t \tilde{K}_{t-1} - P_t a(u_t) \tilde{K}_{t-1} + D_{t-1}$$

(3.2)

where $P_t$ is the price of the consumption good, $I_t$ is the expenditure on physical investment, $H_t$ is holdings of one-period nominal bonds, $D_t$ is cash holdings, $i_t$ is the nominally riskless interest rate, $W_t$ is the wage, and $\Pi_t$ represents profits and other lump-sum transfers paid to the household. $R_{k,t}$ is the rental rate on capital, $\tilde{K}$ the quantity of capital the household owns, and $u_t$ the fraction it chooses to rent (with associated costs $a(u_t)$). The dynamics of investment and capital accumulation will be discussed in more detail below. For now it is sufficient to simply note that the household sells labor and capital and allocates the proceeds between consumption and saving.

For the sake of simplicity, I study the so-called cashless economy described in Woodford (2003). The monetary authority is able to control the interest rate because money enters the household’s utility function, but the effect of money on total utility is sufficiently small that we can ignore it when writing $V$ (i.e. we take the limit where the relative importance of money goes to zero). I do not discuss money any further and from now on drop it from the household’s budget constraint.
3.2.2 The stochastic discount factor

In general, the stochastic discount factor under recursive preferences involves transformations of the household’s value function. It is often practically difficult to directly solve for the value function. As usual with Epstein–Zin preferences, it is possible in this setting to obtain an expression for the stochastic discount factor (SDF) involving consumption growth and an asset return. However, the asset whose return enters the SDF is no longer the household’s total wealth portfolio: it is now an asset that pays a dividend depending on the period utility function $U_t$ and the marginal utility of consumption.

The intertemporal marginal rate of substitution of consumption between neighboring dates is

$$M_{t+1} \equiv \frac{\partial V_t}{\partial C_t} = \beta B_t \frac{V_{t+1}^{\rho - \alpha_t}}{V_t^{\rho - \alpha_t}} \left( \frac{V_{t+1}^{\rho - \alpha_t}}{V_t^{\rho - \alpha_t}} - 1 \right)$$

where $U_{C,t} \equiv \partial U_t / \partial C_t$ is the marginal (period) utility of consumption. $M_{t+1}$ denotes the SDF between dates $t$ and $t + 1$.

In the case where $U_t = C^{1-\rho}_t$ and $B_t$ is constant, $M_{t+1}$ reduces to the usual formula for the SDF when utility depends only on consumption (e.g. Epstein and Zin, 1991). If the (period) marginal utility of consumption depends on labor, then the SDF will be distorted in the usual ways through $U_{C,t+1}$. Even if $U_C$ only depends on consumption, though (i.e. if period utility is separable between consumption and leisure), variation in labor will still affect the SDF through $V_{t+1}$: with recursive preferences, it is not generally possible to separate labor supply decisions from asset prices, unlike the case where preferences are separable between consumption and labor and over time.

3.2.2.1 Substituting in an asset return

Now consider an asset that pays $U_t U_{C,t}^{1-\rho}$ as its dividend in each period. In the usual analysis of Epstein–Zin preferences, one substitutes the return on an asset that pays consumption as its dividend into the SDF. In the present case, dividing period utility, $U_t$, by the marginal utility of consumption intuitively converts $U_t$ from utility units into consumption units.
We now derive the price of a claim to $U_t U_{C,t}^{-1}$. Denote the cum-dividend price of this asset as $W_{U,t}$. The appendix confirms the guess that

$$W_{U,t} = V_{t}^{1-\rho} B_t^{-1} U_{C,t}^{-1} / (1 - \beta B_t)$$

and that we can substitute the return on this asset into the SDF to obtain

$$M_{t+1} = \beta^{1-\rho} \left( \frac{U_{C,t+1}}{U_{C,t}} B_t \frac{1 - \beta B_{t+1}}{1 - \beta B_t} \right)^{\frac{\rho}{1-\rho}} R_{U,t+1}^{\rho/\rho}$$

where

$$R_{U,t+1} \equiv \frac{W_{U,t+1}}{W_{U,t} - U_t U_{C,t}^{-1}}$$

### 3.2.3 Period utility

The period utility function, $U (C_t, \bar{C}_{t-1}, N_t, Z_t)$ is motivated as a reduced form of a model of household production as in Rudebusch and Swanson (2010). Suppose households have power utility over both market goods and goods produced at home.

$$U_t = \frac{\left(C_t^{\eta} \bar{C}_{t-1}^{1-\eta}\right)^{1-\rho}}{1-\rho} + \varphi_1 \frac{C_{H,t}^{1-\rho}}{1-\rho}$$

where $C_{H,t}$ is consumption of the home good. Households do not derive utility directly from leisure, but rather from what they are able to produce in their non-market-work time (as in Campbell and Ludvigson, 2001). The home production function is $Z_t N_{H,t}^{\alpha_H}$ for hours worked at home $N_{H,t}$ and a coefficient $0 < \alpha_H < 1$. The level of labor-neutral technology in the economy is assumed to be equal (up to a constant of proportionality) in the home and market production sectors.\(^2\)

\(^2\) Note that in the household sector, an exogenous shift in $Z_t$, all else equal, raises output one-for-one, whereas below we will see that in the market sector it will raise output less than proportionally. The reason is that in the market sector, an increase in $Z_t$ also leads to an identical increase in the size of the capital stock. So, ultimately, the marginal product of labor in both sectors is proportional to $Z_t$.

One way to rationalize this slight elision would be if the household accumulates durable goods at home that aid household production. That feature of the model is left out for simplicity.
The period utility function can then be written as
\[
U_t = \left( \frac{\bar{C}_t^{\eta}C_{t-1}^{1-\eta}}{1-\rho} \right)^{1-\rho} + Z_t^{1-\rho} \phi_1 \left( \bar{H} - N_t \right)^{\alpha_H (1-\rho)}
\]
where \(N_{H,t} = \bar{H} - N_t\). \(\bar{H}\) denotes the maximum number of hours that the household can work, either at home or in the market, and \(N_t\) is market labor. If sleep is part of home production, then \(\bar{H}\) might equal 8760 hours for annual data. More generally, though, \(\bar{H}\) might be smaller. As a practical matter, \(\bar{H}\) affects both the elasticity of utility with respect to market labor and the Frisch elasticity. The three parameters \(\phi_1\), \(\bar{H}\), and \(\alpha_H\) jointly determine three primary features of household behavior: hours worked, the Frisch elasticity, and the elasticity of utility with respect to market labor.

The first term in (3.8) gives the utility that comes from consumption. The household has power utility over a Cobb–Douglas aggregate of current and (aggregate) past consumption. This formulation differs from the standard recent implementation in the macro literature in that I assume a multiplicative instead of additive habit. Campbell and Cochrane (1999) show that an additive habit can induce time-varying risk aversion, whereas the multiplicative habit will have no affect on risk aversion; that way, variation in risk preferences is driven purely by \(\alpha_t\). The key feature of the additive habit is simply that the marginal utility of current consumption is increasing in last period’s consumption, which induces consumers to try to smooth consumption growth, as observed in the data. To obtain that result in this setting (assuming \(0 < \eta < 1\)), we need \(\rho < 1\).

3.3 Aggregate supply

For the supply side of the model, I follow exactly the setup in Justiniano, Primiceri, and Tambalotti (JPT; 2010). JPT is a standard medium-scale New-Keynesian model. It has 7 fundamental shocks—price and wage markups, labor-augmenting technical change, investment-specific productivity, monetary policy, discount rates \((B_t)\), and government spending. In JPT’s formulation, the monetary authority’s inflation target is constant. I allow it to vary to help match the movements in the long end of the yield curve. Other
than that and the preference specification, my model is identical to theirs.

The model is also highly similar to Smets and Wouters (SW; 2003). The critical difference between the present setup and SW is that technology is difference-stationary rather than trend-stationary, where the former is standard in the production-based asset pricing literature.\(^3\) The difference-stationarity assumption helps generate large risk premia: when technology is trend-stationary, there is very little overall risk in the economy, so households must have an implausibly high coefficient of relative risk aversion in order to generate realistic asset prices.\(^4\)

Since the model is standard and laid out in JPT and the main contribution of this paper is the preference specification and bond pricing, the remainder of this section gives a relatively short description of the production setup. The appendix gives a full derivation of the model, and the reader is referred to JPT for a more detailed analysis. My description follows theirs closely.

### 3.3.1 Producers of physical goods

Final-good producers are competitive in both input and output markets and have a CES production function,

\[
Y_t = \left[ \int_0^1 Y_t(i)^{1+\lambda_p} \, di \right]^{1+\lambda_p},
\]

(3.9)

where \(i\) indexes the types of intermediate goods, \(Y_t\) is output of the final good, which can be used for either consumption or investment, \(Y_t(i)\) is the use of intermediate of type \(i\), and the elasticity of substitution across the intermediates, which determines markups in the intermediate-goods sector, varies over time.

Intermediate-good producers are monopolists for their own goods with production function

\[
Y_t(i) = \max \left\{ K_t(i)^{\gamma} Z_t^{1-\gamma} N_t(i)^{1-\gamma} - Z_t \bar{F}, 0 \right\}
\]

(3.10)

\(^3\) A difference-stationary process has first-differences that follow a stationary process, so it is integrated of order one. A trend-stationary process, on the other hand, is a process that has random stationary deviations around a non-stochastic trend (where the trend is generally unmodeled and taken as exogenous).

\(^4\) Below, I estimate average risk aversion to be 18.7 (ignoring the correction from Swanson, 2011). Rudebusch and Swanson (2011), who use stationary technology (with a slightly different preference specification) choose an analogous parameter to be 149.
where $\bar{F}$ is a fixed cost of production that ensures that profits are zero in steady state. $K_t(i)$ and $N_t(i)$ are intermediate-good producer’s $i$ purchases of capital and labor services, and $Z_t$ is the level of labor-augmenting technology.

### 3.3.2 Price setting

We assume Calvo pricing. In every period, a fraction $1 - \xi_p$ of intermediate good producers can change their prices, while the remainder index their prices following the rule,

$$ P_t(i) = P_{t-1}(i) \pi_{t-1}^{i_p} \pi_t^{1-i_p} \quad (3.11) $$

where $P_t(i)$ is the price of good $i$ in terms of the numeraire, $\pi_t \equiv P_t/P_{t-1}$ is aggregate inflation, and

$$ P_t = \left[ \int_0^1 P_t(i)^{\lambda_{p,t}} di \right]^{\lambda_{p,t}} \quad (3.12) $$

is the aggregate price index (equal to the marginal cost of a unit of the final good). $\pi$ is the steady-state inflation rate, and the parameter $\lambda_{p}$ determines the degree of indexation to lagged versus average inflation.

The firms that can choose their prices freely in a given period set them to maximize the present discounted value of profits over the period before they are allowed to choose a new price

$$ E_t \left\{ \sum_{s=0}^{\infty} \xi_p^s M_{t,t+s} \left[ P_t(i) \left( \prod_{k=1}^{s} \pi_{t+k-1}^{i_p} \pi_t^{1-i_p} \right) Y_{t+s}(i) - W_{t+s} N_{t+s}(i) - R_{k,t+s} K_{t+s} \right] \right\} \quad (3.13) $$

where $M_{t,t+s} \equiv \prod_{j=1}^{s} M_{t+j}$, $W_{t+s}$ is the wage rate, and $R_{t+s}^k$ is the rental rate for capital.

### 3.3.3 Employment agencies and wage setting

Each household is a monopolistic supplier of specialized labor, $N_t(j)$. Competitive employment agencies aggregate labor supply into a homogeneous labor input (just as the
final good producers aggregate intermediate goods) with the production function,

$$N_t = \left[ \int_0^1 N_t (j)^{1+\lambda_{w,t}} dj \right]^{1+\lambda_{w,t}} (3.14)$$

where, as with prices, $\lambda_{w,t}$ determines the elasticity of demand and hence markups in the labor market. $\lambda_{w,t}$ acts as a labor-supply shock. Since the employment agencies are competitive, the price of a unit of the homogeneous labor input is

$$W_t = \left[ \int_0^1 W_t (j)^{1+\lambda_{w,t}} dj \right]^{1+\lambda_{w,t}} (3.15)$$

The labor demand function is then

$$N_t (j) = \left( \frac{W_t (j)}{W_t} \right)^{1+\lambda_{w,t}} N_t (3.16)$$

As with prices, wages can only be changed intermittently, with probability $(1 - \xi_w)$. If a household cannot change its wage, it indexes according to the rule

$$W_t (j) = W_{t-1} (j) \left( \pi_{t-1} \frac{Z_{t-1}}{Z_{t-2}} \right)^{1/\xi_{w}} \left( \pi \exp (\gamma) \right)^{1-1/\xi_{w}} (3.17)$$

where $\gamma$ is the average growth rate of technology. The household will choose its wage in a manner similar to how the intermediate-good firms set prices: it maximizes expected utility over the period that the wage will remain unchanged.

### 3.3.4 Capital and investment

Intermediate-good firms rent capital from the households at rate $R_{k,t}$. Households own a stock of capital $\bar{K}_t$ and choose a utilization rate $u_t$. The effective quantity of capital rented to firms in period $t$ is

$$K_t = u_t \bar{K}_t (3.18)$$
The household pays a cost of utilization \( a (u_t) \) per unit of capital, with \( u = 1 \) in steady state, \( a (1) = 0 \) and \( \chi \equiv a'' (1) / a' (1) \).\(^5\)

Households accumulate capital according to the rule,

\[
\dot{K}_t = (1 - \delta) \dot{K}_{t-1} + \mu_t \left( 1 - S \left( \frac{I_t}{I_{t-1}} \right) \right) I_t
\]

where \( \delta \) is the depreciation rate and the function \( S \) incorporates adjustment costs in the rate of investment. In steady state, \( S = S' = 0 \) and \( S'' > 0 \). \( \mu_t \) is a shock to the cost of investment at date \( t \).

### 3.3.5 Government policy

The central bank follows a Taylor rule taking the form

\[
\frac{R_t}{R} = \left( \frac{R_{t-1}}{R} \right)^{\rho_R} \left[ \frac{\pi_t^*}{\pi_t} \right]^{\phi_{\pi}} \left( \frac{X_t}{\bar{X}_t} \right)^{\phi_X} \left\{ 1 - \rho_R \left[ \frac{X_t / X_{t-1}}{\bar{X}_t / \bar{X}_{t-1}} \right] \right\}^{\phi_{\mu_X}} \eta_{mp,t}
\]

where \( R_t \) is the gross nominal interest rate, \( R \) is its steady-state value, \( X_t \) is total output, \( X_t^* \) is the level of output that would prevail if prices had always been flexible, and \( \pi_t^* \) is the inflation target at date \( t \). The central bank is allowed to respond to both the level and change in the output gap. This flexibility helps ensure the model can match the dynamics of short-term interest rates, which is obviously critical for capturing the dynamics of the term structure. \( \eta_{mp,t} \) is an exogenous monetary policy shock.

\( \pi_t^* \) is a time-varying inflation target, which can potentially help match the high inflation and long-term interest rates seen in the early part of the sample. More generally, \( \pi_t^* \) induces a level factor in the term structure.

The government finances public spending by selling single-period bonds. Government expenditures, \( G_t \), are a time-varying fraction of total output,

\[
G_t = \left( 1 - \frac{1}{g_t} \right) Y_t
\]

\(^5\) As usual, in the log-linear approximation, the conditions on the first and second derivatives in steady state are sufficient to describe the dynamics of the model.
where $g_t$ follows an exogenous process defined below. Households receive no utility from government expenditures. As long as the share of output consumed by the government is stationary, that assumption will have minimal effects on asset prices.

### 3.3.6 Market clearing

The aggregate resource constraint is

$$C_t + I_t + G_t + a(u_t) \bar{K}_{t-1} = Y_t \quad (3.22)$$

### 3.3.7 Exogenous processes

The price and wage markup shocks follow ARMA(1,1) processes,

$$\log(1 + \lambda_{p,t}) = (1 - \rho_p) \log(1 + \lambda_p) + \rho_p \log(1 + \lambda_{p,t-1}) + \varepsilon_{p,t} - \theta_p \varepsilon_{p,t-1} \quad (3.23)$$

$$\log(1 + \lambda_{w,t}) = (1 - \rho_w) \log(1 + \lambda_w) + \rho_w \log(1 + \lambda_{w,t-1}) + \varepsilon_{w,t} - \theta_w \varepsilon_{w,t-1} \quad (3.24)$$

where $\varepsilon_{p,t} \sim N\left(0, \sigma_{p}^2\right)$ and $\varepsilon_{w,t} \sim N\left(0, \sigma_{w}^2\right)$. The ARMA(1,1) form potentially helps match both the high and low-frequency features of inflation.

Productivity has a unit root and its growth rate follows an AR(1) process,

$$\Delta z_t = (1 - \rho_z) \bar{z} + \rho_z \Delta z_{t-1} + \varepsilon_{z,t} \quad (3.25)$$

where $\varepsilon_{z,t} \sim N\left(0, \sigma_{z}^2\right)$. The AR(1) setup potentially allows the model to incorporate the long-run risks studied by Bansal and Yaron (2004).

The level of investment-specific productivity is assumed to be a stationary AR(1) process,

$$\log \mu_t = \rho_{\mu} \log \mu_{t-1} + \varepsilon_{\mu,t} \quad (3.26)$$

where $\varepsilon_{\mu,t} \sim N\left(0, \sigma_{\mu}^2\right)$. Note that $\mu_t$ simply determines the efficiency of the transformation of the final output good into the investment good, so investment still benefits from the unit-root innovations to $Z_t$.

The government’s share of output, the monetary policy shock, and the time-preference
The shock follow AR(1) processes,

\[
\log g_t = (1 - \rho_g) \log g + \rho_g \log g_{t-1} + \epsilon_{g,t}
\]

(3.27)

\[
\eta_{mp,t} = \rho_{mp} \eta_{mp,t-1} + \epsilon_{mp,t}
\]

(3.28)

\[
\log B_t = \rho_b \log B_{t-1} + \epsilon_{b,t}
\]

(3.29)

where \(\epsilon_{g,t} \sim N\left(0, \sigma^2_g\right)\), \(\epsilon_{mp,t} \sim N\left(0, \sigma^2_{mp}\right)\), and \(\epsilon_{b,t} \sim N\left(0, \sigma^2_b\right)\).

The two exogenous processes that are added to JPT’s original model are the inflation target and risk aversion. As in Dew-Becker (2011a), I allow the innovations to risk aversion to be correlated with \(\epsilon_{z,t}\). Intuitively, this means that risk aversion depends on innovations to the permanent component of consumption. There are also exogenous innovations to risk aversion. We thus have

\[
\alpha_t = \rho_a \alpha_{t-1} + (1 - \rho_a) \bar{\alpha} + \theta_{a,z} \epsilon_{z,t} + \epsilon_{a,t}
\]

(3.30)

with \(\epsilon_{a,t} \sim N\left(0, \sigma^2_a\right)\).

While a number of recent papers have studied models with time-varying inflation targets (e.g. Gurkaynak, Sack, and Swanson, 2005; Doh, 2010), there is little understanding of what actually drives the inflation target. Because the inflation target has a very strong impact on long-term bond prices, the relationship between the inflation target and the other innovations is a key determinant of the prices of long-term bonds. I therefore consider a loose specification where innovations to the inflation target may be correlated with all of the other fundamental shocks (excluding risk aversion). We thus have

\[
\log \pi^*_t = (1 - \rho_\pi) \log \pi + \rho_\pi \log \pi^*_{t-1} + \theta_{\pi, g} \epsilon_{g,t} + \theta_{\pi, z} \epsilon_{z,t} + \theta_{\pi, p} \epsilon_{p,t} + \theta_{\pi, w} \epsilon_{w,t} + \theta_{\pi, b} \epsilon_{b,t} + \theta_{\pi, \mu} \epsilon_{\mu,t} + \theta_{\pi, mp} \epsilon_{mp,t} + \epsilon_{\pi*,t}
\]

(3.31)

with \(\epsilon_{\pi*,t} \sim N\left(0, \sigma^2_{\pi*}\right)\). All of the shocks \(\epsilon\) are also assumed to be independent.

The \(\theta\) parameters in equations (3.30) and (3.31) are somewhat difficult to interpret and
choose priors for. I therefore transform these parameters so that they can be interpreted as variance shares. Define

\[ \sigma^2_{\alpha} \equiv \theta^2_{\alpha,z} \sigma^2_z + \hat{\sigma}^2_{\alpha} \] (3.32)

\[ \sigma^2_{\pi^*} \equiv \theta^2_{\pi^*,g} \sigma^2_g + \theta^2_{\pi^*,z} \sigma^2_z + \theta^2_{\pi^*,p} \sigma^2_p + \theta^2_{\pi^*,\mu} \sigma^2_\mu + \theta^2_{\pi^*,mp} \sigma^2_{mp} + \hat{\sigma}^2_{\pi^*} \] (3.33)

\( \sigma^2_{\alpha} \) and \( \sigma^2_{\pi^*} \) are the variances of the total innovations to risk aversion and the inflation target, respectively. Next, define

\[ \sigma_{\alpha,z} \equiv \text{sign} (\theta_{\alpha,z}) \frac{\theta^2_{\alpha,z} \sigma^2_z}{\sigma^2_{\alpha}} \] (3.34)

\( \sigma_{\alpha,z} \) is the share of the total variance of the innovations to risk aversion that is accounted for by labor-neutral technology shocks. The sign of \( \sigma_{\alpha,z} \) determines whether the effect of technology shocks on risk aversion is positive or negative. Similarly, for \( \pi^*_t \) we can define

\[ \sigma_{X,z} \equiv \text{sign} (\theta_{\pi^*,X}) \frac{\theta^2_{\pi^*,X} \sigma^2_X}{\sigma^2_{\pi^*}} \] (3.35)

for \( X \in \{g, z, p, w, b, \mu, mp\} \). The parameters \( \sigma_{\alpha,z}, \sigma_{X,z}, \sigma^2_{\alpha}, \) and \( \sigma^2_{\pi^*} \) map uniquely into the original parameters \( \theta_{\alpha,z}, \theta_{\pi^*,X}, \sigma^2_{\alpha}, \) and \( \sigma^2_{\pi^*} \) are more easily interpreted as they represent variances and signed variance shares.

3.4 Model solution

The standard method for approximating models of the form studied here is perturbation. The drawback of perturbation methods for our purposes is that if we want time-variation in risk aversion to have any effect on the dynamics of the model, we need to take a third-order approximation to the model. Since the solution would be non-linear, we would have to use the particle filter or some other nonlinear method in order to calculate the marginal likelihood of the model. I have found, though, that it is in general very difficult to find the peak of the likelihood function for this model, and it would be infeasible with a method as slow as the particle filter. This is a common problem in models of the
I therefore use the essentially affine approximation method described in Dew-Becker (2011b). The essentially affine method delivers an approximation to the equilibrium dynamics of the model that is linear in the state variables but still allows time-varying risk aversion to affect the behavior of the endogenous variables. Dew-Becker (2011b) describes the method in detail and show that Euler equation errors in simulated models are competitive with third-order perturbations. Local to the non-stochastic steady-state, the essentially affine approximation is as accurate as a first-order perturbation (in a Taylor sense), and hence less accurate than higher-order perturbations. However, in a stochastic setting, it performs well. This section gives a short overview of the method, and the appendix provides further details.

Denote the vector of the variables in the model (including the exogenous processes) as \( X_t \) and the vector of fundamental shocks as \( \varepsilon_t \equiv [\varepsilon_{mp,t}, \varepsilon_{z,t}, \varepsilon_{b,t}, \varepsilon_{g,t}, \varepsilon_{w,t}, \varepsilon_{a,t}, \varepsilon_{\pi,t}] \). The equations determining the equilibrium of the model take the form

\[
0 = G (X_t, X_{t+1}, \sigma \varepsilon_{t+1}) \tag{3.36}
\]

where the expectation operator may appear in the function \( G \). There is one equation for each variable. \( \sigma \) is a parameter controlling the variance of the shocks. We will approximate around the point \( \sigma = 0 \), with the non-stochastic steady-state defined as the point \( \bar{X} \) such that

\[
0 = G (\bar{X}, \bar{X}, 0)
\]

The equations \( G \) can be divided into two types: those that do not involve taking expectations over the SDF and those that do.

\[
G (X_t, X_{t+1}, \varepsilon_{t+1}) = \begin{bmatrix}
D (X_t, X_{t+1}, \sigma \varepsilon_{t+1}) \\
\varepsilon_t [M (X_t, X_{t+1}, \sigma \varepsilon_{t+1}) F (X_t, X_{t+1}, \sigma \varepsilon_{t+1})]
\end{bmatrix} \tag{3.37}
\]

where \( D \) and \( F \) are vector-valued functions and \( M \) is the (scalar-valued) stochastic discount
For the equations that do not involve the SDF, I use standard perturbation methods and simply take a log-linear approximation. We approximate $D$ as

$$
0 = \log \left( D \left( \exp \left( x_t \right), \exp \left( x_{t+1} \right), \sigma \epsilon_{t+1} \right) + 1 \right)
$$

(3.38)

$$
0 \approx d_0 + d_x \hat{x}_t + d_{x'} \hat{x}_{t+1} + d_{\epsilon} \sigma \epsilon_{t+1}
$$

(3.39)

where the terms $d_0$, $d_x$, $d_{x'}$, and $d_{\epsilon}$ are coefficients from a Taylor approximation and

$$
x_t \equiv \log X_t
$$

$$
\hat{x}_t \equiv \log X_t - \log \bar{X}
$$

$D$ will include equations such as budget constraints.

The second set of equations is dynamic and involves expectations. In many economic models, including the present one, equations involving expectations take the form

$$
1 = E_t \left[ M \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right) F \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right) \right]
$$

(3.40)

where $M \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right)$ is the stochastic discount factor induced by the household’s intertemporal optimization condition. The key source of non-linearity in the model is the time-variation in risk aversion, which induces heteroskedasticity in the SDF. It is therefore natural to deal with $M$ and $F$ separately to isolate the relevant non-linearity.

I now show that if we log-linearize $F$, we can transform (3.40) into a linear condition that can be solved alongside the remaining equations. $M \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right)$ will not even be log-linear in the state variables, but we will be able to state the equilibrium conditions in as a set of linear expectational difference equations.

---

Note that this formulation does not actually restrict $F$. Specifically, suppose there were a set of equilibrium conditions $1 = E_t h \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right)$, i.e., that do not involve the SDF. We could simply say that $F \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right) \equiv h \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right) / M \left( X_t, X_{t+1}, \sigma \epsilon_{t+1} \right)$. 

---
First, guess that the equilibrium dynamics of the model take the form

\[ \hat{x}_{t+1} = C + \Phi \hat{x}_t + \Psi \epsilon_{t+1} \]  

(3.41)

We confirm in the end that the solution is actually in this form.

The next step then is to take log-linear approximations to \( M \) and \( F \) separately. Log-linearizing \( F \) is straightforward, and we obtain,

\[ \log F (x_t, x_{t+1}, \sigma \epsilon_{t+1}) \approx f_0 + f_x x_t + f_{x'} x_{t+1} + f_\epsilon \sigma \epsilon_{t+1} \]  

(3.42)

For \( M \), in the case of the preferences laid out in section 3.2, the appendix shows that it is possible to derive a first-order accurate expression of the form

\[ m^{(1)}_{t+1} = m_0 + m_x \hat{x}_t + (\kappa_0 + \alpha_t \kappa_1) \sigma \epsilon_{t+1} - \frac{1}{2} \alpha_t^2 \kappa_1 \Sigma \kappa_1' \]  

(3.43)

where \( \Sigma \) is the variance matrix of \( \epsilon_t \). The superscript (1) indicates that \( m^{(1)}_{t+1} \) is first-order accurate for the true SDF. (3.43) is the essentially affine form from Duffee (2002).

Taking the expectation of the approximated Euler equation yields,

\[ 0 = \log E_t \exp \left( m_0 + m_x \hat{x}_t + (\kappa_0 + \alpha_t \kappa_1) \sigma \epsilon_{t+1} - \frac{1}{2} \alpha_t^2 \kappa_1 \Sigma \kappa_1' \right) \]

\[ + f_0 + f_x x_t + f_{x'} x_{t+1} + f_\epsilon \sigma \epsilon_{t+1} \]

\[ = m_0 + m_x \hat{x}_t + f_0 + f_x x_t + f_{x'} (C + \Phi \hat{x}_t) \]

\[ + \frac{1}{2} \sigma^2 (f_{x'} + f_\epsilon) \Psi \Sigma \Psi' (f_{x'} + f_\epsilon) + \alpha_t \sigma^2 \kappa_1 \Sigma \Psi' (f_{x'} + f_\epsilon) \]

Since every equation in the system is now linear in the variables of the model, we can solve the system for the parameters \( \Phi \) and \( \Psi \) from (3.41). Specifically, we solve the
following system,

\[
0 = d_0 + d_x \hat{x}_t + d_x' \hat{x}_{t+1} + d_\varepsilon \varepsilon_{t+1} \\
0 = m_0 + m_x \hat{x}_t + f_0 + f_x x_t + f_x' (C + \Phi \hat{x}_t) \\
+ \frac{1}{2} \sigma^2 (f_x' + f_x) \Psi \Sigma \Psi' (f_x' + f_x') + \alpha_t \sigma^2 \kappa \Sigma \Psi' (f_x' + f_x')
\]  

(3.44)  

(3.45)

at the point \( \sigma = 1 \). The reason that the essentially affine SDF is useful is that the expectation in (3.45) will be linear in the state variables, so we have a simple linear system to solve. This system can be solved through, for example, Sims’ (2001) Gensys algorithm.

Dew-Becker (2011b) shows that the transition function for the model obtained through the essentially affine method is first-order accurate for the true transition function and first-order equivalent to a first-order perturbation. Clearly, though, the approximation includes higher-order terms that account for movements in risk aversion. \( \alpha_t \) will affect not only asset prices but also the dynamics of real variables. Dew-Becker (2011b) calibrates a simple version of the RBC model with time-varying risk aversion and finds that the essentially affine approximation has accuracy between that of second and third-order perturbations.

Standard results derived in the appendix also deliver real and nominal zero-coupon bond prices.

3.5 Empirics

I estimate the model using standard Bayesian methods. The observable data is the same as in JPT, but with bond prices added. Both real variables and bond prices are linear functions of the underlying state variables contained in the vector \( x_t \), so we can write the model in state-space form and measure the likelihood using the Kalman filter. I proceed by finding the posterior mode and running a monte carlo chain from that point to sample from full posterior distribution. The appendix describes the details of the estimation.
3.5.1 Data

The sample is 1983q1 to 2004q4. I do not include the financial crisis in the sample because the zero lower bound on nominal interest rates becomes binding, a phenomenon that the model is not designed to capture. The sample is cut off in 1983 in order to ensure that monetary policy is consistent over the estimation period.

The observable variables are real GDP, consumption, and investment growth, hours worked per capita, wage and price inflation, and yields on three-month, 1, 2, 3, 5, and 10-year Treasury bonds. The 1 through 5-year yields are obtained from the Fama-Bliss CRSP files, the 10-year yield from Gurkaynak, Sack, and Wright (2006), and the three-month yield from the Fama risk-free rate CRSP file. The bond yields and inflation rates are always reported in annualized percentage points, unless otherwise noted. The real variables are all obtained from the BEA and the BLS. Consumption is defined as expenditures on non-durables and services, while investment is the sum of residential and non-residential fixed investment and consumer durables expenditures. Real wages are calculated as nominal compensation per hour in the non-farm business sector (from the BLS) divided by the GDP deflator. The change in the log GDP deflator is the measure of inflation. Hours worked per capita in the non-farm business sector are obtained from Francis and Ramey (2009) as updated on Valerie Ramey’s website. None of the variables are detrended.

Figure 3.1 plots the data used in the estimation (with the exception of the intermediate-term bond yields). Output, consumption, and investment growth all look stationary over the sample and relatively homoskedastic. Hours worked per capita has a strong upward trend in this sample. Interest rates decline significantly over the sample, even though inflation only declines marginally. The short-term interest rate is substantially more volatile than the long-term rate, and the term spread is clearly countercyclical.

The model has 9 fundamental shocks, but we have 13 observable variables. I follow JPT in assuming that the 6 macro variables plus the short-term interest rate are observed without error. I also assume that the 10-year bond yield is measured without error, which will help identify the inflation target. For the remaining bonds, I assume that the yields have i.i.d. measurement errors with identical standard deviations. The standard deviation
Figure 3.1: Data series for estimation

Note: No variables are detrended. GDP, consumption, and investment are obtained from the BEA. Compensation per hour, and inflation are obtained from the BLS. Hours worked is obtained from Valerie Ramey’s website. The one-quarter yield is the Fama risk-free rate. The ten-year yield is from Gurkaynak, Sack, and Wright (2006).
of these measurement errors is another parameter that will be estimated. The assumption of zero measurement error for the long and short ends of the yield curve forces the model to focus on matching the term spread, while leaving some flexibility in matching curvature.

### 3.5.2 Priors

Table 3.1 lists the parameters and priors. For all of the parameters that I share with JPT, I choose the same priors. The remaining parameters are listed in the bottom section of the table. Many of them have uniform priors since I do not have strong a priori views about, for example, the fraction of the variance of the Federal Reserve’s inflation target that is driven by shocks to government spending.

For the volatility of risk aversion, I choose a beta distribution over the ratio of the unconditional standard deviation of risk aversion to its mean. This means that average risk aversion is forced to be at least one standard deviation above zero. This prior could potentially be tightened to enforce a stronger restriction. As a practical matter, the data tends to push for a high volatility for risk aversion, and average risk aversion in the estimation simply rises high enough to accommodate the unconditional standard deviation.

I constrain the persistence of the inflation target to follow nearly a random walk with $\rho_{\pi} = 0.99$, consistent with the idea that the target is highly persistent. The assumption that $\rho_{\pi} < 1$ ensures that inflation is stationary so that there is a steady-state around which we can approximate. The priors over the shares of the variances of the inflation target and risk aversion coming from the other shocks are uniform.

### 3.5.3 Posterior modes

Table 3.1 lists the posterior modes for the parameters along with the 5th and 95th percentiles of the posterior distribution. Many of the posterior modes are reasonably close to the corresponding prior means. I focus mainly on those parameters that differ from the prior or are unique to this model.

The prior for the variance of the innovations to the inflation target favors a reasonably low standard deviation, but the posterior seems to want a highly volatile target—the
### Table 3.1: Priors and posterior modes

<table>
<thead>
<tr>
<th>Description</th>
<th>Distribution</th>
<th>Priors</th>
<th>Posterior</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Capital share</td>
<td>Normal</td>
<td>Mean: 0.3</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>Wage indexation</td>
<td>Beta</td>
<td>Mean: 0.15</td>
<td>0.59</td>
</tr>
<tr>
<td>3</td>
<td>Price markup</td>
<td>Beta</td>
<td>Mean: 0.15</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>Mean technology growth</td>
<td>Normal</td>
<td>Mean: 0.25</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>Price adjustment frequency</td>
<td>Beta</td>
<td>Mean: 0.5</td>
<td>0.52</td>
</tr>
<tr>
<td>6</td>
<td>Mean wage markup</td>
<td>Normal</td>
<td>Mean: 0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>7</td>
<td>Mean log hours per capita</td>
<td>Normal</td>
<td>Mean: 6.7</td>
<td>0.65</td>
</tr>
<tr>
<td>8</td>
<td>Mean quarterly inflation</td>
<td>Normal</td>
<td>Mean: 0.5</td>
<td>-0.20</td>
</tr>
<tr>
<td>9</td>
<td>Discount factor</td>
<td>Gamma</td>
<td>Mean: 0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>10</td>
<td>Inverse Frisch elasticity</td>
<td>Gamma</td>
<td>Mean: 1.2</td>
<td>1.83</td>
</tr>
<tr>
<td>11</td>
<td>Taylor rule inflation</td>
<td>Normal</td>
<td>Mean: 1.7</td>
<td>0.89</td>
</tr>
<tr>
<td>12</td>
<td>Technology shock AR</td>
<td>Beta</td>
<td>Mean: 0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>13</td>
<td>Government spending AR</td>
<td>Beta</td>
<td>Mean: 0.6</td>
<td>0.99</td>
</tr>
<tr>
<td>14</td>
<td>Investment technology AR</td>
<td>Beta</td>
<td>Mean: 0.6</td>
<td>0.50</td>
</tr>
<tr>
<td>15</td>
<td>Neutral tech. shock vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.80</td>
</tr>
<tr>
<td>16</td>
<td>Gov't spending vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.29</td>
</tr>
<tr>
<td>17</td>
<td>Investment tech. vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.29</td>
</tr>
<tr>
<td>18</td>
<td>Price markup AR</td>
<td>Beta</td>
<td>Mean: 0.6</td>
<td>0.99</td>
</tr>
<tr>
<td>19</td>
<td>Wage markup AR</td>
<td>Beta</td>
<td>Mean: 0.6</td>
<td>0.77</td>
</tr>
<tr>
<td>20</td>
<td>Monetary policy AR</td>
<td>Beta</td>
<td>Mean: 0.4</td>
<td>0.19</td>
</tr>
<tr>
<td>21</td>
<td>Price markup MA</td>
<td>Beta</td>
<td>Mean: 0.5</td>
<td>0.16</td>
</tr>
<tr>
<td>22</td>
<td>Wage markup MA</td>
<td>Beta</td>
<td>Mean: 0.5</td>
<td>0.98</td>
</tr>
<tr>
<td>23</td>
<td>MP shock vol.</td>
<td>IG(1)</td>
<td>Mean: 0.1</td>
<td>0.14</td>
</tr>
<tr>
<td>24</td>
<td>Neutral tech. shock vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.80</td>
</tr>
<tr>
<td>25</td>
<td>Gov't spending vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.29</td>
</tr>
<tr>
<td>26</td>
<td>Investment tech. vol.</td>
<td>IG(1)</td>
<td>Mean: 0.5</td>
<td>0.29</td>
</tr>
<tr>
<td>27</td>
<td>Price markup vol.</td>
<td>IG(1)</td>
<td>Mean: 0.1</td>
<td>0.12</td>
</tr>
<tr>
<td>28</td>
<td>Wage markup vol.</td>
<td>IG(1)</td>
<td>Mean: 0.1</td>
<td>0.35</td>
</tr>
<tr>
<td>29</td>
<td>Demand shock vol.</td>
<td>IG(1)</td>
<td>Mean: 1</td>
<td>0.41</td>
</tr>
<tr>
<td>30</td>
<td>Inflation target vol.</td>
<td>IG(1)</td>
<td>Mean: 0.1</td>
<td>0.33</td>
</tr>
<tr>
<td>31</td>
<td>MP var. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.39</td>
</tr>
<tr>
<td>32</td>
<td>gvar. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.00</td>
</tr>
<tr>
<td>33</td>
<td>mu var. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.01</td>
</tr>
<tr>
<td>34</td>
<td>Price shock var. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>-0.04</td>
</tr>
<tr>
<td>35</td>
<td>Wage shock var. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>-0.06</td>
</tr>
<tr>
<td>36</td>
<td>bvar. shr. in π*</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.28</td>
</tr>
<tr>
<td>37</td>
<td>RRA volatility/RRA mean</td>
<td>Beta</td>
<td>Mean: 0.5</td>
<td>0.95</td>
</tr>
<tr>
<td>38</td>
<td>zvar. shr. in RRA</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.00</td>
</tr>
<tr>
<td>39</td>
<td>RRA persistence</td>
<td>U[1,1]</td>
<td>Mean: 0.58</td>
<td>0.77</td>
</tr>
<tr>
<td>40</td>
<td>Inverse EIS</td>
<td>U[1,1]</td>
<td>Mean: 0.5</td>
<td>0.76</td>
</tr>
<tr>
<td>41</td>
<td>Mean risk aversion</td>
<td>Normal</td>
<td>Mean: 15</td>
<td>18.70</td>
</tr>
<tr>
<td>42</td>
<td>Bond measurement errors (bp)</td>
<td>IG(1)</td>
<td>Mean: 13</td>
<td>8.20</td>
</tr>
</tbody>
</table>

Note: Priors, posterior mode, and percentiles of the posterior distribution from the benchmark model. The far-right column reports the parameters from JPT, where applicable.
estimated standard deviation of the innovations to the annualized inflation target is 1.3 percent. This helps the model capture the observed volatility of the level factor in bond yields, but it is implausibly high.

The shock to the level of labor-neutral technology has an important effect on the inflation target, accounting for 16 percent of the variance of its innovations. Following a positive innovation to technology, the central bank is estimated to lower its inflation target, consistent with the idea that following beneficial supply shocks that drive inflation downward, the central bank takes the opportunity to drive inflation lower persistently (e.g. Gurkaynak, Sack, and Swanson, 2005). This mechanism will turn out to be critical to the main results.

The labor-neutral technology shock has a standard deviation of 0.72 and an autocorrelation of 0.22. The permanent component of the technology process (the Beveridge–Nelson trend) thus has a standard deviation of 1.00, which is similar to the values often calibrated in the production-based asset pricing literature (e.g. Gourio, 2010, and Dew-Becker, 2011a). The estimated long-run variance of technology growth is far smaller than the value calibrated in the long-run risks literature (e.g. Bansal and Yaron, 2004, and Kaltenbrunner and Lochstoer, 2010), but it is consistent with estimates obtained in JPT and SW and with simple univariate estimates from consumption and output data.

The standard deviation of the investment technology shock is far larger than the prior, which is also consistent with JPT, showing that innovations that are isolated to the investment sector may play a large role in fluctuations. Alternatively, it could mean that the model simply matches investment poorly.

The estimates imply that there is essentially zero correlation between innovations to technology and innovations to risk aversion. This runs against the theory from Dew-Becker (2011a), and implies that the price of risk in bond markets is driven by some factor other than permanent innovations to household consumption. Intuitively, part of the source of this result is that the price of risk is measured to be high in recessions, but over the 1983–2004 period, productivity growth has been only weakly procyclical, and in fact rose substantially in the 2000 recession.

As in JPT, the government spending shock is estimated to follow nearly a unit root,
explaining the trend in the consumption-output ratio over the sample. The wage markup shock also follows nearly a unit root, which helps capture the strong trend in hours worked per capita seen in figure 3.1.

In general, my parameter estimates are reassuringly similar to the values from JPT reported in the far-right column of table 3.1, even though I use a different sample period (post-1983 versus post-war) and extra data on bond yields. The main place where my estimates seem to differ from JPT is in price and wage determination. My estimates imply that wages are strongly indexed to inflation, whereas JPT estimate little indexation. The wage-markup shock is also substantially more volatile under my estimates than JPT. Interestingly, when bond prices are dropped from the estimation, I obtain values for wage indexation and the volatility of the wage markup shock that are much closer to JPT. This suggests that wage dynamics are important for matching the term structure. I also estimate a smaller inverse Frisch elasticity and a lower average price markup. In both cases, my estimates bring the model in closer to the priors and micro estimates.

3.6 Asset pricing

This section studies the asset-pricing implications of the model. I first analyze the fit of the model to the term structure and show that it is competitive with a non-structural model. Next, I decompose the variance of the SDF to understand the source of the positive term premium in the model. I then analyze the prices of other assets, including the aggregate capital stock and a claim to aggregate profits.

3.6.1 Bond prices

3.6.1.1 Fitted yields

Figure 3.2 plots the deviations of the fitted yields from their actual values for the five yields that are assumed to be measured with error (reported in annualized basis points). The estimated standard deviation of these fitting errors is 8 basis points, which is economically small compared to the overall variation of the yields that is on the order of hundreds
of basis points. The errors are all centered around zero, meaning that the model can capture the shape of the term structure on average. The volatility of the errors looks somewhat higher for the 1 and 5-year yields and in the earlier part of the sample. There is clearly some autocorrelation in the errors; the fitted value for the 3-year yield is consistently too high in the first half of the sample, and the 4-year fitted yield is consistently too low in the second half, for example. And there is also some cross-correlation in the errors; the first principal component explains 37 percent of the total variance of the errors (twice what it would if the errors were orthogonal). These are thus clearly not classical (i.i.d.) measurement errors, but their small mean and volatility shows that the model does a reasonable job of fitting the data, and they are not disturbingly far from white noise.

While there nine unobservable shock processes that can help us match the data, the model is asked to fit 13 data series, so obtaining a good fit for the bond yields is not trivial. Loosely, we have 6 macro variables that identify 6 shock processes, plus three extra processes (the monetary policy shock, the inflation target, and risk aversion) that can be used to fit the bond yields. The degrees of freedom here are thus comparable to a non-structural bond-pricing model with three unobservable factors, but we also have numerous constraints on dynamics and risk prices. Table 3.2 lists the standard deviation of the yield errors in basis points obtained from regressing the bond yields on their first three principal components. The standard deviations are all between 4 and 8 basis points. I force the structural model to match the 1-quarter and 10-year yields exactly, and the remaining yields have errors with standard deviations of 8 basis points. The fit of the model to the yields is thus comparable to a non-structural model with three unobservable factors that have completely unrestricted dynamics.

The third and fourth rows of table 3.2 report the measurement errors in constrained models that assume constant relative risk aversion and a constant inflation target (the other parameters are reesetimated). In both cases, the measurement errors have standard deviations roughly three times larger than the benchmark model, giving a measure of the improvement in fit generated by the benchmark model.

Figure 3.2 and table 3.2 show that the model is able to provide a very close fit to bond yields in the data. The quality of the fit is essentially identical to the that of a purely non-
Figure 3.2: Bond yield errors

Note: each axis plots the measurement errors in basis points for one of the bond yields. Errors are measured from the Kalman-filtered estimates at the posterior mode.
Table 3.2: Fitting Errors

<table>
<thead>
<tr>
<th></th>
<th>1-quarter</th>
<th>1-year</th>
<th>2-year</th>
<th>3-year</th>
<th>4-year</th>
<th>5-year</th>
<th>10-year</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>4.54</td>
<td>7.32</td>
<td>6.29</td>
<td>4.26</td>
<td>6.20</td>
<td>7.98</td>
<td>6.53</td>
</tr>
<tr>
<td>Benchmark model</td>
<td>0</td>
<td>8.12</td>
<td>8.12</td>
<td>8.12</td>
<td>8.12</td>
<td>8.12</td>
<td>0</td>
</tr>
<tr>
<td>Constant RRA</td>
<td>0</td>
<td>23.30</td>
<td>23.30</td>
<td>23.30</td>
<td>23.30</td>
<td>23.30</td>
<td>0</td>
</tr>
<tr>
<td>Constant π*</td>
<td>0</td>
<td>23.87</td>
<td>23.87</td>
<td>23.87</td>
<td>23.87</td>
<td>23.87</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Fitting errors measured in annualized basis points. The model-based estimates use the posterior modal estimate for the standard deviation. The 1-quarter and 10-year errors are constrained to equal zero in the structural model. The errors from PCA are the standard deviations of the residuals from regressions on the bond yields on their first three principal components.
Another way to evaluate the fit of the model is to ask whether the steady state of the model matches the average term structure in the data. Looking at the steady state keeps the Kalman filter from using large deviations in the unobservable state variables to fit the term structure. Figure 3.3 plots the average term structure in the sample along with its model-implied steady state. The solid black line gives the steady-state term structure in the model, renormalized so that the ten-year yield matches the empirical ten-year yield. To capture the uncertainty in the empirical term structure, the grey area gives the 95-percent confidence intervals for the means of the empirical yields relative to the ten-year yield (i.e. the confidence intervals for the spreads; the intervals are calculated using the Newey-West method with lag a 6-quarter lag window). What figure 3.3 shows is that the model matches the spread between the 10 and 2-year yields, but it does not match the curvature of the term structure below two years. However, all of the model-implied yields are within the 95 confidence intervals. One potential explanation at the very short end of the yield curve is that there is a small liquidity premium that the model is not incorporating.

We will see that two features of the model are critical for generating the large steady-state term premium: first, following a positive shock to technology, the Fed’s inflation target falls; second, variation in risk aversion raises the premia on risky assets. To see the prima facie evidence that these two effects are key, figure 3.3 includes two lines giving the steady-state term structure in constrained models. The first line assumes that innovations to the inflation target are uncorrelated with the permanent technology shock, while the second line assumes that risk aversion is fixed. Neither line reestimates the other parameters, so they simply isolate the effects of those two features of the model.

The line exiting the top of the chart is for the model when shocks to technology are assumed to have no impact on the inflation target. We then obtain the usual result that the term structure is downward-sloping, and the steady-state term spread is -261 basis points.

---

7 I use this normalization because the estimated inflation target is above zero through most of the sample. The unconditional variance of the inflation target is sufficiently high that its average level is not well identified.
Figure 3.3: Steady-state nominal bond yields

Note: The solid black line gives the yield curve at the model’s steady state; the grey lines are for the model with constant risk aversion and where the inflation target is unaffected by shocks to labor-neutral technology. The other parameters are not reestimated. All the linear are normalized to match the 10-year yield exactly, so the plot measures steady-state spreads. Boxes are average sample yields. The grey area is the 95% confidence band for the average yields relative to the 10-year yield, calculated using the Newey–West method with 6 lags. The solid black line gives the yield curve at the model’s steady state, normalized to match the 10-year yield exactly.
Time-varying risk aversion also turns out to be important, though. When risk aversion is fixed, the term structure is still upward-sloping, but the spread is quantitatively small—only 63 basis points in steady-state, compared to 207 in the data and 152 in the benchmark model.

3.6.1.3 Term premia

The size of the steady-state term spread in the model can be interpreted as the average term premium—it is the excess return (in logs) that an investor earns in expectation by buying a long-term bond and holding it to maturity instead of buying short-term bonds and rolling them over for the same amount of time. An important feature of this model is that risk aversion varies over time, which should make the term premium also vary over time.

The top panel of figure 3.4 plots the expected annualized excess return on holding a ten-year nominal bond (over a one-quarter bond) from the benchmark model against the expected excess return from a regression of bond returns on the Cochrane–Piazzesi (CP) factor. Cochrane and Piazzesi (2005) argue that a tent-shaped factor in forward yields summarizes the price of risk in the term structure, so their factor can be viewed as a simple non-structural benchmark for return forecasting.

The structural forecast is highly correlated (34 percent) with the fitted value using the CP factor, and its standard deviation is roughly 20 percent larger. The two series rise by similar amounts in the two recessions in the sample, but the benchmark model also implies that the term premium rose in 1988 and 1999, whereas the CP factor is stable in those episodes.

The bottom panel of figure 3.4 plots the term premium against the term spread. The term premium is defined as the spread between the 10-year yield and the average of the expected 1-quarter yields over the life of a 10-year bond. The variance of the term premium is non-trivial in comparison to the term spread. In the two recessions in the sample, the increases in the term spread are substantially larger than the movements in the term premium, but the term premium does rise in both episodes. Interestingly, the movements
Figure 3.4: Expected returns and the term premium

Note: Top panel gives expected excess returns on a 10-year bond over the following quarter, annualized. Values for the Cochrane–Piazzesi are from a linear regression. The term premium in the bottom panel is defined as the gap between the 10-year nominal yield and the mean of expected 1-quarter yields over the following 10 years.

125
in the term spread outside of the two recessions seem almost entirely driven by movements in the term premium. In particular, the rises in the term spread in 1984, 1985, 1987, 1996, and 1999 are all associated with increases in the term premium of equal magnitudes. On the other hand, the inversions of the yield curve in 1989 and 2000, both just prior to recessions, are associated with only minor declines in risk aversion, and the subsequent rises in the term spread with similarly small rises in risk aversion.

3.6.1.4 Variation in interest rates

To show how the bond yields respond to the various shocks, figure 3.5 plots responses of a level, slope, and curvature factor to the 9 fundamental shocks. Following Bekaert, Cho, and Moreno (2010), the level factor is defined as the average of the 1-quarter and 5 and 10-year yields; the slope factor is the 10-year/1-quarter term spread; and curvature is the sum of the 5 and 1-year yields minus twice the 3-year yield. The shocks are orthogonalized in the sense that the interactions between the inflation target and risk aversion and the other shocks are switched off. So figure 3.5 shows, for example, the effect of a pure increase in the level of technology, holding the inflation target fixed. The shocks are all unit standard deviations.

For the level factor, a number of shocks, the monetary policy and time preference shocks in particular, have important effects at high frequencies. The low-frequency movements, as we would expect, are mainly driven by shifts in the inflation target, while risk aversion also plays a role. Somewhat surprisingly, positive monetary policy shocks, which raise the short-term interest rate above its Taylor-rule value, are actually associated with declines in the level factor. The reason is that these shocks drive down expected inflation. So a positive monetary policy shock drives the real interest rate up, but nominal interest rates actually fall. The response to the time-preference shock is more intuitive: an increase in $B_t$ is analogous to an increase in patience, so interest rates fall.

The determinants of the slope factor are similar to those for the level factor: monetary policy and time-preference matter at high frequencies, while risk aversion determines the dynamics at lower frequencies. An increase in risk aversion increases the term spread,
Figure 3.5: Responses of term structure factors to orthogonalized shocks

Note: responses of each of the term structure factors to the orthogonal structural shocks. Specifically, risk aversion and the inflation target are only affected by their own shocks, not the shocks to the other exogenous processes. The level factor is the average of the 1-quarter, 5-year, and 10-year yields. The slope factor is the gap between the 10-year and 1-year yields. Curvature is the sum of the 5-year and 1-year yields minus twice the 3-year yield. The shocks are all unit standard deviations. All scales in each row are identical and are measured in annualized percentage points.
which fits with results on bond return forecasting (Campbell and Shiller, 1988) and the fact that the term spread forecasts high equity returns (Fama and French, 1989).

3.6.2 Determinants of asset prices

3.6.2.1 The variance of the SDF

An asset’s expected excess return over the real riskless interest rate is determined by its covariance with the stochastic discount factor. One of the more interesting outputs of a model as rich as this one is the variance decomposition for the SDF. Table 3.3 reports a variance decomposition for the SDF at the one-quarter horizon. The variance of the SDF is essentially entirely driven by the neutral technology and risk aversion shocks. The bar chart in the bottom panel of table 3.3 decomposes the variance of the SDF into components coming from the neutral technology shock, the risk aversion shock, and the remaining shocks combined. The lines at the top of each bar give the 2.5 and 97.5 percentiles of the posterior distribution. The 97.5 percentile for the variance share in the SDF for non-technology and non-risk-aversion shocks is less than 2 percent.

On first glance this result might be somewhat surprising, but it is in fact a deep characteristic of models with Epstein–Zin preferences with a high EIS and high risk aversion. One way to see the source of this finding is to simply look at the household’s SDF,

\[ M_{t+1} = \beta B_t \frac{1 - \beta B_{t+1}}{1 - \beta B_t} \frac{U_{C,t+1}}{U_{C,t}} \frac{V_{t+1}^{\rho - \alpha_t}}{E_t V_{t+1}^{1-\alpha_t}} \frac{\rho - \alpha_t}{\rho - \alpha_t} \]

For a household with a large EIS, the variance of \( \frac{U_{C,t+1}}{U_{C,t}} \) is generally small (at least with standard preferences). In the case where the household does not have a habit \( (\eta = 0) \), this term is equal to \( (C_{t+1}/C_t)^{-\rho} \). If the household has an EIS greater than 1, then \( \rho \) is less than 1 and the variance of \( \frac{U_{C,t+1}}{U_{C,t}} \) will be less than the variance of log consumption growth. A one-percent permanent decline in consumption will raise this term by the factor \( 1.01^\rho \).

The majority of the variance of the SDF is driven by the term \( \frac{V_{t+1}^{\rho - \alpha_t}}{E_t V_{t+1}^{1-\alpha_t}} \frac{\rho - \alpha_t}{\rho - \alpha_t} \). Here, a one-percent permanent decline in consumption will make this term (approximately) equal to...
Table 3.3: One-quarter ahead variance decompositions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monetary policy</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.34</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>Neutral tech.</td>
<td>0.57</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.15</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>Gov't spending</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>4</td>
<td>Investment tech.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>Price markup</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>Wage markup</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.01</td>
</tr>
<tr>
<td>7</td>
<td>Time preference</td>
<td>0.01</td>
<td>0.94</td>
<td>0.94</td>
<td>0.91</td>
<td>0.36</td>
<td>0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>Inflation target</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.02</td>
<td>0.03</td>
<td>0.52</td>
</tr>
<tr>
<td>9</td>
<td>Risk aversion</td>
<td>0.41</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.46</td>
<td>0.23</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Moments:

<table>
<thead>
<tr>
<th></th>
<th>10 Standard deviation</th>
<th>11 Correl. w/ SDF</th>
<th>12 Expected return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.47</td>
<td>-0.14</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>4.06</td>
<td>-0.17</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>4.06</td>
<td>-0.09</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>8.67</td>
<td>-0.80</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>6.54</td>
<td>-0.53</td>
<td>2.47</td>
</tr>
<tr>
<td></td>
<td>19.65</td>
<td>0.01</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.04</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1.07</td>
<td>-0.12</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Variance decompositions and 95% credible intervals

Note: decompositions of one-quarter ahead forecast error. Levered returns for capital and consumption claims assume that the investor finances half the purchase price of the given claim with a 10-year nominally riskless bond. The 10-year return is the one-quarter return from holding a 10-year nominally riskless bond. The moments in rows 10-12 are annualized. The black bars in the figure give the 95 percent credible region based on random draws from the posterior density.
$1.01^{\alpha_t - \rho}$. When $\alpha_t \gg \rho$, as estimated here and in most models with Epstein–Zin preferences, it is therefore the variation in $V_{t+1}$ that determines the behavior of the SDF.

So what determines movements in $V_{t+1}$? One way to think about it is to split movements in consumption into permanent and temporary components. A purely temporary shock to consumption will have a relatively small impact on $V_{t+1}$ because the household is not very averse to shifting consumption over time. A permanent shock to consumption, on the other hand, will tend to shift $V_{t+1}$ by an equal amount. Risk aversion shocks also affect $V_t$. The reason is simply that increases in risk aversion directly drive $V_{t+1}$ down because households are more averse to the future uncertainty that they face. Since it is the shocks to $V_{t+1}$ that determine the movements in $M_{t+1}$, we should thus not be surprised that it is mainly the neutral technology and risk aversion shocks that drive the variance of the SDF.

### 3.6.2.2 Impulse responses

Since the technology and risk aversion shocks are the key to understanding the pricing kernel, it is natural to ask how they affect the economy. Figure 3.6 plots impulse responses to an increase in labor-neutral technology and a decrease in risk aversion. Dotted lines give 95-percent credible intervals (the range between the 2.5 and 97.5 percentiles in the posterior distribution). These impulse responses are different from those in figure 3.5 because they do not turn off the interactions between the inflation target and risk aversion and the other shocks. The idea is that we want to see what happens on average following these two shocks, since that behavior is what is relevant for understanding the correlations with the SDF.

Following the technology shock, inflation falls, while output and real interest rates rise: a standard positive supply shock. The declines in inflation and nominal interest rates are especially pronounced and persistent. This behavior is the result of the fact that the estimates imply that the inflation target falls following an increase in technology. However, note that inflation falls more than the inflation target does, so the effect is not entirely driven by the inflation target.

A similar result is obtained in JPT and SW, even though they do not have time-varying
Figure 3.6: Responses to technology and risk-aversion shocks

Note: responses in percentage points to a unit standard deviation positive shock to labor-neutral technology and a negative shock to risk aversion. Interest rate, inflation, and inflation target are annualized. Dotted lines give 2.5 and 97.5 percentiles from the posterior distribution.
inflation targets. In all of these models, because prices are sticky, when there is a positive supply shock, rather than cutting prices, firms simply produce the same quantity as previously, thereby reducing employment and hence demand. Positive technology shocks are thus associated with small increases in output and large declines in the output gap (defined as the difference between output and the level it would take if prices were flexible). There is also a small empirical literature that provides more reduced-form evidence on effects of this sort using direct measures of technology (e.g. Basu, Fernald, and Kimball, 2006).

The slow response of output to a technology shock is particularly notable. The impulse response function in figure 3.6 suggests that output and perhaps also consumption growth should be predictable and positively autocorrelated, though other shocks could obscure those relationships. Figure 3.7 therefore plots the empirical and model-predicted autocorrelation functions for consumption and output growth. For output, the model implies that the 1-quarter autocorrelation should be 0.50, while it is only 0.33 in the data. But 0.50 is well within the 95 percent confidence interval for the empirical value. In fact, nearly the entire autocorrelation function for output in the model is captured within the 95 percent confidence interval in the data.

The model also implies strong one-quarter autocorrelation in consumption growth, and the autocorrelation is in fact higher than the upper end of the 95 percent confidence interval. However, this autocorrelation dies out very rapidly, and at lags longer than one quarter the autocorrelation of consumption growth in the model matches the behavior in the data well. The model implies little or no serial correlation in consumption and output growth at horizons longer than two quarters.

Figure 3.6 also reports impulse responses for a decline in risk aversion. Inflation, interest rates, and the output gap all rise. This shock therefore takes the form of a classic demand shock. The effects are far smaller than those for the technology shock. Risk aversion has two main channels through which it affects the real economy. First, to the extent that physical investment is risky, a decline in risk aversion makes households more willing to purchase physical capital. Second, a decline in precautionary saving demand makes households want to consume more for any given level of interest rates. For this model,
Figure 3.7: Empirical and model-implied autocorrelations

Notes: Empirical and model-implied autocorrelation functions. Gray regions are 95-percent confidence intervals.
the increase in consumption demand dominates, which is why the shock is slightly expansionary.

3.6.2.3 Other asset prices

Variance decompositions  After the SDF, table 3.3 reports variance decompositions for the returns on a number of assets. The bottom panel of table 3.3 reports the fraction of the variance of the one-quarter innovation to each return coming from the neutral technology shock, the risk aversion shock, and all other shocks combined.

Column 2 reports the variance decomposition for the return on the utility portfolio. 87 percent of its variance comes from the time-preference shock. The reason is simply that the utility claim has a relatively long duration, like that of a consol with a coupon that grows at the average rate of the economy, so shifts in real interest rates have a large effect on its price. The time-preference shock mainly affects real interest rates, so it drives the variance of the utility claim. Row 11 shows that the correlation of the utility return with the SDF is -0.14

More interestingly, the third and fourth columns of table 3.3 report variance decompositions for a claim on aggregate profits and the same claim levered two to one on short-term nominal debt. Once again, little of the variance of the return is driven by the technology or risk aversion shocks. While the neutral technology shock does play a role, it is relatively small; the vast majority of the variation is driven by shifts in the rate of time preference due to its effects on real interest rates. Assuming that equity is levered on nominal riskless debt does not change this result because the returns on nominal bonds are generally unaffected by the time-preference shock. However, column 5 shows that if equity claims are levered on real debt, then they because highly correlated with the pricing kernel, and the model can actually generate a large equity premium (though one that is still too small by half). The reason is that when the firm is levered on real bonds, the effects of time-preference shocks on the unlevered dividend claim and real bond prices cancel each other out.

---

8 The profits are those earned by the intermediate-good producers, $Y_t - W_t N_t$. 

134
Zero-coupon claims  Lettau and Wachter (2007, 2011) argue that the value premium can be explained if value stocks have relatively short durations and the term structure of zero-coupon consumption (or dividend) claims is downward-sloping. Figure 3.8 plots the steady-state term structures for zero-coupon nominal bonds, inflation indexed bonds, and consumption claims (with the yields normalized to zero at the 1-quarter horizon). For the zero-coupon consumption claims, figure 3.8 plots the steady-state values of

\[- \log \left( \frac{P_{C_j, t}}{E_t C_{t+j}} \right) / j,\]

where \(P_{C_j, t}\) is the price on date \(t\) of an asset that pays one unit of consumption date \(t + j\).\(\left( - \log \left( \frac{P_{C_j, t}}{E_t C_{t+j}} \right) / j \right)\) is the average per-period discount rates applied to assets that pay a unit of consumption at date \(t + j\).

To understand the results in figure 3.8, first note that the real term structure is downward sloping for the usual reason: a positive technology shock (which drives the majority of the variance of the SDF) drives the marginal product of capital upward and raises real interest rates. So the prices of real bonds are low in good times, which induces a downward-sloping real term structure. This effect will also apply to the consumption claims since they are real claims and are discounted (partly) with real interest rates. However, there is a second effect – positive technology shocks raise the expected level of consumption in the future, which means that the consumption claims have a high payoff in good times. In the present setting, the latter effect is slightly stronger, inducing the small upward slope in the term structure for consumption claims.

3.6.2.4 The Hansen–Jagannathan bound

Figure 3.9 plots the fitted quarterly Hansen–Jagannathan (HJ) bound against the nominal term spread. The estimated steady-state level of the HJ bound – 0.24 – is almost identical to the observed Sharpe ratio on the aggregate stock market in this sample of 0.26. This is notable given that equities were not included in the estimation and nominal bonds do not achieve the HJ bound. Perhaps more importantly, though, the price of risk in this model is highly volatile. The estimated standard deviation of the Hansen–Jagannathan bound is 95 percent of its mean. The level of variability here is somewhat higher than but still similar to that used in Dew-Becker (2011a) to match the degree of predictability observed
Figure 3.8: Steady-state term structures

Notes: term structures for nominal and real zero-coupon bonds and zero coupon consumption claims. Horizons are measured in quarters.
for aggregate stock returns in the post-war sample.

3.7 The real economy

Up to now, the analysis has focused mainly on asset pricing. But the model gives a rich description of the real side of the economy. While I leave a deeper analysis of New Keynesian models to papers focused on those models for their own sake, the interaction of the real side of the economy with asset prices is important to this paper.

Figure 3.10 gives a variance decomposition for the variables used in the estimation. The figure decomposes the variance of each variable at frequencies of 6 to 32 quarters into components coming from each of the structural shocks.\(^9\) Except for investment growth, for which the investment-specific shock is completely dominant, none of the other variables examined in figure 3.10 are dominated by any particular shock. Notably, the shock to risk aversion has almost no effect on the variance of any of the real variables at business-cycle frequencies. Its largest effect is on consumption growth, for which the variance share is still only 4 percent. The far-right bar, though, shows that risk aversion has a large effect on the term spread, as we saw in figure 3.5; it explains roughly 1/3 of the variance of the term spread at business-cycle frequencies.

The variance decomposition reported in the top panel of figure 3.10 is rather different from that reported by JPT. They found that the investment shock was an important determinant of not only investment, but also output and consumption growth at business-cycle frequencies. Their model differs from mine in three ways: risk aversion and the inflation target are constant, the preference specification is slightly different (log utility and additive habits), and the data sample covers the entire post-war period and only includes the one-quarter nominal interest rate. The bottom panel of figure 3.8 removes a number of those differences. It drops bond yields (except the short rate) from the estimation and assumes

\(^9\) The variance decomposition is calculated using a spectral decomposition of the state-space form of the model. Specifically, since the structural shocks are orthogonal, the spectral density of the endogenous variables is equal to the sum of the densities obtained when each shock is turned on individually. Calculating variance shares over certain frequencies then simply requires integrating the density over those frequencies. I numerically integrate by calculating the spectral density at 100 increments between wavelengths of 6 and 32 quarters.
Figure 3.9: The Hansen–Jagannathan bound and the term spread.
Figure 3.10: Variance decompositions at business-cycle frequencies

Note: each section of a bar represents a fraction of the variance of one of the series at frequencies between 6 and 32 quarters. The top panel gives results for the benchmark model. The bottom panel reports results where the inflation target and risk aversion are fixed and bond prices are dropped from the estimation. That latter model is identical to JPT except that it uses post-1983 data and uses Epstein–Zin preferences.
the inflation target and coefficient of relative risk aversion are constant. The model is then reestimated.

While I do not replicate JPT’s results exactly, I do also find that the investment technology shock is important for more variables than just investment itself. It accounts for 30 percent of the variation in output growth and 20 percent of consumption growth. Moreover, it accounts for 41 and 48 percent of the variance of the one-quarter nominal interest rate and the term spread, respectively. It is this latter result that explains the divergence between the two models. Since long-term bond prices are included in the estimation of the benchmark model, that model is forced to match the relationship between investment and the term spread. The difference in the variance decompositions suggests that JPT gets this relationship wrong. What is interesting, though, is that when the model is forced to match long-term bond yields, that also changes the decomposition for other variables.

Table 3.3 gives one-quarter-ahead variance decompositions for output, consumption, and investment growth. This decomposition is useful for understanding whether any of these variables would be powerful asset pricing factors. Specifically, in a world where consumption followed a random walk and households had Epstein–Zin preferences with constant relative risk aversion, consumption growth would be perfectly correlated with the SDF, so it would price assets in the economy.

What Table 3.3 shows is that consumption, output, and investment growth are all only weakly correlated with the SDF at the one-quarter horizon—they all have correlations less than 16 percent. So asset pricing with only consumption growth will not work well in this economy. Many asset-pricing studies with Epstein–Zin preferences include both consumption growth and the return on the stock market as pricing factors. If we believe that the stock market is a claim on aggregate capital, then Table 3.3 shows that it will do little to help with asset pricing as it is also only weakly correlated with the SDF.

3.7.1 Model comparison

The primary difference between the model studied here and JPT is the addition of time-varying risk aversion and the time-varying inflation target. An important question,
then, is the extent to which those two factors improve the fit of the model to the data. Clearly, allowing a time-varying inflation target will help increase the volatility of long-term bond yields, and time-varying risk aversion will induce time-varying risk premia. But it is possible that the data can be well explained with a constant risk premium, or perhaps there is no need to have a time-varying inflation target.

Table 3.4 considers two alternatives to the benchmark model: a version where the inflation target is constant, and a version where the coefficient of relative risk aversion is constant. It lists two statistics for each model. First, it gives the standard likelihood ratio used in MLE, i.e. based on \( \log f(y|\theta, M) \), the likelihood of the data conditional on the model and parameters, where \( y \) represents the data and \( \theta \) the parameter vector. \( M \) denotes the model choice, i.e. the full model or one of the two restricted versions. \( \log f(y|\theta, M) \) is closely related to the one-step-ahead forecast error of the model. The likelihood ratio test favors the benchmark over each alternative by a wide margin. One way to see the source of this rejection is to note that in table 3.2, the measurement errors for bond yields, essentially a residual variance, are three times higher in the alternative models than the benchmark. This difference alone is more than sufficient to explain the magnitude of the likelihood ratios.

The second statistic is the Bayes factor, which is based on the marginal likelihood conditional only on the model, \( \log f(y|M) \) (in the economics literature, see Fernandez-Villaverde and Rubio-Ramirez, 2004). I calculate \( \log f(y|M) \) using the monte carlo chain as in Fernandez-Villaverde and Rubio-Ramirez (2004).\(^{10}\)

The difference in the Bayes factors listed in the bottom row of table 3.4 is similar to the values obtained under the usual likelihood ratio test. In order to accept the model with constant risk aversion or a constant inflation target, the ratio of the prior probability of either of those models to that of the benchmark model would have to be greater than \( \exp(140) \). In a statistical sense, then, both the time-varying inflation target and time-varying risk aversion substantially improve the fit of the model.

The last exercise I perform is to compare forecasting power between the structural

\(^{10}\) See also Gelfand and Dey (1994) and Geweke (1999)
Table 3.4: Model comparison statistics

<table>
<thead>
<tr>
<th></th>
<th>Constant RRA</th>
<th>Constant π*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood ratio</td>
<td>185.6</td>
<td>183.8</td>
</tr>
<tr>
<td>p-value</td>
<td>5.0E-41</td>
<td>1.6E-35</td>
</tr>
<tr>
<td>Bayes factor</td>
<td>141.50</td>
<td>142.50</td>
</tr>
</tbody>
</table>

Note: Row 1 gives the marginal likelihood of the data given the model and parameters. Row 2 gives the \( p \)-value for the frequentist LR test. Row 3 is the Bayes factor, the marginal probability of the data conditional on the model (i.e. integrated over the parameter space).
model and a VAR. Rather than include all 13 of the endogenous variables, I use the 6 macro variables and the first three principal components of the bond yields (i.e. I estimate a FAVAR). I estimate a VAR(2) with the restriction that the macro variables and the interest rate factors for not interact with each other (though their innovations may be correlated). This assumption partly helps to limit the number of free parameters in the VAR, but it also means that we can compare the forecasting performance of the structural model to the benchmark models from the macro and asset-pricing literatures.

Figure 3.11 plots root mean squared errors for the forecasts from the VAR(2) and the structural model. All of the variables are forecasted in levels, rather than differences. Because the structural model is so difficult to estimate, I only consider in-sample forecasting performance. For the macro variables, the structural model has somewhat weaker forecasting power than the VAR. The place where the divergence is notable is for inflation. At long horizons, the structural model essentially says that the level factor in interest rates should equal expected inflation since they are both driven by the Fed’s inflation target. Over this sample, however, the level factor fell slowly, while inflation fell rapidly in the early 1980’s and then was low and stable. Forecasting long-term inflation with the level factor in this sample was thus unsuccessful compared to a simple random-walk forecast for inflation (which is basically what the VAR implies).

For the bond yields, the structural model forecasts well. At long horizons, the RMSE generated by the structural model is smaller than the VAR by 10–50 percent, with the largest improvements coming for long-term bonds.

3.8 Conclusion

This paper studies bond pricing in a medium-scale New-Keynesian model with a time-varying price of risk. I show that the model can generate a large and volatile term premium. The term premium is driven by the combination of two factors—a negative response of interest rates to positive technology shocks and variation in risk aversion. Removing either of these effects eliminates the model’s ability to match the magnitude of the term premium.
Figure 3.11: Forecast RMSE

Note: Root mean squared error from forecasts of the endogenous variables. All variables are forecasted in levels. The horizontal axis gives the forecast horizon in quarters.
While shocks to risk aversion and technology determine average asset returns, they have only weak effects on real variables at business-cycle frequencies. The covariance of asset returns with real variables over the business cycle is therefore unimportant for determining average returns. It is true that the Federal Reserve tends to cut interest rates in recessions, but the model shows that most recessions are not high-marginal-utility states of the world. So the usual intuition that the Taylor rule should lead to a downward-sloping yield curve is inaccurate since it does not take into the difference between the shocks that have high risk prices and the shocks that drive the business cycle.

Furthermore, while risk aversion is estimated to be highly volatile and to be an important determinant of the dynamics of the term spread, it has almost no effects on the real economy. This model thus suggests that there is a separation between the price of risk in financial markets and the real economy.
APPENDIX
A. APPENDIX TO CHAPTER 1

A.1 The approximation for average duration

Now using the investment function from the text, we have

\[ I_{it} = \exp(\gamma \log (\eta_{1i}) + \gamma (\mu_{t+1} - r_{t+1}) + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i))) \]  
(A.1)

\[ = \exp (\gamma (\mu_{t+1} - r_{t+1})) \exp (\gamma \log (\eta_{1i}) + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i))) \]  
(A.2)

Using the definition of \( \bar{D}_t \) and the investment function, we have

\[ \bar{D}_t = \sum_i \frac{\exp (\gamma \log \eta_{1i} + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i)))}{\sum_i \exp (\gamma \log \eta_{1i} + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i)))} D (\delta_i) \]  
(A.3)

Now we take a linear approximation to

\[ \exp (\gamma \log \eta_{1i} + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i))) \]  
(A.4)

around the point \( \{ \log \eta_i, \log B_i, \delta \} \) where \( \log \eta_i \equiv \sum_i \log \eta_{1i}, \log B_i \equiv \sum_i \log B_{it} \), and \( \delta \equiv \sum_i \delta_i \). Also define

\[ \log (\eta_{1i}) \equiv \log (\eta_{1i}) - \log \eta_i \]
\[ \log (B_{it}) \equiv \log (B_{it}) - \log B_i \]
\[ \delta_i \equiv \delta_i - \delta \]
We have

\[
\exp \left( \gamma \log \eta_{1i} + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta_i)) \right) 
\approx \exp \left( \gamma \log \eta_{1i} + \gamma \log B_{it} + \log (1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta)) \right) 
\times \left( 1 + \gamma \log (\eta_{1i}) + \gamma \log (B_{it}) \right)
\]

\[
- \hat{\delta}_i \exp (\mu_{t+2} - r_{t+2}) 
\frac{1}{1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta)}
\]

(A.5)

Simple algebra and the observation that

\[
\sum_{i} \left( 1 + \gamma \log (\eta_{1i}) + \gamma \log (B_{it}) - \hat{\delta}_i \exp (\mu_{t+2} - r_{t+2}) \right) \frac{\exp (\mu_{t+2} - r_{t+2})}{1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta)} = N
\]

(A.6)

yields the result

\[
\bar{D}_t \approx N^{-1} \sum_{i} \left( 1 + \gamma \log (\eta_{1i}) + \gamma \log (B_{it}) \right) D_i - \hat{\delta}_i \exp (\mu_{t+2} - r_{t+2}) \frac{1}{1 + \exp (\mu_{t+2} - r_{t+2}) (1 - \delta)} N^{-1} \sum_{i} \hat{\delta}_i D_i
\]

(A.7)

(A.8)

The term \( N^{-1} \sum_{i} \left( 1 + \gamma \log (\eta_{1i}) \right) D_i \) is constant over time, and we thus have the desired result from the text.

A.2 Further robustness tests

This section describes table A.1, which has extra robustness tests for the regressions of \( \bar{D} \) on the term spread and other controls. Column 1 includes up to three lags of the term spread. The second lag enters significantly, and with a coefficient slightly larger than the first lag. This sort of lagged response to price changes is commonly found in the literature. It is generally interpreted as being due to planning and delivery lags. The second column includes every one of the other controls simultaneously, instead of individually as in tables 2 and 3. The result is that the coefficients on the various other controls all become marginally significant at best, while the term spread is still highly significant. There is thus
something different about the term spread from all of the other business cycle, volatility, and investment controls. Column 3 is identical to column 2 except it only uses one lag of the term spread, and the coefficient is still significant at the one percent level.

Finally, column 4 includes the current and lagged values of volatility instead of just the leading value. The leading value has a negative sign, indicating that high future volatility lowers duration today (consistent with a model of irreversible investment). The current value has a positive sign. One way to reconcile this is if the volatility variable should actually enter as a first difference: an increase in volatility lowers average duration, instead of a high value by itself. The lagged level of volatility is uncorrelated with $D_t$. 

---

**Figure A.1: Further robustness tests**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term Spread(t-1)</td>
<td>-0.23 **</td>
<td>-0.40 ***</td>
<td>-0.46 ***</td>
<td>-0.57 ***</td>
</tr>
<tr>
<td></td>
<td>[0.10]</td>
<td>[0.06]</td>
<td>[0.08]</td>
<td>[0.09]</td>
</tr>
<tr>
<td>Term Spread(t-2)</td>
<td>-0.24 ***</td>
<td>-0.27 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.08]</td>
<td>[0.08]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Spread(t-3)</td>
<td>0.11</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.09]</td>
<td>[0.07]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment(t-1)</td>
<td>0.18</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.19]</td>
<td>[0.20]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP(t)</td>
<td>-0.05</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.12]</td>
<td>[0.13]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP(t-1)</td>
<td>0.36 ***</td>
<td>0.55 *</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.10]</td>
<td>[0.26]</td>
<td>[0.22]</td>
<td></td>
</tr>
<tr>
<td>Investment(t)</td>
<td>-0.08</td>
<td>-0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.17]</td>
<td>[0.16]</td>
</tr>
<tr>
<td>Investment(t-1)</td>
<td>-0.12</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.21]</td>
<td>[0.19]</td>
</tr>
<tr>
<td>SD_returns(t+1)</td>
<td>-0.23 **</td>
<td>-0.19 **</td>
<td>-0.28 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.09]</td>
<td>[0.09]</td>
<td>[0.07]</td>
<td></td>
</tr>
<tr>
<td>SD_returns(t)</td>
<td></td>
<td>0.15 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.07]</td>
<td></td>
</tr>
<tr>
<td>SD_returns(t-1)</td>
<td></td>
<td></td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.05]</td>
</tr>
<tr>
<td>N</td>
<td>59</td>
<td>47</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>R2</td>
<td>0.48</td>
<td>0.74</td>
<td>0.69</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Note: See tables 2 and 3.
B. APPENDIX TO CHAPTER 2

B.1 The certainty equivalent

This section looks at the relationship between the certainty equivalents using $G^{\text{habit}}$ and $G^{TV}$. I first show that the two certainty equivalents are equal up to a second order approximation around the non-stochastic version of the model. Next, I show that in the continuous-time limit, the preferences associated with the two certainty equivalents are identical.

B.1.1 Second-order approximation

This section approximates the certainty equivalent $G^{-1}(E_t(G(V_{t+1})))$ where $V_{t+1} = V_t \times (1 + \sigma \epsilon_{t+1})$ around the point $\sigma = 0$. We assume that $E_t \epsilon_{t+1} = 0$ and $E_t \epsilon_{t+1}^2 = 1$.

Now consider the derivative of $G^{-1}(E_t(G(V_{t+1})))$ with respect to $\sigma$,

$$
\frac{d}{d\sigma} G^{-1}(E_t(G(V_{t+1}))) = \frac{d}{d\sigma} E_t(G(V_{t+1})) \frac{G'(G^{-1}(E_t(G(V_{t+1}))))}{G'(G^{-1}(E_t(G(V_{t+1}))))}
$$

(B.1)

We have

$$
\frac{d}{d\sigma} E_t(G(V_{t+1})) = \int G'(V_t (1 + \sigma \epsilon_{t+1})) \epsilon_{t+1} V_t dF(\epsilon_{t+1})
$$

(B.2)

where $F$ is the cdf of $\epsilon_{t+1}$. Evaluated at $\sigma = 0$, $\frac{d}{d\sigma} E_t(G(V_{t+1})) = 0$, and therefore

$$
\frac{d}{d\sigma} G^{-1}(E_t(G(V_{t+1}))) = 0
$$

(B.3)

So all certainty equivalents taking this form are identical up to the first order in approximations around $\sigma$. 

150
Next, consider the second derivative,

\[
\frac{d^2}{d\sigma^2} G^{-1} (E_t (G (V_{t+1}))) = \frac{\left[ G' (G^{-1} (E_t (G (V_{t+1}))))\right]^2}{[G' (G^{-1} (E_t (G (V_{t+1}))))]^2} \left[ \frac{d}{d\sigma} G (V_{t+1}) \right] \left[ \frac{d}{d\sigma} G' (G^{-1} (E_t (G (V_{t+1}))))\right] \]

(B.4)

Since \( \frac{d}{d\sigma} G (V_{t+1}) \) is equal to zero at \( \sigma = 0 \), we can ignore the second term in the numerator. The second derivative of the expectation is

\[
\frac{d^2}{d\sigma^2} E_t (G (V_{t+1})) = \int G'' (V_t (1 + \sigma \epsilon_{t+1})) \epsilon_{t+1}^2 V_t^2 dF (\epsilon_{t+1})
\]

(B.5)

At \( \sigma = 0 \), \( \frac{d^2}{d\sigma^2} E_t (G (V_{t+1})) \mid_{\sigma=0} = G'' (V_t) V_t^2 \). We also have \( G' (G^{-1} (E_t (G (V_{t+1})))) \mid_{\sigma=0} = G' (V_t) \), and hence \( \frac{d^2}{d\sigma^2} G^{-1} (E_t (G (V_{t+1}))) \mid_{\sigma=0} = \frac{G''(V_t)V_t^2}{G'(V_t)} \). So any two choices of \( G \), say \( G_1 \) and \( G_2 \) are equivalent up to the second order if \( \frac{G''(V_t)}{G'(V_t)} = \frac{G''(V)}{G'(V)} \) for any \( V_t \). That relationship holds for \( G^{\text{habit}} \) and \( G^{TV} \).

### B.1.2 Continuous time

Duffie and Epstein (1992) show how to extend Epstein–Zin preferences to continuous time. They derive a utility function following the process

\[
dV_t = \mu_t + \sigma_t dB_t
\]

(B.6)

\[
= \left( -f (\epsilon_t, V_t) - \frac{1}{2} A (V_t) \sigma_t \epsilon_t^2 \right) dt + \sigma_t dB_t
\]

(B.7)

for a Wiener process \( dB_t \).

As in the main text, suppose the household’s certainty equivalent under discrete-time Epstein–Zin preferences is \( G^{-1} (E_t (G (V_{t+1}))) \). Duffie and Epstein (1992) show that the analogous choice of \( A \), obtained as a limiting case as the length of time periods approaches zero, is \( A (V_t) = \frac{G''(V_t)}{G'(V_t)} \). In the case where \( G^{\text{power}} (V_t) = V_t^{1-\alpha} \), we have

\[
A^{\text{power}} (V_t) = \frac{G^{\text{power}''}(V_t)}{G^{\text{power}'}(V_t)} = \frac{-\alpha}{V_t}
\]

(B.8)
and for $G^{habit} = (V_t - H_t)^{1-\alpha}$

$$A^{habit} (V_t) = \frac{-\alpha}{V_t - H_t} \quad \text{(B.9)}$$

For $G^{TV} = V_t^{1-\alpha_t}$,

$$A^{TV} (V_t) = \frac{-\alpha_t}{V_t} \quad \text{(B.10)}$$

So $G^{TV}$ and $G^{habit}$ are identical if $\alpha_t = \alpha \frac{V_t}{V_t - H_t}$, which is what is used in the text.

For all three choices of the certainty equivalent $G$, we can use the standard choice for $f$, $f (c_t, V_t) = \beta \frac{c_t^{1-\rho} - V_t^{1-\rho}}{V_t^{\rho}}$. $\rho$ then determines the elasticity of intertemporal substitution, while $\Lambda$ determines risk aversion.

**B.2 Derivation of the SDF**

We can obtain the stochastic discount factor (SDF) by calculating the intertemporal marginal rate of substitution. We calculate two derivatives. First,

$$\frac{\partial V_t}{\partial C_t} = V_t^{\rho} \left( 1 - \exp (-\beta) \right) C_t^{-\rho} \quad \text{(B.11)}$$

Next, we differentiate $V_t$ with respect to $C_{t+1} (w)$, where $w$ denotes one state of the world, and $\pi_w$ is the probability of that state,

$$\frac{\partial V_t}{\partial C_{t+1} (w)} = \pi_w V_t^{\rho} \exp (-\beta) R_t^{-\rho} G_t^{(-1)\psi}$$

$$\times \left( E_t [ G_t (V_{t+1} (w))] \right) G_t' (V_{t+1} (w)) V_t^{\rho} (w) (1 - \beta) C_t^{-\rho} (w) \quad \text{(B.12)}$$

where $G_t'$ is the derivative of $G_t$ and $G_t^{(-1)\psi}$ the derivative of $G_t^{-1}$. $R_t \equiv G^{-1} (E_t G_t (V_{t+1}))$.

The subscripts on $G_t$ refer to the fact that $G_t$ depends on the potentially-time-varying parameter $\alpha_t$. The assumption that $\alpha_t$ is exogenous to the household is necessary for this formula for the derivative to be correct (in the same way that external habits lead to a more tractable formula for the SDF than do internal habits).

The SDF can be derived from a consumer’s first order conditions for optimization as...
$M_{t+1}(w) = \frac{1}{\pi w} \frac{\partial V_t}{\partial C_{t+1}(w)}$. We then have

$$M_{t+1}(w) = \exp\left(-\beta\right) \frac{G'_t(V_{t+1}(w)) V^\rho_{t+1}(w) C^{-\rho}_{t+1}(w)}{R^\rho_t C^{-\rho}_t}$$

(B.13)

where the last line follows from the fact that $G_t^{(-1)t}(x) = 1/G'_t(x)$.

In the case of $G_t(V) = V^{1-\alpha_t}$, the SDF becomes

$$M_{t+1} = \exp\left(-\beta\right) \frac{V^\rho_{t+1}(w) C^{-\rho}_{t+1}(w)}{R^\rho_t C^{-\rho}_t}$$

(B.14)

B.2.1 Substituting in an asset return

Consider an asset that pays $C_t$ as its dividend. We guess that its cum-dividend price is $W_t = V_t^{1-\rho} C^{-\rho}_t (1 - \exp(-\beta))^{-1}$. This guess can be confirmed by simply inserting it into the household’s Euler equation.

The return on the consumption claim is

$$R_{w, t+1} = \frac{W_{t+1}}{W_t - C_t} = \frac{V^{1-\rho}_{t+1}}{\exp(-\beta) R_t (V^{1-\rho}_{t+1} C^{-\rho}_t)}$$

(B.15)

Which yields

$$\frac{V^\rho_{t+1}(w)}{R^\rho_t} = (R_{w, t+1} \exp(-\beta))^{\rho-\alpha_t} \left(\frac{C_{t+1}}{C_t}\right)^{-\rho \frac{\rho-\alpha_t}{1-\rho}}$$

(B.16)

We can then insert this into the SDF to yield

$$M_{t+1} = \exp\left(-\beta\right) \frac{1}{\pi w} \left(\frac{C_{t+1}}{C_t}\right)^{-\rho \frac{\rho-\alpha_t}{1-\rho}} R_{w, t+1}^{\rho-\alpha_t}$$

(B.17)

B.3 The log-linear model with production

B.3.1 Steady state

In the nonstochastic steady state, the interest rate earned by all assets, $r$, is equal to

$$r = \beta + \rho \mu$$

(B.18)
Standard manipulations show that the steady-state ratio of capital to technology is then

$$
\bar{K} = \left( \frac{\exp(\beta + \rho \mu) - 1 + \delta}{\gamma} \right)^{1/(\gamma - 1)} \quad (B.19)
$$

We can obtain the steady-state consumption-output ratio by using the budget constraint,

$$
\bar{C} = \bar{Y} + (1 - \delta) \bar{K} - \exp(\mu) \bar{K} \quad (B.20)
$$

$$
1 - \frac{\bar{C}}{\bar{Y}} = - (1 - \delta - \exp(\mu)) \frac{\bar{K}}{\bar{Y}} \quad (B.21)
$$

### B.3.2 The budget constraint

The approximation I use for the budget constraint is identical to Campbell (1994). The budget constraint is

$$
K_{t+1} = A_t K_t^{1-\gamma} - C_t + (1 - \delta) K_t \quad (B.22)
$$

I look for a log-linear approximation taking the form

$$
k_{t+1} = \lambda_0 + \lambda_k k_t + \lambda_a a_t + \lambda_c c_t, \text{ where the } \lambda \text{ terms are coefficients from the approximation.}
$$

The budget constraint can be rewritten as

$$
\log [\exp(\Delta k_{t+1}) - (1 - \delta)] = y_t - k_t + \log (1 - \exp(c_t - y_t)) \quad (B.22)
$$

Taking a log-linear approximation to the left-hand side around the point $\Delta k_{t+1} = \mu$, we have

$$
\log [\exp(\Delta k_{t+1}) - (1 - \delta)] \approx \log [\exp(\mu) - (1 - \delta)] + \frac{\exp(\mu)}{\exp(\mu) - (1 - \delta)} (\Delta k_{t+1} - \mu) \quad (B.23)
$$

To approximate the right-hand side of (B.22), we approximate $\log (1 - \exp(c_t - y_t))$ around the steady state $cy$,

$$
\log (1 - \exp(c_t - y_t)) \approx \log (1 - \exp(cy)) + \frac{-\exp(cy)}{1 - \exp(cy)} (c_t - y_t - cy) \quad (B.24)
$$
This implies
\[
\log [\exp (\mu) - (1 - \delta)] + \frac{\exp (\mu)}{\exp (\mu) - (1 - \delta)} (\Delta k_{t+1} - \mu) \approx \log (1 - C/Y) + y_t - k_t \\
+ \frac{- \exp (cy)}{1 - \exp (cy)} (c_t - y_t - cy)
\]
(B.25)

Now we can find the coefficients in the linear approximation to the budget constraint. The constant term is
\[
\lambda_0 = \frac{\exp (\mu) - (1 - \delta)}{\exp (\mu)} \left( \log (1 - C/Y) - \left( \frac{- \exp (cy)}{1 - \exp (cy)} \right) cy \right)
\]
(B.26)

The coefficients on \( k, a, \) and \( c, \) are then
\[
\lambda_k = \frac{\exp (\mu) - (1 - \delta)}{\exp (\mu)} \left( \gamma - 1 - \gamma \left( \frac{- \exp (cy)}{1 - \exp (cy)} \right) \right) + 1
\]
(B.27)
\[
\lambda_a = \frac{\exp (\mu) - (1 - \delta)}{\exp (\mu)} (1 - \gamma) - (1 - \gamma) \frac{\exp (\mu) - (1 - \delta)}{\exp (\mu)} \left( \frac{- \exp (cy)}{1 - \exp (cy)} \right)
\]
(B.28)
\[
\lambda_c = \frac{\exp (\mu) - (1 - \delta)}{\exp (\mu)} \left( \frac{- \exp (cy)}{1 - \exp (cy)} \right)
\]
(B.29)

Now note that \( \lambda_k + \lambda_a + \lambda_c = 1, \) So we have
\[
k_{t+1} = \lambda_0 + \lambda_k k_t + \lambda_a a_t + (1 - \lambda_k - \lambda_a) c_t
\]
(B.30)

**B.3.3 Capital return**

To approximate the return on capital, we say
\[
r_{k,t+1} = \log (\gamma \exp ((1 - \gamma) (a_{t+1} - k_{t+1})) + 1 - \delta)
\]
\[
\approx \log (\gamma \exp ((\gamma - 1) \tilde{k}) + 1 - \delta) + \frac{(\gamma - 1) \gamma \exp ((\gamma - 1) \tilde{k}) (k_{t+1} - a_{t+1} - \tilde{k})}{\gamma \exp ((\gamma - 1) \tilde{k}) + 1 - \delta}
\]
(B.31)
\[
r_{k,t+1} \approx r + r_{kk} (\tilde{k}_{t+1} - \tilde{k})
\]
(B.32)
where \( r_{kk} \equiv (\gamma - 1)(\exp (r) - 1 + \delta) / \exp (r) \)

### B.3.4 Risk aversion

I guess that the innovation to the value function van be written as \( \kappa_v \epsilon_{t+1} \), so that

\[
\alpha_{t+1} = \phi \alpha_t + (1 - \phi) \bar{\alpha} + \lambda \kappa_v \epsilon_{t+1}
\]  

(B.34)

I confirm this guess below.

### B.3.5 Consumption dynamics

Writing \( \tilde{k}_t \equiv k_t - a_t \) and \( \tilde{c}_t \equiv c_t - a_t \), we have

\[
\tilde{k}_{t+1} = \lambda_0 + \lambda_k \tilde{k}_t + \lambda_c \tilde{c}_t - \sigma_a \epsilon_{a,t+1} - \mu
\]  

(B.35)

Now we guess that the consumption function is \( \tilde{c}_t = \eta_{c0} + \eta_{ck} \tilde{k}_t + \eta_{ca} \alpha_t \) (note here that I use \( \lambda \) terms for the budget constraint, which are terms depending only on the underlying parameters of the model; the \( \eta \) terms are coefficients from the optimal consumption rule). Then we have

\[
\tilde{k}_{t+1} = \lambda_0 + \lambda_k \tilde{k}_t + \lambda_c \tilde{c}_t - \sigma_a \epsilon_{a,t+1} - \mu
\]  

(B.36)

\[
\eta_{k0} \equiv \lambda_0 + \lambda_c \eta_{c0} - \mu
\]  

(B.37)

\[
\eta_{kk} \equiv \lambda_k + \lambda_c \eta_{ck} \quad \eta_{ka} \equiv \lambda_c \eta_{ca}
\]

This equation specifies the dynamics of capital conditional on the underlying parameters of the model and the two unknown coefficients determining the dynamics of consumption.

For consumption growth, we say \( \Delta c_{t+1} = \eta_{d0} + \eta_{dk} \tilde{k}_t + \eta_{da} \alpha_t + \kappa_d \epsilon_{t+1} \), where

\[
\eta_{d0} \equiv \eta_{ck} \eta_{k0} - \eta_{ca} (\phi \alpha - 1) \bar{\alpha} + \mu \quad \eta_{dk} \equiv \eta_{ck} (\eta_{kk} - 1) \\
\kappa_d \equiv \sigma_a (1 - \eta_{ck}) + \eta_{ca} \lambda \kappa_v \quad \eta_{da} \equiv \eta_{ck} \eta_{ka} + \eta_{ca} (\phi \alpha - 1)
\]
The remainder of the appendix confirms that our guesses for the form of the consumption and value functions are correct.

B.3.6 Wealth return

In the presence of balanced growth, the long-run response of consumption to an innovation of $\sigma_a \varepsilon_t$ to technology must be exactly $\sigma_a \varepsilon_t + 1$. This is equivalent to saying that

$$\Delta E_{t+1} \sum_{j=0}^{\infty} \Delta c_{t+j+1} = \sigma_a \varepsilon_{t+1}$$

(B.38)

In the case where $\theta$ approaches 1 (the steady-state dividend/price ratio approaches zero) or the consumption response only takes one period, $\Delta E_{t+1} \sum_{j=0}^{\infty} \Delta c_{t+j+1} = \Delta E_{t+1} \sum_{j=0}^{\infty} \theta^j \Delta c_{t+j+1}$.

We therefore have the approximation,

$$r_{w,t+1} = E_t r_{w,t+1} + \sigma_a \varepsilon_{t+1} - \Delta E_{t+1} \sum_{j=1}^{\infty} \theta^j r_{w,t+j+1}$$

(B.39)

This extra approximation is not strictly necessary, and the model is straightforward to solve without it. However, it substantially simplifies many of the formulas and makes them more transparent. The results reported below on the accuracy of the log-linear solution apply to the solution using this approximation.

Now note that

$$\Delta E_{t+1} \sum_{j=1}^{\infty} \theta^j r_{w,t+j+1} = \Delta E_{t+1} \sum_{j=1}^{\infty} \theta^j (\alpha c_{t+j} \eta_{wa} + \rho E_{t+j} \Delta c_{t+j})$$

(B.40)

$$= \eta_{wa} \frac{\theta}{1 - \theta \phi} \sigma_{aa} + \rho (\eta_{ca} \sigma_a - \eta_{ca} \sigma_{aa})$$

(B.41)

The second term follows from the approximation

$$\Delta E_{t+1} \sum_{j=1}^{\infty} \theta^j \rho E_{t+j} \Delta c_{t+j} \approx \Delta E_{t+1} \sum_{j=1}^{\infty} \rho E_{t+j} \Delta c_{t+j}$$

The right hand side of this equation is simply $\rho$ multiplied by the total amount of consumption growth expected following period $t + 1$. Since we know that in the long run, the
consumption-technology ratio is stationary, we just need to know how much consumption declines relative to technology at period \( t + 1 \). That’s going to be exactly \( \eta_c k \sigma_a \varepsilon_{t+1} - \eta_c a \sigma_{aa} \varepsilon_{t+1} \) (since capital falls by \( \sigma_a \) and \( \alpha \) falls by \( -\sigma_{aa} \)).

We then have

\[
\kappa_r = \sigma_a - \eta_w a \frac{\theta}{1 - \theta \phi} \sigma_{aa} - \rho (\eta_c k \sigma_a - \eta_c a \sigma_{aa})
\]

(B.42)

The return is thus

\[
r_{t+1} = \eta_w 0 + \rho E_t \Delta c_{t+1} + \eta_{wa} \alpha_t + \sigma_a \varepsilon_{t+1} + \left( -\eta_w a \frac{\theta}{1 - \theta \phi} \sigma_{aa} + \frac{\theta}{1 - \theta k_k \eta} \eta_d k \sigma_a \right) \varepsilon_{t+1}
\]

(B.43)

### B.3.7 The Euler equation for wealth

The asset pricing equation gives us

\[
1 = E_t \left[ \exp \left( \rho \frac{1 - \alpha_t}{1 - \rho} \left( \frac{C_{t+1}}{C_t} \right)^{-\rho \frac{1 - \alpha_t}{1 - \rho}} \frac{1 - \alpha_t}{1 - \rho} \right) R_{w,t+1} \right]
\]

(B.44)

The log of the term inside the expectation is

\[
\log \left[ \exp \left( \rho \frac{1 - \alpha_t}{1 - \rho} \left( \frac{C_{t+1}}{C_t} \right)^{-\rho \frac{1 - \alpha_t}{1 - \rho}} \frac{1 - \alpha_t}{1 - \rho} \right) R_{w,t+1} \right] = -\frac{1 - \alpha_t}{1 - \rho} \beta - \rho \frac{1 - \alpha_t}{1 - \rho} \Delta c_{t+1}
\]

\[
+ \frac{1 - \alpha_t}{1 - \rho} (\eta_w 0 + \rho E_t \Delta c_{t+1} + \eta_{wa} \alpha_t + \kappa_r \varepsilon_{t+1})
\]

(B.45)

\[
= -\frac{1 - \alpha_t}{1 - \rho} \beta - \rho \frac{1 - \alpha_t}{1 - \rho} \kappa_d \varepsilon_{t+1}
\]

\[
+ \frac{1 - \alpha_t}{1 - \rho} (\eta_w 0 + \eta_{wa} \alpha_t + \kappa_r \varepsilon_{t+1})
\]

(B.46)

Now taking the log of the expectation and dividing by \( \frac{1 - \alpha_t}{1 - \rho} \) gives

\[
0 = -\beta + \eta_w 0 + \eta_{wa} \alpha_t + \frac{11 - \alpha_t}{2 \frac{1 - \rho}{1 - \rho}} (-\rho \kappa_d + \kappa_r)^2
\]

(B.47)
which implies

\[ \eta_{w0} = \beta - \frac{1}{2} \frac{(-\rho \kappa_d + \kappa_r)^2}{1 - \rho} \]  \hspace{1cm} (B.48)

\[ \eta_{wa} = \frac{1}{2} \frac{1}{1 - \rho} (-\rho \kappa_d + \kappa_r)^2 \]  \hspace{1cm} (B.49)

\[ \kappa_r = \sigma_a - \eta_{wa} \frac{\rho}{1 - \theta \phi} \sigma_{aa} - \rho (\eta_{ck} \sigma_a - \eta_{ca} \sigma_{aa}) \]  \hspace{1cm} (B.50)

B.3.8 The Euler equation for capital

The stochastic discount factor follows

\[ m_{t+1} = -\beta - \rho E \Delta c_{t+1} - \rho \frac{1 - \alpha_t}{1 - \rho} \kappa_d \epsilon_{t+1} + \frac{\rho - \alpha_t}{2} \frac{1}{1 - \rho} (-\rho \kappa_d + \kappa_r)^2 + \kappa_r \epsilon_{t+1} \]  \hspace{1cm} (B.51)

For the capital return we have

\[ 1 = E_t \left[ \exp \left( m_{t+1} + r + r_{kk} (\eta_{k0} + \eta_{kk} \tilde{k}_t + \eta_{ka} \alpha_t - \sigma_a \epsilon_{a,t+1} - \bar{k}) \right) \right] \]  \hspace{1cm} (B.52)

which implies

\[ 0 = -\beta - \rho E \Delta c_{t+1} + \frac{1}{2} \frac{(1 - \alpha_t)}{(1 - \rho)} (-\rho \kappa_d + \kappa_r)^2 \]

\[ + r + r_{kk} (\eta_{k0} + \eta_{kk} \tilde{k}_t + \eta_{ka} \alpha_t - \bar{k}) \]

\[ + \frac{1}{2} (r_{kk} \sigma_a + \kappa_r)^2 - (r_{kk} \sigma_a + \kappa_r) \left( \frac{1 - \alpha_t}{1 - \rho} \right) (-\rho \kappa_d + \kappa_r) \]  \hspace{1cm} (B.53)

Note that all of the nonlinearities disappear (i.e. the \( \alpha_t^2 \) terms), and this equation is linear in the state variables. We can thus solve through the method of undetermined coefficients as usual.

For the coefficients on capital,

\[ 0 = -\rho \left( \lambda_k \eta_{ck} + \lambda_c \eta_{ck}^2 - \eta_{ck} \right) + r_{kk} (\lambda_k + \lambda_c \eta_{ck}) \]  \hspace{1cm} (B.54)
This is quadratic in $\eta_{kk}$, and we have

$$
\eta_{ck} = \frac{\rho \left(1 - \lambda_k\right) + r_{kk}\lambda_c \pm \sqrt{\left(\rho \left(1 - \lambda_k\right) + r_{kk}\lambda_c\right)^2 + 4\rho\lambda_c r_{kk}\lambda_k}}{2\rho\lambda_c}
$$

(B.55)

Now note that $\lambda_c < 0$, $\lambda_k > 0$, and $r_{kk} < 0$. This implies that

$$
\sqrt{\left(\rho \left(1 - \lambda_k\right) + r_{kk}\lambda_c\right)^2 + 4\rho\lambda_c r_{kk}\lambda_k} > \rho \left(1 - \lambda_k\right) + r_{kk}\lambda_c
$$

(B.56)

and hence $\eta_{ck}$ has a positive and a negative root. The root where $\eta_{ck} < 0$ violates the transversality condition (high capital implies low consumption), so we choose the root with $\eta_{ck} > 0$. Note that the formula for $\eta_{ck}$ does not involve $\bar{\alpha}$ or $\sigma_{aa}$, which confirms remark 1.

For the coefficients on $\alpha_t$, we have

$$
0 = -\rho\eta_{da} - \frac{1}{2} \left(-\rho\kappa_d + \kappa_r\right)^2 + r_{kk}\eta_{ka} + (r_{kk}\sigma_a + \kappa_r) \left(-\rho\kappa_d + \kappa_r\right) \frac{1}{1 - \rho}
$$

(B.57)

**B.3.9 Other parameters**

To solve for $(-\rho\kappa_d + \kappa_r)$, simply combine the equations for $\eta_{iva}$ and $\kappa_r$, yielding

$$
-\rho\kappa_d + \kappa_r = \frac{-1 + \sqrt{1 + 2\frac{\sigma}{1 - \rho\sigma_a^2}\sigma_a^2}}{1 - \rho\frac{\sigma}{1 - \rho\sigma_a^2}}
$$

(B.58)

We choose the root for this equation that has the property that it approaches zero as $\sigma_a$ approaches zero. That is, we know that when the shocks have zero variance, all assets have the same return, and so $\eta_{iva} = 0$. 

160
B.3.10  Excess returns and the risk-free rate (result 2)

To calculate excess returns, we can simply calculate the covariance of the wealth return with the SDF. The Sharpe ratio of the wealth portfolio is

\[
\frac{E_t r_{w,t+1} - r_{f,t+1} + \frac{1}{2} \kappa_r^2}{\kappa_r} = \alpha_t - \rho \left( -\rho \kappa_d + \kappa_r \right) + \rho \kappa_d \kappa_r
\]

(B.59)

\[
= (\alpha_t - \rho) \frac{-1 + \sqrt{1 + 2 \frac{\theta}{1 - \theta} \sigma_a \sigma_d}}{\theta} + \rho \kappa_d \kappa_r
\]

(B.60)

This also immediately gives a formula for the risk-free rate

\[
E_t r_{w,t+1} - r_{f,t+1} + \frac{1}{2} \kappa_r^2 = \alpha_t - \rho \left( -\rho \kappa_d + \kappa_r \right) \kappa_r + \rho \kappa_d \kappa_r \quad (B.61)
\]

The Campbell–Shiller approximation for the wealth-consumption ratio is

\[
\bar{w}_t - c_t = \frac{z}{1 - \theta} + E_t \sum_{j=0}^{\infty} \theta^j \left( \Delta c_{t+j+1} - r_{t+j+1} \right)
\]

(B.63)

for a constant \(z\) depending on the average consumption-wealth ratio (i.e. related to \(\theta\)), \(z \equiv -\log \theta - (1 - \theta) \log (\frac{1}{\theta} - 1)\). Now

\[
E_t \sum_{j=0}^{\infty} \theta^j \Delta c_{t+j+1} = -\eta_{ck} (k_t - \bar{k}) - \eta_{ca} (\alpha_t - \bar{\alpha})
\]

(B.64)

under the approximation \(\theta = 1\) from above, and

\[
E_t \sum_{j=0}^{\infty} \theta^j r_{w,t+j+1} = E_t \sum_{j=0}^{\infty} \theta^j \left( \alpha_{t+j} \eta_{wa} + \rho E_{t+j} \Delta c_{t+j+1} \right)
\]

(B.65)
So

\[ w_t - c_t = \frac{z}{1 - \theta} + (1 - \rho) (-\eta_{ck} (k_t - \bar{k}) - \eta_{ca} (\alpha_t - \bar{\alpha})) - \eta_{wa} \frac{\theta}{1 - \theta \phi} (\alpha_t - \bar{\alpha}) \quad (B.66) \]

B.3.12 The value function and risk aversion (result 4)

At any time, household value is

\[ W_t = V_t^{1-\rho} C_t^{\rho} / (1 - \exp (-\beta)) \quad (B.67) \]

\[ v_t = \frac{(w_t - c_t)}{1 - \rho} + c_t + \frac{\log (1 - \exp (-\beta))}{1 - \rho} \quad (B.68) \]

The innovation to the value function, \( v_{t+1} - E_t v_{t+1} \), is equal to the sum of the innovations to \( \frac{(w_{t+1} - c_{t+1})}{1 - \rho} \) and \( c_{t+1} \), which are

\[ \Delta E_{t+1} \frac{(w_{t+1} - c_{t+1})}{1 - \rho} + \Delta E_{t+1} c_{t+1} = \sigma_a \varepsilon_{t+1} - \eta_{wa} \frac{\theta}{1 - \rho} \frac{1}{1 - \theta \phi} \sigma_a \varepsilon_{t+1} \]

\[ \kappa_v = \sigma_a \varepsilon_{t+1} - \frac{\theta}{1 - \theta \phi} \frac{1}{2} \left( \frac{1}{1 - \rho} \right)^2 (-\rho \kappa_d + \kappa_r)^2 \sigma_a \varepsilon_{t+1} \]

(B.69)

(B.70)

Using the formula from above that defines \((-\rho \kappa_d + \kappa_r)\), we have \( \kappa_v = \frac{(-\rho \kappa_d + \kappa_r)}{1 - \rho} \) and

\[ \sigma_a = \lambda \kappa_v = \lambda \frac{-1 + \sqrt{1 + 2 \frac{\theta}{1 - \theta \phi} \sigma_a \varepsilon_a}}{\frac{\theta}{1 - \theta \phi} \sigma_a} \]

(B.71)

B.3.13 Affine bond pricing (result 5)

\[ m_{t+1} = -\beta - \rho E \Delta c_{t+1} - \rho \frac{1 - \alpha_t}{1 - \rho} \kappa_d \varepsilon_{t+1} + \frac{\rho - \alpha_t}{1 - \rho} \left( \frac{1}{2} \frac{\theta}{1 - \rho} \right) (-\rho \kappa_d + \kappa_r)^2 + \kappa_r \varepsilon_{t+1} \]

(B.72)
\[ \text{var} (m_{t+1}) = \left( \frac{1 - \alpha_t}{1 - \rho} (-\rho \kappa_d + \kappa_r) - \kappa_r \right)^2 \sigma^2 \]  
(B.73)

\[ = \left[ \left( \frac{1 - \alpha_t}{1 - \rho} \right)^2 (-\rho \kappa_d + \kappa_r)^2 + \kappa_r^2 - 2 \frac{1 - \alpha_t}{1 - \rho} (-\rho \kappa_d + \kappa_r) \kappa_r \right] \sigma^2 \]  
(B.74)

And hence

\[ r_{f,t+1} = -\left( E_t m_{t+1} + \frac{1}{2} \text{var} (m_{t+1}) \right) \]  
(B.75)

\[ = -\left( -\beta - \rho \left( \eta_{d0} + \eta_{dk} \tilde{k}_t + \eta_{da} \alpha_t \right) + \frac{1}{2} \frac{(1 - \alpha_t)}{(1 - \rho)} (-\rho \kappa_d + \kappa_r)^2 + \frac{1}{2} \kappa_r^2 - \frac{1 - \alpha_t}{1 - \rho} (-\rho \kappa_d + \kappa_r) \kappa_r \right) \]  
(B.76)

It is straightforward to show that

\[ m_{t+1} = -r_{f,t+1} + \left( \frac{1 - \alpha_t}{1 - \rho} (-\rho \kappa_d + \kappa_r) - \kappa_r \right) \varepsilon_{t+1} - \frac{1}{2} \left( \frac{1 - \alpha_t}{1 - \rho} (-\rho \kappa_d + \kappa_r) + \kappa_r \right)^2 \sigma^2 \]  
(B.77)

\[ = -r_{f,t+1} - \frac{1}{2} (\omega_0 + \omega_1 \alpha_t)^2 \sigma^2 + (\omega_0 + \omega_1 \alpha_t) \varepsilon_{t+1} \]  
(B.78)

So the SDF takes the essentially affine form with

\[ \omega_0 = \frac{(-\rho \kappa_d + \kappa_r)}{1 - \rho} - \kappa_r \quad \omega_1 = \frac{(-\rho \kappa_d + \kappa_r)}{1 - \rho} \]

B.3.14 Accuracy of the approximation

Table B.1 reports simple statistics summarizing the relationship between the projection solution and the log-linear approximation to the model. The first column lists the mean difference between the solutions, the second column the standard deviation of the gap, and the third column the standard deviation of the gap scaled by the standard deviation of the variable in the projection solution. I report deviations for log capital, log consumption growth, the coefficient of relative risk aversion, and the Sharpe ratio of the wealth portfolio. For the simulations, both models start with the same initial levels of capital and risk aversion and use the same technology shocks. I then simulate the models for 20,000 periods.
**Figure B.1:** Comparison of results from simulations of projection and log-linear model solutions

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Scaled std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>-8.72E-04</td>
<td>0.00073</td>
<td>0.020</td>
</tr>
<tr>
<td>Cons. Growth</td>
<td>3.16E-08</td>
<td>0.00013</td>
<td>0.028</td>
</tr>
<tr>
<td>RRA</td>
<td>1.11E-02</td>
<td>0.75344</td>
<td>0.116</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>-2.15E-03</td>
<td>0.00892</td>
<td>0.123</td>
</tr>
</tbody>
</table>

Note: Comparison of the projection and log-linear solutions. The two simulations use the same shocks but different policy functions. The first column is the mean difference between the simulations, the second column the standard deviation, and the third column the standard deviation of the difference scaled by the standard of the variable in the projection solution. **RRA** is relative risk aversion.
Table B.1 shows that for capital and consumption growth, the log-linear approximation is nearly identical to the projection solution. The mean differences are essentially zero, and the standard deviations of the errors are both less than 3 percent of the standard deviations of the variables themselves. For risk aversion and the Sharpe ratio, the log-linear approximation is essentially identical to the projection solution on average, but the standard deviation of the differences is now roughly 12 percent of the standard deviation of the variables themselves.

An alternative method of checking the accuracy of the approximation is to look at Euler equation errors. Figure A.1 plots histograms of the log₁₀ Euler equation errors, $\log_{10}\left(E_t \left[M_{t+1} \left(\alpha K_{t+1}^{\alpha-1} + 1 - \delta \right) \right] - 1\right)$ under the projection solution and the log-linear solution at each date in the simulation.

### B.4 Details of return forecasting

#### B.4.1 The method from Lettau and Ludvigson (2001)

If consumption and wealth are cointegrated, then we have the relationship

$$c_t = \zeta w_t + \xi_t$$  \hspace{1cm} (B.79)

where $\zeta$ is a parameter, and $\xi_t$ is a mean-zero, stationary, and not necessarily i.i.d. error term. If we observed wealth, $\zeta$ and $\xi_t$ could be directly estimated. We do not observe wealth, though, especially the human component. Lettau and Ludvigson (2001) therefore use the approximation

$$w_t = \omega a_t + (1 - \omega) h u_t$$  \hspace{1cm} (B.80)

where $a_t$ is asset wealth and $h u_t$ human wealth. This equation simply says that log aggregate wealth is equal to the sum of log asset and human wealth. Since the level of aggregate wealth is equal to the sum of the levels of asset and human wealth, the approximation is valid as long as the shares of asset and human wealth in aggregate wealth are stationary not too variable. The fact that labor’s share of income has been stationary in the post-war US data makes this assumption reasonable.
Figure B.2: $\log_{10}$ Euler equation error densities

Note: Densities of Euler equation errors under the two solution methods. The log errors are defined as $\log_{10}(\left|E[M_{t+1}\Delta R_{t+1}]\right|^{-1})$. Densities are estimated using a kernel smoother on simulated data. In both cases, the model used is the benchmark single-shock model with EZ-habit preferences and constant labor supply.
Finally, we assume that labor income, $y_t$, can be viewed as the dividend from human wealth and that the dividend/price ratio for human wealth is stationary. That is,

$$y_t = g + hu_t - \mu_t$$  \hspace{1cm} (B.81)

where $g$ is a parameter and $\mu_t$ is a mean-zero stationary $b_{z,1}$ term. This implies that

$$w_t = \omega a_t + (1 - \omega) y_t + (1 - \omega) g + \mu_t$$  \hspace{1cm} (B.82)

$$c_t = \zeta \omega a_t + \zeta (1 - \omega) y_t + \zeta (1 - \omega) g + \zeta \mu_t + \xi_t$$  \hspace{1cm} (B.83)

since $\zeta + \zeta \mu_t$ is mean-zero and stationary, regardless of any correlation between $\zeta_t$ and $\mu_t$, the variables $c_t$, $a_t$, and $y_t$ are jointly cointegrated. The parameters $\zeta$, $\omega$, and $g$ can be estimated through standard methods for cointegrated models. As Lettau and Ludvigson point out, the estimation is of these parameters is superconsistent, converging linearly with sample size, so these parameters can be taken as known with certainty in any subsequent analyses (in particular, stock return forecasts).

I follow Lettau and Ludvigson in referring to the cointegrating residual, $\zeta \mu_t + \xi_t = c_t - \zeta \omega a_t - \zeta (1 - \omega) y_t - \zeta (1 - \omega) g$ as $cay_t$, and I refer to $\omega a_t + (1 - \omega) \times y_t$ as $ay_t$. $ay_t$ is an estimate of total wealth derived from data on consumption, asset wealth, and labor income, taking advantage of an assumed cointegrating relationship between the three variables. I estimate the parameters using standard maximum likelihood methods.

**B.4.2 Sensitivity analysis for return forecasting**

The results in section 2.4.2 depend on choices for two parameters – the EIS and the persistence of risk aversion. Tables B.2 and B.3 report the ratio of the $R^2$ for excess value to $cay$ for 1, 5, 10, and 20-quarter returns across a variety of choices for the EIS and the persistence of risk aversion.

Table B.2 varies the EIS between 0.75 and 10. The numbers in bold represent points where $cay$ outperforms $\hat{\alpha}$. When the EIS is greater than 1, $cay$ only ever outperforms at the 1-quarter horizon, and then only if the EIS is set to 10. With an EIS less than 1, though,
always has an R\textsuperscript{2} substantially larger that of \( \hat{\alpha} \). Moreover, the sign on \( \hat{\alpha} \) in the return regressions flips. Intuitively, this is because in the construction of \( \hat{v} \), when the EIS is less than 1, the weight on aggregate wealth is negative. The theory would predict that high risk aversion is associated with low returns, but with the EIS less than 1, \( \hat{\alpha} \) and future returns are actually positively correlated.

Table B.3 presents R\textsuperscript{2} ratios for the same set of regressions, but now varying the persistence of risk aversion. Across a fairly wide range of autocorrelations, \( \hat{\alpha} \) outperforms \textit{cay} at most horizons. The best performance is found with an annual autocorrelation of 0.9, which corresponds to \( \phi = 0.974 \). Even with an autocorrelation as low as 0.65 (\( \phi = 0.9 \)), though, \( \hat{\alpha} \) performs nearly as well as \textit{cay}. As with the EIS, the place where \textit{cay} is most likely to outperform is with 1-quarter returns. Table B4 lists R\textsuperscript{2}s for \textit{cay}, PE, and \( \hat{\alpha} \) for pre and post-1980 samples.

### B.4.3 Out-of-sample forecasting regressions

An alternative to the in-sample regressions studied in the main text is out-of-sample tests of forecasting power. I consider the mean squared forecast \( b_{z,1} \) (MSFE) based tests from analyzed in Clark and McCracken (2001, 2005) and Clark and West (2007).

Suppose we want to test whether a single variable, \( x_t \), forecasts stock returns, \( r_t \), against the null that \( r_t \) is i.i.d. (the methods used here apply to any null model that is nested; i.e. they are appropriate for asking whether \( x_t \) has marginal forecasting power when added to some other model). The forecast horizon can be any length. Therefore, denote \( r_{t,t+j} \equiv \sum_{\tau=t}^{t+j} r_{\tau} \).

We compare the residuals from the null model, \( e_{1t} \equiv r_{t,t+j} - \hat{\beta}_{0,t} \) (for an estimated constant mean \( \hat{\beta}_{0,t} \) using data prior to date \( t \)) to the residuals from the alternative model, \( e_{2t,t+j} = r_{t,t+j} - \hat{\beta}_{0,t} - \hat{\beta}_{1t} x_{t+j-1} \) (where \( \hat{\beta}_{1t} \) is a constant regression coefficient estimates on the data from \( \tau = 0 \) to \( \tau = t-1 \)). The samples for the regressions are begun after the first 20 percent of the sample.

The measure of the difference in MSFE is

\[
 f_{t,t+j} \equiv e_{1t,t+j}^2 - e_{2t,t+j}^2 + (e_{1t,t+j} - e_{2t,t+j})^2 \tag{B.84}
\]
Figure B.3: Relative $R^2$'s for varying EIS

<table>
<thead>
<tr>
<th>Span</th>
<th>EIS=0.1</th>
<th>0.25</th>
<th>0.75</th>
<th>1.25</th>
<th>1.5</th>
<th>2</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 quarter</td>
<td>0.39</td>
<td>0.26</td>
<td>0.29</td>
<td>1.07</td>
<td>1.10</td>
<td>1.08</td>
<td><strong>0.96</strong></td>
</tr>
<tr>
<td>5 quarters</td>
<td>0.44</td>
<td>0.28</td>
<td>0.29</td>
<td>1.16</td>
<td>1.21</td>
<td>1.21</td>
<td>1.09</td>
</tr>
<tr>
<td>10 quarters</td>
<td>0.61</td>
<td>0.42</td>
<td>0.31</td>
<td>1.41</td>
<td>1.49</td>
<td>1.49</td>
<td>1.37</td>
</tr>
<tr>
<td>20 quarters</td>
<td>1.28</td>
<td>1.02</td>
<td><strong>0.21</strong></td>
<td>1.92</td>
<td>2.08</td>
<td>2.15</td>
<td>2.09</td>
</tr>
</tbody>
</table>

Note: This table lists the ratio of the $R^2$ for a univariate regression of long-horizon returns on estimated risk aversion to the $R^2$ for $\sigma$. Values less than 1 are in bold. The span in quarters is listed in the left hand column. The top row gives the EIS. The EIS is used to calculate household value and risk aversion.
**Figure B.4:** Relative $R^2$s for varying persistence of risk aversion

<table>
<thead>
<tr>
<th>Span</th>
<th>Autocorr.=0.95</th>
<th>0.9</th>
<th>0.85</th>
<th>0.8</th>
<th>0.75</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 quarter</td>
<td>0.86</td>
<td>1.27</td>
<td>1.10</td>
<td>0.90</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>5 quarters</td>
<td>1.03</td>
<td>1.44</td>
<td>1.21</td>
<td>0.97</td>
<td>0.78</td>
<td>0.64</td>
</tr>
<tr>
<td>10 quarters</td>
<td>1.17</td>
<td>1.73</td>
<td>1.49</td>
<td>1.20</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>20 quarters</td>
<td>1.24</td>
<td>2.24</td>
<td>2.08</td>
<td>1.76</td>
<td>1.48</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Note: This table lists the ratio of the $R^2$ for a univariate regression of long-horizon returns on estimated risk aversion to the $R^2$ for $a_{yt}$. Values less than 1 are in bold. The span in quarters is listed in the left hand column. The top row gives the annual autocorrelation of risk aversion.
Figure B.5: $R^2$ from pre and post-1980 univariate return forecasting regressions

<table>
<thead>
<tr>
<th>pre-1980</th>
<th>Estim. RRA</th>
<th>$\sigma_Y$</th>
<th>P/D</th>
<th>post-1980</th>
<th>Estim. RRA</th>
<th>$\sigma_Y$</th>
<th>P/D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q</td>
<td>0.10</td>
<td>0.10</td>
<td>0.03</td>
<td>1q</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>5q</td>
<td>0.28</td>
<td>0.22</td>
<td>0.25</td>
<td>5q</td>
<td>0.18</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>10q</td>
<td>0.27</td>
<td>0.16</td>
<td>0.27</td>
<td>10q</td>
<td>0.48</td>
<td>0.36</td>
<td>0.19</td>
</tr>
<tr>
<td>20q</td>
<td>0.38</td>
<td>0.13</td>
<td>0.39</td>
<td>20q</td>
<td>0.56</td>
<td>0.33</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Note: $R^2$s from univariate regressions of long-horizon stock returns on estimated risk aversion, $\sigma_Y$, and the price/dividend ratio. The highest value for each horizon and sample is listed in bold.
Under the null, the MSFE for the \( e_1 \) model tends to be smaller than the MSFE for the \( e_2 \) model because the \( e_2 \) model has added noise due to the extraneous predictor. Intuitively, model \( e_1 \) correctly imposes the constraint that \( \beta_1 = 0 \) under the null. The term \((e_{1t,t+j} - e_{2t,t+j})^2\) is essentially a correction for this effect.

When the forecast horizon is more than a single observation, \( f_{t,t+j} \) is serially correlated. To correct for this, we divide by a consistent estimate of its long-run variance (spectral density at frequency zero). Following Clark and West (2007), I use the Newey–West measure with a lag window of \( 1.5 \times j \). Denote this measure of the long-run variance as \( S(f_{t,t+j}) \). The long-run variance corrects for the fact that the forecast \( b_{z,1} \) from overlapping samples will be serially correlated. Clark and McCracken tabulate the critical values of the statistic \( \left( \frac{(T - j) \sum_{t=1}^{T-j} f_{t,t+j}}{S(f_{t,t+j})} \right) \).

In the main text, \( \hat{\alpha} \) is calculated using full-sample information. In particular, we need to calculate the cointegrating relationship between consumption, labor income, and financial wealth. We also need to know the average growth rate of value. For the out-of-sample forecasts, all of those parameters are estimated using only backward-looking information. The only possible source of look-ahead bias here would be data revisions.

The top panel of figure B.2 plots the values of the statistics using \( \hat{\alpha}_t \) as the predictor against a null of a constant expected equity for horizons from 1 to 20 quarters. We can easily reject the null at the 5 percent level at all horizons and at the 1% level for 2–13 quarter horizons.

### B.4.3.1 Bootstrapping

A major concern with predictive regressions is that asymptotic distribution theory is often a poor guide to small-sample behavior. A simple way to deal with that concern is to use a bootstrap to construct confidence intervals for the test statistics. I construct bootstrap samples in the following way. I select bootstrap samples of stock returns and growth rates of consumption, asset wealth, and labor income. I then construct level series for consumption, wealth, and income, and calculate \( \hat{\alpha} \) using purely backward-looking information as above. Finally, I construct the test statistic from above for each bootstrapped sample at

172
Estimated risk aversion depends on the cointegrating model used to estimate $\gamma$. The top panel tests whether estimated risk aversion has marginal forecasting power against a null of a constant-mean model for returns. The cointegrating vector is reestimated in each period using only backward-looking information. The bottom panel tests adding estimated risk aversion to a null model including a constant and $\gamma$ and vice versa.

Note: Out-of-sample test statistics from Clark and McCracken (2001, 2005) based on the reduction in out-of-sample RMSE.
each horizon from 1 to 20 quarters. I bootstrap 10,000 samples of data. The top panel of
diagram B.2 plots the 95th and 99th percentiles of the bootstrapped test statistics, and the
out-of-sample forecasting power is still significant at the 5 percent level.

B.4.3.2 $\hat{\alpha}$ versus $cay$

We can also use the out-of-sample test to ask whether estimated risk aversion forecasts
stock returns better than $cay$. The null model is now one where stock returns depend on
a constant and the lagged value of $cay$, and the encompassing alternative adds the lagged
value of $\hat{\alpha}$. The bottom panel of diagram A2 plots the test statistics. At every horizon, we can
reject the null that $\hat{\alpha}$ does not improve the forecast using $cay$ at the 5 percent level, and we
can reject the null at the 1 percent level at every horizon longer than 1 quarter.

Figure B.2 also plots the statistic for a test of whether $cay$ has any marginal predictive
power above that of $\hat{\alpha}$. At horizons shorter than 8 quarters, we cannot reject the null
that it does not. At longer horizons, though, there is evidence that both variables contain
important information for forecasting stock returns.
C. APPENDIX TO CHAPTER 3

C.1 Results from the text

C.1.1 Price of a utility claim and the SDF

The utility claim pays \( U_{t-1} \) as its dividend. We confirm that its cum-dividend price is \( W_{U,t} = V_t^{1-\rho} B_t^{-1} U_{C,t}^{-1} / (1 - \beta) \) by simply inserting this guess into the Euler equation,

\[
W_{U,t} = U_t U_{C,t}^{-1} + E_t \left[ W_{U,t+1} M_{t+1} \right]
\]  \hspace{1cm} (C.1)

\[
V_t^{1-\rho} B_t^{-1} U_{C,t}^{-1} / (1 - \beta) = U_t U_{C,t}^{-1} + E_t \left[ \frac{V_t^{1-\rho} B_t^{-1} U_{C,t+1}^{-1}}{(1 - \beta)} \beta B_t U_{C,t+1} \frac{V_{t+1}^{\rho-\alpha}}{E_t V_{t+1}^1 (1-\alpha)} \right]
\]  \hspace{1cm} (C.2)

\[
V_t^{1-\rho} = (1 - \beta) B_t U_t + \beta E_t \left[ \frac{V_t^{1-\rho}}{(E_t V_{t+1}^{1-\alpha})^{1-\rho}} \right]
\]  \hspace{1cm} (C.3)

\[
V_t^{1-\rho} = (1 - \beta) B_t U_t + \beta E_t \left[ V_{t+1}^{1-\alpha} \right]^{1-\rho}
\]  \hspace{1cm} (C.4)

The last line shows that the guess for the price of the utility claim was in fact correct, since the Euler equation is guaranteed to hold.
Next, we consider the return on the utility claim. We have

\[
R_{U,t+1} = \frac{V_{t+1}^{1-\rho} B_{t+1}^{-1} U_{C,t+1}^{-1} / (1 - \beta)}{V_t^{1-\rho} B_t^{-1} U_{C,t}^{-1} / (1 - \beta) - U_t U_{C,t}}
\]  

(C.5)

\[
= \frac{V_t^{1-\rho}}{V_t^{1-\rho} - (1 - \beta) B_t U_t} \left( \frac{B_{t+1} U_{C,t+1}^{-1}}{B_t U_{C,t}} \right)^{-1}
\]  

(C.6)

\[
= \frac{V_t^{1-\rho}}{\beta \left( E_t V_t^{1-a_t} \right)^{1-\rho}} \left( \frac{B_{t+1} U_{C,t+1}^{-1}}{B_t U_{C,t}} \right)^{-1}
\]  

(C.7)

\[
\left( \frac{V_t^{1-\rho}}{(E_t V_t^{1-a_t})^{1-\rho}} \right)^{\frac{\rho-a_t}{1-\rho}} = R_{U,t+1}^{\frac{\rho-a_t}{1-\rho}} B_t \left( \frac{U_{C,t+1}}{U_{C,t}} \right)^{\frac{\rho-a_t}{1-\rho}} \left( \frac{B_{t+1}}{B_t} \right)^{\frac{\rho-a_t}{1-\rho}}
\]  

(C.8)

Now substitute the return into the SDF:

\[
M_{t+1} = \beta \frac{B_{t+1} U_{C,t+1}}{B_t U_{C,t}} \frac{R_{U,t+1}^{\frac{\rho-a_t}{1-\rho}}}{U_{C,t}} \left( \frac{U_{C,t+1}}{U_{C,t}} \right)^{\frac{\rho-a_t}{1-\rho}} \left( \frac{B_{t+1}}{B_t} \right)^{\frac{\rho-a_t}{1-\rho}}
\]  

(C.9)

\[
= \beta \frac{1-a_t}{1-\rho} \left( \frac{U_{C,t+1}}{U_{C,t}} B_{t+1} \right)^{\frac{1-a_t}{1-\rho}} \left( \frac{B_{t+1}}{B_t} \right)^{\frac{1-a_t}{1-\rho}} R_{U,t+1}^{\frac{\rho-a_t}{1-\rho}}
\]  

(C.10)

This is the formula from the text.

C.1.2 First-order condition for wage setting

\[
0 = E_t^* \sum_{k=0}^{\infty} \xi_k \frac{\psi_{t+k}}{\theta_w} \left[ \frac{1 + \theta_w}{\theta_w} \left( \frac{W_{t+k}}{W_{t+k}} \right)^{-\frac{1+a_t}{\rho}} \frac{N_{t+k}}{W_{t+1}} \right]^{\frac{\rho-a_t}{1-\rho}} - \lambda_{t+k} N_{t+k} (j)
\]  

(C.11)

\[
= E_t^* \sum_{k=0}^{\infty} \xi_k \left[ \left( \frac{dV_t}{dN_{t+k}} (j) + \frac{dV_t}{dC_{t+k}} \frac{W_{t+k} (j)}{P_t} \right) \left( 1 + \theta_w \right) - \theta_w \frac{dV_t}{dC_{t+k}} \frac{W_{t+k} (j)}{P_t} \right] N_{t+k} (j)
\]  

(C.12)

\[
= E_t^* \sum_{k=0}^{\infty} \xi_k \left[ \frac{dV_t}{dN_{t+k}} (j) \left( 1 + \theta_w \right) + \frac{dV_t}{dC_{t+k}} \frac{W_{t+k} (j)}{P_t} \right] N_{t+k} (j)
\]  

(C.13)

\[
= E_t^* \sum_{k=0}^{\infty} \xi_k \frac{dV_t}{dC_{t+k}} \left[ \frac{dV_t}{dN_{t+k}} (j) \left( 1 + \theta_w \right) + \frac{W_{t+k} (j)}{P_t} \right] N_{t+k} (j)
\]  

(C.14)
C.2 Approximation method

This section proceeds as follows. First, I fix notation for a general set of equilibrium conditions. Next, I describe the specifics of how to solve for the model’s equilibrium dynamics. Third, I show that the essentially affine approximation method delivers bond-pricing formulas in the essentially affine class of Duffee (2002).

C.2.1 Equilibrium conditions

Denote the vector of the variables in the model (including the exogenous processes) as $X_t$. In addition to the various variables described above that track the state of the economy and the shocks, $X_t$ will include the price/dividend ratio of the utility portfolio, the return on the utility portfolio, and the marginal utility of consumption. The vector of fundamental shocks in the model is denoted $\varepsilon_t \equiv [\varepsilon_{mp,t}, \varepsilon_{z,t}, \varepsilon_{b,t}, \varepsilon_{\mu,t}, \varepsilon_{\pi,t}, \varepsilon_{w,t}, \varepsilon_{a,t}, \varepsilon_{\pi+1,t}]$.

The equations determining the equilibrium of the model take the form

$$0 = G(X_t, X_{t+1}, \sigma \varepsilon_{t+1}) \quad (C.15)$$

where the expectation operator may appear in the function $G$. There is one equation for each variable in the model. The new parameter $\sigma$ is the usual parameter used in perturbation approximations that controls the variance of the shocks. In the true model, $\sigma = 1$, but we will consider an approximation around the point $\sigma = 0$.

The equations $G$ can be divided into two types: those that do not involve taking expectations over the SDF and those that do.

$$G(X_t, X_{t+1}, \varepsilon_{t+1}) = \begin{bmatrix} D(X_t, X_{t+1}, \sigma \varepsilon_{t+1}) \\ E_t [M(X_t, X_{t+1}, \sigma \varepsilon_{t+1}) F(X_t, X_{t+1}, \sigma \varepsilon_{t+1})] \end{bmatrix} \quad (C.16)$$

where $D$ and $F$ are vector-valued functions and $M$ is the (scalar-valued) stochastic discount factor. Note that this formulation does not actually restrict $F$. Specifically, suppose there were a set of equilibrium conditions $1 = E_t h(x_t, x_{t+1}, \sigma \varepsilon_{t+1})$, i.e. that do not involve the SDF. We could simply say that $F(x_t, x_{t+1}, \sigma \varepsilon_{t+1}) \equiv h(x_t, x_{t+1}, \sigma \varepsilon_{t+1}) / M(x_t, x_{t+1}, \sigma \varepsilon_{t+1})$. 

177
Note, though, that in this model, all expectational equations involve the SDF.

For the equations that do not involve the SDF, I use standard perturbation methods and simply take a log-linear approximation. We approximate $D$ as

$$0 = \log (D (\exp (x_t), \exp (x_{t+1}), \sigma \epsilon_{t+1}) + 1)$$  \hspace{1cm} (C.17)

$$0 \approx d_0 + d_x \hat{x}_t + d_x' \hat{x}_{t+1} + d_e \sigma \epsilon_{t+1}$$  \hspace{1cm} (C.18)

where the terms $d_0, d_x, d_x', d_e$ are coefficients from a Taylor-series approximation and

$$x_t \equiv \log X_t$$

$$\hat{x}_t \equiv \log X_t - \log \bar{X}$$

### C.2.2 Linearizing the Euler equations

I now show that if we log-linearize the function $F$, we can transform (3.40) into a linear condition that can be solved alongside the remaining equations. $M_{t+1}$ will not even be log-linear in the state variables, but we will be able to state the equilibrium conditions as a set of linear expectational difference equations.

First, guess that the approximated equilibrium dynamics of the model take the form

$$\hat{x}_{t+1} = C + \Phi \hat{x}_t + \Psi \sigma \epsilon_{t+1}$$  \hspace{1cm} (C.19)

where $\hat{x}_t \equiv \log (X_t) - \log (\bar{X})$. We confirm in the end that the solution to the approximated equilibrium conditions actually takes this form. Next, define the matrices $\Gamma_{UC}, \Gamma_b,$ and $\Gamma_r$ as matrices that select individual elements of $x_t$ such that

$$\hat{a}_{t,C} = \Gamma_{UC} \hat{x}_t, \quad \hat{b}_t = \Gamma_b \hat{x}_t, \quad \hat{r}_{t,U} = \Gamma_r \hat{x}_t$$  \hspace{1cm} (C.20)

where lower-case letters with circumflexes denote log deviations from non-stochastic steady-state values. That is, $\hat{a}_{t,C} \equiv \log (U_{C,t}) - \log (\bar{U}_C)$, etc. For the sake of arithmetical convenience, also define an auxiliary variable $\zeta_t \equiv \frac{1 - \alpha_t}{1 - \rho}$. 

178
C.2.2.1 The essentially affine SDF

\[ U = \left( \frac{C^\eta C_{t-1}^{1-\eta}}{1 - \rho} \right)^{1-\rho} + Z_t^{1-\rho} \varphi_1 \frac{(\bar{H} - N_t)^{\alpha_1(1-\rho)}}{1 - \rho} \]

\[ U_C = \eta C_t^{\eta(1-\rho)} - 1_c(1-\eta)(1-\rho) Z_t^{\eta(1-\rho)} - 1 - \eta(1-\rho) Z_t^{1-\rho} \]

\[ = \eta C_t^{\eta(1-\rho)} - 1_c(1-\eta)(1-\rho) Z_t^{\eta(1-\rho)} - 1 - \eta(1-\rho) Z_t^{1-\rho} \]

The SDF is

\[ M_{t+1} = \beta^{1-\beta} \left( \frac{U_{C,t+1} - 1_c - \beta B_{t+1}}{U_{C,t} - 1_c - \beta B_t} \right)^{1-\alpha} \varphi_1 \frac{\zeta_t}{\gamma_{U,t+1}} \quad (C.21) \]

\( M_{t+1} \) is completely log-linear in the endogenous variables \( U_{C,t} \) and \( R_{U,t+1} \), but it is not log-linear in \( B_t \) and \( B_{t+1} \). If not for these terms, we could use the exact formula for \( M_{t+1} \) in what follows. Since those terms are non-linear, I use the approximations,

\[ \log \frac{1 - \beta \exp(b_{t+1})}{\exp(-b_t) - \beta} \approx \frac{1}{1 - \beta} b_t - \frac{\beta}{1 - \beta} b_{t+1} \quad (C.22) \]

\[ m_{t+1}^{(1)} \equiv \hat{m} + \zeta_t \left( \Delta \hat{u}_{C,t+1} + \frac{1}{1 - \beta} b_t - \frac{\beta}{1 - \beta} b_{t+1} \right) + (\zeta_t - 1) \hat{r}_{U,t+1} \quad (C.23) \]

\( m_{t+1}^{(1)} \) is first-order accurate for \( m_{t+1} \), hence the superscript. In the continuous time limit, this formula becomes exact.

The Euler equation for the return on the utility portfolio is

\[ 1 = E_t \exp \left( \zeta_t \left( \Delta \hat{u}_{C,t+1} + \frac{1}{1 - \beta} b_t - \frac{\beta}{1 - \beta} b_{t+1} \right) + \zeta_t \hat{r}_{U,t+1} \right) \quad (C.24) \]

Taking logs of both sides and taking advantage of log-normality gives

\[ 0 = E_t \Delta \hat{u}_{C,t+1} - \hat{u}_{C,t} + \frac{1}{1 - \beta} b_t - \frac{\beta}{1 - \beta} E_t b_{t+1} + E_t \hat{r}_{U,t+1} \]

\[ + \frac{1}{2} \zeta_t \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1 - \beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' \left( \Gamma_{UC} - \frac{\beta}{1 - \beta} \Gamma_b + \Gamma_r \right)' \quad (C.25) \]
Moreover, this implies that the approximated SDF can be written as

\[ m_{t+1}^{(1)} = \zeta_t \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \sigma_{\epsilon_{t+1}} - \Gamma_r (\Phi x_t + \Psi \sigma_{\epsilon_{t+1}}) - \bar{r} \]

\[ - \frac{1}{2} \zeta_t^2 \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right)' \]  

(C.26)

which is the essentially affine form from the text. The reason that this form is useful is that any time the SDF is essentially affine, we can obtain an exact expression for \( E_t \exp (m_{t+1} + f_0 + f_x x_t + f_{x'} x_{t+1}) \). It also means that we can price any asset whose payoffs are linear in the endogenous variables, including real and nominal bonds.

C.2.2.2 Approximation to F

Next, we take a first-order Taylor approximation to \( \log F \) such that \( \log F(x_t, x_{t+1}, \epsilon_{t+1}) \approx f_0 + f_x x_t + f_{x'} x_{t+1} \), giving

\[ 1 = E_t \exp \left( \hat{m}_{t+1}^{(1)} + f_x x_t + f_{x'} x_{t+1} \right) \]  

(C.27)

Taking logs and evaluating the expectation yields

\[ 0 = -E_t \hat{r}_{U_{t+1}} + f_x x_t + f_{x'} x_{t+1} + \frac{1}{2} \sigma^2 (-\Gamma_r + f_{x'}) \Psi \Psi' (-\Gamma_r + f_{x'})' \]

\[ + \zeta_t \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' (-\Gamma_r + f_{x'})' \]  

(C.28)

(C.28) is the equation that we ultimately place into the system to be solved. It is completely linear in both \( x_t \) and \( \zeta_t \) (equivalently, \( \alpha_t \)).

C.2.2.3 Solution

Since every equation in the system is now linear in the variables of the model, we can solve the system for the parameters \( \Phi \) and \( \Psi \) from (C.19). Specifically, we solve the
following system,

\[
0 = -E_t \hat{r}_{t,t+1} + f_x x_t + f_x' E_t x_{t+1} + \frac{1}{2} \sigma^2 (-\Gamma_r + f_{x'}) \Psi \Psi' (-\Gamma_r + f_{x'})' \\
+ \xi_t \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1 - \beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' (-\Gamma_r + f_{x'})' \\
0 = d_0 + d_x x_t + \hat{d}_x x_{t+1} + d_\epsilon \sigma \epsilon_{t+1}
\]  

(C.29)

where \( \sigma = 1 \) in the stochastic equilibrium that we approximate. This system can be solved using, for example, Sims’ (2001) Gensys algorithm.

The last wrinkle here is that we cannot simply insert (C.28) into the set of equations to be solved since it involves the matrix \( \Psi \), which is one of the unknown structures we are solving for. I deal with this with a simple fixed-point iteration: I begin with the equations that we obtain from perturbation,

\[
0 = -\Gamma_r E_t x_{t+1} + f_x x_t + f_x' E_t x_{t+1} \\
0 = d_0 + d_x \hat{x}_t + \hat{d}_x x_{t+1} + d_\epsilon \sigma \epsilon_{t+1}
\]

which will deliver an initial value of \( \Psi \), denoted \( \Psi^{(1)} \). We then use \( \Psi^{(1)} \) to change the equilibrium condition to take the form

\[
0 = -\Gamma_r \Phi x_t - \bar{r} + f_x x_t + f_x' \Phi x_t + \frac{1}{2} \sigma^2 (-\Gamma_r + f_{x'}) \Psi^{(1)} \Psi^{(1)'} (-\Gamma_r + f_{x'})' \\
+ \xi_t \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1 - \beta} \Gamma_b + \Gamma_r \right) \Psi^{(1)} \Psi^{(1)'} (-\Gamma_r + f_{x'})' \\
0 = d_0 + d_x \hat{x}_t + \hat{d}_x x_{t+1} + d_\epsilon \sigma \epsilon_{t+1}
\]

(C.31)

which delivers a value \( \Psi^{(2)} \). Then simply iterate to convergence. I treat parameter sets for which the iteration diverges as inadmissible, setting the marginal likelihood to zero.
C.2.3 Bond pricing

To solve for bond prices, we guess that bond prices are log-linear in the vector of state variables, so that

\[
p_{n,t} = A_n + B_n x_t \tag{C.33}
\]

where \( p_{n,t} \) is the price of a zero-coupon bond that matures on date \( t + n \) and pays 1 unit of consumption. We can also write the price of a nominal bond as \( p_{n,t}^S \). Using the formula for the SDF from above, we have

\[
\exp \left( p_{n,t}^S \right) = E_t \exp \left( \frac{\zeta_t}{\Gamma} \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \varepsilon_{t+1} - \Gamma_r (\Phi x_t + \Psi \varepsilon_{t+1}) - \tilde{r} \right) + A_{n-1} + (B_{n-1} - \Gamma_\pi) (\Phi x_t + \Psi \varepsilon_{t+1}) - \frac{1}{2} \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' \left( \Gamma_{UC} + \Gamma_b + \Gamma_r \right)' \tag{C.34}
\]

\[
= -\tilde{r} - \Gamma_r \Phi x_t + A_{n-1} + (B_{n-1} - \Gamma_\pi) \Phi x_t + \frac{1}{2} \sigma^2 (B_{n-1} - \Gamma_\pi - \Gamma_r) \Psi \Psi' (B_{n-1} - \Gamma_\pi - \Gamma_r)' \tag{C.35}
\]

\[
+ \sigma^2 \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' \left( B_{n-1} - \Gamma_\pi - \Gamma_r \right)' \Gamma \zeta x_t \tag{C.36}
\]

Matching coefficients gives

\[
A_n = -\tilde{r} + A_{n-1} + \frac{1}{2} \sigma^2 (B_{n-1} - \Gamma_\pi - \Gamma_r) \Psi \Psi' (B_{n-1} - \Gamma_\pi - \Gamma_r)' \tag{C.37}
\]

\[
B_n = -\Gamma_r \Phi + (B_{n-1} - \Gamma_\pi) \Phi + \left( \Gamma_{UC} - \frac{\beta}{1-\beta} \Gamma_b + \Gamma_r \right) \Psi \Psi' (B_{n-1} - \Gamma_\pi - \Gamma_r)' \Gamma \zeta \tag{C.38}
\]

C.3 Estimation

Much of the analysis discusses the behavior of the model around the posterior mode (i.e. the peak of the posterior distribution; it is also a maximum-likelihood estimate penalized by the prior distribution). I also start the Metropolis–Hastings chain from that point. To search for the posterior mode, I begin by running a genetic algorithm on a population of 60 points drawn from the prior distribution. The genetic algorithm searches the parameter space by mixing parameter sets in the population and also allowing random mutations. After 30 iterations of the genetic algorithm, I take the point in the population with the
highest posterior density and use it as the starting point for Chris Sims’ CSMINWEL algorithm, which is a derivative-based hill-climbing algorithm that is designed for DSGE models. When CSMINWEL gets stuck, I also try the standard simplex algorithm.

I ran this combined search 2500 times (each search takes roughly an hour, so access to a large computing cluster was essential). The point that I am calling the posterior mode was found to be the peak in fewer than 100 of the searches. In other words, it is extremely difficult to find the peak of the posterior likelihood. In general, I found that it was far easier to find the posterior mode when risk aversion or the inflation target was held fixed, and easier still when bond prices were also dropped from the estimation. Even though the priors help to add curvature to the posterior likelihood surface, I still find many local maxima, a problem that plagues the bond-pricing literature. Furthermore, since the model is so highly constrained there is no straightforward way to use the more recent estimation algorithms for affine term structure models proposed by, for example, Hamilton and Wu (2012) and Joslin, Singleton, and Zhu (2010).

I simulate the posterior distribution using the adaptive random-walk Metropolis–Hastings algorithm of Haario, Saksman, and Tamminen (2001). I initialize the chain at the posterior mode. For the proposal distribution, I begin with a normal distribution whose variance matrix is equal to that of the prior, multiplied by $(2.38^2)/d$, where $d$ is the dimension of the parameter vector (49), which is the optimal scaling factor of Gelman, Roberts, and Gilks (1996). After 10,000 iterations of the algorithm, I update the variance matrix of the proposal distribution to be equal to the observed variance matrix for the first 10,000 iterations of the chain. Subsequently, the variance is updated each on each iteration using the sample variance of the chain up to the current iteration.1 I achieve relatively rapid mixing this way. The full chain has 1,000,000 draws, but it mixes well even by 100,000 draws.

---

1 A more common method in the DSGE literature is to use the hessian of the posterior around the posterior mode to determine the variance of the proposal distribution. I have difficulty calculating the hessian due to numerical instability.


