Returns, Dividends, and Optimal Portfolios
RETURNS, DIVIDENDS, AND OPTIMAL PORTFOLIOS

By Magnus Sander

A PhD thesis submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Business Economics

October 2016
Acknowledgements

First, I would like to thank the assessment committee consisting of Thomas Quistgaard Pedersen, Lasse Bork, and Hans Dewachter for their many useful comments and suggestions. Some of them have been incorporated into the present version of the dissertation while others remain for future research.

The dissertation was written during my graduate studies at the Department of Economics and Business Economics at Aarhus University from September 2012 to July 2016. I am thankful to the department and to the Center for Research in Econometric Analysis of Time Series (CREATES) for providing research facilities as well as financial support. Further, I am grateful for financial support from Hugo Evers & Co’s Studiefond, Augustinus Fonden, and Direktør Einar Hansen og hustru fru Vera Hansens Fond.

I had the good fortune to get Stig Vinther Møller as my thesis advisor. Stig is one of the most positive and encouraging people I have met, and it is a luxury for a young researcher to have a mentor like him. He has taught me so much about empirical finance and how to write papers. Moreover, I would like to thank my co-supervisor Tom Engsted for helping me visit UCSD and Allan Timmermann for hosting me there from September 2014 to April 2015. I was also fortunate to work with Martin M. Andreasen, who taught me about interest rate modeling.

Further, I enjoyed sharing an office with Jonas and Palle, and Niels always ensured a great social environment at the university. Finally, I am grateful to my family, who has always been supportive and shown an interest in my work.

Magnus Sander

Oslo, October 2016
# Contents

Summary vi

Danish Summary vii

1 Market Timing in Recessions and Expansions 1
   1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
   1.2 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
   1.3 Return predictability across the business cycle . . . . . . . . . . . . 4
   1.4 Gains from switching predictability . . . . . . . . . . . . . . . . . . 6
   1.5 Longer horizon . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
   1.6 Markov switching . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
   1.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
Acknowledgements . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

2 Dividends, Earnings, and Predictability 36
   2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
   2.2 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
   2.3 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
   2.4 Evidence from the United States . . . . . . . . . . . . . . . . . . . . 41
   2.5 International evidence . . . . . . . . . . . . . . . . . . . . . . . . . . 46
Summary

The price of a financial asset equals the expected value of its discounted cash flows. Understanding asset prices is therefore closely linked to understanding the time-series behavior of discount rates and cash flows, which is the overall topic of this dissertation. Chapter 1 demonstrates the practical value of time-series models by using them to create investment strategies for the S&P 500 equity index and U.S. government bonds. Chapters 2 and 3 analyze the same assets but from a more fundamental angle by exploring the sources of price variation.

Chapter 1 investigates whether stock and bond return forecasts can be monetized. Amit Goyal and Ivo Welch wrote an influential paper in 2008, arguing that stock returns cannot be predicted. This finding provoked empirical finance scholars across the globe to develop new and better models to prove them wrong. Several studies have since found that stock returns are indeed predictable, however, primarily in bad states of the economy. See, e.g., Sam James Henkel, J. Spencer Martin, and Federico Nardari (2011). Interestingly, the opposite is true for bond markets, as bond returns are primarily predictable in expansions (see Chapter 3). It seems that return predictability is intimately linked to the business cycle. Chapter 1 designs portfolio management strategies that exploit this link. The chapter finds that accounting for switches in return predictability across states of the economy notably increases risk-adjusted returns. This gain is robust over time and prevails using real-time recession indicators. However, using inaccurate recession signals can severely hurt investment performance.

Chapter 2 is joint work with Stig V. Møller. In the 1970s, the prevailing understanding was that stock price variation comes from variation in expected dividends. Since then, a large literature has argued that, in fact, the evidence suggests the opposite, i.e., all variation in prices comes from expected return (discount rate) variation. See, for instance, John Cochrane’s 2011 presidential address to the American Finance Association. The implication is that dividend growth cannot be predicted. In recent years, new research has challenged this idea. In Chapter 2, we add to the debate by showing how dividend growth is predictable once you control for earnings. We show how the conventional approach to forecasting dividends fails because it omits earnings, and that this finding is consistent with John Lintner’s (1956) classic model of dividend behavior. Several studies have reported how dividend growth predictability seems to have disappeared after the Second World War. However, once earnings are included in the forecasting model, dividend growth predictability is strong in both pre-war and post-war data as well as internationally.
Chapter 3 is joint work with Martin M. Andreasen, Tom Engsted, and Stig V. Møller. The interest rate on a long-term government bond equals the average future short rate plus a risk premium (expected excess return). In an important paper, John Campbell and Robert Shiller (1991) find that variation in this risk premium is a key component of variation in the long interest rate. Standard bond models often find a close link between risk premia and the business cycle, and they generally predict risk premia to increase throughout recessions as the spread between long and short interest rates increases. We challenge this conventional wisdom. The standard models perform poorly in recessions as risk premia tend to fall during these periods, along with short interest rates, and thus move in the opposite direction of the slope of the yield curve. We show how a small tweak to canonical models of the term structure can reproduce this result. Letting the market price of risk in the Gaussian affine term structure model switch across the business cycle gives excess return forecasts with the correct sign during recessions.

Danish Summary


References


MARKET TIMING IN RECESSIONS AND EXPANSIONS

Magnus Sander
Aarhus University and CREATEs

Abstract

This paper analyzes the economic value of linking return predictability to the business cycle. Recent studies show that stock returns are predictable in recessions while bond returns are predictable in expansions. I examine whether these findings can be exploited in real-time trading by letting the coefficients of popular return regressions switch across states of the economy. These switching models provide meaningful economic gains relative to their constant coefficient versions, and the gains are robust over time. However, choosing a good recession signal is important as inaccurate business cycle turning points corrupt the switching extensions.
1.1 Introduction

This paper measures the economic value of conditioning stock and bond return forecasts on recession dummies. Motivated by recent studies, I restrict stock returns to only be predictable in recessions and bond returns to only be predictable in expansions. These restrictions generate notable increases in risk-adjusted performance in a standard setup with power utility and a simple univariate regression model. I also let regression coefficients switch across recessions and expansions without restricting them. This extension performs particularly well when combined with forecast averaging or Bayesian shrinkage. The economic gains are robust over time and prevail using real-time recession signals. However, high accuracy of business cycle turning points is essential and seemingly small tweaks to the classification rule can have large economic implications.

Goyal and Welch (2008) forcefully argue that simple regression models cannot forecast stock returns out of sample. Thornton and Valente (2012) reach similar conclusions for bonds. However, these studies do not condition on the business cycle. Henkel et al. (2011) let coefficients be regime-dependent and document that stock returns are predictable but only in economic recessions.¹ Andreasen et al. (2016) find the mirror image for bond returns as they are only forecastable in expansions. Thus, return predictability seems to be highly dependent on the business cycle and, moreover, it is asymmetric across stocks and bonds. This motivates why I extend standard regression models to take macroeconomic conditions into account.

Several other papers allow predictability to change over time. Timmermann (2008) argues that profit seeking traders cause predictability to only be present in pockets of time and proposes an adaptive forecast combination approach. Pesaran and Timmermann (2002) suggest a two-step procedure to forecasting in the presence of structural breaks. Dangl and Halling (2012) find that time-varying parameter models improve stock market timing over models with constant coefficients. In contrast to these studies, I specifically impose that predictability is linked to the business cycle.

Guidolin and Timmermann (2007) and Guidolin and Hyde (2012) show that Markov switching models have the potential to capture state-dependence in returns. Markov switching models let the return distribution depend on unobservable realizations of a Markov chain. The inferred states may or may not be related to the business cycle. Instead, I use observable recession dummies. This choice is natural since most of the studies that identify differences in predictability across states of the economy have

¹See also Rapach et al. (2010), Neely et al. (2014), Dangl and Halling (2012), and Rapach and Zhou (2013). Further, Kacperczyk et al. (2014) examine the time-varying skill of mutual fund managers and conclude that they are only good at market timing in contractions.
done so using recession dummies (see citations above). Further, I can estimate the
models using simple linear regressions rather than relying on potentially unstable
numerical estimation methods. I do also show results for Markov switching models
but find that, in contrast to the dummy switching strategies, they do not improve
on constant coefficient models for real-time trading.

A large part of the literature on real-time asset allocation investigates a single risky
asset (either stocks or bonds) and only uses an investment horizon of one period. I
allow for investments in both stocks and bonds and also report results for a longer
investment horizon. Conditioning predictability on the business cycle appears less
important for long-horizon investors but this result is not due to intertemporal
hedge demands. Specifically, strategic and myopic long-horizon investors achieve
similar risk-adjusted performance indicating that hedge demands do not matter
out of sample. This finding is an extension of the work by Diris et al. (2015) who
report similar results without taking regime-switching into account.

The rest of the paper is organized as follows. Section 1.2 presents data on returns,
the main predictors, and recession indicators. Section 1.3 motivates the switching
strategies. Section 1.4 implements the switching strategies in a cross-asset setup.
Section 1.5 shows results for a long-horizon investor and provides details on the
calculation of portfolio weights. Section 1.6 implements Markov switching models.
Section 1.7 concludes.

1.2 Data

I use S&P 500 stock returns and long-term U.S. government bond returns. The
bond returns are from Ibbotson’s Stocks, Bonds, Bills and Inflation Yearbook. I
compute excess returns on stocks and bonds by subtracting the risk-free return,
which is the lagged three-month T-bill rate. All returns are in logs. The main
predictors are the log dividend-price ratio of the S&P 500 for stocks and the term
spread for bonds. The term spread is computed as the difference between the yield
on Ibbotson’s long-term bond and the three-month T-bill rate. The dividend-price
ratio and the term spread are among the most popular predictors in the literature,
see, e.g., Campbell and Shiller (1988b) and Fama and French (1989). All returns
and predictors are from the updated Goyal and Welch (2008) data set. Unless
otherwise specified, all data are collected on a monthly frequency from 1926:1 to

---

3See Campbell and Thompson (2008), Rapach et al. (2010), Neely et al. (2014), and Dangl
and Sarno et al. (2016) for the case of bonds.

3Using 10-year U.S. government bonds from Global Financial Data gives qualitatively similar
results.

4I thank Amit Goyal for making it available on his website.
To identify recessions, I consider four different dummy variables. The first uses the recession dates from the National Bureau of Economic Research (NBER) as they are the standard choice in the literature. The second is based on the Aruoba-Diebold-Scotti Business Conditions Index, see Aruoba et al. (2009). I classify recessions as observations below the threshold value -0.80 from Berge and Jordà (2011). The next two dummies are based on the Purchasing Managers’ Index. Christiansen et al. (2014) show that this index is a strong predictor of U.S. recessions. The first dummy uses a threshold value of 44.5, which is also from Berge and Jordà (2011), and the second dummy uses a threshold value of 50 from the Institute for Supply Management.\footnote{www.instituteforsupplymanagement.org.} I classify months in which the PMI is below these thresholds as recessions. I collect NBER dates and the Purchasing Managers’ Index from St. Louis Fed’s FRED database and the Aruoba-Diebold-Scotti Business Conditions Index from the Philadelphia Fed’s website.\footnote{The NBER dummy is available from 1926:1, the Aruoba-Diebold-Scotti Business Conditions Index is available from 1960:3, and the Purchasing Managers’ Index is available from 1948:1. I use the latest vintage of all data as real-time vintages are not available for my sample period. Berge and Jordà (2011) argue that data revisions should affect indices much less than single series, and Chauvet and Piger (2008) find that business cycle turning points are quite robust to data revisions.}

1.3 Return predictability across the business cycle

I create real-time forecasts of stock and bond returns from the following regression

\[
r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1},
\]

where \( r_{t+1} \) is the excess log return on a risky asset from time \( t \) to \( t + 1 \), and \( x_t \) is the associated predictor observed at time \( t \). I use data from 1926:1 to 1960:12 to estimate (1) using OLS and create the first return forecast for 1961:1. My benchmark is to impose no predictability, \( \beta = 0 \), and compute forecasts from

\[
r_{t+1} = \alpha_0 + \varepsilon_{0, t+1}.
\]

I repeat this exercise every month until 2013:12 using an expanding estimation window. The choice of evaluation period ensures that all recession dummies are available for the first forecast.

To measure forecasting performance from a mean squared error perspective, I com-
pute the out-of-sample $R^2$:

$$R^2_{\text{oos}} = 1 - \frac{\sum_{t=T_0}^{T-1} \hat{\varepsilon}_{t+1}^2}{\sum_{t=T_0}^{T-1} \hat{\varepsilon}_{0,t+1}^2},$$  \hspace{1cm} (3)$$

where $\hat{\varepsilon}_{t+1}$ and $\hat{\varepsilon}_{0,t+1}$ are the recursive residuals from the candidate model and the benchmark, respectively, and $T_0$ is the time of the first forecast.\(^7\) I evaluate $H_0$: $R^2_{\text{oos}} \leq 0$ against $H_A$: $R^2_{\text{oos}} > 0$ using the Clark and West (2007) test for equal predictive power which takes into account that the benchmark is nested in the candidate model.

To also assess forecasting performance from an economic perspective, I follow Pettenuzzo et al. (2014) and calculate the Certainty Equivalent Return (CER) for a power utility investor. The utility function is

$$U_{t+1} = W_{t+1}^{1-\gamma} / (1 - \gamma),$$  \hspace{1cm} (4)$$

where

$$W_{t+1} = \omega_t \exp (r_{t+1} + r_{f,t+1}) + (1 - \omega_t) \exp (r_{f,t+1}),$$

and $\omega_t$ is the optimal risky asset weight and $r_{f,t+1}$ is the log risk-free return. CER is the fixed fee that equates expected utility from using the no-predictability benchmark (2) with expected utility from using the regression model (1):

$$\text{CER} = \left[ \frac{\sum_{t=T_0}^{T-1} U_{t+1}}{\sum_{t=T_0}^{T-1} U_{0,t+1}} \right]^{1/(1-\gamma)} - 1,$$  \hspace{1cm} (5)$$

where $\gamma$ is the coefficient of relative risk aversion and $U_{t+1}$ and $U_{0,t+1}$ are realized utilities from the candidate model and the benchmark, respectively.\(^8\) The coefficient of relative risk aversion $\gamma$ is 5 and the investment horizon is one month. The investor is allowed to vary his weight in the single risky asset between -50% and 150% with the rest placed in the risk-free asset. I compute the optimal real-time weights using the across-path regression methodology of Brandt et al. (2005). For details see Section 1.5. To evaluate $H_0$: CER $\leq$ 0 against $H_A$: CER $>$ 0, I follow Diris et al. (2015) and use a Diebold and Mariano (1995) test for the difference in average utility.

Table 1 shows forecasting performance of the univariate regression versus the no-

---

\(^7\)I compute the recession $R^2_{\text{oos}}$ as $R^2_{\text{oos,rec}} = 1 - \left( \sum_{t=T_0}^{T-1} \hat{\varepsilon}_{t+1}^2 \times I_t \right) / \left( \sum_{t=T_0}^{T-1} \hat{\varepsilon}_{0,t+1}^2 \times I_t \right)$, where $I_t$ is the recession dummy. The expansion values are computed similarly.

\(^8\)To compute the recession CERs, I use $\text{CER}_{\text{rec}} = \left[ \left( \sum_{t=T_0}^{T-1} U_{t+1} \times I_t \right) / \left( \sum_{t=T_0}^{T-1} U_{0,t+1} \times I_t \right) \right]^{1/(1-\gamma)} - 1$. Expansion values are computed similarly.
predictability benchmark in NBER recessions and expansions. Panel A presents results for stock returns using the log dividend-price ratio as a predictor and panel B presents results for bond returns using the term spread as a predictor. P-values are in parentheses. We see evidence that stock returns are only predictable in recessions, while bond returns are only predictable in expansions. This is true both when using statistical and economic evaluation criteria. In recessions, timing the stock market beats the no-predictability benchmark reflected by a positive $R^2_{oos}$ and CER, whereas timing the bond market produces a negative $R^2_{oos}$ and CER. In expansions, the opposite is true. These findings echo those of Henkel et al. (2011) for stocks and Andreasen et al. (2016) for bonds.

Figure 1 provides more intuition. I modify (1) by conditioning the right-hand side on the NBER dummy $I_t$:

$$r_{t+1} = \left( \alpha^{REC} + \beta^{REC} x_t \right) I_t + \left( \alpha^{EXP} + \beta^{EXP} x_t \right) (1 - I_t) + \tilde{\epsilon}_{t+1}. \quad (6)$$

I plot recursive OLS estimates of (6) and (1) showing $\beta^{REC}$ in recessions (grey bars) and $\beta^{EXP}$ in expansions. For stocks, the coefficient jumps up in recessions reflecting stronger predictability of returns in bad states of the economy. The switch in stock coefficients is consistent with countercyclical equity risk premia, which is a feature of several established asset pricing models such as Campbell and Cochrane (1999) and Bansal and Yaron (2004). However, generating the strong nonlinearity shown in Figure 1 could pose a challenge to these models. Gargano (2013) investigates the issue and indeed finds that adjustments are needed and proposes a new long-run risk model which has the ability to match the time-varying return predictability in the data.

For bonds, the coefficients in Figure 1 even switch sign being positive in expansions but negative in recessions. Andreasen et al. (2016) show how this pattern is consistent with bond risk premia decreasing during recessions and suggest that it could be fueled by accommodating monetary policy in these periods. The dramatic switches in bond and stock return coefficients we see in Figure 1 explain why the standard forecasting models without switching tend to be unreliable and motivate why I propose new models which allow for asymmetric predictability across the business cycle.

### 1.4 Gains from switching predictability

Based on the distinctive patterns in return predictability documented in Section 1.3, I propose two adjustments of the standard regression model. Motivated by
Table 1, the restricted coefficient strategy (REST) uses the standard model in (1) but shuts down market timing when it does not work. For stocks, the strategy is

\[ r_{t+1} = \begin{cases} \alpha + \beta x_t + \varepsilon_{t+1}, & \text{if } I_t = 1 \\ \alpha_0 + \varepsilon_{0,t+1}, & \text{if } I_t = 0, \end{cases} \]  

where \( I_t \) is a recession dummy. For bonds, market timing is only activated in expansions. REST exploits how predictability is only there in pockets of time and disappears for stocks in expansions and bonds in recessions.

Motivated by Figure 1, the unrestricted coefficient strategy (UNREST) is more ambitious by allowing for market timing in both states of the economy. UNREST lets the coefficients switch freely using (6):

\[ r_{t+1} = \begin{cases} \alpha^{REC} + \beta^{REC} x_t + \varepsilon_{t+1}, & \text{if } I_t = 1 \\ \alpha^{EXP} + \beta^{EXP} x_t + \varepsilon_{t+1}, & \text{if } I_t = 0. \end{cases} \]  

Both strategies require a recession signal, \( I_t \). The NBER classification is the most popular choice, but it is only available with a significant publication lag. Ng (2012) reports an average delay of nine months for the period 1980 to 2008. However, Ng also argues that due to the availability of real-time business cycle indicators, it is not very realistic that people first know that they are in a recession nine months ex-post. I follow Ng by assuming a three month publication lag for the NBER dates, see also Kauppi (2008) and Christiansen et al. (2014). For the current state of the economy (the most recent three months), I consider the Aruoba-Diebold-Scotti index (ADS) and the Purchasing Managers’ Index (PMI). For details on recession thresholds, see Section 1.2. For comparison, I also show results assuming no publication lag for the NBER dummy.

Table 2 presents CER values for a power utility investor comparing the REST and UNREST strategies with the no-predictability benchmark. I also show results for the constant coefficient model in (1). To get an aggregate measure of the economic value of predictability, I consider a joint setup with both stocks, bonds, and the risk-free asset. Wealth is therefore now calculated as

\[ W_{t+1} = \left( \omega_t^t \exp \left( r_{t+1} + r^{f_{t+1}} t_2 \right) + \left( 1 - \omega_t^t t_2 \right) \exp(r f_{t+1}) \right), \]

where \( r_t = (r_{t,s}, r_{t,b})' \) is the vector of excess log returns on stocks and bonds, and \( \omega_t = (\omega_{t,s}, \omega_{t,b})' \) is the vector of portfolio weights.

Panel A presents results with weights on the two risky assets restricted to the interval \([0,1]\), and panel B presents results for the interval \([-0.5,1.5]\). The weight
restrictions apply to each asset individually and to the sum of weights. The weights are computed using the across-path regression methodology of Brandt et al. (2005). All models are estimated recursively to simulate a real-time trading environment. In parentheses are Diebold and Mariano p-values. The coefficient of relative risk aversion is 5, but setting it to other reasonable values such as, e.g., 2 or 10 shows a similar picture.

The CER figures in Table 2 indicate that a power utility investor would be willing to pay a fixed annual fee of 0.6% to market time using the standard regression model rather than assume zero predictability. The utility gain is statistically insignificant with a p-value of around 30%. However, the REST strategy using NBER dates increases the fee to 1.2% without short selling and 1.9% with short selling and lowers the p-value to 4% in both cases. Further, the performance improvement is not limited to the NBER turning points. Especially ADS and PMI seem to be good real-time indicators of recessions.

The UNREST strategy also provides economic value, however, the utility gains are not statistically significant using this simple univariate specification. UNREST introduces extra parameter uncertainty by partitioning the observations according to $I_t$ when estimating the model. Further, the choice of recession indicator seems quite important. The strategy even loses versus the no-predictability benchmark for the PMI(50) dummy. Figure 2 provides an explanation. It shows the time-series of the indicators along with dummy thresholds and NBER recession dates. The plot reveals that a cutoff value of 50 for PMI identifies recessions too often. Table 3 quantifies the effect. For 16% of the months in the evaluation period, PMI(50) signals a recession while, at the same time, the NBER committee signals an expansion. Imprecise turning points spoil business cycle strategies such as REST and UNREST.

The two most robust recession indicators are PMI and ADS with cutoff points from Berge and Jordà (2011). ADS is a daily index based on a dynamic factor model of economic variables and is updated continuously as new data are released, which is at least once a week. It has a mean of zero. PMI is released on the first business day of the month and is based on a survey of the manufacturing sector. It is an index from 0 to 100 with values below 50 indicating a recession in the manufacturing economy. In principle, both series are available in real time. However, I emphasize that the recession thresholds of -0.80 for ADS and 44.5 for PMI from Berge and Jordà (2011) are based on full-period analysis. Further, the PMI results presented above show how changing the threshold slightly can have a non-negligible effect on performance. I therefore estimate real-time thresholds for my evaluation period using an expanding window of PMI data going back to the first observation in 1948:1. Each period, I simply recalculate the optimal threshold following Berge
and Jordà (2011). Applying these thresholds with the univariate REST strategy gives CERs of 1.22% without short selling and 1.91% with short selling. These results are very similar to the ones for PMI(44.5) in Table 2.

1.4.1 Portfolio weights and robustness over time

Figure 3 shows stock and bond weights over time for the REST strategy (using NBER dates), the constant coefficient model, and the no-predictability benchmark. Weights for the UNREST strategy are in the Appendix. As expected, the REST weights bounce between the constant coefficient weights and the no-predictability weights depending on the state of the economy. One interesting period in time is the bull-market of the 1990s during which REST places a higher weight on stocks than the constant coefficient model. This is because the low level of dividend-price ratios after 1990 pulls down the stock weight in the standard regression model through a low forecast of the equity premium, see also Campbell and Thompson (2008). Meanwhile, since the U.S. economy is expanding for most of this period, REST adheres to the no-predictability benchmark. The no-predictability forecasts are detached from the dividend-price ratio and therefore more optimistic about equity returns, which pulls up the weight on stocks.

Overall, the weights of the two market timing strategies display more variability than the no-predictability benchmark. The impact of transaction costs on relative performance could therefore be a concern. However, the Appendix shows how transaction costs would need to be unrealistically high given the portfolio turnover reflected in Figure 3 to eliminate the gains associated with market timing.

Figure 4 sheds more light on the robustness of the REST strategy over time. In the left side of the plot, I show the cumulative return and CER (both in logs) computed relative to the no-predictability benchmark for the case of short selling. The cumulative CER measures the compounded risk-adjusted gain from using REST rather than the benchmark. The improved performance from (occasionally) timing the market with REST versus assuming zero predictability is clearly not confined to any period. Both the cumulative CER and the cumulative return gain are increasing rather steadily over time.

In the right side of Figure 4, I instead compute CERs and returns relative to the constant coefficient model. The REST model deactivates market timing when it does not work. The figure shows how ignoring the low dividend-price ratio in the 1990s leads to serious performance gains during this period. Still, the performance improvement relative to the standard model is not explained by any single period. However, the choice of recession signal is important. For PMI, simply going from the optimal cutoff point of 44.5 from Berge and Jordà (2011) to the widely used
cutoff point 50 makes the gains much less robust over time.

1.4.2 More predictors

So far, I have only presented results using univariate regressions. One worry is that switching only matters if you start out with a poor forecasting model. To address this worry, I expand the information set to \( i = 1, \ldots, N \) predictors. The issue with using many predictors is the risk of in-sample overfitting. I deal with this issue using three popular approaches from the literature on return forecasting. For each approach I implement multivariate versions of REST and UNREST as well as a standard constant coefficient model.

The first is a shrinkage vector autoregression (VAR), see, e.g., Diris et al. (2015). Consider the multivariate regression

\[
r_{t+1} = \alpha + \beta_1 x_{1,t} + \ldots + \beta_N x_{N,t} + \varepsilon_{t+1}.
\]

(9)

For each OLS estimate \( \hat{\beta}_i \), computed using a sample of \( T \) observations I follow Connor (1997) and calculate Bayesian coefficients using the shrinkage formula

\[
\beta^s_i = \frac{T}{T + (1/\rho_i)} \hat{\beta}_i,
\]

(10)

where \( \rho_i = E \left( \frac{R^2_i}{1-R^2} \right) \) and \( R^2_i \) is the marginal decrease in the full fit of the model, \( R^2 \), from dropping variable \( i \). I adjust the intercept accordingly

\[
\alpha^s_i = \bar{r}_{t+1} - (\beta^s_1 \bar{x}_{1,t} + \ldots + \beta^s_N \bar{x}_{N,t}),
\]

(11)

where bars denote sample averages. I take a simple approach and set \( 1/\rho_i = 1200 \) for \( i = 1, \ldots, N \) corresponding to a monthly \( E \left( R^2_i \right) \) of roughly 0.1% for each variable.

I also implement the mean forecast combination (FC):

\[
r_{t+1} = \frac{1}{N} \sum_{i=1}^{N} (\alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}),
\]

(12)

where \( (\alpha_i, \beta_i) \) are univariate OLS estimates for predictor \( i \). See, for instance, Rapach et al. (2010).

Finally, I use a diffusion index (DI):

\[
r_{t+1} = \alpha + \beta f_t + \varepsilon_{t+1},
\]

(13)
where $f_t$ is the first common factor from principal component analysis of the $N$ predictors in the sample of $T$ observations. See Ludvigson and Ng (2007, 2009).

The first predictors are the dividend-price ratio, the term spread, and the three return series. In addition, I consider the following variables: the yield on the long Ibbotson bond, the credit spread (Keim and Stambaugh (1986)), the log earnings-price ratio (Campbell and Shiller (1988a)), the book-to-market ratio (Pontiff and Schall (1998)), and lagged inflation (Fama and Schwert (1977)). All predictors are from the Goyal and Welch (2008) data set. Rapach and Zhou (2013) provide evidence of switching return predictability for several of these predictors for the case of stocks.

Table 4 shows results for the REST and UNREST strategies. They are both straightforward extensions of the univariate case. REST only activates the given forecasting model in recessions for stocks and in expansions for bonds. It uses the no-predictability benchmark for stocks in expansions and bonds in recessions. UNREST interacts predictors and intercepts with recession and expansion dummies. All models are estimated recursively. I also include results for the usual constant coefficient versions of the models.

The UNREST strategy seems to interact particularly well with the shrinkage and combination models, which are both designed to reduce forecast variability. The REST strategy provides good results for the diffusion index. Looking at forecast combinations, the performance fee without short selling increases from roughly 1% to 1.6% using the UNREST model with NBER dates rather than the standard model. With short selling, the fee increases from around 1.1% to 2%. While meaningful CERs are also realized using the most precise turning points from ADS and PMI, we again see how using popular but less accurate recession signals can turn the gains into losses. As an example, forecast combinations with the UNREST strategy and short selling provides a risk-adjusted return of roughly 2.1% using ADS but -0.7% using PMI(50). Finally, comparing Table 2 and 4 even the best performing multivariate strategies do not seem to improve on the best univariate strategies which suggests that the more parsimonious choice is sufficient to capture economic gains from switching predictability.

---

9 The credit spread is defined as the yield on BAA rated corporate bonds minus the T-bill rate.
10 I exclude the long yield from the VAR model to avoid multicollinearity problems due to also having the term spread and the risk-free return among the predictors.
1.5 Longer horizon

So far, I have only presented results for a one-period investment horizon. This choice is natural since recessions are infrequent and short-lived. NBER dated recessions account for approximately 13% of the evaluation period. On the other hand, a long-horizon investor who rebalances his portfolio every month could still benefit from switching predictability.

Let $y_t$ be a vector of returns and state variables. The optimization problem of a $K$-period investor is

$$V(K, y_t, W_t) = \max_{\{\omega_t, \ldots, \omega_{t+K-1}\}} E_t [U(W_{t+K})],$$

(14)

where

$$W_{t+K} = W_t \prod_{j=t}^{t+K-1} (\omega_j' \exp(r_{j+1} + rf_{j+1} + 1) + (1 - \omega_j' t_j) \exp(rf_{j+1})),$$

(15)

and $U(W_{t+K}) = W_t^{1-\gamma} (1 - \gamma)$. I standardize initial wealth, $W_t = 1$. To compute optimal weights, I follow Diris et al. (2015) by using a refined version of the across-path regression methodology developed by Brandt et al. (2005).\(^{11}\) The steps are as follows:

1. Generate $S$ sample paths with length $K$ of returns and predictors using a forecasting model:

$$y_{t+1} = g(y_t, I_t, \eta_{t+1}),$$

(16)

where $\eta_{t+1}$ is a vector of innovations. I simulate 50,000 paths of excess log returns using either a constant coefficient model, a REST model, or an UNREST model. For the state variables (including the risk-free return), I use a vector autoregression with one lag.\(^{12}\) Further, instead of imposing a distribution on the innovations, I use a standard residual bootstrap resampling the entire vector of residuals $\hat{\eta}_{t+1}$ to take cross-correlation into account.\(^{13}\) Finally, I simulate the recession dummy $I_t$ from a two-state Markov chain based on the Maximum Likelihood estimates of transition

\(^{11}\)Diris et al. (2015) implement refinements from Binsbergen and Brandt (2007) as well as Koijen et al. (2010).

\(^{12}\)The only restriction I impose on the vector autoregression is to exclude the long yield on the right-hand side to avoid problems with singularity due to also having the term spread and the risk-free rate in the system. Also, for the UNREST strategies I interact the right-hand side of the vector autoregression with recession and expansion dummies to be consistent with the return dynamics.

\(^{13}\)Letting the error distribution switch across states of the economy by drawing residuals conditional on the recession dummy does not affect my conclusions.
2. Construct a two-dimensional grid of stock and bond weights while taking upper and lower bounds on the weights into account. Using a very coarse grid does not affect accuracy, so I choose increments of 25% in both dimensions. The residual is invested in the risk-free asset.

Now, recursively determine the optimal asset allocation for each simulated path. Start with the final portfolio weights in period \( t + K - 1 \) and iterate back to period \( t \) by repeating steps 3-5:

3. For a set of portfolio weights on the grid (see step 2), calculate realized utilities from realized end-of-period wealth using (15) across all \( S \) simulated paths. Except for the final period, the computation will involve path-specific weights from previous iterations.

4. Run an across-path simple regression of the \( S \) realized utilities on a constant and lagged state variables and returns. We now have an expression for conditional expected utility:

\[
E_t \left[ U \left( W_{t+K} \right) \right] = \theta_0 \left( \omega_{t,s}; \omega_{t,b} \right) + \theta_1 \left( \omega_{t,s}; \omega_{t,b} \right)' y_t. \tag{17}
\]

For \( K = 1 \) this regression only involves the constant.

5. The coefficients in (17) turn out to be roughly quadratic in the grid of weights:

\[
\begin{align*}
\theta_0 \left( \omega_{t,s}; \omega_{t,b} \right) & \approx \gamma_{00} + \gamma_{01} \omega_{t,s} + \gamma_{02} \omega_{t,b} + \gamma_{03} \omega_{t,s}^2 + \gamma_{04} \omega_{t,b}^2 + \gamma_{05} \omega_{t,s} \omega_{t,b} \\
\theta_1 \left( \omega_{t,s}; \omega_{t,b} \right) & \approx \gamma_{10} + \gamma_{11} \omega_{t,s} + \gamma_{12} \omega_{t,b} + \gamma_{13} \omega_{t,s}^2 + \gamma_{14} \omega_{t,b}^2 + \gamma_{15} \omega_{t,s} \omega_{t,b}.
\end{align*}
\tag{18}
\]

The parameters in (18) are estimated using simple regressions across the grid of weights and result in \( R^2 \)'s close to one. By plugging the approximation in (18) into (17), conditional expected utility becomes a function of \( y_t \) and portfolio weights. Along each path, we observe \( y_t \), and so the only unknowns are the weights. Thus, finding the optimal portfolio weights amounts to solving a simple quadratic optimization problem (given short-selling constraints). For \( K = 1 \) there is no dependence on \( y_t \) since (17) corresponds to a simple average across paths, and we therefore only have to run an optimizer once.

\[\text{14} \text{ The estimates of the diagonal elements in the transition probability matrix are } \tilde{p}_{ii} = \frac{n_{ii}}{\sum_{j=1}^{n_{ij}}} \text{ for } i = 1, 2, \text{ where } n_{ij} \text{ is the number of transitions from state } i \text{ to state } j \text{ in the sample. The off-diagonal elements follow directly given that each row must sum to one.}\]

\[\text{15} \text{ For the univariate model, I only include the dividend-price ratio and the term spread here. Further, as I assume that asset dynamics depend on the business cycle, I interact the constant as well as the state variables with business cycle dummies for the periods where less than 90% of the paths are in the same state.}\]
I compute CER for $K$-month investors that start one month apart. The first investor uses data up to period $T_0$ to estimate his chosen model and commit to his initial portfolio weights $\omega_{T_0}$. In the next period, $T_0 + 1$, he reestimates his model and chooses new portfolio weights given that his investment horizon is now $K - 1$ months. This continues up until the final portfolio choice, $\omega_{T_0 + K - 1}$. His string of chosen portfolio weights results in one terminal wealth value $W_{T_0 + K}$ and an associated realized utility $U(W_{T_0 + K})$. The next investor starts in period $T_0 + 1$ and performs similar choices ending up with his realized utility, $U(W_{T_0 + K + 1})$. The CER versus the no-predictability benchmark is computed as

$$CER_K = \left[ \frac{\sum_{t=T_0}^{T-K} U_{t+K}}{\sum_{t=T_0}^{T-K} U_{0,t+K}} \right]^{1/(1-\gamma)} - 1. \quad (19)$$

Table 5 presents results without short selling.\(^{16}\) In the interest of brevity, I focus on the univariate model and the diffusion index. Panel A shows CERs for $K = 24$ with $p$-values are from a Diebold and Mariano (1995) test for the difference in average utility using Newey-West standard errors with $K - 1$ lags to take the overlap of subsequent $K$-period investors into account. This choice of investment horizon reflects a tradeoff between being able to mimic a strategic investor and the vast increase in computational cost as well as the reduction in non-overlapping observations. The annualized CER levels are generally higher than for the $K = 1$ case (see Table 2 and Table 4). Although there is still an improvement for the REST and UNREST strategies relative to the constant coefficient models, the impact of allowing for switching is now smaller, especially for the univariate model.

The investor in panel A solves the problem in (14) with $K = 24$ and therefore takes into account how returns and state variables are expected to fare in the distant future through the hedge component of his portfolio. Panel B shows results for a repeated myopic investor. The myopic investor only looks one period ahead when choosing his optimal weights and thus ignores that he will be investing 23 periods after that. However, the repeated myopic investor attains CERs that are very similar to the long-horizon investor.\(^{17}\) Consequently, the difference between the short-horizon performance fees in Table 2 and 4 and the long-horizon fees in Table 5 are not due to the hedge component of the portfolio. Rather, they are due to evaluating (utility of) wealth aggregated over 1 period versus 24 periods. Intuitively, since the average recession lasts around 12 months, differences between recessions and expansions are less important for a long horizon investor than for a short horizon investor.

---

\(^{16}\)I find that the approximation in (18) has reduced accuracy when allowing for short selling at long horizons.

\(^{17}\)Diris et al. (2015) reach similar conclusions for a setup without regime switching.
1.6 Markov switching

The Markov switching vector autoregression (MS-VAR) is an alternative way of capturing regime switching. Notable examples from finance include Henkel et al. (2011), Guidolin and Timmermann (2007), and Guidolin and Hyde (2012). MS-VARs are an appealing way of modeling business cycles, because the states are treated as unobservable. In this way, MS-VARs naturally incorporate uncertainty (even ex-post) about which economic regime each period belongs to. On the other hand, the models are highly nonlinear and sometimes produce wild forecasts. The MS-VAR is

$$y_{t+1} = A_{s_t} + B_{s_t} y_t + \eta_{t+1},$$

(20)

where $\eta_{t+1} \sim N(0, \Omega_{s_t})$, and the latent state variable $s_t$ follows a two-state Markov chain with a fixed transition probability matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

(21)

where $p_{ij}$ refers to the probability of switching from state $i$ at time $t$ to state $j$ at time $t+1$. I impose restrictions on $B_{s_t}$ to reduce the parameter space. The first specification of returns is a univariate specification in which stock returns are predicted only by the dividend-price ratio and bond returns are predicted only by the term spread. This is directly related to the dummy regression in (6), only that the states are now unobserved. The two predictors only load on lagged values of themselves. The second specification I consider uses the diffusion index of the $N$ predictors in Section 1.4.2 for both stock and bond returns. The diffusion index only loads on its lagged value. I estimate the models using an expectation maximization algorithm from Krolzig (1997).

I report full-sample estimation results in the Appendix. One regime has lower persistence and much higher volatility than the other. I identify this as the recession regime. Figure 5 shows smoothed state probabilities. The recession regime has a relatively large degree of overlap with NBER dates. Table 6 reports CERs versus the no-predictability benchmark. For comparison, I also report results for the constant coefficient versions of the univariate and diffusion index models in (1) and (13). All models are estimated recursively. The MS-VAR models offer no improvement on the constant coefficient specifications. With short selling, the MS-VAR models even perform worse than the no-predictability benchmark. While the full-period estimates of the MS-VARs seemed to produce regimes similar to the NBER

---

18 See Section 1.5 for details on the calculation of optimal weights. To make results comparable I use the same residual based bootstrap. I compute residuals by weighting conditional means across states using smoothed probabilities. Further, the risk-free rate is a standard AR(1) to reduce the number of parameters in the MS-VAR.
classification, the out-of-sample performance of the MS-VARs is underwhelming. The simpler REST and UNREST strategies using dummy variables appear to be a more attractive way of capturing regime switches in predictability.\textsuperscript{19}

1.7 Conclusion

I measure the economic value of allowing return predictability to switch across the business cycle. I do this for a power utility investor choosing between stocks, bonds, and a risk-free asset. First, I let stocks only be predictable in recessions and bonds only be predictable in expansions. Second, I let the regression intercept and slope coefficients change freely across recessions and expansions. Both strategies combine recession dummies with standard return forecasting methods such as univariate regressions and diffusion indices. For the most accurate recession signals, these dummy based strategies considerably improve risk-adjusted returns when compared to constant coefficient forecasts. However, the gains depend strongly on the choice of recession indicator, and I show how even some popular turning points are not accurate enough to justify the dummy based extension of standard models.

Acknowledgements

I thank Allan Timmermann, Chang Lee, participants at the 2015 MFA conference, Stig Vinther Møller, Thomas Quistgaard Pedersen, and Tom Engsted for helpful comments. I acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

\textsuperscript{19}Guidolin and Timmermann (2007) and Guidolin and Hyde (2012) find economic gains to using Markov Switching models with four and three states, respectively.
References


A Appendix

A.1 Estimation results

Table A.1 shows OLS estimates and White standard errors for the univariate and diffusion index models in panels A and B. In panels C and D are estimates of the corresponding Markov switching models. I estimate the Markov switching models using the expectation maximization algorithm of Krolzig (1997) and compute standard errors using the outer product method.

Table A.1: Full period estimates

This table reports estimation results for forecasting models of stock and bond returns. I show results for a univariate regression (Univariate) and a diffusion index (DI) regression. Univariate uses the dividend-price ratio to predict stock returns and the term spread to predict bond returns. DI uses the first common factor from principal component analysis of ten predictors to forecast both returns. In panels A and B, the parameters switch according to the NBER dummy. The MS Univariate and MS DI models in panels C and D are Markov switching versions of the models in panels A and B. Slope is the estimated slope coefficient, Variance is the estimated error variance, and Stay Probability (Stay Prob.) is the estimated probability of staying in a given state from month to month. The sample period is 1926:01-2013:12.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Univariate</th>
<th>Panel B: DI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recessions</td>
<td>Expansions</td>
</tr>
<tr>
<td><strong>Stock Slope</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.018</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Bond Slope</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.210</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td>(0.264)</td>
<td>(0.068)</td>
</tr>
<tr>
<td><strong>Stock Variance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.704×10⁻²</td>
<td>0.138×10⁻²</td>
</tr>
<tr>
<td></td>
<td>(0.043×10⁻²)</td>
<td>(0.007×10⁻²)</td>
</tr>
<tr>
<td><strong>Bond Variance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.113×10⁻²</td>
<td>0.034×10⁻²</td>
</tr>
<tr>
<td></td>
<td>(0.007×10⁻²)</td>
<td>(0.002×10⁻²)</td>
</tr>
<tr>
<td><strong>Stay Prob.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.83</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>
A.2 UNREST weights

Figure A.1 shows stock and bond weights for the no-predictability benchmark, the univariate constant coefficient model, and the univariate UNREST model.

Figure A.1: Stock and bond weights of the UNREST strategy
The plot shows recursive stock and bonds weights for the CONST and UNREST models in Table 2. UNREST uses the NBER recession dates (grey bars). I also show weights from the no-predictability benchmark (NOPRED).
A.3 Transaction costs

Table A.2 reports break-even transaction costs measuring the unit fee that would make the investor indifferent between using a candidate forecasting model and the benchmark. Following Han (2006) and Thornton and Valente (2012), I calculate average traded value (turnover) as

\[
V = \frac{1}{T - T_0 - 1} \sum_{t = T_0 + 1}^{T-1} \left( \omega_{t,s} - \omega_{t-1,s} \frac{\exp(r_{t,s} + r_f)}{W_t} \right) \\
+ \left( \omega_{t,b} - \omega_{t-1,b} \frac{\exp(r_{t,b} + r_f)}{W_t} \right).
\]

I then measure break-even cost using

\[
\tau^{BE} = \frac{1}{T - T_0} \sum_{t = T_0}^{T-1} \left( \frac{W_{t+1} - W_{0,t+1}}{V - V_0} \right),
\]

where \(V\) and \(V_0\) are the portfolio and benchmark average traded value, respectively. Table A.2 both reports break-even costs in percent and annualized traded value in percent. The costs are all much higher than the normal range considered in the literature. As an example, Balduzzi and Lynch (1999) use 0.50% for individual stocks and 0.01% for S&P 500 futures.

Table A.2: Break-even transaction cost

This table reports unit transaction cost, \(\tau^{BE}\), that would make a power utility investor indifferent between the candidate model and the benchmark. \(12 \times V\) measures the (annualized) traded value for each strategy. Both are in percent. The numbers correspond to the no-predictability benchmark (NOPRED), CONST, REST, and UNREST models in Table 2. REST and UNREST use the NBER recession dates.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: (LB = 0), (UB = 1)</th>
<th>Panel B: (LB = -0.5), (UB = 1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau^{BE})</td>
<td>(12 \times V)</td>
<td>(\tau^{BE})</td>
</tr>
<tr>
<td>NOPRED</td>
<td>3.09</td>
<td>59</td>
</tr>
<tr>
<td>CONST</td>
<td>3.09</td>
<td>94</td>
</tr>
<tr>
<td>REST</td>
<td>4.01</td>
<td>97</td>
</tr>
<tr>
<td>UNREST</td>
<td>2.30</td>
<td>123</td>
</tr>
</tbody>
</table>

23
Table 1: Out-of-sample return predictability

The table reports out-of-sample $R^2 (R_{oos}^2)$ and annualized certainty equivalent returns (CERs) relative to the no-predictability benchmark. Both are in percent. The metrics are based on recursive estimates of $r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}$, where $r_{t+1}$ is either excess log stock returns (panel A) or excess log bond returns (panel B). The predictor $x_t$ is the log dividend-price ratio for stocks and the term spread for bonds, and the benchmark assumes no predictability. The evaluation period is 1961:1 to 2013:12, and I report results for both the full period and conditioning on NBER dated recessions and expansions. CER is computed for a one-month power utility investor with a coefficient of relative risk aversion of 5. The investor has access to a single risky asset (either stocks or bonds) and T-bills. The weight on the risky asset can vary between -50% and 150%. See Section 1.5 for details on how the weights are computed. The p-values in parentheses are from the Clark and West (2007) test for $R_{oos}^2$ and the Diebold and Mariano (1995) test for CER.

<table>
<thead>
<tr>
<th></th>
<th>Full period</th>
<th>Recessions</th>
<th>Expansions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Stocks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{oos}^2$</td>
<td>-0.20 (0.14)</td>
<td>1.36 (0.06)</td>
<td>-0.75 (0.28)</td>
</tr>
<tr>
<td>CER</td>
<td>-0.46 (0.85)</td>
<td>1.60 (0.08)</td>
<td>-0.78 (0.95)</td>
</tr>
<tr>
<td><strong>Panel B: Bonds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{oos}^2$</td>
<td>1.47 (0.01)</td>
<td>-1.70 (0.67)</td>
<td>2.74 (0.00)</td>
</tr>
<tr>
<td>CER</td>
<td>1.49 (0.13)</td>
<td>-4.41 (0.82)</td>
<td>2.45 (0.03)</td>
</tr>
</tbody>
</table>
Table 2: Economic value of switching predictability

This table reports annualized certainty equivalent returns (CERs) relative to the no-predictability benchmark for a power utility investor with access to stocks, bonds, and T-bills. The CERs are in percent. CONST uses a univariate regression: \( r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1} \), where \( r_{t+1} \) is either excess log stock returns or excess log bond returns. The predictor \( x_t \) is the log dividend-price ratio for stocks and the term spread for bonds. REST only uses the univariate regression for stocks in recessions and bonds in expansions. It shifts to the no-predictability benchmark for stocks in expansions and bonds in recessions. UNREST uses \( r_{t+1} = (\alpha^{REC} + \beta^{REC} x_t)I_t + (\alpha^{EXP} + \beta^{EXP} x_t)(1-I_t) + \epsilon_{t+1} \), where \( I_t \) is a recession dummy. To identify recessions, I use either NBER dates, the Aruoba-Diebold-Scotti index (ADS) and a cutoff of -0.8, or the Purchasing Managers’ Index (PMI) with a cutoff of 44.5 or 50. The weights on stocks and bonds are restricted to be between 0% and 100% (panel A) or -50% and 150% (panel B). The restrictions apply to both risky assets individually and to the sum of weights. See Section 1.5 for details on how the weights are computed. The models are estimated recursively and the evaluation period is 1961:01-2013:12. The investment horizon is one month and the coefficient of relative risk aversion is 5. The p-values in parentheses are from a Diebold and Mariano test of equal average utility versus the no-predictability benchmark.

<table>
<thead>
<tr>
<th>Panel A: LB = 0, UB = 1</th>
<th>Panel B: LB = -0.5, UB = 1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONST</strong></td>
<td><strong>CONST</strong></td>
</tr>
<tr>
<td>0.59 (0.26)</td>
<td>0.56 (0.34)</td>
</tr>
<tr>
<td><strong>REST</strong></td>
<td><strong>REST</strong></td>
</tr>
<tr>
<td>1.16 (0.04)</td>
<td>1.91 (0.04)</td>
</tr>
<tr>
<td>1.14 (0.04)</td>
<td>1.82 (0.05)</td>
</tr>
<tr>
<td>1.27 (0.03)</td>
<td>2.03 (0.03)</td>
</tr>
<tr>
<td>1.00 (0.06)</td>
<td>1.58 (0.05)</td>
</tr>
<tr>
<td><strong>UNREST</strong></td>
<td><strong>UNREST</strong></td>
</tr>
<tr>
<td>0.92 (0.17)</td>
<td>1.43 (0.18)</td>
</tr>
<tr>
<td>0.60 (0.26)</td>
<td>0.98 (0.26)</td>
</tr>
<tr>
<td>0.43 (0.32)</td>
<td>0.66 (0.33)</td>
</tr>
<tr>
<td>-0.77 (0.78)</td>
<td>-1.86 (0.86)</td>
</tr>
</tbody>
</table>
Table 3: Recession indicators’ overlap with NBER dates
The table reports the overlap between the NBER dummy and other recession dummies as a percentage of number of observations in the evaluation period: 1961:1-2013:12. The other dummy is either the Aruoba-Diebold-Scotti index (ADS) using a cutoff of -0.8, or the Purchasing Managers’ Index (PMI) with a cutoff of 44.5 or 50. The dummies are all lagged to be consistent with the regression models in other tables.

<table>
<thead>
<tr>
<th>Other Indicator</th>
<th>Expansion</th>
<th>Recession</th>
<th>Expansion</th>
<th>Recession</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBER</td>
<td>86.5</td>
<td>13.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADS</td>
<td>85.4</td>
<td>11.2</td>
<td>1.1</td>
<td>2.4</td>
</tr>
<tr>
<td>PMI(44.5)</td>
<td>85.1</td>
<td>8.2</td>
<td>1.4</td>
<td>5.3</td>
</tr>
<tr>
<td>PMI(50)</td>
<td>70.6</td>
<td>11.5</td>
<td>15.9</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Table 4: More predictors
This table reports annualized certainty equivalent returns (CERs) relative to the no-predictability benchmark for a power utility investor with access to stocks, bonds, and T-bills. The CERs are in percent. CONST uses either a shrinkage vector autoregression (VAR), forecast combination (FC), or a diffusion index (DI) to forecast returns. REST only uses the constant coefficient model for stocks in recessions and bonds in expansions. It shifts to the no-predictability benchmark for stocks in expansions and bonds in recessions. UNREST interacts predictors and intercepts with recession and expansion dummies. To identify recessions, I use either NBER dates, the Aruoba-Diebold-Scotti index (ADS) and a cutoff of -0.8, or the Purchasing Managers' Index (PMI) with a cutoff of 44.5 or 50. The weights on stocks and bonds are restricted to be between 0% and 100% (panel A) or -50% and 150% (panel B). The restrictions apply to both risky assets individually and to the sum of weights. See Section 1.5 for details on how the weights are computed. The models are estimated recursively and the evaluation period is 1961:01-2013:12. The investment horizon is one month and the coefficient of relative risk aversion is 5. The p-values in parentheses are from a Diebold and Mariano test of equal average utility versus the no-predictability benchmark.

<table>
<thead>
<tr>
<th>Panel A: ( LB = 0 ), ( UB = 1 )</th>
<th>VAR</th>
<th>FC</th>
<th>DI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONST</td>
<td>CONST</td>
<td>CONST</td>
</tr>
<tr>
<td>REST</td>
<td>0.54 (0.23)</td>
<td>1.02 (0.01)</td>
<td>0.72 (0.18)</td>
</tr>
<tr>
<td>UNREST</td>
<td>0.49 (0.18)</td>
<td>1.49 (0.01)</td>
<td>0.96 (0.00)</td>
</tr>
<tr>
<td>NBER</td>
<td>0.39 (0.23)</td>
<td>1.59 (0.01)</td>
<td>0.95 (0.00)</td>
</tr>
<tr>
<td>ADS</td>
<td>0.33 (0.26)</td>
<td>1.12 (0.05)</td>
<td>0.92 (0.00)</td>
</tr>
<tr>
<td>PMI(44.5)</td>
<td>0.31 (0.27)</td>
<td>0.29 (0.36)</td>
<td>0.75 (0.01)</td>
</tr>
<tr>
<td>PMI(50)</td>
<td>0.49 (0.18)</td>
<td>1.49 (0.01)</td>
<td>0.96 (0.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ( LB = -0.5 ), ( UB = 1.5 )</th>
<th>VAR</th>
<th>FC</th>
<th>DI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONST</td>
<td>CONST</td>
<td>CONST</td>
</tr>
<tr>
<td>REST</td>
<td>0.19 (0.43)</td>
<td>1.11 (0.06)</td>
<td>0.48 (0.34)</td>
</tr>
<tr>
<td>UNREST</td>
<td>1.05 (0.10)</td>
<td>2.02 (0.05)</td>
<td>1.71 (0.00)</td>
</tr>
<tr>
<td>NBER</td>
<td>0.94 (0.13)</td>
<td>2.01 (0.05)</td>
<td>1.70 (0.00)</td>
</tr>
<tr>
<td>ADS</td>
<td>0.92 (0.13)</td>
<td>1.42 (0.11)</td>
<td>1.71 (0.00)</td>
</tr>
<tr>
<td>PMI(44.5)</td>
<td>0.71 (0.17)</td>
<td>-0.12 (0.54)</td>
<td>1.18 (0.00)</td>
</tr>
<tr>
<td>PMI(50)</td>
<td>0.71 (0.20)</td>
<td>-0.87 (0.71)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Long horizon investor

This table reports annualized certainty equivalent returns (CERs) relative to the no-predictability benchmark for a power utility investor with access to stocks, bonds, and T-bills. The CERs are in percent. CONST uses either a univariate regression (Univariate) or a diffusion index (DI) to forecast returns. REST only uses the constant coefficient model for stocks in recessions and bonds in expansions. It shifts to the no-predictability benchmark for stocks in expansions and bonds in recessions. UNREST interacts predictors and intercepts with recession and expansion dummies. To identify recessions, I use either NBER dates, the Aruoba-Diebold-Scotti index (ADS) and a cutoff of -0.8, or the Purchasing Managers' Index (PMI) with a cutoff of 44.5 or 50. The weights on stocks and bonds are restricted to be between 0% and 100%. The restrictions apply to both risky assets individually and to the sum of weights. The models are estimated recursively and the evaluation period is 1961:01-2013:12. The investment horizon is 24 months and the coefficient of relative risk aversion is 5. The long-horizon investor in panel A takes hedge-demands into account whereas the myopic investor in panel B only looks one month ahead when choosing weights. The p-values in parentheses are from a Diebold and Mariano test of equal average utility versus the no-predictability benchmark. I use Newey-West standard errors with 23 lags.

<table>
<thead>
<tr>
<th>Panel A: $K = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Univariate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DI</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CONST</td>
</tr>
<tr>
<td>1.52 (0.02)</td>
</tr>
<tr>
<td>REST</td>
</tr>
<tr>
<td>1.68 (0.00)</td>
</tr>
<tr>
<td>UNREST</td>
</tr>
<tr>
<td>1.68 (0.02)</td>
</tr>
<tr>
<td>REST</td>
</tr>
<tr>
<td>1.47 (0.00)</td>
</tr>
<tr>
<td>UNREST</td>
</tr>
<tr>
<td>1.48 (0.03)</td>
</tr>
<tr>
<td>NBER</td>
</tr>
<tr>
<td>1.75 (0.00)</td>
</tr>
<tr>
<td>ADS</td>
</tr>
<tr>
<td>1.69 (0.00)</td>
</tr>
<tr>
<td>PMI(44.5)</td>
</tr>
<tr>
<td>1.61 (0.00)</td>
</tr>
<tr>
<td>PMI(50)</td>
</tr>
<tr>
<td>1.61 (0.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: <em>Myopic</em> $K = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Univariate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DI</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CONST</td>
</tr>
<tr>
<td>1.54 (0.02)</td>
</tr>
<tr>
<td>REST</td>
</tr>
<tr>
<td>1.73 (0.00)</td>
</tr>
<tr>
<td>UNREST</td>
</tr>
<tr>
<td>1.62 (0.02)</td>
</tr>
<tr>
<td>REST</td>
</tr>
<tr>
<td>1.40 (0.01)</td>
</tr>
<tr>
<td>UNREST</td>
</tr>
<tr>
<td>1.44 (0.03)</td>
</tr>
<tr>
<td>NBER</td>
</tr>
<tr>
<td>1.78 (0.00)</td>
</tr>
<tr>
<td>ADS</td>
</tr>
<tr>
<td>1.65 (0.00)</td>
</tr>
<tr>
<td>PMI(44.5)</td>
</tr>
<tr>
<td>1.65 (0.00)</td>
</tr>
<tr>
<td>PMI(50)</td>
</tr>
<tr>
<td>1.65 (0.00)</td>
</tr>
</tbody>
</table>
Table 6: Markov switching

This table reports annualized certainty equivalent returns (CERs) relative to the no-predictability benchmark for a power utility investor with access to stocks, bonds, and T-bills. The CERs are in percent. CONST uses either a univariate regression (Univariate) or a diffusion index (DI) to forecast returns. The MS-VARs are Markov switching versions of these models. The weights on stocks and bonds are restricted to be between 0% and 100% (panel A) or -50% and 150% (panel B). The restrictions apply to both risky assets individually and to the sum of weights. See Section 1.5 for details on how the weights are computed. The models are estimated recursively and the evaluation period is 1961:01-2013:12. The investment horizon is one month and the coefficient of relative risk aversion is 5. The p-values in parentheses are from a Diebold and Mariano test of equal average utility versus the no-predictability benchmark.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: $LB = 0$, $UB = 1$</th>
<th>Panel B: $LB = -0.5$, $UB = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONST</td>
<td>MS-VAR</td>
</tr>
<tr>
<td>Univariate</td>
<td>0.59 (0.26)</td>
<td>0.59 (0.26)</td>
</tr>
<tr>
<td>DI</td>
<td>0.72 (0.18)</td>
<td>0.61 (0.23)</td>
</tr>
</tbody>
</table>
Figure 1: Slope coefficients

The plot shows recursive estimates of the constant coefficient model $r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}$ and the switching coefficient model

$$r_{t+1} = (\alpha^{REC} + \beta^{REC} x_t) I_t + (\alpha^{EXP} + \beta^{EXP} x_t) (1 - I_t) + \varepsilon_{t+1}.$$  

$r_{t+1}$ is either excess log stock returns (upper part of the plot) or excess log bond returns (lower part of the plot). The predictor $x_t$ is the log dividend-price ratio for stocks and the term spread for bonds. $I_t$ is the NBER dummy. The first estimates use data from 1926:1 to 1960:12 and the windows is expanding thereafter. I show $\beta^{REC}$ in NBER recessions (grey bars) and $\beta^{EXP}$ in expansions for the switching coefficient model.
Figure 2: Recession indicators
The plot shows the Aruoba-Diebold-Scotti index (ADS) and the Purchasing Managers’ Index (PMI). NBER dated recessions are marked with grey bars.
Figure 3: Stock and bond weights

The plot shows recursive stock and bonds weights for the CONST and REST models in Table 2. REST uses the NBER recession dates (grey bars). I also show weights from the no-predictability benchmark (NOPRED). Short selling is allowed.
Figure 4: Cumulative CER and return

The plot shows cumulative log certainty equivalent return (CER) and log portfolio return of the univariate REST strategy in Table 2. I compute both measures relative to the no-predictability benchmark (NOPRED) in the left side of the plot and relative to the CONST model in the right side of the plot. I compute the monthly log CERs using an expanding window and then sum them over time. The cumulative returns are the monthly difference in realized log wealth summed over time. I show results using the Aruoba-Diebold-Scotti index (ADS) with a cutoff of -0.8 and the Purchasing Managers’ Index (PMI) using cutoff values of 44.5 and 50. Grey bars are NBER recession dates.
Figure 5: Smoothed MS-VAR recession probabilities
The plot shows smoothed recession probabilities using full-sample estimates of the MS-VARs in Section 1.6. The grey bars are NBER recessions.
DIVIDENDS, EARNINGS, AND PREDICTABILITY

Stig V. Møller  
_Aarhus University and CREATES_

Magnus Sander  
_Aarhus University and CREATES_

Abstract

We show that the dividend yield and earnings yield jointly are strong predictors of dividend growth. We motivate the joint specification with a theoretical model and show how omitting the earnings yield biases the dividend yield coefficient towards zero, explaining why the dividend yield by itself is a poor predictor of dividend growth. Our empirical results are robust in pre- and post-war U.S. data, in recessions and expansions, in international data, and when controlling for additional predictors.
2.1 Introduction

Several papers have shown that the ability of the dividend yield to predict dividend growth is weak, see Campbell and Shiller (1988), Cochrane (2008, 2011), among others. We argue that the weak predictive power of the dividend yield for future dividend growth can be explained by a missing variable problem. The background is Lintner’s (1956) dividend model from which it can be derived that the dividend yield and the earnings yield should be used jointly as explanatory variables of future dividend growth. Omitting the earnings yield from the equation causes the coefficient on the dividend yield to be biased towards zero. Together with the earnings yield, however, the dividend yield is a strong predictor of dividend growth. We show that this result is remarkably robust in both U.S. and international data.

We are not the first to show that the dividend yield predicts dividend growth when including the earnings yield in the specification. This intriguing result was first discovered by Ang and Bekaert (2007). They find that the dividend yield on its own contains only weak predictive power of dividend growth, but once they control for the earnings yield, the dividend yield is a significant predictor of dividend growth (and so is the earnings yield). Building on the work of Ang and Bekaert (2007), we show that their finding can be explained within the framework of Lintner’s (1956) dividend model. Omitting the earnings yield conceals predictability by biasing the dividend yield coefficient towards zero.

We analyze and compare the strong predictive power of the dividend yield-earnings yield ($dy - ey$) model against a number of new findings in the literature. In recent years, an increasing number of papers have challenged the view that dividend yields do not predict dividend growth. Chen (2009) and Golez and Koudijs (2014) show that dividend yields predict dividend growth in the pre-war years, but not in the post-war years. This pre-war vs. post-war effect does not show up when using the $dy - ey$ model to predict dividend growth. While the dividend yield by itself has no predictive power for dividend growth in post-war data, the $dy - ey$ model contains substantial predictive power for dividend growth in both pre- and post-war data. We find that the omitted variable bias from not including the earnings yield is less severe in the pre-war period in part due to a lower correlation between $dy$ and $ey$. In addition, as Chen et al. (2012) also show, there is less dividend smoothing in pre-war data. These two effects help explain the reversal in dividend growth predictability when using the dividend yield as the only predictor.

Engsted and Pedersen (2010) and Rangvid et al. (2014) show that dividend growth is predictable from dividend yields in countries with small market capitalizations but not in large markets such as the U.S. We examine a cross section of 14 developed countries and show that the $dy - ey$ model contains much more predictive power for
dividend growth than does univariate models with either the dividend yield or earnings yield as the only predictor. In line with the previous literature, we find that the $dy$-$ey$ model has more predictive power for dividend growth in countries with small market capitalizations compared to countries with large market capitalizations.

Recent literature has produced robust evidence that equity returns in the U.S. are substantially more predictable during economic downturns than during economic expansions, see Rapach et al. (2010) and Henkel et al. (2011). Hence, a natural question is whether the predictive power of the $dy$-$ey$ model varies with the state of the economy. Unlike return predictability, which seems to be restricted to a few periods around recessions, the $dy$-$ey$ model strongly predicts dividend growth in both expansions and recessions.

We compare the $dy$-$ey$ model against other commonly used predictive variables such as the short rate, the term spread, the default spread, and the consumption-wealth ratio ($cay$) of Lettau and Ludvigson (2001). These variables do not add much additional information about expected dividend growth beyond that contained in the $dy$-$ey$ model. However, the short rate and $cay$ both help to predict returns in agreement with the findings of Ang and Bekaert (2007) and Lettau and Ludvigson (2001). Hence, different sets of variables drive expected returns and expected dividend growth.

Overall, despite of its simplicity, the $dy$-$ey$ model contains robust predictive power for dividend growth in different subsamples, in both recessions and expansions, when controlling for additional variables, and across countries.1

The rest of the paper is structured as follows. Section 2.2 motivates why the dividend yield and earnings yield should be used jointly as predictors of dividend growth. Section 2.3 describes the data. Section 2.4 examines the U.S. evidence of dividend growth predictability, while Section 2.5 examines the international evidence. Section 2.6 discusses the omitted variable bias from not including the earnings yield. Section 2.7 concludes.

### 2.2 Motivation

This section motivates why the dividend yield and earnings yield should be used jointly as predictors of dividend growth. Consider Lintner’s (1956) model of dividend growth. 

---

1Our results from simple OLS regressions confirm recent evidence of strong dividend growth predictability based on more advanced methods. Golez (2014) extracts a forward looking measure of expected dividend growth from options and futures and shows that it predicts dividend growth, while Binsbergen and Koijen (2010) use state-space models to show that past values of dividend growth help to forecast both returns and dividend growth.
Dend payout in log form
\[
\Delta d_{t+1} = a + c \left( d^*_t - d_t \right) + u_{t+1},
\]
(1)

where \( d_t \) is the actual log dividend at time \( t \) and \( \Delta d_{t+1} = d_{t+1} - d_t \) is log dividend growth at time \( t + 1 \). We specify the log target dividend as \( d^*_t = r + e_{t+1} \), where \( r \) is the log target payout ratio and \( e_{t+1} \) is actual log earnings at time \( t + 1 \). The non-negative parameter \( c \) measures the speed of adjustment towards the target and reflects the degree of dividend smoothing. The model can be motivated by a quadratic cost function where managers are penalized for deviations of dividend growth from a normal rate as well as deviations of realized dividends from target dividends, see Garrett and Priestley (2000). We next assume that \( e_{t+1} \) is well approximated by a random walk. After rearranging, we then arrive at the following specification for dividend growth
\[
\Delta d_{t+1} = \alpha - c \left( d_t - e_t - r \right) + \nu_{t+1}.
\]
(2)

Lintner’s model can therefore be seen as a theoretical motivation for predicting dividend growth using the log payout ratio, \( d_t - e_t \). If the current payout ratio is above the target level, dividend growth is expected to fall. We also see from (2) that the level of predictability is linked to the degree of dividend smoothing. If we then add and subtract \( c \times p_t \), where \( p_t \) is the log price, and ignore the constant \( r \), we obtain the following model
\[
\Delta d_{t+1} = \alpha - c \times dp_t + c \times ep_t + \nu_{t+1},
\]
(3)

where \( dp_t = d_t - p_t \) is the dividend yield and \( ep_t = e_t - p_t \) is the earnings yield. Consider a miss-specified model that only includes the dividend yield
\[
\Delta d_{t+1} = \alpha + \beta dp_t + \varepsilon_{t+1},
\]
(4)

where \( \beta = -c \) and \( \varepsilon_{t+1} = c \times ep_t + \nu_{t+1} \). If we estimate this model using OLS, we get an omitted variable bias due to not including the earnings yield. The bias is
\[
E \left( \hat{\beta} \right) - (-c) = \gamma c,
\]
(5)

Lintner (1956) originally specified the model with all variables in levels. We follow Garrett and Priestley (2000) and Chen et al. (2012) by specifying the model in logs.

Ever since the work of Ball and Watts (1972), several empirical studies have shown that earnings are close to a random walk, although there are also empirical studies finding some deviations from the random walk, see the review in Kothari (2001).
where $\gamma$ is the slope coefficient from an auxiliary regression of $ep_t$ on $dp_t$ (and a constant). Rearranging, we get

$$E(\hat{\beta}) = -c(1 - \gamma).$$

(6)

If the dividend yield and the earnings yield have a high correlation ($\gamma$ close to 1), regressing dividend growth on the dividend yield could lead us to wrongly conclude that dividend growth is not predictable. The intuition is that $dp_t$ and $ep_t$ have the opposite sign but are positively correlated. Omitting $ep_t$ pulls the estimated coefficient of $dp_t$ towards zero. This point is not restricted to the Lintner model, but extends to other models of dividend behavior where potential omitted variables correlate with the dividend yield. The models of Marsh and Merton (1987) and Garrett and Priestley (2012) are interesting alternatives.4

2.3 Data

2.3.1 U.S. data

We use S&P 500 data to compute returns, dividend growth, dividend yields, and earnings yields. In our main regressions, we use a quarterly sample over the period 1927:1 to 2013:4. We compute returns on the S&P 500 index including dividends. To compute the excess return, we subtract a three-month T-bill rate. We derive monthly dividend payments from returns with and without dividends and compute annual dividends as the sum of dividend payments on the S&P 500 index over the past year. We compute dividend growth as the quarterly growth rate in annual dividends and the quarterly dividend yield is then given by the sum of dividends over the past year divided by the end-of-quarter price. In a similar vein, the quarterly earnings yield is defined as earnings over the past year divided by the end-of-quarter price. All variables are in logs.

We also work with annual data over the period 1871 to 2013. We have obtained both the quarterly and annual data from the updated Goyal and Welch (2008) dataset, which is available on Amit Goyal’s website.5

4The focus of the paper is on predictability of dividend growth from dividend yields and we acknowledge that the dividend behavior model of Lintner (1956) does not provide guidance on predictability of returns from dividend yields. In particular, the model cannot be used to understand the joint dynamics of expected returns and expected dividend growth.

2.3.2 International data

We also carry out an international analysis using data on the following 14 countries: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Japan, the Netherlands, Singapore, South Africa, Switzerland, and United Kingdom. For these countries, we obtain returns, dividend growth, dividend yields, and earnings yields from Datastream over the period 1973:1 to 2013:4. We denote these variables in local currency. Similar to Rangvid et al. (2014), we use the individual country data to generate global aggregate portfolios of returns and dividend growth as well as corresponding global dividend and earnings yields. We use both an equally-weighted global portfolio and a value-weighted global portfolio based on U.S. dollar market capitalizations.

2.4 Evidence from the United States

Table 1 shows results from forecasting regressions

\[ y_{t+k} = \alpha_k + \beta_{dp}^k dp_t + \beta_{ep}^k ep_t + \varepsilon_{t+k}, \]  

(7)

where \( y_{t+k} \) is the \( k \)-period ahead dividend growth rate or excess return. We use the dividend yield \( dp_t \) and the earnings yield \( ep_t \) as regressors, either separately or jointly. The data are quarterly and the sample period is from 1927:1 to 2013:4. Panels A and B report results for \( k = 1 \) and \( k = 4 \), respectively. Below OLS estimates of slope coefficients are \( t \)-statistics in parentheses based on Newey-West standard errors with \( k + 3 \) lags and in brackets are \( t \)-statistics based on standard errors from a stationary bootstrap. We resample the data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. We set the number of replications to 50,000 and compute the bootstrap standard error on \( \beta_k \) as the standard deviation of the 50,000 simulated slope coefficients.

Focusing first on univariate regressions, the table shows that the dividend yield is statistically insignificant in predicting dividend growth at the 5% level. The earnings yield is marginally significant in predicting dividend growth for \( k = 1 \) but turns insignificant for \( k = 4 \). In addition, the two variables only display weak predictive power for returns.

Turning to the bivariate regressions, we find no increase in return predictability by using the dividend yield and earnings yield jointly. However, together they contain substantial predictive power for dividend growth. Consistent with Ang and Bekaert

6 We also report results for non-overlapping annual regressions in Section 2.4.1.1.
(2007), the dividend yield coefficient is negative and the earnings yield coefficient is positive. Comparing the univariate and bivariate one-quarter ahead regressions, the coefficient on $dp_t$ goes from $-0.018$ to $-0.059$ and the coefficient on $ep_t$ goes from $0.013$ to $0.058$. In addition, both variables are now statistically significant at the 1% level. The impact on the adjusted $R^2$ is also striking. $dp_t$ on its own generates an $R^2$ of 3.4%, $ep_t$ an $R^2$ of 1.5%, but together they generate an $R^2$ of 19.5%. Similar patterns appear at the one-year ahead forecast horizon.

These empirical results can be explained within the framework of Lintner’s (1956) dividend model. From equation (3), we see that the model implies a negative dividend yield coefficient and a positive earnings yield coefficient, as we find in the data. The model in (3) restricts the dividend yield and earnings yield coefficients to be equal but of opposite sign. In the bivariate regressions, we have allowed the variables to have separate coefficients. In accordance with the implications of the Lintner model, the estimation results show that the dividend yield and earnings yield coefficients are of opposite sign and of similar magnitude. For instance, with $k = 1$, the dividend yield coefficient is $-0.059$ and the earnings yield coefficient is 0.058. Statistically, we fail to reject the null hypothesis that these two coefficients are equal in absolute value with $t$-statistics of $-0.06$ (Newey-West adjusted) and $-0.04$ (stationary bootstrap). We return to these points in Section 2.6.1 where we in more detail analyze the omitted variable bias from not including the earnings yield in the specification.

### 2.4.1 Pre-war vs. post-war results

Chen (2009) shows that returns are predictable and dividend growth unpredictable from dividend yields in post-war years, but in pre-war years the opposite pattern applies, see also Golez and Koudijs (2014). Hence, one may wonder whether the strong dividend growth predictability we have found for the 1927:1-2013:4 period is driven mainly by the pre-war years. Over the period 1950:1-2013:4, Table 2 shows that the dy-ey specification strongly predicts dividend growth, both at the one-quarter ahead horizon and the one-year ahead horizon. The earnings yield on its own also significantly predicts dividend growth, but the degree of predictive power for future dividend growth increases substantially when the dividend yield and earnings yield are used in a joint specification.

Table 3 shows results for the pre-war period from 1927:1-1949:4. Consistent with Chen (2009), the evidence of dividend growth predictability is stronger in this period. The dividend yield on its own significantly predicts dividend growth and generates an $\tilde{R}^2$ of 25.1% at the one-quarter ahead horizon. However, the $\tilde{R}^2$ increases from 25.1% to 56.1% by controlling for the earnings yield. That is, the
missing variable problem is also present in the pre-war period.

2.4.1.1 Annual frequency

To further judge the pre-war vs. post-war effect, we analyze annual data and run one-year ahead regressions. With annual data, we can extend the sample period back to 1871, as in, e.g., Chen (2009). We can also make sure that any evidence of predictability is not attributable to the seasonality of dividend payments.

The annual results in Table 4 clearly illustrate the tale-of-two-periods finding of Chen (2009). The dividend yield on its own significantly predicts dividend growth over the 1871-1949 period, but not afterwards. Conversely, it significantly predicts returns over the 1950-2013 period, but not before. In the 1871-1949 period, the dividend yield coefficient is $-0.435$ and highly significant, while in the 1950-2013 period, it drops to $-0.014$ and turns insignificant. In a similar vein, the $R^2$ drops from 47.1% to $-0.6%$. However, once we control for the earnings yield, the dividend yield becomes a highly significant predictor of dividend growth in the post-war period and so does the earnings yield. The joint specification gives an $R^2$ of 60.1% and 45.7% in the 1871-1949 and 1950-2013 periods, respectively. Hence, the large and puzzling reversal in dividend growth predictability does not show up when using the $dy$-$ey$ model.

2.4.2 Economic significance

To judge the economic significance of the above regression results, we analyze standard deviations of expected dividend growth and expected returns. Cochrane (2008) finds standard deviations of about 5% for expected returns and about 0% for expected dividend growth when using $dp_t$ as a sole predictor in annual regressions. As we show in Table 5, this roughly matches our results for the post-war period (see Panel B). However, when using $dp_t$ and $ep_t$ together, the standard deviation of expected dividend growth rises from close to 0% to 3.8% on an annual horizon, which is about 70% of the average level of dividend growth. That is, expected dividend growth varies almost as much as its level.

If we consider the full sample from 1927:1 to 2013:4, the economic significance of dividend growth predictability is even stronger (see Panel A). The standard deviation of expected dividend growth using only $dp_t$ is 3.6% at the annual horizon and it increases to 6.8% when including $ep_t$ in the regression, which implies that expected dividend growth actually varies by more than its own level. These results show that the economic effect of controlling for $ep_t$ is sizeable. However, this only holds when forecasting dividend growth as the standard deviation of expected returns does not
change much by controlling for $ep_t$.

### 2.4.3 Additional variables

Movements in expected returns that have positive correlation with movements in expected dividend growth will have offsetting effects on the dividend yield. In particular, other variables may predict both returns and dividend growth but in an offsetting way so that the dividend yield does not change, which will make the dividend yield highly persistent and unable to uncover return and dividend growth predictability, see Menzly et al. (2004), Lettau and Ludvigson (2005), Binsbergen and Koijen (2010), and Golez (2014).

To examine the importance of including additional variables, Table 6 reports results from forecasting regressions with quarterly observations

$$y_{t+1} = \alpha + \beta^{dp} dp_t + \beta^{ep} ep_t + \beta^z z_t + \epsilon_{t+1}, \quad (8)$$

where $z_t$ is either the short rate, the term spread, the default spread, or $cay$. These variables are among the most commonly used predictive variables in the literature. Comparing the results from Table 1 and 6, we see that controlling for these variables adds little additional predictive power for future dividend growth. However, especially the short rate and $cay$ help in predicting future returns, which is in line with the findings of Ang and Bekaert (2007) and Lettau and Ludvigson (2001). These results highlight that it is not the same variables that drive expected dividend growth and expected returns.

### 2.4.4 Recessions vs. expansions

In a recent study, using the dividend yield among other predictive variables, Henkel et al. (2011) show that return predictability exists in recessions only. Similarly, Rapach et al. (2010), Dangl and Halling (2012), and Rapach and Zhou (2013) also find that return predictability is primarily present in recessions. Zhu (2015) introduces regime switching into to the present value model of Binsbergen and Koijen (2010) and finds evidence of time-varying return and dividend growth predictability.

To investigate whether the predictive power of the $dy-ey$ specification varies across

---

7The short rate is proxied by the three-month T-bill rate, the term spread is defined as the long-term yield on government bonds minus the short rate, and the default spread is the difference in yields between BAA and AAA rated corporate bonds. We use same data and definitions as in Goyal and Welch (2008). The consumption-wealth ratio ($cay$) of Lettau and Ludvigson (2001, 2005) is obtained from the website of Martin Lettau.
the business cycle, we run simple OLS switching regressions of the following form

\[ y_{t+1} = \left( \alpha_R + \beta_{R}^{dp} dp_t + \beta_{R}^{ep} ep_t \right) \times I_t + \left( \alpha_E + \beta_{E}^{dp} dp_t + \beta_{E}^{ep} ep_t \right) \times (1 - I_t) + \varepsilon_{t+1}. \] (9)

To classify business cycles, we use the NBER recession indicator, which we denote by \( I_t \) in (9). The use of NBER regime dates is a natural choice given the fact that the NBER methodology is the most common definition of recessions and expansions. Investors do not know NBER recession dates in real time but this is not an issue. Our goal is not to set up an implementable investment strategy.

Table 7 shows the results of estimating (9) with quarterly observations. We classify recessions using the quarterly turning points from the NBER. Consistent with Henkel et al. (2011), the dividend yield significantly predicts returns in recessions, but not in expansions. However, this result only holds for the post-war period, as the dividend yield is insignificant in both expansions and recessions over the full period from 1927:1 to 2013:4.

Interestingly, conditioning on the state of business cycle does not matter much with respect to the predictive power of the \( dy\)-\( ey \) specification for future dividend growth. When the two variables are used together, the dividend yield and earnings yield coefficients are generally highly significant in both states of the economy. Hence, while return predictability seems concentrated to a few recession periods, dividend growth predictability from the \( dy\)-\( ey \) model holds in both recessions and expansions.

### 2.4.5 Overlapping observations

By calculating quarterly growth rates of annual dividends we create overlapping observations. Several studies document that overlapping observations combined with persistent regressors and small samples may lead to wrong conclusions when running predictive regressions, see, e.g., Valkanov (2003) and Ang and Bekaert (2007) for the case of stock returns. To account for the time-overlap in observations, we have computed standard errors from a stationary bootstrap. To further illustrate the robustness of the results obtained with the \( dy\)-\( ey \) model, we now simulate from a system without predictability but with autocorrelated error terms. Specifically, we assume an MA(4) process for dividend growth and a bivariate VAR(1) for \( dp_t \) and \( ep_t \). The system is

\[ \Delta d_{t+1} = \alpha + \varepsilon_{t+1} + \rho_1 \varepsilon_t + \rho_2 \varepsilon_{t-1} + \rho_3 \varepsilon_{t-2} + \rho_4 \varepsilon_{t-3} \] \( (10) \)

\[ x_{t+1} = \mu + \Phi x_t + \eta_{t+1}, \] \( (11) \)
where $x_t = (dp_t, ep_t)$. We estimate the system equation-by-equation using Maximum Likelihood (ML). Using these estimates, we create 50,000 bootstrap samples with $T = 348$ observations corresponding to the length of our full sample period.

To create the samples, we draw $(\hat{\varepsilon}_{t+1}, \hat{\eta}_{t+1})^\prime$ jointly from the ML residuals to preserve cross-correlation. The simulations are initiated at random draws of $x_t$, $\hat{\varepsilon}_t$, $\hat{\varepsilon}_{t-1}$, $\hat{\varepsilon}_{t-2}$, and $\hat{\varepsilon}_{t-3}$. For each simulated sample, we reestimate equation (7) and save $\hat{\beta}_{dp}^k$ (results are similar for $\hat{\beta}_{ep}^k$). The histograms of simulated coefficients for $k = 1$ and $k = 4$ are plotted in Figure 1 along with the original sample estimates for the full period 1927:1-2013:4 (see also Table 1). The simulations show that it is very unlikely to find the levels of dividend growth predictability from $dp_t$ and $ep_t$ we see in the data if dividend growth is truly unpredictable but (artificially) autocorrelated due to overlapping observations. Zero percent of the simulated coefficients are as extreme as the OLS sample estimates. In unreported results, we obtain empirical $p$-values of 1 percent or less for the post-war period as well as if we compute the empirical $p$-value based on Newey-West $t$-statistics rather the OLS coefficients.

\subsection{2.5 International evidence}

To examine whether the US pattern of strong dividend growth predictability from the $dy-ey$ model extends to other countries, we analyze data from 14 developed countries. We start out running panel regressions using all 14 countries to make a statement about the average predictive relationship across countries. We forecast one-quarter ahead and the sample runs from 1973:1 to 2013:4. Similar to the international predictability studies of Ang and Bekaert (2007) and Hjalmarsson (2010), we constrain the slope coefficients to be the same across countries but allow for heterogeneous intercepts. Following the procedure in Thompson (2011), we compute standard errors that are robust to heteroskedasticity as well as correlation in both the country and time dimension. In addition, we compute standard errors from a stationary bootstrap where the random time-series blocks are drawn commonly for all countries to preserve cross-correlation.

Table 8 shows that in international equity markets the dividend yield significantly predicts dividend growth as well as returns, although the explanatory power obtained from the return regression is relatively weak.\footnote{Other variables may help predict international equity returns such as the short interest rate, see Ang and Bekaert (2007) and Hjalmarsson (2010).} The panel regressions also reveal that the $dy-ey$ model gives much stronger evidence of dividend growth predictability than do univariate models with the dividend yield or the earnings yield as the only predictor. As with the U.S. evidence, the magnitudes of the dividend
and earnings yield coefficients increase substantially when the two variables are used jointly: the $dp_t$ coefficient goes from $-0.038$ to $-0.076$ and the $ep_t$ coefficient goes from $-0.011$ to $0.049$. These results again confirm our omitted variable interpretation.

Ang and Bekaert (2007) analyze countries with large market capitalizations (U.S., U.K., Germany, and France) and find that the dividend yield does not contain much predictive power for dividend growth. Similarly, the international studies of Engsted and Pedersen (2010) and Rangvid et al. (2014) also find evidence of weak dividend growth predictability from dividend yields in countries with large market capitalizations but show that dividend yields significantly predict dividend growth in small markets. These findings also relate to the work of Maio and Santa-Clara (2015), who examine the U.S. cross section of equities and show that there is more dividend growth predictability for small firms than for large firms.

To analyze the impact of market capitalization, Table 8 shows results for two global portfolios: an equal-weighted portfolio that gives a weight of 1/14 to each country and a value-weighted global portfolio based on market capitalizations. Unlike the panel regressions, we do not find any evidence of significant return predictability for these global portfolios, which may be attributed to lack of power due to the small sample size. In addition, the dividend yield and earnings yield on their own do not seem to contain much predictive power for dividend growth. In agreement with the results of Rangvid et al. (2014), we find that the dividend yield by itself generates a larger $R^2$ for the equal-weighted portfolio. However, there is plenty of dividend growth predictability for both portfolios when the $dy$-$ey$ model is used, albeit the degree of predictive power is more pronounced for the equal-weighted portfolio.

There is empirical evidence that financial market integration has increased significantly in recent decades, see, e.g., Pukthuanthong and Roll (2009) and Bekaert et al. (2011). Specifically, Bekaert et al. (2011) find that a large set of developed markets has been effectively integrated since 1993. To account for the changes in the level of market integration, Table 9 shows results from two sub-periods: 1973:1-1992:4 and 1993:1-2013:4. In the period were the markets in our sample were less financially integrated (before 1993), we find less support for the $dy$-$ey$ specification. In the 1973:1-1992:4 subperiod, the $dy_t$ coefficient changes from $-0.038$ in the univariate regression to $-0.055$ in the joint specification and the $ey_t$ coefficient changes sign from negative to positive, which lends partial support to the $dy$-$ey$ specification. However, $ey_t$ is not statistically significant. In the period with high market integration (after 1993), we obtain stronger evidence in favor of using $dy_t$ and $ey_t$ jointly

---

9 Ang and Bekaert (2007) show how pooling cross-country data increases the power in international return regressions.

10 Our results are not directly comparable with those of Rangvid et al. (2014): they use an unbalanced panel of 50 countries, while we use a balanced panel of 14 countries.
in forecasting regressions of dividend growth. As Table 9 shows, the magnitudes of the dy and ey coefficients are much larger in the joint specification relative to the univariate regressions. Moreover, the dividend yield by itself generates an explanatory power of 8.0%, but this number increases to 13.2% when controlling for the earnings yield.

To examine the cross-country patterns in dividend growth predictability, we next run forecasting regressions for each of the 14 individual countries (i.e., we do not constrain the slope coefficients to be the same across countries). Figure 2 summarizes the results. Consistent with the U.S. evidence, both the dividend yield coefficient and the earnings yield coefficient increase substantially in bivariate regressions using dp and ep as regressors compared to univariate regressions with either dp or ep as the only regressor. In addition, the explanatory power increases substantially by controlling for the earnings yield. This pattern is observed in the majority of countries. However, there is no obvious pattern in the degree of omitted variable bias across countries. The pattern seems complicated and does not appear to be related to market capitalization. The main conclusion we draw from the international results is that the omitted variable problem is not just restricted to the U.S. but exists in many countries.

2.6 Discussion

2.6.1 Omitted variable bias

In Section 2.2, we showed how an omitted variable bias potentially can explain the non-predictability of dividend growth from dividend yields. From the Lintner (1956) model, we derived that the expected slope estimate from regressing Δd_{t+1} on dp is given by $E(\hat{\beta}) = -c(1 - \gamma)$, where $-c$ is the slope coefficient from regressing Δd_{t+1} on (dp_{t} - ep_{t}) or equivalently on (d_{t} - e_{t}), and $\gamma$ is the slope coefficient from a regression of ep on dp (each regression includes a constant). In the following, we take a closer look at this relation.

Table 10 reports OLS estimates of $\beta, -c$, and $\gamma$. Consistent with the results in Table 1, $\hat{c}$ is much higher in absolute value compared to $\hat{\beta}$. For the full period with a quarterly frequency, the estimates are $-0.059$ and $-0.018$, respectively. Since dp and ep are highly correlated, we expect the dp coefficient to be close to zero when omitting ep. In fact, the $\hat{\gamma}$-value indicates that the dp coefficient will be reduced by 70% of its true value on average in repeated sampling.

---

11 This means that we relax the assumption of a constant payout smoothing coefficient across countries.
If we combine $\gamma$ with the OLS estimate of $c$ (which is our best guess of the true value), we can compute the expected $dp_t$ coefficient implied by the Lintner model, $E(\hat{\beta})$. Table 10 shows that $E(\hat{\beta})$ is roughly equal to the univariate sample estimate. Comparing the full-period results with the post-war results, we see that the higher level of dividend growth predictability using $dp_t$ for the full period is both due to a weaker relationship with $ep_t$ (lower $\gamma$) and less dividend smoothing (higher $c$). Chen et al. (2012) also find that dividends are much more smooth in the post-war period. Our analysis extends their work by revealing a direct link between the level of dividend smoothing and the bias in regressions of dividend growth on $dp_t$.

We also report quantiles of the distribution of $\hat{\beta}$ coefficients implied by the Lintner model. We compute them by simulating from the equation-by-equation ML estimates of the following system:

\[
\Delta d_{t+1} = \alpha - c(d_t - e_t) + \rho_1 \varepsilon_{t+1} + \rho_2 \varepsilon_{t-1} + \rho_3 \varepsilon_{t-2} + \rho_4 \varepsilon_{t-3} \\
x_{t+1} = \mu + \Phi x_t + \eta_{t+1}.
\]

We incorporate uncertainty about the true $c$ by drawing a new value for each simulation from the empirical likelihood. We map out the likelihood by reestimating the model for a grid of $c$-values going from $-2\hat{\gamma}$ to 0 with increments of 0.0005, where $\hat{\gamma}$ is the ML estimate. Cochrane (2008) uses a similar procedure to model the uncertainty about the dividend-price ratio process when simulating returns and dividend growth.

Table 10 shows that the empirical $\hat{\beta}$ value is included in the 95% confidence interval implied by the model. We also report the probability of finding significant dividend growth predictability ("power") from estimating the wrong regression using only $dp_t$ and from our main regression using $dp_t$ and $ep_t$ jointly. The probability of rejecting the null of no predictability increases from 0.49 to 0.98 using $dp_t$ and $ep_t$ together instead of only $dp_t$ for the full period ($k = 1$) and from 0.22 to 0.84 for the post-war period. The probabilities are calculated with a critical value for the Newey-West $t$-statistic of 1.96. The significant increase in power highlights the importance of using dividends and earnings jointly when predicting dividend growth.

---

12 This is because the multivariate regression slopes on $dp_t$ and $ep_t$ are roughly equal in absolute value as predicted by the model, see Table 1 and Section 2.2. The univariate and multivariate estimates are linked algebraically as $\hat{\beta} = \hat{\beta}^{dp} + \hat{\beta}^{ep}\hat{\gamma}$, where $\hat{\gamma}$ is the slope estimate from a univariate regression of $\Delta d_{t+1}$ on $dp_t$, and $\left(\hat{\beta}^{dp}, \hat{\beta}^{ep}\right)'$ are the multivariate slope estimates from a regression of $\Delta d_{t+1}$ on $(dp_t, ep_t)'$.

13 We include lagged MA terms to address the issue of overlapping observations. The results are robust towards the lag length specification.

14 We do not include the MA terms in these regressions.

15 Using a 1% critical value of 2.58 gives similar results.
growth.

### 2.6.2 Other specifications of dividends, earnings, and prices

When predicting dividend growth it is ultimately an empirical question what the optimal combination of dividends, earnings, and prices is. The dividend-price ratio \((d_t - p_t)\), the payout ratio \((d_t - e_t)\), and our main \(dy-ey\) specification \((d_t - p_t, e_t - p_t)\) all correspond to different sets of weights on \(d_t\), \(p_t\), and \(e_t\). Garrett and Priestley (2012) instead use cointegration analysis to estimate the weights. Ignoring the constant, their \(dpe_t\) cointegration residual is computed as \((d_t - \delta_1 p_t - \delta_2 e_t)\), where the coefficients are estimated from a regression of \(d_t\) on a constant, \(p_t\), and \(e_t\). An alternative approach is to impose no structure a priori and let the data decide by including all three variables on their own in a multivariate regression. Lettau and Ludvigson (2005) use this approach in the context of their cointegration analysis of consumption, dividends, and labor income.

In Table 11 we report results for the various alternatives described above (excluding \(dp_t\), which has no predictive power). For the unrestricted \(d, p, e\) regression, the right hand side variables are not stationary. However, conditional on the three variables being cointegrated, we can rely on standard asymptotics, see Sims et al. (1990). We show results for 1927:1-2013:4 and for 1950:1-2013:4 using a quarterly and an annual horizon. For the full period, all the models we consider provide similar explanatory power with \(R^2\)s around 20% at the quarterly horizon reflecting strong dividend growth predictability. For the post-war period 1950:1-2013:4, the explanatory power remains roughly intact for the \(dy-ey\) model. However, the \(R^2\) and \(t\)-statistic for the \(dpe\) cointegration residual drops substantially. In Table 12 we report results for the long annual sample and the conclusions are similar.

Interestingly, for the free specification, in which we do not impose structure on the weights of dividends, prices, and earnings, we get slopes and \(R^2\)s that are not too different from the main \(dy-ey\) specification. This is true for both pre- and post-war data even though the loadings change significantly across the two subsamples. The \(dy-ey\) model thus provides flexibility when modeling dividend growth and it does not rely on estimating cointegration vectors. Looking at the pre-war period, we see how the less restricted \(dy-ey\) model and the free multivariate specification give slope estimates on dividends, earnings, and prices that are similar to those

---

16 In our quarterly sample 1927:1-2013:4, we estimate the cointegration coefficients to be \((d_t - 0.40p_t - 0.43e_t)\) and we get similar values for the post-war period. Garrett and Priestley (2012) use annual real data, whereas we use quarterly nominal data. In unreported results, we find that the \(dp-ep\) model also strongly predicts real dividend growth.

17 As none of the models have significant predictive power for returns, we only report results for dividend growth.
of the $dpe$ cointegration residual, which has a nonzero weight on prices. The $\bar{R}^2$s are also quite close. The post-war period shows a different pattern, as the $dy$-$ey$ model and the free specification give results that mimic the payout ratio, $d - e$. For the $dy$-$ey$ model, the post-war coefficients on the dividend yield and earnings yield are very close in absolute value canceling out the effect of prices. For the free specification, the post-war price coefficient is very close to zero and is insignificant. These results suggest that dividend payments have become less responsive to price levels over time, which goes hand in hand with the finding of Chen et al. (2012) that dividends have become more smooth in the post-war period. That is, as prices are more volatile than earnings, we should expect to see more smooth dividends payments if they respond more to earnings than to prices, other things being equal.

2.7 Conclusion

We argue that the dividend yield by itself is a poor predictor of dividend growth due to an omitted variable bias. We show that dividend growth is strongly predictable when combining the dividend yield and earnings yield in a joint specification. Omitting the earnings yield conceals predictability by biasing the dividend yield coefficient towards zero. This result is consistent with theoretical models of dividend behavior such as the seminal model of Lintner (1956). In a simulation study, we show how omitting the earnings yield leads to a substantial reduction in power. Our empirical results are robust in pre- and post-war U.S. data, in recessions and expansions, when controlling for additional predictors, and in international data.

Acknowledgements

We would like to thank seminar participants, participants at the 2013 OMI-SoFiE Financial Econometrics Summer School, Tom Engsted, Lasse Bork, and, in particular, Geert Bekaert and two anonymous referees for useful comments and suggestions. We acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.
References


Table 1: Dividend growth and return predictability from dividend yields and earnings yields

The table reports results from forecasting regressions, \( y_{t+k} = \alpha_k + \beta_k^{dp} dp_t + \beta_k^{ep} ep_t + \varepsilon_{t+k} \), where \( y_{t+k} \) is the \( k \)-period ahead dividend growth rate or excess return. We use the dividend yield and the earnings yield as regressors, either separately or jointly. The frequency is quarterly and the sample period is from 1927:1 to 2013:4. Panels A and B report results for \( k = 1 \) and \( k = 4 \), respectively. Below OLS estimates of \( \beta_k \) are \( t \)-statistics in parentheses based on Newey-West standard errors with \( k + 3 \) lags and in brackets are \( t \)-statistics based on standard errors from a stationary bootstrap. We resample data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. \( \bar{R}^2 \) is the adjusted \( R \)-squared.

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>( dp )</td>
<td>( ep )</td>
</tr>
<tr>
<td>Panel A: ( k = 1 )</td>
<td></td>
</tr>
<tr>
<td>-0.018</td>
<td>0.021</td>
</tr>
<tr>
<td>(-1.12)</td>
<td>(1.13)</td>
</tr>
<tr>
<td>0.013</td>
<td>0.028</td>
</tr>
<tr>
<td>(2.09)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>[1.98]</td>
<td>[1.81]</td>
</tr>
<tr>
<td>-0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>(-3.36)</td>
<td>(2.13)</td>
</tr>
<tr>
<td>[2.42]</td>
<td>[3.11]</td>
</tr>
<tr>
<td>Panel B: ( k = 4 )</td>
<td></td>
</tr>
<tr>
<td>-0.081</td>
<td>0.093</td>
</tr>
<tr>
<td>(-1.56)</td>
<td>(1.97)</td>
</tr>
<tr>
<td>[-1.14]</td>
<td>[1.97]</td>
</tr>
<tr>
<td>0.029</td>
<td>0.102</td>
</tr>
<tr>
<td>(0.98)</td>
<td>(2.08)</td>
</tr>
<tr>
<td>[1.08]</td>
<td>[2.08]</td>
</tr>
<tr>
<td>-0.220</td>
<td>0.197</td>
</tr>
<tr>
<td>(-3.07)</td>
<td>(3.85)</td>
</tr>
<tr>
<td>[-2.10]</td>
<td>[3.16]</td>
</tr>
</tbody>
</table>
Table 2: Dividend growth and return predictability from dividend yields and earnings yields: Post-war evidence
The sample period is from 1950:1 to 2013:4. Otherwise see notes to Table 1.

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>dp</td>
<td>ep</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Panel A: $k = 1$</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td>(0.04)</td>
</tr>
<tr>
<td>[0.04]</td>
<td>[0.04]</td>
</tr>
<tr>
<td>-0.029</td>
<td>(7.94)</td>
</tr>
<tr>
<td>[-3.05]</td>
<td>[5.36]</td>
</tr>
<tr>
<td>-0.007</td>
<td>(-0.31)</td>
</tr>
<tr>
<td>[-0.32]</td>
<td>[-0.32]</td>
</tr>
</tbody>
</table>
Table 3: Dividend growth and return predictability from dividend yields and earnings yields: Pre-war evidence

The sample period is from 1927:1 to 1949:4. Otherwise see notes to Table 1.

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp$ $ep$ $R^2$</td>
<td>$dp$ $ep$ $R^2$</td>
</tr>
</tbody>
</table>

**Panel A: $k = 1$**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.113$</td>
<td>$0.051$</td>
<td>$25.1%$</td>
<td>$(0.72)$</td>
<td>$-0.2%$</td>
<td>$(0.64)$</td>
<td>$25.1%$</td>
</tr>
<tr>
<td>$(-4.16)$</td>
<td>$(-0.72)$</td>
<td>$[0.64]$</td>
<td>$0.101$</td>
<td>$3.3%$</td>
<td>$[1.98]$</td>
<td>$0.101$</td>
</tr>
<tr>
<td>$0.013$</td>
<td>$0.013$</td>
<td>$0.101$</td>
<td>$0.013$</td>
<td>$0.101$</td>
<td>$0.013$</td>
<td>$0.101$</td>
</tr>
<tr>
<td>$(0.56)$</td>
<td>$0.8%$</td>
<td>$(2.05)$</td>
<td>$3.3%$</td>
<td>$(2.05)$</td>
<td>$3.3%$</td>
<td>$(2.05)$</td>
</tr>
<tr>
<td>$[0.52]$</td>
<td>$[2.05]$</td>
<td>$[1.98]$</td>
<td>$[2.05]$</td>
<td>$[1.98]$</td>
<td>$[2.05]$</td>
<td>$[1.98]$</td>
</tr>
</tbody>
</table>

**Panel B: $k = 4$**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.529$</td>
<td>$0.220$</td>
<td>$54.6%$</td>
<td>$(1.58)$</td>
<td>$3.4%$</td>
<td>$(1.46)$</td>
<td>$3.4%$</td>
</tr>
<tr>
<td>$(-5.08)$</td>
<td>$(-5.08)$</td>
<td>$[1.46]$</td>
<td>$0.338$</td>
<td>$9.4%$</td>
<td>$[2.36]$</td>
<td>$9.4%$</td>
</tr>
<tr>
<td>$-0.100$</td>
<td>$-0.100$</td>
<td>$0.338$</td>
<td>$0.338$</td>
<td>$0.338$</td>
<td>$0.338$</td>
<td>$0.338$</td>
</tr>
<tr>
<td>$(-0.96)$</td>
<td>$-0.84$</td>
<td>$2.28$</td>
<td>$9.4%$</td>
<td>$[2.36]$</td>
<td>$9.4%$</td>
<td>$[2.36]$</td>
</tr>
<tr>
<td>$[-0.84]$</td>
<td>$[2.36]$</td>
<td>$[2.36]$</td>
<td>$[2.36]$</td>
<td>$[2.36]$</td>
<td>$[2.36]$</td>
<td>$[2.36]$</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.856$</td>
<td>$0.484$</td>
<td>$-0.014$</td>
<td>$0.347$</td>
<td>$0.347$</td>
<td>$0.347$</td>
<td>$0.347$</td>
</tr>
<tr>
<td>$(-13.39)$</td>
<td>$(-13.39)$</td>
<td>$[9.74]$</td>
<td>$[0.05]$</td>
<td>$1.21$</td>
<td>$8.3%$</td>
<td>$8.3%$</td>
</tr>
<tr>
<td>$[-8.98]$</td>
<td>$[9.74]$</td>
<td>$[0.05]$</td>
<td>$1.21$</td>
<td>$8.3%$</td>
<td>$1.21$</td>
<td>$8.3%$</td>
</tr>
</tbody>
</table>
Table 4: Long annual sample

The table reports results from forecasting regressions, \( y_{t+1} = \alpha + \beta dp_t + \beta ep_t + \varepsilon_{t+1} \), where \( y_{t+1} \) is the one-year ahead dividend growth rate or excess return. We use the dividend yield and the earnings yield as regressors, either separately or jointly. Below OLS estimates of \( \beta \) are \( t \)-statistics in parentheses based on Newey-West standard errors with one lag and in brackets are \( t \)-statistics based on standard errors from a stationary bootstrap. We resample the data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. \( R^2 \) is the adjusted \( R \)-squared. We report results from two sample periods: 1871-1949 (Panel A) and 1950-2013 (Panel B).

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>( dp )</td>
<td>( ep )</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Panel A: 1871-1949</td>
<td></td>
</tr>
<tr>
<td>-0.435</td>
<td>0.057</td>
</tr>
<tr>
<td>(-6.62)</td>
<td>[0.69]</td>
</tr>
<tr>
<td>-0.047</td>
<td>0.036</td>
</tr>
<tr>
<td>(-0.76)</td>
<td>[0.53]</td>
</tr>
<tr>
<td>-0.593</td>
<td>0.229</td>
</tr>
<tr>
<td>(-7.46)</td>
<td>[3.29]</td>
</tr>
<tr>
<td>Panel B: 1950-2013</td>
<td></td>
</tr>
<tr>
<td>-0.014</td>
<td>0.100</td>
</tr>
<tr>
<td>(-0.62)</td>
<td>(1.96)</td>
</tr>
<tr>
<td>[0.52]</td>
<td>[2.02]</td>
</tr>
<tr>
<td>-0.133</td>
<td>0.151</td>
</tr>
<tr>
<td>(-3.91)</td>
<td>(4.91)</td>
</tr>
<tr>
<td>[2.73]</td>
<td>[4.24]</td>
</tr>
</tbody>
</table>

58
We run forecasting regressions, \( y_{t+k} = \alpha_k + \beta_k^{dp} dp_t + \beta_k^{ep} ep_t + \varepsilon_{t+k} \), where \( y_{t+k} \) is the \( k \)-period ahead dividend growth rate or excess return. We use the dividend yield and the earnings yield as regressors, either separately or jointly. We report \( E(y_{t+k}) \), \( \sigma(\hat{y}_{t+k}) \), and \( \sigma(\hat{y}_{t+k})/E(y_{t+k}) \), where \( \hat{y}_{t+k} \) is the fitted values from the regressions. All values are in percent. We report results from two sample periods: 1927:1-2013:4 (Panel A) and 1950:1-2013:4 (Panel B).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(y_{t+k}) )</td>
<td>( k = 1 )</td>
<td>1.11 1.11 1.11</td>
<td>1.47 1.47 1.47</td>
</tr>
<tr>
<td>( \sigma(\hat{y}_{t+k}) )</td>
<td>( k = 1 )</td>
<td>0.81 0.57 1.88</td>
<td>0.94 1.18 1.19</td>
</tr>
<tr>
<td>( \sigma(\hat{y}<em>{t+k})/E(y</em>{t+k}) )</td>
<td>( k = 1 )</td>
<td>0.73 0.51 1.70</td>
<td>0.64 0.81 0.81</td>
</tr>
<tr>
<td>( E(y_{t+k}) )</td>
<td>( k = 4 )</td>
<td>4.40 4.40 4.40</td>
<td>5.66 5.66 5.66</td>
</tr>
<tr>
<td>( \sigma(\hat{y}_{t+k}) )</td>
<td>( k = 4 )</td>
<td>3.63 1.25 6.79</td>
<td>4.16 4.38 4.60</td>
</tr>
<tr>
<td>( \sigma(\hat{y}<em>{t+k})/E(y</em>{t+k}) )</td>
<td>( k = 4 )</td>
<td>0.83 0.28 1.54</td>
<td>0.74 0.77 0.81</td>
</tr>
</tbody>
</table>

59
Table 6: Controlling for additional variables

The table reports results from forecasting regressions, $y_{t+1} = \alpha + \beta^d p_t + \beta^p e_{p,t} + \beta^z z_t + \varepsilon_{t+1}$, where $y_{t+1}$ is the one-quarter ahead dividend growth rate or excess return. $z_t$ is either the short rate, the term spread, the default spread, or $cay$. We report results from two sample periods: 1927:1-2013:4 (Panel A) and 1950:1-2013:4 (Panel B). $cay$ is available from 1952:1. Below OLS estimates of slope coefficients are $t$-statistics in parentheses based on Newey-West standard errors with 4 lags and in brackets are $t$-statistics based on standard errors from a stationary bootstrap. We resample the data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. $\hat{R}^2$ is the adjusted $R$-squared.

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp$ $ep$ $z$</td>
<td>$\hat{R}^2$</td>
</tr>
<tr>
<td>Short rate</td>
<td>&amp; -0.063 &amp; 0.064 &amp; -0.146 &amp; -0.006 &amp; 0.038 &amp; -0.349</td>
</tr>
<tr>
<td>$(-3.54)$</td>
<td>$(4.12)$</td>
</tr>
<tr>
<td>[-2.12]</td>
<td>[2.62]</td>
</tr>
<tr>
<td>Term spread</td>
<td>&amp; -0.059 &amp; 0.050 &amp; 0.052 &amp; 0.000 &amp; 0.035 &amp; 0.801</td>
</tr>
<tr>
<td>$(-3.42)$</td>
<td>$(4.17)$</td>
</tr>
<tr>
<td>[-2.39]</td>
<td>[2.97]</td>
</tr>
<tr>
<td>Default spread</td>
<td>&amp; -0.041 &amp; 0.046 &amp; -1.462 &amp; 0.004 &amp; 0.031 &amp; 0.659</td>
</tr>
<tr>
<td>$(-3.76)$</td>
<td>$(4.79)$</td>
</tr>
<tr>
<td>[-2.13]</td>
<td>[3.07]</td>
</tr>
<tr>
<td><strong>cay</strong></td>
<td>&amp; -0.030 &amp; 0.035 &amp; 0.065 &amp; 0.020 &amp; 0.006 &amp; 0.653</td>
</tr>
<tr>
<td>$(-3.63)$</td>
<td>$(8.47)$</td>
</tr>
<tr>
<td>[-3.01]</td>
<td>[5.88]</td>
</tr>
</tbody>
</table>
Table 7: Recessions vs. expansions

The table reports results from running forecasting regressions of the following form: \( y_{t+1} = (\alpha_R + \beta_{dp}^R dp_t + \beta_{ep}^R ep_t) \times I_t + (\alpha_E + \beta_{dp}^E dp_t + \beta_{ep}^E ep_t) \times (1 - I_t) + \varepsilon_t \), where \( I_t \) is the NBER recession indicator and \( y_{t+1} \) is the one-quarter ahead dividend growth rate or excess return. We report results from two sample periods: 1927:1-2013:4 (Panel A) and 1950:1-2013:4 (Panel B). Below OLS estimates of slope coefficients are \( t \)-statistics in parentheses based on Newey-West standard errors with 4 lags and in brackets are \( t \)-statistics based on standard errors from a stationary bootstrap. We resample the data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. \( R^2 \) is the adjusted \( R \)-squared.

<table>
<thead>
<tr>
<th>Dividend growth</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>( dp \times I )</td>
<td>( dp \times (1-I) )</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Panel A: 1927:1-2013:4</td>
<td></td>
</tr>
<tr>
<td>-0.044</td>
<td>-0.003</td>
</tr>
<tr>
<td>(-1.63)</td>
<td>(-0.34)</td>
</tr>
<tr>
<td>[-1.19]</td>
<td>[-0.31]</td>
</tr>
<tr>
<td>0.018</td>
<td>0.013</td>
</tr>
<tr>
<td>(2.28)</td>
<td>(1.80)</td>
</tr>
<tr>
<td>[1.35]</td>
<td>[2.04]</td>
</tr>
<tr>
<td>-0.099</td>
<td>-0.036</td>
</tr>
<tr>
<td>(-3.35)</td>
<td>(-2.48)</td>
</tr>
<tr>
<td>[-2.36]</td>
<td>[-2.08]</td>
</tr>
<tr>
<td>0.009</td>
<td>0.002</td>
</tr>
<tr>
<td>(1.21)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>[0.82]</td>
<td>[0.31]</td>
</tr>
<tr>
<td>0.018</td>
<td>0.015</td>
</tr>
<tr>
<td>(3.85)</td>
<td>(2.20)</td>
</tr>
<tr>
<td>[2.82]</td>
<td>[2.34]</td>
</tr>
<tr>
<td>-0.026</td>
<td>-0.028</td>
</tr>
<tr>
<td>(-4.68)</td>
<td>(-2.76)</td>
</tr>
<tr>
<td>[-1.40]</td>
<td>[-2.57]</td>
</tr>
<tr>
<td>Panel B: 1950:1-2013:4</td>
<td></td>
</tr>
<tr>
<td>0.018</td>
<td>0.015</td>
</tr>
<tr>
<td>(3.85)</td>
<td>(2.20)</td>
</tr>
<tr>
<td>[2.82]</td>
<td>[2.34]</td>
</tr>
<tr>
<td>-0.026</td>
<td>-0.028</td>
</tr>
<tr>
<td>(-4.68)</td>
<td>(-2.76)</td>
</tr>
<tr>
<td>[-1.40]</td>
<td>[-2.57]</td>
</tr>
</tbody>
</table>
Table 8: International evidence
The table reports results from international one-quarter ahead forecasting regressions of dividend growth and returns using the dividend yield and earnings yield as predictive variables. We use data from 14 developed countries over the period 1973:1 to 2013:4. We forecast equal-weighted (EW) and value-weighted (VW) global portfolios. For each regression, we report slope estimates, Newey-West $t$-statistics with 4 lags in parentheses, and bootstrap $t$-statistics in brackets. We also run cross-section fixed-effect panel regressions from which we report Thompson (2011) two-way clustered robust $t$-statistics with 4 lags in parentheses and bootstrap $t$-statistics in brackets. We use a stationary bootstrap where the random time-series blocks are drawn commonly for all countries to preserve cross-correlation. For the panel regressions, the $\bar{R}^2$ column reports the within $R^2$.

<table>
<thead>
<tr>
<th></th>
<th>Dividend growth</th>
<th></th>
<th></th>
<th></th>
<th>Returns</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$dp$</td>
<td>$\bar{R}^2$</td>
<td>$ep$</td>
<td>$\bar{R}^2$</td>
<td>$dp$</td>
<td>$ep$</td>
<td>$\bar{R}^2$</td>
</tr>
<tr>
<td>Equal-weighted</td>
<td>-0.024</td>
<td>-0.008</td>
<td>-0.131</td>
<td>0.129</td>
<td>-0.004</td>
<td>0.002</td>
<td>-0.073</td>
</tr>
<tr>
<td></td>
<td>(-1.94) 4.7%</td>
<td>(-0.75) -0.1%</td>
<td>(-3.50) (3.39)</td>
<td>19.9%</td>
<td>(-0.63) 0.1%</td>
<td>(0.30) -0.5%</td>
<td>(-3.39) (3.83)</td>
</tr>
<tr>
<td></td>
<td>[-1.55] [-0.73]</td>
<td>[-3.23] [3.29]</td>
<td></td>
<td></td>
<td>[-0.56] [0.30]</td>
<td>[-2.56] [3.07]</td>
<td></td>
</tr>
<tr>
<td>Value-weighted</td>
<td>-0.004</td>
<td>0.002</td>
<td>-0.073</td>
<td>0.078</td>
<td>-0.038</td>
<td>-0.011</td>
<td>-0.076</td>
</tr>
<tr>
<td></td>
<td>(-0.63) 0.1%</td>
<td>(0.30) -0.5%</td>
<td>(-3.39) (3.83)</td>
<td>15.2%</td>
<td>(-3.49) 3.7%</td>
<td>(-2.39) 0.3%</td>
<td>(-4.13) (3.20)</td>
</tr>
<tr>
<td></td>
<td>[-0.56] [0.30]</td>
<td>[-2.56] [3.07]</td>
<td></td>
<td></td>
<td>[-3.59] [-1.99]</td>
<td>[-3.95] [3.03]</td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>-0.038</td>
<td>-0.011</td>
<td>-0.076</td>
<td>0.049</td>
<td>-0.040</td>
<td>0.044</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(-3.49) 3.7%</td>
<td>(-2.39) 0.3%</td>
<td>(-4.13) (3.20)</td>
<td>6.1%</td>
<td>(1.68) 1.0%</td>
<td>(1.55) 1.0%</td>
<td>(0.28) (0.22)</td>
</tr>
<tr>
<td></td>
<td>[-3.59] [-1.99]</td>
<td>[-3.95] [3.03]</td>
<td></td>
<td></td>
<td>[1.51]  [1.46]</td>
<td>[0.31] [0.25]</td>
<td></td>
</tr>
<tr>
<td>Equal-weighted</td>
<td>0.017</td>
<td>0.014</td>
<td>0.056</td>
<td>-0.044</td>
<td>0.043</td>
<td>0.039</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.96) 0.1%</td>
<td>(0.76) -0.2%</td>
<td>(1.04) (-0.77)</td>
<td>-0.2%</td>
<td>(2.86) 1.8%</td>
<td>(2.50) 1.5%</td>
<td>(2.10) (0.66)</td>
</tr>
<tr>
<td></td>
<td>[0.89] [0.72]</td>
<td>[0.95] [-0.73]</td>
<td></td>
<td></td>
<td>[2.55]  [2.38]</td>
<td>[1.87] [0.74]</td>
<td></td>
</tr>
</tbody>
</table>
Table 9: International evidence: subsample analysis
The table reports results from international one-quarter ahead forecasting regressions of dividend growth using the dividend yield and earnings yield as predictive variables. We use data from 14 developed countries and show results for two subperiods: 1973:1 to 1992:4 and 1993:1 to 2013:4. We estimate cross-section fixed-effect panel regressions from which we report Thompson (2011) two-way clustered robust t-statistics with 4 lags in parentheses and bootstrap t-statistics in brackets. We use a stationary bootstrap where the random time-series blocks are drawn commonly for all countries to preserve cross-correlation.

<table>
<thead>
<tr>
<th></th>
<th>dp</th>
<th>$R^2_{within}$</th>
<th>ep</th>
<th>$R^2_{within}$</th>
<th>dp</th>
<th>ep</th>
<th>$R^2_{within}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973:1-1992:4</td>
<td>-0.038</td>
<td>-0.022</td>
<td>-0.055</td>
<td>0.021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.41)</td>
<td>(-2.21)</td>
<td>1.4%</td>
<td>(-2.62)</td>
<td>(1.04)</td>
<td></td>
<td>4.8%</td>
</tr>
<tr>
<td></td>
<td>[-5.08]</td>
<td>[-3.11]</td>
<td>[-3.68]</td>
<td>[1.44]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993:1-2013:4</td>
<td>-0.077</td>
<td>-0.015</td>
<td>-0.143</td>
<td>0.093</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.35)</td>
<td>(-1.50)</td>
<td>0.3%</td>
<td>(-5.36)</td>
<td>(6.35)</td>
<td></td>
<td>13.2%</td>
</tr>
<tr>
<td></td>
<td>[-4.44]</td>
<td>[-0.93]</td>
<td>[-5.86]</td>
<td>[4.72]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10: Explaining the non-predictability of dividend growth as an omitted variable bias

$\hat{c}$ is the OLS estimate from regressing $\Delta d_{t+k}$ on $d_t - c_t$, $\hat{\gamma}$ is the OLS estimate from regressing $ep_t$ on $dp_t$, and $\hat{\beta}$ is the OLS estimate from regressing $\Delta d_{t+k}$ on $dp_t$. $E(\hat{\beta})$ is calculated as $-\hat{c}(1 - \hat{\gamma})$. The quantiles are from simulations of the system in equations (12)-(13). For the coefficients, we use equation-by-equation Maximum Likelihood estimates. For each simulation, we draw a value from the empirical likelihood of $c$ where the likelihood is calculated with the other parameters maxed out. We use 50,000 simulations and draw $(\hat{\varepsilon}_{t+1}, \hat{\eta}_{t+1}^{dp}, \hat{\eta}_{t+1}^{ep})'$ jointly to preserve cross-correlation. In each simulated sample we regress $\Delta d_{t+k}$ on $dp_t$ and report quantiles of the resulting distribution of slope coefficients. We also report probabilities of rejecting the null of no predictability (power) for the univariate regression of $\Delta d_{t+k}$ on $dp_t$ ($t$-statistic is $t^\hat{\beta}$) and for the multivariate regression of $\Delta d_{t+k}$ on $dp_t$ and $ep_t$ ($t$-statistic is $t^{\hat{\beta}^*}$).

<table>
<thead>
<tr>
<th>Data</th>
<th></th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{c}$</td>
<td>$\hat{\gamma}$</td>
</tr>
<tr>
<td>1927:1-2013:4</td>
<td>-0.059</td>
<td>0.698</td>
</tr>
<tr>
<td>1950:1-2013:4</td>
<td>-0.034</td>
<td>0.794</td>
</tr>
<tr>
<td>$k = 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1927:1-2013:4</td>
<td>-0.210</td>
<td>0.702</td>
</tr>
<tr>
<td>1950:1-2013:4</td>
<td>-0.121</td>
<td>0.799</td>
</tr>
<tr>
<td>$k = 4$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11: Other specifications of dividends, earnings, and prices

The table reports results from $k$-quarter ahead forecasting regressions, $\Delta d_{t+k} = \alpha + \beta' x_t + \varepsilon_{t+k}$. We report results for $x_t = (dp_t, ep_t)'$, $x_t = d_t - e_t$, $x_t = dpe_t = (d_t - \delta_0 - \delta_1 p_t - \delta_2 e_t)$, and $x_t = (d_t, p_t, e_t)’$. The cointegration coefficients $(\delta_0, \delta_1, \delta_2)'$ are estimated from a regression of $d_t$ on a constant, $p_t$, and $e_t$. In parentheses are $t$-statistics using Newey-West standard errors with $k + 3$ lags and in brackets are $t$-statistics using standard errors from a stationary bootstrap. We resample the data in blocks of random size and determine the average block-size using the Politis and White (2004) automatic selection procedure. $\bar{R}^2$ is the adjusted $R$-squared.

<table>
<thead>
<tr>
<th></th>
<th>$k = 1$</th>
<th></th>
<th>$k = 4$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>$\beta_2$</td>
<td>$\beta_3$</td>
<td>$\bar{R}^2$</td>
</tr>
<tr>
<td>Panel A: 1927:1-2013:4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dp, ep$</td>
<td>-0.059</td>
<td>0.058</td>
<td></td>
<td>19.5%</td>
</tr>
<tr>
<td></td>
<td>(-3.36)</td>
<td>(4.05)</td>
<td></td>
<td>[2.42]</td>
</tr>
<tr>
<td>$d - e$</td>
<td>-0.059</td>
<td></td>
<td></td>
<td>19.7%</td>
</tr>
<tr>
<td></td>
<td>(-3.69)</td>
<td></td>
<td></td>
<td>[-2.78]</td>
</tr>
<tr>
<td>$dpe$</td>
<td>-0.089</td>
<td></td>
<td></td>
<td>17.1%</td>
</tr>
<tr>
<td></td>
<td>(-3.69)</td>
<td></td>
<td></td>
<td>[-2.52]</td>
</tr>
<tr>
<td>$d, p, e$</td>
<td>-0.089</td>
<td>0.017</td>
<td>0.063</td>
<td>22.5%</td>
</tr>
<tr>
<td></td>
<td>(-3.25)</td>
<td>(1.59)</td>
<td>(3.35)</td>
<td>[-3.14]</td>
</tr>
<tr>
<td>Panel B: 1950:1-2013:4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dp, ep$</td>
<td>-0.029</td>
<td>0.037</td>
<td></td>
<td>15.4%</td>
</tr>
<tr>
<td></td>
<td>(-3.53)</td>
<td>(7.94)</td>
<td></td>
<td>[-3.05]</td>
</tr>
<tr>
<td>$d - e$</td>
<td>-0.034</td>
<td></td>
<td></td>
<td>14.3%</td>
</tr>
<tr>
<td></td>
<td>(-5.95)</td>
<td></td>
<td></td>
<td>[-4.27]</td>
</tr>
<tr>
<td>$dpe$</td>
<td>-0.032</td>
<td></td>
<td></td>
<td>4.7%</td>
</tr>
<tr>
<td></td>
<td>(-1.99)</td>
<td></td>
<td></td>
<td>[-1.74]</td>
</tr>
<tr>
<td>$d, p, e$</td>
<td>-0.032</td>
<td>-0.006</td>
<td>0.037</td>
<td>15.0%</td>
</tr>
<tr>
<td></td>
<td>(-3.14)</td>
<td>(-0.89)</td>
<td>(7.66)</td>
<td>[-2.90]</td>
</tr>
</tbody>
</table>
Table 12: Other specifications of dividends, earnings, and prices: evidence from long annual sample

We use annual observations. Otherwise see notes for Table 11.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$R^2$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: 1871-1949</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dp$, $ep$</td>
<td>-0.593</td>
<td>0.229</td>
<td></td>
<td></td>
<td>-0.133</td>
<td>0.151</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-7.46)</td>
<td>(3.40)</td>
<td>60.1%</td>
<td></td>
<td>(-3.91)</td>
<td>(4.91)</td>
<td>45.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-7.09]</td>
<td>[3.29]</td>
<td></td>
<td></td>
<td>[-2.73]</td>
<td>[4.24]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d - e$</td>
<td>-0.331</td>
<td></td>
<td></td>
<td></td>
<td>-0.142</td>
<td></td>
<td>44.9%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.92)</td>
<td></td>
<td>29.6%</td>
<td></td>
<td>(-4.25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-3.85]</td>
<td></td>
<td></td>
<td></td>
<td>[-3.82]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dpe$</td>
<td>-0.578</td>
<td></td>
<td></td>
<td></td>
<td>-0.146</td>
<td></td>
<td>19.6%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-6.51)</td>
<td></td>
<td>52.1%</td>
<td></td>
<td>(-1.82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-6.06]</td>
<td></td>
<td></td>
<td></td>
<td>[-1.74]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d$, $p$, $e$</td>
<td>-0.577</td>
<td>0.383</td>
<td>0.219</td>
<td>60.1%</td>
<td>-0.146</td>
<td>-0.012</td>
<td>0.153</td>
<td>45.2%</td>
</tr>
<tr>
<td></td>
<td>(-6.84)</td>
<td>(6.88)</td>
<td>(3.19)</td>
<td></td>
<td>(-2.96)</td>
<td>(-0.45)</td>
<td>(5.11)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Distribution of $dp_t$ regression coefficients assuming no predictability

We simulate 50,000 bootstrap samples of length $T = 348$ using $\Delta d_{t+1} = 0.012 + \hat{\xi}_{t+1} + 0.489\xi_t + 0.499\xi_{t-1} + 0.433\xi_{t-2} - 0.286\xi_{t-3}$, $dp_{t+1} = -0.096 + 0.951dp_t + 0.026ep_t + \hat{\eta}_{t+1}^{dp}$, and $ep_{t+1} = -0.162 + 0.016dp_t + 0.922ep_t + \hat{\eta}_{t+1}^{ep}$. The coefficients are the full-sample (1927:1-2013:4) equation-by-equation Maximum Likelihood estimates. We draw $(\hat{\xi}_{t+1}, \hat{\eta}_{t+1}^{dp}, \hat{\eta}_{t+1}^{ep})$ jointly to preserve cross-correlation. In each simulated sample, we regress $\Delta d_{t+k}$ on $(dp_t, ep_t)^\prime$ for $k = 1$ and $k = 4$. The histograms show the distribution of OLS estimates of the slope coefficients on $dp_t$. The black lines are the original full-sample estimates and the percentages indicate the fraction of simulations more extreme than the original estimates.
Figure 2: Dividend growth regressions for individual countries

The figure shows results from forecasting regressions, \( y_{i,t+1} = \alpha_i + \beta_{i,t} d_{i,t} + \beta_{i,t} e_{i,t} + \varepsilon_{i,t+1} \), where \( y_{i,t+1} \) is the one-period ahead dividend growth rate for country \( i \). We use the dividend yield and the earnings yield as regressors, either separately or jointly. The frequency is quarterly and the sample period is from 1973:1 to 2013:4. Panel A shows the estimates of \( \beta_{i,t} d_{i,t} \) obtained from univariate regressions using only \( d_{i,t} \) as regressor (white bars) and from bivariate regressions using \( d_{i,t} \) and \( e_{i,t} \) as regressors (black bars). Similarly, Panel B shows estimates of \( \beta_{i,t} e_{i,t} \) obtained from univariate regressions using only \( e_{i,t} \) as regressor (white bars) and from bivariate regressions using \( d_{i,t} \) and \( e_{i,t} \) as regressors (black bars). Finally, Panel C shows adjusted \( R^2 \)s from the univariate dividend yield regressions (white bars) and bivariate regressions (black bars). For illustrative purposes, we have divided the slope coefficients for South Africa (ZAF) by two.
BOND MARKET ASYMMETRIES ACROSS RECESSIONS AND EXPANSIONS: NEW EVIDENCE ON RISK PREMIA

Martin M. Andreasen  
Aarhus University and CREATES

Tom Engsted  
Aarhus University and CREATES

Stig V. Møller  
Aarhus University and CREATES

Magnus Sander  
Aarhus University and CREATES

Abstract

This paper provides new evidence on bond risk premia by conditioning the classic Campbell-Shiller regressions on the business cycle. In expansions, we find mostly positive intercepts and negative slopes, but the results are completely reversed in recessions with negative intercepts and positive slopes. The pattern in these coefficients is explained by a term structure model with business cycle dependent loadings in the market price of risk. This model also predicts realized excess returns with the right sign during expansions and recessions, whereas the Gaussian ATSM is unable to explain realized excess returns on medium- and long-term bonds during recessions.
3.1 Introduction

The expectations hypothesis of the term structure specifies that any long-term interest rate is given by the average of expected future short rates and that excess returns on any long-term bond therefore are unpredictable. Despite its appealing intuition, the expectations hypothesis is often rejected empirically, perhaps most forcefully in the seminal study of Campbell and Shiller (1991). This paper provides new evidence on bond risk premia by conditioning the classic Campbell-Shiller regressions on the business cycle through separate intercepts and slope coefficients in expansions and recessions. This extension is motivated by previous research showing that interest rates are more persistent in expansions than in recessions and that two-state models describe interest rate dynamics much better than single-state models, see Hamilton (1988), Gray (1996), Ang and Bekaert (2002), Bansal and Zhou (2002), Hamilton and Okimoto (2011), among others.

Our modified Campbell-Shiller regressions reveal a unique business cycle dependent pattern in the relation between the slope of the yield curve (i.e., the yield spread) and subsequent yield changes. In expansions, we find mostly positive intercepts and negative regression slopes that decrease with maturity, similar to what typically is reported for ordinary Campbell-Shiller regressions, i.e. without conditioning on the business cycle (see, e.g., Bekaert, Hodrick and Marshall (1997) and Dai and Singleton (2002)). By contrast, in recessions, we obtain negative intercepts and positive slope coefficients that generally increase with maturity. This switch in the regression coefficients is significant in statistical tests. Given that the slope coefficients are closer to the expectations hypothesis in recessions than in expansions, this novel result suggests excess returns and hence bond risk premia are more predictable by the yield spread during expansions than during recessions. We also show that the evidence of asymmetric return predictability extends to other prominent yield-based predictors such as the forward spread (Fama and Bliss (1987)) and the return forecasting factor introduced by Cochrane and Piazzesi (2005).

To provide an explanation of these new empirical findings, we propose a dynamic term structure model (DTSM) with business cycle dependent loadings in the market price of risk. Contrary to previous DTSMs with regime switching, we discipline our model to have observed regimes of either expansions or recessions but include time-varying physical transition probabilities between regimes, which may depend on the business cycle and the current yield curve. To ensure fast and reliable inference of our model, we extend the sequential regression (SR) approach of Andreasen and

\[1\] Hence, forecasting bond returns during recessions from macro variables as in Ludvigson and Ng (2009) seems more promising, although Duffee (2013) and Bauer and Hamilton (2015) question the ability of macro variables to predict bond returns beyond the information contained in the yield curve when accounting for small-sample distortions.

71
Christensen (2015) to estimate all parameters in the regime-dependent market price of risk in closed form by a modified OLS regression. This implies that our regime-switching model can be estimated in a few minutes with no additional computational costs compared to the Gaussian affine term structure model (ATSM).

The main results from estimating our regime-switching model with three pricing factors on monthly U.S. data (1961:6 to 2013:12) are as follows. First, our model is able to match evidence from ordinary Campbell-Shiller regressions, but also the empirical intercepts and slope coefficients in the modified Campbell-Shiller regressions conditioning on the business cycle. As a consequence, model-implied expected excess returns correlate positively with realized excess returns during both expansions and recessions. In contrast, the inability of the Gaussian ATSM to match the switch in the Campbell-Shiller regression slopes implies that this model is unable to explain realized excess returns for medium- and long-term bonds during recessions. Second, our model also replicates the observed asymmetry in return predictability, i.e. that the yield spread, the forward spread, and the Cochrane-Piazzesi factor have strong predictive power in expansions but not in recessions. Third, to account for these asymmetric patterns in the yield curve across the business cycle, our model generates a negative relation between the short rate and excess returns in expansions (as in the Gaussian ATSM), but a positive relation in recessions. This means that our model displays a tendency for expected excess returns to peak at the start of a recession and then mean-revert during the middle and last part of a recession when the Federal Reserve starts to lower its policy rate. Thus, our model suggests that the Federal Reserve is able to remove some of the risks attached to recessions, as accommodating monetary policy reduces the required risk compensation in the bond market. This effect of monetary policy is not present in the Gaussian ATSM, which generally predicts that excess returns increase throughout the entire recession. We finally show that our fully flexible regime-switching model is robust to accounting for the zero lower bound and may be simplified to only having two parameters in the market price of risk that switch between expansions and recessions.

The rest of the paper is organized as follows. Section 3.2 presents new empirical evidence on bond risk premia from modified Campbell-Shiller regressions and their related return regressions. We then propose a new DTSM with regime-switching in Section 3.3 and describe how to estimate this model by the SR approach. Section 3.4 explores the ability of this model to replicate the empirical findings from Section 3.2, while Section 3.5 examines the robustness of our model and relates it to the existing literature. Section 3.6 concludes. The Appendix contains data descriptions and some econometric details related to our analysis.\textsuperscript{2}

\textsuperscript{2}In addition, robustness checks and all model derivations are provided in an Online Appendix, which is available from the authors' homepages or upon request.
3.2 New empirical evidence on bond risk premia

This section presents new evidence on the business cycle properties of bond risk premia. We present our main empirical finding in Sections 3.2.1 to 3.2.3 and explore its robustness in Section 3.2.4. The main implication of this finding for DTSMs is finally discussed in Section 3.2.5.

3.2.1 Modified Campbell-Shiller regressions

To motivate our new empirical finding, consider the ordinary Campbell and Shiller (1991) regression

\[
y_{t+m,k-m} - y_{t,k} = \alpha_k + \beta_k \frac{m}{k-m} (y_{t,k} - y_{t,m}) + u_{t+m,k},
\]

where \( y_{t,k} \) refers to the \( k \)-period bond yield in period \( t \). We set one period equal to one month and implement (1) by running the regressions with \( m = 3 \) for \( k = 6, 9, 12, \ldots, 120 \), implying a total of 39 regressions. That is, \( y_{t,m} \) in (1) corresponds to the three month interest rate. The upper part of Figure 1 shows the results of estimating (1) from 1961:6 to 2013:12, with intercepts on the left and slope estimates on the right. The full and dashed lines are based on Fama and Bliss (1987) and Gürkaynak, Sack and Wright (2007) bond yields, respectively.\(^3\) According to the expectations hypothesis, we should see \( \beta_k = 1 \) for all \( k \). However, as in Campbell and Shiller (1991), the slope estimates are negative and decreasing with maturity, constituting a clear violation of the expectations hypothesis.

Now consider what happens when conditioning on the state of the business cycle. That is, we run modified Campbell-Shiller regressions interacted with business cycle dummies, i.e.

\[
y_{t+m,k-m} - y_{t,k} = \alpha_k^{\text{EXP}} 1_{\{z_t \geq c\}} + \beta_k^{\text{EXP}} \frac{m}{k-m} 1_{\{z_t \geq c\}} (y_{t,k} - y_{t,m}) + \alpha_k^{\text{REC}} (1 - 1_{\{z_t \geq c\}}) + \beta_k^{\text{REC}} \frac{m}{k-m} (1 - 1_{\{z_t \geq c\}}) (y_{t,k} - y_{t,m}) + u_{t+m,k},
\]

for \( k = 6, 9, 12, \ldots, 120 \). Here, \( 1_{\{z_t \geq c\}} \) is the indicator function with a value of one for expansions when \( z_t \geq c \) and zero otherwise. Recessions are measured by letting \( z_t \) refer to the Purchasing Managers’ Index (PMI), where the threshold value \( c = 44.5 \) from Berge and Jordà (2011) is used to identify recessions and expansions.

\(^3\) Some of the long-term bond yields from Fama and Bliss (1987) are missing in the 1960s and the start of the 1970s due to lack of long-term coupon bonds, and our estimation sample is therefore reduced accordingly. See Appendix A.1 for further details on the data.
The PMI is a widely watched indicator of business cycle activity and has the advantage of being available in real time without publication lags or subsequent data revisions. Furthermore, Christiansen, Eriksen and Møller (2014) demonstrate that the PMI is the single best recession indicator among a large panel of economic variables. As shown below, our results are robust to using standard NBER recessions, although they are subject to publication lags and therefore not our preferred recession indicator.

The lower part of Figure 1 summarizes the results of estimating the modified Campbell-Shiller regression in (2). In expansions, the slope coefficients $\beta_k^{EXP}$ and intercepts $\alpha_k^{EXP}$ are broadly similar to those obtained in the ordinary Campbell-Shiller regressions. In recessions, however, the slope coefficients $\beta_k^{REC}$ are positive and mostly increasing with maturity. At long maturities they are even above one. In addition, the intercepts $\alpha_k^{REC}$ are negative at all maturities, which under the expectations hypothesis is consistent with an upward sloping yield curve. We also emphasize that these differences appear both when using Fama and Bliss (1987) and Gürkaynak, Sack and Wright (2007) bond yields. Hence, in recessions the slope of the yield curve predicts future long-term bond yields in the direction implied by the expectations hypothesis, whereas the opposite holds during expansions.

Recessions are typically short lived. Nevertheless, allowing for the sign switch has a large effect on the goodness-of-fit in the Campbell-Shiller regressions. This is illustrated in Figure 2, which shows that the $R^2$ statistic increases from around 3% in the ordinary Campbell-Shiller regressions to around 5% to 6% when conditioning on the business cycle.

### 3.2.2 Statistical significance

Having outlined economically interesting differences in the Campbell-Shiller coefficients across recessions and expansions, a natural next step is to test whether these differences are statistically significant. Table 1-2 therefore report estimation results from ordinary and modified Campbell-Shiller regressions for the most commonly considered maturities of $k = 24, 48, 72, 96, 120$. To facilitate hypothesis testing, output from the modified Campbell-Shiller regressions is reported as the difference in intercepts and slope coefficients across regimes, i.e. $\alpha^\Delta_k = \alpha_k^{REC} - \alpha_k^{EXP}$ and $\beta^\Delta_k = \beta_k^{REC} - \beta_k^{EXP}$, and associated $t$-statistics are robust to heteroskedasticity and autocorrelation.\footnote{We use a three-lag Newey-West adjustment, but the results are robust towards other reasonable choices of the lag length.}

Table 1-2 show that $\alpha^\Delta_k$ is nearly statistically significant across all reported maturities, both when using Fama and Bliss (1987) and Gürkaynak, Sack and Wright

74
We also see that estimates of $\beta_k^A$ are large in economic terms but not always statistically significant in these single-maturity regressions. Still, at long maturities, we tend to reject that $\beta_k^A = 0$, at least when using a 10% significance level (in two-sided tests). The table also reveals that the slope coefficients in expansions typically are larger in magnitude than the corresponding estimates in ordinary Campbell-Shiller regressions. For instance, using Fama-Bliss bonds yields at the six-year maturity ($k = 72$), the slope coefficient is $-2.91$ when solely focusing on expansions but only $-2.43$ in the ordinary Campbell-Shiller regression. The related t-statistics are $-3.62$ and $-2.51$, implying that the evidence against the expectations hypothesis strengthens in expansions.

Rather than restricting our attention to a single maturity at a time, we next turn to more powerful Wald tests examining the joint significance of the switch in all intercepts and all slope coefficients, respectively. We start by considering the commonly used five maturities to represent the 10-year maturity spectrum, before gradually including more bond yields to increase the power of the tests. The results are provided in Table 3 using Fama-Bliss bond yields. Although we may be unable to reject $\beta_k^A = 0$ for certain maturities when considered one at the time, the Wald tests strongly reject the null hypothesis of equality between recession and expansion slopes. This even holds when using the standard five maturities $k = 24, 48, 72, 96, 120$ to represent the 10-year maturity spectrum. Similarly, the Wald tests also strongly reject that $\alpha_k^A = 0$.

Overall, we therefore conclude that the reported differences in intercepts and slope coefficients across expansions and recessions in Figure 1 are statistically significant.

### 3.2.3 Modified return regressions

The modified Campbell-Shiller regressions in (2) show that the slope coefficients during recessions generally are closer to the expectations hypothesis than the corresponding loadings in expansions. As excess returns are unpredictable under the expectations hypothesis, we therefore find higher return predictability by the yield spread during expansions than in recessions. But does this pattern also hold for other classic yield-based predictors of bond returns such as the forward spread (Fama (1976), Fama and Bliss (1987)) and the Cochrane and Piazzesi (2005) (CP) return forecasting factor?

We address this question by running standard univariate return regressions for

---

5 Here, we exploit the one-to-one relation between Campbell-Shiller regressions and regressing excess returns on the yield spread. The exact expression for this relation is provided in our Online Appendix.
three-months ahead excess returns

\[ xhpr_{t+m,k} = \mu_k + \theta_k x_{t,k} + e_{t+m,k} \]  \hspace{1cm} (3)

with \( m = 3 \), and its modified version conditioning on the business cycle, i.e.

\[ xhpr_{t+m,k} = \mu_k^{\text{EXP}} 1_{\{z_t \geq c\}} + \theta_k^{\text{EXP}} 1_{\{z_t \geq c\}} x_{t,k} + \mu_k^{\text{REC}} (1 - 1_{\{z_t \geq c\}}) x_{t,k} + \tilde{e}_{t+m,k} \]  \hspace{1cm} (4)

Here, \( xhpr_{t+m,k} \equiv hpr_{t+m,k} - \frac{m}{12} y_{t,m} \) and the holding period return is \( hpr_{t+m,n} = \frac{k-m}{12} y_{t+m,k-m} + \frac{k}{12} y_{t,k} \), whereas \( x_{t,k} \) may either refer to the yield spread \( y_{t,k} - y_{t,m} \), the forward spread, or the CP factor.\(^6\)

The upper panel of Figure 3 uses the yield spread in (3) and (4) and shows the expected pattern in bond return predictability across the business cycle. That is, \( \theta_k^{\text{EXP}} \) is positive in expansions and generally increases with maturity, whereas \( \theta_k^{\text{REC}} \) is negative in recessions and generally decreases with maturity. The middle and lower part of Figure 3 show that the forward spread and the CP factor, respectively, also imply different degrees of predictability for bond returns across the business cycle, with regression slopes mostly having the ‘wrong’ sign in recessions.

To summarize, the return regressions reveal a striking pattern in the slope coefficients across the business cycle. In expansions we find the typical pattern of positive slope coefficients for the yield spread, the forward spread, and the CP factor, while in recessions the slope coefficients are either close to zero or even negative.

### 3.2.4 Robustness

This subsection shows that the reported asymmetric pattern in bond yields during expansions and recessions is robust to a wide range of alternative specifications of our modified Campbell-Shiller regressions and their related return regressions.

#### 3.2.4.1 Recession indicator

The commonly used NBER recession indicator is often announced with a substantial time delay and subject to revisions, unlike recessions identified from the PMI.\(^7\)

\(^6\)The forward spread is \( f_t^{(k-m,k)} = \frac{m}{12} y_{t,m} \), where \( f_t^{(k-m,k)} = \frac{k}{12} y_{t,k} - \frac{k-m}{12} y_{t,k-m} \) is the forward rate between time \( t + k - m \) and \( t + k \). The CP factor is given by \( cp_t = \hat{c}^T F_t \), where \( F_t = \left[ 1, y_{t,12}, f_t^{(12,24)}, f_t^{(24,36)}, f_t^{(36,48)}, f_t^{(48,60)} \right]^T \) and \( \hat{c} \) is obtained by regressing one-year ahead excess bond returns on \( F_t \), i.e. \( \frac{1}{2} \sum_{i=2}^{5} x_{t+12,i} = \gamma^T F_t + \varepsilon_{t+12} \).

\(^7\)Other leading indicators are typically published with a time delay and often undergo considerable revisions following the initial publication.
To examine the relationship between these two recession indicators, the upper part of Figure 4 plots NBER recessions together with the PMI in excess of its threshold value, meaning that negative realizations correspond to recessions. This figure shows a great overlap between the two recession indicators, although the PMI recessions in the 1970s and in 2008 start somewhat later than those of the NBER. To study whether the switch in the Campbell-Shiller regressions is sensitive to this difference, we next re-run our modified Campbell-Shiller regressions in (2) with NBER recessions. The lower part of Figure 4 shows that we obtain exactly the same asymmetric pattern in intercepts and slope estimates as reported in Figure 1, implying that our main finding is robust to using NBER recessions.

3.2.4.2 Forecasting horizon

We have so far used a forecast horizon of three months \((m = 3)\), as this short horizon helps to reduce the potential problem of forecasting into an expansion at the end of a recession, and vice versa. However, as \(m = 6\) and \(m = 12\) are also commonly considered in the literature, we next re-estimate the modified Campbell-Shiller regressions with these slightly longer forecast horizons. Figure 5 clearly shows that our main empirical finding is robust to this modification, as the asymmetric pattern in the Campbell-Shiller coefficients also appears with \(m = 6\) and \(m = 12\). The Online Appendix provides the corresponding Wald tests with \(m = 6\) and \(m = 12\), showing that there also for these alternative forecasting horizons is a significant switch in the intercepts and in the slope coefficients across recessions and expansions.

3.2.4.3 The sample period

It is obviously also interesting to explore whether our main finding is robust across various subsamples. The two most obvious subsamples are probably from 1972:1 to 2013:12, as the panel of Fama-Bliss bond yields is balanced from 1972:1, and from 1983:1 to 2013:12, which corresponds to the post-Volcker period. Table 4 shows that the Wald tests also reject the null hypothesis of coefficient equality across recessions and expansions in these subsamples.

To increase the number of observations where the economy is in recession, we next extend our analysis back to 1926:1 by using the 20-year government bond return series from Ibbotson and the corresponding yield spread. With only one maturity available for this extended sample period, we exploit the one-to-one relation between the Campbell-Shiller regression and the return regression with the yield spread and explore the robustness of our main finding by re-estimating (3) and (4). The PMI is not available back to 1926:1 and we therefore implement (4) with NBER
recessions. Table 5 presents results for the full period from 1926:1 to 2013:12 and for two subsamples before and after 1961:6, implying that the early period from 1926:1 to 1961:5 has no overlap with the sample analyzed so far. Across all three samples, the slope coefficient is significantly more negative in recessions than in expansions and the intercept is significantly more positive. In addition, there is a substantial increase in the $R^2$ by allowing the coefficients to change across expansions and recessions. These results highlight that the systematic difference in bond predictability across the business cycle is not limited to the 1961:6-2013:12 period, although this period is the main focus of our paper.

Accordingly, the asymmetric pattern in bond yields between expansions and recessions reported in Section 3.2.1 does not appear to be restricted to certain subsamples and hence certain monetary policy regimes.

3.2.4.4 Speed of transition, small-sample biases, and out-of-sample evidence

We have also investigated whether the switch in (2) and (4) between recessions and expansions is best modeled using a smoothly changing function or the binary specification used in (2) and (4). In smooth transition regressions with the logistic function, we find that the transition is practically instant. We have also checked that our results are robust to small-sample biases using a stationary bootstrap and find that the differences in recession and expansion coefficients are not an artifact of small-sample biases. In addition, out-of-sample regressions show that all the classic yield-based predictors (i.e., the yield spread, the forward spread, and the Cochrane-Piazzesi factor) only imply significant bond return predictability in expansions. The details of these robustness checks are delegated to our Online Appendix.

3.2.5 Implication for DTSMs

There is a strong conclusion to draw from the above empirical evidence: Bond market asymmetries across the business cycle are substantial and economically important. In particular, the predictive power of the yield spread for future bond yields and bond returns crucially depends on the state of the business cycle. To reproduce the negative slope coefficients in ordinary Campbell-Shiller regressions, most DTSMs require a negative correlation between the short rate and excess returns according to Dai and Singleton (2002). Thus, to explain the sign switch in the slope coefficients within our modified Campbell-Shiller regressions, a successful model should generate a negative relation between the short rate and excess returns in expansions, but a positive relation in recessions.
3.3 A DTSM with a regime-dependent market price of risk

We next explore whether a DTSM can reproduce the documented asymmetries in the U.S. yield curve across the business cycle, while at the same time matching properties of bond risk premia unrelated to the business cycle as given by (1) and (3). Dai and Singleton (2002) show that the Gaussian ATSM can reproduce loadings from ordinary Campbell-Shiller regressions, and this model therefore serves as a natural starting point. We proceed by motivating our extension of the Gaussian ATSM in Section 3.3.1, before formally presenting the proposed model in Section 3.3.2. Estimation of this model by the SR approach is briefly described in Section 3.3.3. We finally relate the proposed DTSM to the existing literature in Section 3.3.4.

3.3.1 Motivation

This section explores whether a further generalization of the market price of risk in the Gaussian ATSM has the potential to reproduce the documented asymmetries in the U.S. yield curve. To introduce our notation for this model, the short rate is given by

\[ r_t = \alpha + \beta' x_t, \]  

where \( \alpha \) is a scalar and \( \beta \) is an \( n_x \times 1 \) vector. The dynamics of the \( n_x \) pricing factors under the risk-neutral measure \( Q \) are specified as

\[ x_{t+1} = (I - \Phi) x_t + \Sigma x \varepsilon_{x,t+1}^Q \]  

with \( \varepsilon_{x,t+1}^Q \sim \mathcal{N}(0, I) \). The \( P \) dynamics follow from an essential affine market price of risk as in Duffee (2002), i.e. \( \lambda_t = \Sigma^{-1}_x (\lambda_0 + \lambda_x x_t) \). In the absence of arbitrage, the price of a zero-coupon bond in period \( t \) with maturity \( k \) is then given by \( P_{t,k} = \exp \{ A_k + B'_k x_t \} \), where the recursive expressions for \( A_k \) and \( B_k \) are easily derived. Provided that the \( P \) distribution for \( x_t \) is stationary, the ordinary Campbell-Shiller coefficients in the Gaussian ATSM are given by (with \( m = 1 \)):  

\[ \beta_k = 1 + \frac{\left( \frac{1}{k} B'_k + \beta' \right) \mathbb{V}[x_t] \lambda_x B_{k-1}}{\left( \frac{1}{k} B'_k + \beta \right) \mathbb{V}[x_t] \left( \frac{1}{k} B_k + \beta \right)} \]  

\[ \varepsilon_{x,t+1}^Q \sim \mathcal{N}(0, I). \]  

\[ \varepsilon_{x,t+1}^Q \sim \mathcal{N}(0, I). \]  

The intercept in the \( Q \) distribution for \( x_t \) is normalized to zero.

The derivations are provided in our Online Appendix.
\[ \alpha_k = -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \beta_k \mathbb{E}\left[ \frac{y_{t,k} - y_{t,1}}{k-1} \right] \]
\[ -\frac{1}{k-1} \left( B'_{k-1} (\lambda_0 + \lambda_x \mathbb{E}[x_t]) + \left( \frac{1}{k} B' + \beta' \right) \mathbb{E}[x_t] \right). \] (8)

Now suppose risk is re-priced across the business cycle in such a way that \( \lambda_x \) has different values in expansions and recessions. According to (7), this modification has the potential to generate negative values of \( \beta_k \) in expansions and positive values in recessions. Equation (8) shows that a switch in \( \lambda_x \) between expansions and recessions also affects the intercept \( \alpha_k \) even if \( \lambda_0 \) is constant. However, it may be necessary to also let \( \lambda_0 \) switch between recessions and expansions to match the observed difference in \( \alpha_k^{REC} \) and \( \alpha_k^{EXP} \).

Thus, considering a regime-dependent market price of risk based on economic activity has the potential to explain the observed asymmetries in the modified Campbell-Shiller regressions or, equivalently, the asymmetric behavior of bond risk premia across the business cycle.

### 3.3.2 Model description

We next formally present a DTSM with a regime-dependent market price of risk based on economic activity. As in the Gaussian ATSM, the dynamics of the short rate under \( \mathbb{Q} \) are given by (5) and (6), implying that zero-coupon bond prices and the yield curve have the same expression as in the Gaussian ATSM, i.e.

\[ y_{t,k} = \tilde{A}_k + \tilde{B}_k x_t, \] (9)

with \( \tilde{A}_k \equiv \frac{1}{k} A_k \) and \( \tilde{B}_k \equiv \frac{1}{k} B_k \) for \( k = 1, 2, ..., K \).

We now deviate from the Gaussian ATSM by assuming that the market price of risk is piece-wise affine in \( x_t \), with loadings depending on whether the economy is in expansion or recession. That is, we let

\[ \lambda_t = 1_{\{z_t \geq c\}} \Sigma^{-1}_x \left( \lambda^{(1)}_0 + \lambda^{(1)}_x x_t \right) + \left( 1 - 1_{\{z_t \geq c\}} \right) \Sigma^{-1}_x \left( \lambda^{(2)}_0 + \lambda^{(2)}_x x_t \right) \] (10)

but assume otherwise the standard expression for the stochastic discount factor, i.e. \( M_{t,t+1} = \exp\left\{ -r_t - \frac{1}{2} \lambda'_t \mathbf{1}_t - \varepsilon_{x,t+1} \lambda_t \right\} \). To discipline our model, the regimes are taken to be observed, whereas previous work mainly considers unobserved regimes. As in Section 3.2, \( z_t \) refers to the PMI and recessions are identified when \( z_t \) is below its threshold value of \( c = 44.5 \). Given that our model remains conditional Gaussian,
a simple change of measure gives the following $\mathbb{P}$ dynamics

$$x_{t+1} = 1\{z_t \geq c\} \lambda_0^{(1)} + \left(1 - 1\{z_t \geq c\}\right) \lambda_0^{(2)} + \left(I - \Phi + 1\{z_t \geq c\} \lambda_x^{(1)} + \left(1 - 1\{z_t \geq c\}\right) \lambda_x^{(2)}\right)x_t + \Sigma_x e_{x,t+1}^P. \tag{11}$$

The model is closed by letting the $\mathbb{P}$ dynamics of $z_t$ evolve as

$$z_{t+1} = \gamma_0 + \gamma_z z_t + \gamma_x' x_t + \Sigma_{zz} e_{z,t+1}^P, \tag{12}$$

where $e_{z,t+1}^P \sim \mathcal{N}(0, 1)$ and is independent of $e_{x,t+1}^P$. That is, $z_t$ may depend on its own lag if $\gamma_z \neq 0$ and lagged values of the pricing factors if $\gamma_x \neq 0_{n_x \times 1}$. The latter implies that yield curve dynamics affect economic activity and hence introduce a feedback effect from financial markets to the real economy. If $\gamma_x = \beta \times \kappa$, where $\kappa$ is some scalar, then (12) reduces to $z_{t+1} = (\gamma_0 - \kappa \alpha) + \gamma_z z_t + \kappa r_t + \Sigma_{zz} e_{z,t+1}^P$, implying that the short rate is a sufficient statistic for economic activity, as assumed in the standard New Keynesian model, see Woodford (2003). In the general case where $\gamma_x \neq 0_{n_x \times 1}$, all yield curve dynamics as captured by the pricing factors $x_t$ may matter for economic activity.\(^\text{10}\) For instance, changes in long-term bond yields may have an independent effect on economic activity beyond variation in the short rate, as considered in some of the recent macroeconomic literature (see, for instance, Andrés, López-Salido and Nelson (2004) and Gerler and Karadi (2013)). Finally, if $\gamma_x = 0_{n_x \times 1}$, then yield curve dynamics do not affect economic activity as assumed in Ang and Piazzesi (2003).

We also note from (11) that the switch in $\lambda_t$ between recessions and expansions generates an instantaneous switch in the pricing factors and hence in the yield curve. This property of our model is thus consistent with the modified Campbell-Shiller and return regressions in Section 3.2.1 and 3.2.3.

An interesting aspect of our model relates to the fact that $z_t$ does not enter in the $\mathbb{Q}$ distribution of the short rate and hence does not appear as a priced factor in (9). This greatly simplifies the model and implies that $z_t$ is a 'hidden' factor as in Duffee (2011), meaning that our model displays unspanned macroeconomic risk similar to the work of Joslin, Priebsch and Singleton (2014). Hence, our macro variable $z_t$ only affects yield curve dynamics indirectly by improving the forecast distribution of the pricing factors. Such improved predictions may be useful when forecasting the yield curve but also when computing expected excess returns or other measures of bond risk premia. Joslin, Priebsch and Singleton (2014) consider a setup where the macro variables enter linearly in the law of motion for the pricing factors under the $\mathbb{P}$ measure, implying that their model may be characterized as displaying 'linear' unspanned macroeconomic risk. Our setup differs slightly from the one considered

\(^{10}\) A similar specification is adopted in Diebold, Rudebusch and Aruoba (2006).
by Joslin, Priebsch and Singleton (2014), because the macro variable $z_t$ in our model has a non-linear effect on the $\mathbb{P}$ distribution for the pricing factors, meaning that our model may be described as having 'nonlinear' unspanned macroeconomic risk.

Finally, given that the $\mathbb{Q}$ distribution of the short rate is identical to the one in the Gaussian ATSM, the proposed model implies the same identifying assumptions as the Gaussian ATSM with latent pricing factors. We therefore require $i)$ $\beta = 1$, $ii)$ $\Phi$ to be diagonal with increasing eigenvalues, and $iii)$ $\Sigma_x$ to be triangular.\(^{11}\) This identification scheme constrains the $\mathbb{Q}$ dynamics for the pricing factors, whereas the $\mathbb{P}$ dynamics are unrestricted to simplify estimation by the SR approach.

### 3.3.3 Model estimation by the SR approach

It is well known that DTSMs may be challenging to estimate, mainly because the parameters describing the market prices of risk can be hard to identify with highly persistent pricing factors. One may therefore encounter numerical instability and problems with local optima when estimating DTSMs (see, for instance, Duffee (2002)). Compared to the Gaussian ATSM, our model has potentially twice the number of parameters for the market prices of risk, suggesting that it may be quite demanding to estimate. To overcome this limitation and avoid problems with numerical instability, we draw on recent innovations in estimation methods for DTSMs, starting with the pioneering work of Joslin, Singleton and Zhu (2011). More specifically, we extend the SR approach of Andreasen and Christensen (2015) to estimate all parameters in the regime-dependent market price of risk by a modified OLS regression, even when allowing for measurement errors in all bond yields, as recommended by Hamilton and Wu (2014). In other words, all parameters in the market price of risk are obtained instantaneously within the SR approach, meaning that our extension of the Gaussian ATSM comes at no additional computational costs.\(^{12}\)

We next describe the SR approach when adopted to the proposed DTSM with a regime-dependent market price of risk. In the interest of space, we only present the three steps in the SR approach and refer to Andreasen and Christensen (2015) for technical details and how to obtain standard errors.

---

\(^{11}\)There exist other normalization schemes; for instance, the one recently suggested by Joslin, Singleton and Zhu (2011).

\(^{12}\)If some linear combinations of bonds yields are assumed to be perfectly priced by the model, the linear regression methods of Joslin, Singleton and Zhu (2011) and Hamilton and Wu (2012) may also be modified to estimate our regime-switching model with no additional computational costs compared to the Gaussian ATSM. However, we prefer the adopted approach because it also enables us to estimate the shadow rate extension of our regime-switching model considered below in Section 3.5.1, whereas the estimators of Joslin, Singleton and Zhu (2011) and Hamilton and Wu (2012) do not apply to nonlinear DTSMs.
Step 1: The first step of the SR approach estimates the latent pricing factors \( \mathbf{x}_t \) and the risk-neutral coefficients by a sequence of cross-section regressions. More formally, the risk-neutral coefficients are denoted by \( \boldsymbol{\theta}_1 \equiv \left[ \theta_{11} \: \theta_{12} \right]' \), where \( \theta_{11} \equiv \left[ \alpha \: \text{vec}(\Phi)' \right]' \) and \( \theta_{12} \equiv \text{vech}(\Sigma_\mathbf{x}) \). Selecting \( n_y \) maturities along the yield curve and accounting for measurement errors \( v_{t,k} \), we then express (9) in stacked form as

\[
y_t = \tilde{\mathbf{A}}_k (\theta_1) + \tilde{\mathbf{B}}_k (\theta_1) \mathbf{x}_t + \mathbf{v}_t,
\]

where \( y_t \equiv \left[ y_{t,1} \: y_{t,2} \: \ldots \: y_{t,n_y} \right]' \) and similarly for \( \tilde{\mathbf{A}}_k (\theta_1), \tilde{\mathbf{B}}_k (\theta_1), \) and \( \mathbf{v}_t \).

When \( n_y \) is large relative to the number of pricing factors, the SR approach estimates \( \mathbf{x}_t \) by minimizing the squared distance between observed and model-implied bond yields, i.e. by regressing \( y_t - \tilde{\mathbf{A}}_k (\theta_1) \) on \( \tilde{\mathbf{B}}_k (\theta_1) \). For our ATSM in (13), this implies

\[
\hat{x}_t (\theta_1) = \left( \tilde{\mathbf{B}}_k (\theta_1)' \tilde{\mathbf{B}}_k (\theta_1) \right)^{-1} \tilde{\mathbf{B}}_k (\theta_1)' \left( y_t - \tilde{\mathbf{A}}_k (\theta_1) \right)
\]

for \( t = 1, 2, \ldots, T \). The estimated factors are denoted \( \{ \hat{x}_t (\theta_1) \}_{t=1}^T \) because they are computed for a given \( \theta_1 \). We then estimate \( \theta_1 \) by pooling all squared residuals from these cross-section regressions and minimize their sum with respect to \( \theta_1 \), i.e.

\[
\hat{\theta}_1^{\text{step1}} = \arg \min_{\theta_1 \in \Theta_1} \frac{1}{T n_y} \sum_{t=1}^T \left\| y_t - \tilde{\mathbf{A}}_k (\theta_1) - \tilde{\mathbf{B}}_k (\theta_1) \hat{x}_t (\theta_1) \right\|^2,
\]

where \( \Theta_1 \) denotes the feasible domain of \( \theta_1 \) and \( \| \mathbf{a} \| \equiv \sqrt{\sum_{i=1}^n a_i^2} \) for any \( \mathbf{a} \in \mathbb{R}^n \).

Step 2: The second step estimates the \( \mathbb{P} \) dynamics of \( z_t \) and \( \mathbf{x}_t \). To describe the procedure, let

\[
\theta_2^x \equiv \left[ \left( \lambda_0^{(1)} \right)' \: \left( \lambda_0^{(2)} \right)' \: \text{vec} \left( \lambda_1^{(1)} \right)' \: \text{vec} \left( \lambda_1^{(2)} \right)' \: \text{vech} \left( \Sigma_\mathbf{x} \right)' \right]'
\]

and \( \theta_2^z \equiv \left[ \gamma_0 \: \gamma_x \: \Sigma_z \right]' \) contain all the parameters governing the \( \mathbb{P} \) dynamics of \( \mathbf{x}_t \) and \( z_t \), respectively. Replacing the unobserved \( \mathbf{x}_t \) in (11) by \( \hat{x}_t (\hat{\theta}_1^{\text{step1}}) \) from the first step, we then estimate \( \theta_2^x \) by extending the SR approach of Andreasen and Christensen (2015) to \( \mathbb{P} \) dynamics with regime-switching. That is, we run a modified regression based on (11) that accounts for estimation uncertainty in \( \hat{x}_t (\hat{\theta}_1^{\text{step1}}) \), as described in Appendix A.4. The elements in \( \theta_2^x \) are estimated in a similar fashion based on (12) using the regression provided in Andreasen and Christensen (2015). Importantly, both \( \left( \theta_2^x \right)^{\text{step2}} \) and \( \left( \theta_2^z \right)^{\text{step2}} \) are given in closed form, meaning that all of the parameters in the \( \mathbb{P} \) distribution of \( z_t \) and \( \mathbf{x}_t \) are obtained instantaneously, including all coefficients in the market price of risk.
To ensure stationarity of $y_{t,k}$ we require $x_t$ to be stationary under the $\mathbb{P}$ measure. This condition holds if the loading matrices on $x_t$ in (11) are stable in recessions and in expansions, i.e. if all eigenvalues of the matrices $I - \Phi + \lambda^{(i)}_x$ for $i = \{1, 2\}$ are inside the unit circle.\(^{13}\) If one of these conditions are not satisfied, then we downscale $I - \Phi + \lambda^{(i)}_x$ by $\delta_i$ for $i = \{1, 2\}$ using the data-driven procedure of Andreasen and Meldrum (2014), which is described in Appendix A.5.

**Step 3:** The matrix $\Sigma_x$ is estimated in both the first and second step. As noted by Andreasen and Christensen (2015), $\hat{\Sigma}_x^{\text{step}1}$ is estimated very inaccurately compared to $\hat{\Sigma}_x^{\text{step}2}$, which therefore is our preferred estimate.\(^{14}\) Given this more efficient estimate of $\Sigma_x$, we then condition on the value of $\hat{\Sigma}_x^{\text{step}2}$ and re-estimate $\theta_{11}$, i.e.

$$\hat{\theta}_{11}^{\text{step}3} = \arg \min_{\theta_{11} \in \Theta_{11}} \frac{1}{T n_y} \sum_{t=1}^{T} \left\| y_t - \tilde{A}_k \left( \theta_{11}; \hat{\Sigma}_x^{\text{step}2} \right) - \tilde{B}_k \left( \theta_{11}; \hat{\Sigma}_x^{\text{step}2} \right) \tilde{x}_t \left( \theta_{11} \right) \right\|^2. $$

Given the estimated factors $\left\{ \hat{x}_t \left( \hat{\theta}_{11}^{\text{step}3}; \hat{\Sigma}_x^{\text{step}2} \right) \right\}_{t=1}^{T}$, we finally update the estimates of $\theta_2^\circ$ and $\theta_2^z$ by re-running step 2.

### 3.3.4 Comparing to existing DTSMs with regime-switching

When formulating DTSMs with regime-switching there is an inherent trade-off between the richness of a given model and the computational complexity related to bond pricing and estimation. We have therefore chosen to consider the most parsimonious model capable of matching the asymmetric properties of bond yields documented in Section 3.2. This implies that we omit regime-switching in $\Sigma_x$ as in Ang, Bekaert and Wei (2008) or variation in $\Sigma_x$ within regimes as in Bansal and Zhou (2002), because time-varying second moments are not required to reproduce the new properties of bond risk premia provided in Section 3.2.\(^{15}\) We also restrict the flexibility of our model by having observed regimes of either expansions or recessions, whereas most DTSMs with regime-switching consider unobserved regimes, although their estimated values often are closely related to the business cycle as in Bansal and Zhou (2002) and Dai, Singleton and Yang (2007). Similar to these two papers, the market price of factor risk in our model is allowed to change freely across regimes, whereas Ang, Bekaert and Wei (2008) consider a somewhat more restricted formulation. As in Dai, Singleton and Yang (2007), we also accommodate

---

\(^{13}\)A formal proof of this result is provided in Theorem 6.12 of Pötscher and Prucha (1997).

\(^{14}\)See Andreasen and Christensen (2015) for how to combine the two estimates in an optimal way.

\(^{15}\)Furthermore, a direct extension of our estimator for $\Sigma_x$ (appearing in Appendix A.4) to accommodate regime-switching reveals only minor changes in the estimated value of $\Sigma_x$ between recessions and expansions.
time-varying transition probabilities between regimes under the $\mathbb{P}$ measure when $\gamma_z \neq 0$ or $\gamma_x \neq 0$, whereas these probabilities are constant in Bansal and Zhou (2002) and Ang, Bekaert and Wei (2008).

### 3.4 Empirical findings

This section estimates our DTSM with regime-switching in the case of three pricing factors ($n_x = 3$), implying a model with four state variables when accounting for $z_t$. Here, we use the same Fama-Bliss dataset as applied in Section 3.2. We first discuss the estimated coefficients in Section 3.4.1, before studying the ability of this regime-switching model to match the observed asymmetry in the U.S. yield curve in Section 3.4.2. The implied estimates of expected excess returns from this regime switching model and the Gaussian ATSM are finally compared in Section 3.4.3.

#### 3.4.1 Estimation results

The proposed ATSM with regime switching in the market price of risk is indexed by $\mathcal{M}^{ATSM}_{\lambda_0, \lambda_x}$, where subscripts indicate whether $\lambda_0$ and/or $\lambda_x$ are allowed to switch between expansions and recessions. Table 6 shows that the first factor is very persistent under the $\mathbb{Q}$ measure with $\Phi(1,1)\approx 1.2 \times 10^{-8}$ in all models and may therefore be interpreted as a level factor. The low value of $\Phi(1,1)$ implies that the intercept in the short rate is unidentified and we therefore let $\alpha = 0$. The second and third factor display less persistence under $\mathbb{Q}$ in all considered models and may therefore be interpreted as slow and fast decaying factors. Unlike $\Phi$, the estimates of $\Sigma_x$ are somewhat affected by regime switching in the market price of risk, although the sign of the off-diagonal elements in $\hat{\Sigma}_x$ is similar across all models. Table 6 also provides the estimated dynamics for economic activity, which display moderate persistence with $\gamma_z = 0.93$ across all specifications of $\lambda_t$. We also find that each of the pricing factors has a negative effect on economic activity. Importantly, the null hypothesis that $\gamma_x(1,1) = \gamma_x(2,1) = \gamma_x(3,1)$ is clearly rejected using a Wald test with a $p$-value of 0.000 for all models, implying that the policy rate is not a sufficient statistic for economic activity according to our model.

Table 7 provides the estimates of the market price of risk. For the Gaussian ATSM $\mathcal{M}^{ATSM}$, all elements of $\lambda_0$ are significantly different from zero, meaning that investors on average require compensation for exposure to each of the pricing factors. We also find the familiar result that time-variation in level risk through $\lambda_x(1,:)$, $\lambda_x(2,1)$, and $\lambda_x(3,1)$ is significant and controls much of the variability in the market price of risk and hence excess returns (see, for instance, Cochrane and
We next allow the constant $\lambda_0$ in the market price of risk to switch in $\mathcal{M}^{ATS\lambda^M}_{\lambda_0}$. During recessions $\lambda_0(2,1)$ has a higher value than in expansions, whereas the opposite holds for $\lambda_0(3,1)$. These differences are significant as we reject the null hypothesis that $\lambda_0^{(1)} = \lambda_0^{(2)}$ using a Wald test ($p$-value of 0.012). Suppose instead that only $\lambda_x$ is allowed to switch in the market price of risk, i.e. $\lambda_0$ is constant as in the Gaussian ATSM. For $\mathcal{M}^{ATS\lambda^M}_{\lambda_x}$, we find that time-variation in the second factor through $\lambda_x(2,2)$ now carries a negative price, although most predominately in recessions, whereas variation in this second factor is priced positively in $\mathcal{M}^{ATS\lambda^M}$ and $\mathcal{M}^{ATS\lambda^M}_{\lambda_0}$. We also observe that $\lambda_x(3,2)$ and $\lambda_x(3,3)$ have opposite signs in recessions and expansions. Despite the somewhat wide standard errors for $\lambda_x^{(2)}$, the null hypothesis that $\lambda_x^{(1)} = \lambda_x^{(2)}$ is clearly rejected ($p$-value of 0.000).

We finally consider the full model $\mathcal{M}^{ATS\lambda^M}_{\lambda_0,\lambda_x}$, where both $\lambda_0$ and $\lambda_x$ are allowed to switch. We once again find that the value of $\lambda_0(2,1)$ changes between recessions and expansions, and that time-variation in the second factor through $\lambda_x(2,2)$ carries a positive price in expansions but a negative price in recessions. The latter implies that the pricing factors and hence bond yields are more persistent in expansions than in recessions when measured by the largest eigenvalue of $I - \Phi + \lambda_x^{(1)}$, which is 0.9957 in expansions and 0.9605 in recessions. This finding is consistent with the results in Bansal and Zhou (2002). However, most elements in $\lambda_0^{(2)}$ and $\lambda_x^{(2)}$ for the recession regime are estimated somewhat imprecisely, and we are therefore unable to reject the null hypothesis of no regime switching in $\lambda_t$ ($p$-value of 0.265), although we reject it for both $\mathcal{M}^{ATS\lambda^M}_{\lambda_0}$ and $\mathcal{M}^{ATS\lambda^M}_{\lambda_x}$. This indicates that our general specification for regime-switching in $\lambda_0$ and $\lambda_x$ may be simplified without affecting the model’s ability to fit the data. We return to this possibility below in Section 3.5.2.

### 3.4.2 Explaining asymmetries in the yield curve

We next examine whether the full model with switching in $\lambda_0$ and $\lambda_x$ can match the asymmetries in the yield curve from Section 3.2. The model-implied moments are here obtained by running the regressions in (1) to (4) on a simulated sample of 100,000 observations using the estimates of $\mathcal{M}^{ATS\lambda^M}_{\lambda_0,\lambda_x}$ reported above.

The top part of Figure 6 shows that $\mathcal{M}^{ATS\lambda^M}_{\lambda_0,\lambda_x}$ convincingly matches both intercepts and slope coefficients in ordinary Campbell-Shiller regressions and thus preserves the ability of the Gaussian ATSM in matching these unconditional properties of bond risk premia. The next question is whether $\mathcal{M}^{ATS\lambda^M}_{\lambda_0,\lambda_x}$ meets the challenge of replicating the switch in the Campbell-Shiller regressions across expansions and recessions. The bottom part of Figure 6 shows that $\mathcal{M}^{ATS\lambda^M}_{\lambda_0,\lambda_x}$ reproduces the negative
slope coefficients in expansions and the positive slope coefficients in recessions, although the latter are somewhat higher than the empirical moments. In addition, $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ matches almost perfectly the switch in intercepts.

We next explore whether $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ also replicates the predictability in bond returns implied by i) the yield spread, ii) the forward spread, and iii) the CP factor. The left part of Figure 7 shows that $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ matches the slope coefficients from ordinary return regressions using each of these predictors. The ability of our model to reproduce the predictability of the CP factor suggests that the behavior of this factor to a large extent is spanned by the three canonical pricing factors in our model.\textsuperscript{16} The right part of Figure 7 further shows that $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ also generates much of the observed asymmetry in return predictability across business cycles for all three predictors. That is, our model implies strong predictability of bond returns during expansions, whereas recessions are characterized by weak predictability and often with the ‘wrong’ sign.

Combining the results from Figure 6 and 7, we conclude that the proposed regime switching model $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ to a large extent is capable of explaining the asymmetric behavior of bond yields documented in Section 3.2.\textsuperscript{17} Thus, the switch in the Campbell-Shiller and return regressions across expansions and recessions may be rationalized by a re-pricing of risk among bond investors when the U.S. economy enters recessions.

### 3.4.3 Model-implied excess returns

The presence of regime-switching in the market price of risk has profound implications for bond risk premia. Figure 8 therefore plots expected excess returns from $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ and $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ at the 10-year maturity, with NBER recessions indicated by the shaded regions. The two models provide similar dynamics for expected excess returns in expansions. However, while expected excess returns from $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ tend to increase during the course of a recession, expected excess returns from $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ tend to peak at the start of a recession and then decrease during the middle or last phase of a recession.

Any estimate of bond risk premia is model-dependent, and it may therefore in general be challenging to argue in favor of one set of estimates compared to another. However, the regression evidence we provide in Figures 6 and 7 clearly indicates that

\textsuperscript{16}A similar finding is reported in Dai, Singleton and Yang (2004).

\textsuperscript{17}We have also analyzed the ability of $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ and $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ to match the observed asymmetric patterns in the data. $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ matches the sign switch in intercepts but not in the slope coefficients of our modified Campbell-Shiller regressions and it fails to generate low return predictability in recessions. $\mathcal{M}_{\lambda_0,\lambda_x}^{ATSM}$ does not match neither intercepts nor slope coefficients in the modified Campbell-Shiller regressions and it generates too high return predictability in expansions.
our regime-switching model matches important properties of bond returns unlike the Gaussian ATSM, suggesting that $\mathcal{M}^{ATSM}_{\lambda_0 \lambda_x}$ gives more accurate estimates of bond risk premia compared to $\mathcal{M}^{ATSM}$. Another way to illustrate this is to run Mincer-Zarnowitz regressions of realized excess returns on model-implied expected excess returns $\mathbb{E}_t [xhpr_{t+m,k}]$. Conditioning on the business cycle as in Section 3.2, we consider

$$xhpr_{t+m,k} = \delta_k^{\text{EXP}} (1 - 1_{\{NBER\}}) + \gamma_k^{\text{EXP}} (1 - 1_{\{NBER\}}) \mathbb{E}_t [xhpr_{t+m,k}] + \delta_k^{\text{REC}} 1_{\{NBER\}} + \gamma_k^{\text{REC}} 1_{\{NBER\}} \mathbb{E}_t [xhpr_{t+m,k}] + u_{t+m,k};$$

with $m = 3$, where recessions and expansions are identified using the NBER dummy $1_{\{NBER\}}$ for comparability with Figure 8. The slope coefficients $\gamma_k^{\text{EXP}}$ and $\gamma_k^{\text{REC}}$ are reported in Table 8 when pooling excess returns in bins of two years along the maturity range. Consistent with the regression evidence in Figure 6 and 7, $\mathcal{M}^{ATSM}$ performs well in expansions with $\gamma_k^{\text{EXP}}$ close to the desired value of one, but the model struggles during recessions where $\gamma_k^{\text{REC}}$ is substantially below one and even negative for bond yields with six or more years to maturity. That is, the Gaussian ATSM predicts excess returns for medium- and long-term bonds in the wrong direction during recessions. Our regime switching model largely alleviates this shortcoming with estimates of both $\gamma_k^{\text{EXP}}$ and $\gamma_k^{\text{REC}}$ in the range from 0.54 to 0.76, meaning that $\mathcal{M}^{ATSM}_{\lambda_0 \lambda_x}$ predicts excess returns of all bond yields in the right direction during expansions and recessions.

### 3.5 Additional analysis

This section studies in greater detail the ability of regime-switching in the market price of risk to explain the documented asymmetries in the U.S. yield curve from Section 3.2. We first consider the implications of extending $\mathcal{M}^{ATSM}_{\lambda_0 \lambda_x}$ to enforce the zero lower bound (ZLB) in Section 3.5.1. Given our findings in Section 3.4.1, we then present a simplified version of $\mathcal{M}^{ATSM}_{\lambda_0 \lambda_x}$ in Section 3.5.2, which we use to provide an economic explanation for the observed switch in the Campbell-Shiller regressions across expansions and recessions. The ability of DTSMs with linear unspanned and spanned macroeconomic risk to match loadings from the modified Campbell-Shiller regressions are finally explored in Section 3.5.3.

#### 3.5.1 Accounting for the Zero Lower Bound

Given that short rates typically are low during recessions and may be constrained by the ZLB, it seems natural to explore whether the better performance of our
model is robust to enforcing the ZLB. We address this question by presenting a shadow rate extension of our regime-switching model where the short rate is given by 
\[ r_t = \max \{ 0, \alpha + \beta' x_t \} \]
but the model is otherwise identical to the one described in Section 3.3.2. Hence, the proposed shadow rate model (SRM) has Gaussian pricing factors under the \( Q \) measure and bond prices may therefore be computed by the second-order approximation advocated in Priebsch (2013) when formulated in discrete time. This SRM with regime-switching \( M_{\lambda_0, \lambda_x}^{SRM} \) is estimated as described in Section 3.3.3, except (14) is replaced by nonlinear cross-section regressions to extract the pricing factors. The estimates are provided in Tables 9 and 10.

Figure 9 shows that \( M_{\lambda_0, \lambda_x}^{SRM} \) provides an extremely close fit to intercepts in ordinary Campbell-Shiller regressions and even improves upon the slope coefficients in these regressions for medium- and long-term bond yields compared to \( M_{\lambda_0, \lambda_x}^{ATSM} \). This improvement is further seen to carry over to the slope coefficients during expansions, whereas \( M_{\lambda_0, \lambda_x}^{SRM} \) largely generates the same intercepts and slope coefficients for the recession regime as found for \( M_{\lambda_0, \lambda_x}^{ATSM} \). Although not reported below, we also find that \( M_{\lambda_0, \lambda_x}^{SRM} \) matches the return regressions just as well as seen for \( M_{\lambda_0, \lambda_x}^{ATSM} \). Finally, \( M_{\lambda_0, \lambda_x}^{SRM} \) also forecasts realized excess returns in the correct direction during both expansions and recessions, as shown in Table 8.

Thus, the proposed explanation for the documented asymmetries in the U.S. yield curve, i.e. that investors re-price risk during recessions, is robust to accounting for the ZLB.

### 3.5.2 A simplified regime-switching model

To provide a simplified version of \( M_{\lambda_0, \lambda_x}^{ATSM} \) with a clear economic interpretation, it seems beneficial to first impose more structure on the factor loadings than implied by our fully flexible regime-switching model \( M_{\lambda_0, \lambda_x}^{ATSM} \). We therefore first note that the null hypothesis of \( \Phi(1,1) = 0 \) and \( \Phi(2,2) = \Phi(3,3) \) is not rejected using a Wald test (\( p \)-value of 0.641). Given these restrictions, it is straightforward to show that the factor loadings in \( M_{\lambda_0, \lambda_x}^{ATSM} \) simplify to those in the arbitrage-free Nelson-Siegel (AFNS) model of Christensen, Diebold and Rudebusch (2011) when using its discrete-time formulation in Fontaine and Garcia (2012). That is, we let
\[
\beta = \begin{bmatrix} 1 & 1-e^{-\lambda} & 1-e^{-\lambda} - e^{-\tilde{\lambda}} \end{bmatrix}'
\]
and impose \( \Phi(2,2) = \Phi(3,3) = 1 - e^{-\lambda} \) with \( \Phi(2,3) = -\tilde{\lambda}e^{-\tilde{\lambda}} \), whereas all the remaining elements of \( \Phi \) are zero. Given this specification, the pricing factors now have the familiar interpretation as representing the level, slope, and curvature of the yield curve.

The results from Section 3.4.1 indicate that not all elements in \( \lambda_0 \) and \( \lambda_x \) switch between expansions and recessions, and we therefore search for the most parsimo-
nious specification of $\lambda_t$ capable of matching the switch in the Campbell-Shiller regressions across the business cycle. Inspired by the return regressions with the yield spread in Section 3.2.3, the proposed model displays only switching in $\lambda_0(2,1)$ and $\lambda_x(2,2)$, both related to the slope factor at position two in $x_t$. This restricted model is denoted $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$ and has only two additional coefficients in the market price of risk compared to the Gaussian ATSM. The top part of Figure 10 shows that $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$ captures the overall pattern in intercepts and slope coefficients from ordinary Campbell-Shiller regressions. More surprising, perhaps, is the ability of this very parsimonious regime-switching specification to also capture the switch in both intercepts and slope coefficients in the Campbell-Shiller regressions between expansions and recessions. Thus, simply allowing for regime-switching in the dynamics of the slope factor is sufficient to reproduce the loadings in the modified Campbell-Shiller regressions.

To see what drives these results for $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$, we first note from Table 10 that $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$ has a lower value of $\lambda_0(2,1)$ in recessions than in expansions. A standard Wald test for a switch in $\lambda_0(2,1)$ has a $p$-value of 0.069 and is thus significant at the 10% level. We also note that $\lambda_x(2,2)$ switches sign between the two regimes in $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$. The latter arises because $1 - \Phi(2,2) + \lambda_x^{(1)}(2,2) = 0.97$ in expansions, whereas the corresponding estimate in recessions is 0.73. Given that $\Phi(2,2) = 0.0381$ in $M_{AFNS}^{\lambda_0(2,1)\lambda_x(2,2)}$, this change in the persistence of the slope factor then generates a switch in the market price of risk from an insignificant positive value of $\lambda_x^{(1)}(2,2) = 0.0081$ in expansions to a significant negative value of $\lambda_x^{(2)}(2,2) = -0.2363$ in recessions. Furthermore, a standard Wald test for a switch in $\lambda_x(2,2)$ has a $p$-value of 0.036 and is thus significant at the 5% level.

To analyze the effects of these estimates for quarterly excess returns, as considered in our implementation of the Campbell-Shiller regressions in (1) and (2), it is useful to first understand their impact on monthly excess returns which are available in closed form and proportional to $B_{k-1}\lambda_t$. Given that $B_{k-1}(2,1) < 0$, a more negative value of $\lambda_0(2,1)$ in recessions therefore increases monthly excess returns, as investors require a larger compensation for slope risk in this regime. On the other hand, a negative value of $\lambda_x^{(2)}(2,2)$ in recessions has a positive effect on $\lambda_x^{(2)}x_t$ with an upward sloping yield curve, because the slope factor with the AFNS loadings is defined as the short rate minus the long rate (see Diebold and Li (2006)). When accounting for $B_{k-1}(2,1) < 0$, we thus conclude that $\lambda_x^{(2)}(2,2)$ in total has a negative impact on monthly excess returns. Figure 11 shows that these partial effects carry over to the quarterly excess return for a 10-year bond which we compute by Monte Carlo integration. Note also how fears of recessions on its own may affect expected quarterly excess returns (as in the mid 1990s). We finally emphasize that

\[18\] Here, we ignore a minor convexity term in the expression for monthly excess returns. The exact expression is provided in our Online Appendix.
\( \mathcal{M}_{\lambda_0(2,1)\lambda_x(2,2)}^{AFNS} \) predicts realized excess returns with the right sign during expansions and recessions according to the Mincer-Zarnowitz regressions in Table 8, despite \( \mathcal{M}_{\lambda_0(2,1)\lambda_x(2,2)}^{AFNS} \) only having two parameters that switch between regimes.

Thus, the narrative implied by our model is as follows. When the U.S. economy enters a recession, bond investors immediately re-price the risks attached to economic activity with a switch in \( \lambda_0(2,1) \) and therefore require a higher compensation for exposure to the slope of the yield curve. At the start of a recession, this effect generally dominates the one from a switch in \( \lambda_x(2,2) \) according to Figure 11, as the yield curve here tends to be fairly flat, and this increases expected excess returns. However, as monetary policy becomes more accommodating during the course of a recession, the lower short rate generates a steepening of the yield curve and reduces the risks attached to low future economic activity. This in turn strengthens the effect from the switch in \( \lambda_x(2,2) \), which generally has a negative impact on excess returns according to Figure 11. Thus, our model suggests that the Federal Reserve is able to remove some of the risks attached to recessions, as accommodating monetary policy reduces the required risk compensation in the bond market. Importantly, the switch in \( \lambda_x(2,2) \) from positive in expansions to negative in recessions implies that excess returns and the short rate both fall during this phase of a recession and hence become positively correlated, as required to match the positive slope coefficients in the Campbell-Shiller regressions during recessions. The Gaussian ATSM does not allow for a switch in \( \lambda_x(2,2) \), meaning that this model does not imply that accommodating monetary policy is able to remove some of the risks in the bond market during recessions. As a result, excess returns tend to increase throughout recessions in the Gaussian ATSM, as also seen in Figure 8. After a recession, \( \mathcal{M}_{\lambda_0(2,1)\lambda_x(2,2)}^{AFNS} \) reverts back to the familiar setting described in Dai and Singleton (2002), where excess returns fall when the short rate increases, and vice versa. That is, excess returns and the short rate are once again negatively correlated as required to match the negative slope in the Campbell-Shiller regressions during expansions.

### 3.5.3 Models with linear unspanned and spanned macroeconomic risk

The proposed mechanism to explain the documented asymmetries in U.S. bond yields modifies the essentially affine specification for the market price of risk such that it only holds locally for expansions and recessions. Hence, \( \lambda_t \) depends non-linearly on the business cycle, as captured by the state variable \( z_t \) for PMI. To explore whether this nonlinearity is truly necessary to match the switch in the Campbell-Shiller regressions, we next follow Joslin, Priebsch and Singleton (2014) and briefly study the case where \( \lambda_t \) only depends linearly on \( z_t \). That is, we now let the market price of risk be essentially affine in all four state variables. As above, \( z_t \) remains
unspanned by the current yield curve, i.e. $\beta(4,1) = 0$, given our specification with $z_t$ appearing last in the state vector. The estimates for this model $M^{\text{Linear}}$ are provided in Tables 9 and 11.

Figure 12 shows that $M^{\text{Linear}}$ provides a very close fit to both intercepts and slope coefficients in ordinary Campbell-Shiller regressions. This is in line with previous findings in the literature, as this model nests the standard three-factor Gaussian ATSM. The lower part of Figure 12 reveals that $M^{\text{Linear}}$ also matches the regime-dependent intercepts in the Campbell-Shiller regressions but only the negative slope coefficients in expansions. That is, $M^{\text{Linear}}$ cannot generate positive slope coefficients during recessions. Table 8 further shows that $M^{\text{Linear}}$ only has predictive power for realized excess returns in expansions but not in recessions. Hence, it must be the nonlinear effect of $z_t$ on the market price of risk that allows $M^{\text{ATSM}}$, $M^{\text{SRM}}$, and $M^{\text{AFNS}}$ to match the slope coefficients in our modified Campbell-Shiller regression and predict excess returns with the right sign during expansions and recessions.

Following the work of Bauer and Rudebusch (2016), we finally explore whether these results for $M^{\text{Linear}}$ are robust to letting $z_t$ be spanned by the current yield curve. That is, we once again let the market price of risk be essentially affine in all four state variables but omit the zero restriction for $\beta(4,1)$. Figure 12 shows that this model $M^{\text{Spanned}}$ displays broadly the same performance as $M^{\text{Linear}}$ in terms of matching intercepts and slope coefficients from ordinary and modified Campbell-Shiller regressions. As a result, the Mincer-Zarnowitz regressions for excess returns in Table 8 are also very similar for the two models. Hence, the spanning assumption of macroeconomic risk, as measured by the PMI, does not seem essential for the aspects of U.S. bond yields studied in this paper.

### 3.6 Conclusion

By conditioning the classic Campbell-Shiller regressions on the business cycle, we identify a strong asymmetric pattern in the relation between yield spreads and subsequent yield changes. When the economy is expanding, we find the familiar pattern of mostly positive intercepts and negative regression slopes that decrease with maturity. However, when the economy is contracting, we observe negative intercepts and positive regression slopes that mostly increase with maturity, i.e. the complete opposite pattern. We also show that this asymmetric effect has profound implications for bond return predictability, as the classic yield-based predictive variables have strong forecasting power for excess returns in expansions but not in recessions.
To explain these new empirical findings, the Gaussian ATSM is extended with business cycle dependent loadings in the market price of risk. We show that this model reproduces the empirical intercepts and slope coefficients from ordinary Campbell-Shiller regressions and, in addition, matches evidence from our modified Campbell-Shiller regressions conditioning on the business cycle. We also show that our model predicts realized excess returns with the right sign during both expansions and recessions unlike the Gaussian ATSM, which predicts excess returns for medium- and long-term bonds in the wrong direction during recessions. Our model also replicates the observed asymmetry in return predictability, i.e. that the classic yield-based predictors have strong forecasting power in expansions but not in recessions. To account for these asymmetric patterns in the yield curve across the business cycle, our model generates a negative relation between the short rate and excess returns in expansions (as in the Gaussian ATSM), but a positive relation in recessions. Thus, our model suggests that the Federal Reserve is able to remove some of the risks attached to recessions, as accommodating monetary policy reduces the required risk compensation in the bond market. This effect of monetary policy is not present in the Gaussian ATSM, which therefore provides less accurate estimates of bond risk premia and expected future short rates compared to our regime-switching model. Accordingly, our model suggests that the positive slope coefficients in the Campbell-Shiller regressions emerge because accommodating monetary policy is able to eliminate some of the risks in the bond market during recessions.

Our simple, yet powerful, approach of conditioning the study of bond markets on the business cycle obviously goes beyond what is covered in the present paper. Ongoing research by Andreasen, Møller and Sander (2016) shows that bond markets in several other countries display a similar switch in the Campbell-Shiller regressions as found in the present paper, both when conditioning on local recessions and U.S. recessions. Andreasen, Møller and Sander (2016) further show that this finding has important implications for exchange rate dynamics, as the slope coefficients in uncovered interest parity (UIP) regressions also switch between expansions and recessions. Thus, recessions may not only have profound effects for the local U.S. bond investor but also for other bond markets and related exchange rate markets.

Acknowledgements

We thank Jens Christensen, John Cochrane, Hans Dewachter, James Hamilton, Leonardo Iania, Robin Lumsdaine, and Thomas Pedersen for useful comments and discussions. Remarks and suggestions from seminar participants at Louvain School of Management (CORE) are also much appreciated. We acknowledge access to computer facilities provided by the Danish Center for Scientific Computing (DCSC).
We also acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.
References


Gerler, M., Karadi, P., 2013. QE 1 vs. 2. vs. 3. ...: A framework for analyzing large-scale asset purchases as a monetary policy tool. International Journal of Central Banking 9 (S1), 5–53.


Appendix

A.1 Data

We use monthly bond yields from an unsmoothed Fama and Bliss (1987) dataset running from 1961:06-2013:12. We limit the analysis to maturities from \( m/12 \) to 10 years with \( m \)-month increments where \( m = 3 \). A similar panel of yields is used in Adrian, Crump and Moench (2013) to obtain their pricing factors. At the very long end of the yield curve, we do not observe all increments and, hence, we interpolate between bond yields of the two nearest maturities.\(^{19}\) We also note that some long-term bond yields are unavailable from 1961:6, as maturities \( k = \{69, 72\} \) start in 1967:11, \( k = \{75, 78\} \) start in 1969:2, \( k = \{81, 84, ..., 114\} \) start in 1971:8, and \( k = \{117, 120\} \) start in 1971:11. We also report results for the Gürkaynak, Sack and Wright (2007) dataset, where bond yields in the \( 7^+ \) to 10-year maturity spectrum from 1961:6 to 1971:8 are calculated by extrapolation of their estimated curves.

A.2 Wald tests

We use Wald tests to examine whether the coefficients in our modified Campbell-Shiller regressions in (2) change across expansions and recessions. In particular, we test the following joint hypotheses:

\[
H_0: \alpha_{REC}^k = \alpha_{EXP}^k, \quad k = 6, 9, 12, ..., 120
\]

\[
H_0: \beta_{REC}^k = \beta_{EXP}^k, \quad k = 6, 9, 12, ..., 120
\]

The tests are carried out by setting up the regressions in a GMM framework, where all parameters are collected in

\[
\Theta' = [\theta'_6, \theta'_9, \ldots, \theta'_{120}]
\]

with \( \theta'_k = [\alpha_k^{EXP}, \beta_k^{EXP}, \alpha_k^{REC}, \beta_k^{REC}] \). The sample moments for the system of linear regressions are

\[
g_T(\Theta)' = [g_{T,6}', g_{T,9}', \ldots, g_{T,120}']
\]

where \( g_{T,k} = \frac{1}{T} \sum_{t=1}^{T} f_{t,k}, \quad f_{t,k} = \bar{u}_{t+m,k} z_{t,k} \), and

\(^{19}\)In addition, the unsmoothed Fama-Bliss forward rates are known to be somewhat rough and we therefore smooth them using a 12-month equal-weighted window.
Expressing our null hypothesis as \( \mathbf{R}\Theta = \mathbf{c} \), a Wald test is then given by

\[
W = \left( \mathbf{R}\hat{\Theta} - \mathbf{c} \right) \left( \mathbf{R} \left( \mathbf{D}_T^\prime \mathbf{S}_T^{-1} \mathbf{D}_T \right)^{-1} \mathbf{R}^\prime / T \right)^{-1} \left( \mathbf{R}\hat{\Theta} - \mathbf{c} \right) \sim \chi^2_K,
\]

where \( \mathbf{D}_T \) is the Jacobian of the moment conditions and \( \mathbf{S}_T \) is the spectral density matrix obtained by the Newey-West estimator using \( m \) lags. For the hypothesis in (16), we have

\[
\mathbf{R} = \mathbf{I}\mathbf{D}_K \otimes [1, 0, -1, 0],
\]

where \( \mathbf{I}\mathbf{D}_K \) is the identity matrix with dimension \( K \) and \( \mathbf{c} = \mathbf{0} \). The integer \( K \) refers to the number of maturities we examine and is equal to 39 for the full panel. For the hypothesis in (17), we have

\[
\mathbf{R} = \mathbf{I}\mathbf{D}_K \otimes [0, 1, 0, -1],
\]

and \( \mathbf{c} = \mathbf{0} \).

### A.3 Lynch-Wachter estimator

Conducting the tests for the full sample period involves making some implementation choices as the Fama-Bliss panel is unbalanced. We use the adjusted moments estimator of Lynch and Wachter (2013), which is an efficient way to explore all the information in the long sample. For simplicity we partition the moments in two parts: One part where we have bond yields for the full period from 1961:06-2013:12 and one part where we only have bond yields starting from 1971:11. The first group contains moments on maturities up to 66 months, while the second group contains moments related to bond yields with maturities from 69 to 120 months. In principle, we could make many partitions based on the available observations for each moment, but this will heavily complicate the computations. We refer to Lynch and Wachter (2013) and our Online Appendix for further details.

### A.4 Step 2 of the SR approach: regime-switching in the time series regression

This subsection describes how to estimate (11) by GMM when accounting for measurement errors in the estimated pricing factors. For notational convenience, we
express (11) as

$$x_{t+1} = 1_{\{z_t \geq c\}} h_0^{(1)} + (1 - 1_{\{z_t \geq c\}}) h_0^{(2)} + h_x^{(1)} x_t^{(1)} + h_x^{(2)} x_t^{(2)} + w_{t+1},$$

where $h_0^{(1)} \equiv \lambda_0^{(1)}, h_x^{(i)} \equiv I - \Phi + \lambda_x^{(i)}$, $x_t^{(1)} \equiv 1_{\{z_t \geq c\}} x_t$, $x_t^{(2)} \equiv (1 - 1_{\{z_t \geq c\}}) x_t$, and $w_{t+1} \equiv \sum x_t \epsilon_{t+1}$ for $i = \{1, 2\}$. Using the same procedure as in Andreasen and Christensen (2015), we estimate $\theta_x^2$ based on

$$\begin{bmatrix}
E \left[ \hat{\omega}_{t+1} 1_{\{z_t \geq c\}} \right] \\
E \left[ \hat{\omega}_{t+1} (\hat{x}_t^{(1)})' \right] \\
E \left[ \hat{\omega}_{t+1} (\hat{x}_t^{(2)})' \right] \\
\text{Var} (\hat{\omega}_{t+1})
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
\text{Cov} (u_{t+1}, u_t^{(1)}) - h_x^{(1)} \text{Var} (u_t^{(1)}) \\
\text{Cov} (u_{t+1}, u_t^{(2)}) - h_x^{(2)} \text{Var} (u_t^{(2)}) \\
\text{Var} (w_{t+1}) + \Omega_{t+1}
\end{bmatrix},$$

where

$$\Omega_{t+1} \equiv \text{Var} (u_{t+1}) + h_x^{(1)} \text{Var} (u_t^{(1)}) (h_x^{(1)})' + h_x^{(2)} \text{Var} (u_t^{(2)}) (h_x^{(2)})'$$

$$- \text{Cov} (u_{t+1}, u_t^{(1)}) (h_x^{(1)})' - h_x^{(1)} \text{Cov} (u_t^{(1)}, u_{t+1})$$

$$- \text{Cov} (u_{t+1}, u_t^{(2)}) (h_x^{(2)})' - h_x^{(2)} \text{Cov} (u_t^{(2)}, u_{t+1}).$$

Here, $u_t$ refers to the estimation uncertainty in the estimated pricing factors, i.e. $\hat{x}_t = x_t + u_t$, where $u_t^{(1)} \equiv 1_{\{z_t \geq c\}} u_t$ and $u_t^{(2)} \equiv (1 - 1_{\{z_t \geq c\}}) u_t$. As in Andreasen and Christensen (2015), all required moments of $u_t$ follow from the first step of the SR approach. The solution to the first four moments conditions in (18) is given in closed form by

$$\begin{bmatrix}
\hat{h}_0^{(1)} \\
\hat{h}_0^{(2)} \\
\hat{h}_x^{(1)} \\
\hat{h}_x^{(2)}
\end{bmatrix}
= \left( \sum_{t=1}^{T-1} \hat{x}_{t+1} a_t' - \sum_{t=1}^{T-1} \hat{A}_{t+1} \right) \left( \sum_{t=1}^{T-1} a_t a_t' - \sum_{t=1}^{T-1} \text{Var} (u_{a,t}) \right)^{-1},$$

where

$$a_t = \begin{bmatrix}
1_{\{z_t \geq c\}} \\
1 - 1_{\{z_t \geq c\}} \\
\hat{x}_t^{(1)} \\
\hat{x}_t^{(2)}
\end{bmatrix}, \quad u_{a,t} = \begin{bmatrix}
0 \\
0 \\
u_t^{(1)} \\
u_t^{(2)}
\end{bmatrix}$$

and $\hat{A}_{t+1} = \begin{bmatrix}
0 & 0 & \text{Cov} (u_{t+1}, u_t^{(1)}) & \text{Cov} (u_{t+1}, u_t^{(2)})
\end{bmatrix}$. The solution to the last moment condition in (18) is also given in closed form and implies the following estimator

$$\text{Var} (w_{t+1}) = \frac{1}{T - 1 - 2 (n_x + 1)} \sum_{t=1}^{T-1} \hat{w}_{t+1} \hat{w}_{t+1}' - \frac{1}{T - 1} \sum_{t=1}^{T-1} \hat{\Omega}_{t+1},$$

(19)
where

\[
\hat{\Omega}_{t+1} = \bar{Var}(u_{t+1}) + \hat{\xi}_{t} \bar{Var}(u^{(1)}_{t}) \left( \hat{\lambda}^{(1)}_{x} \right) \left( \hat{\lambda}^{(2)}_{x} \right) \quad
\]

\[
- \bar{Cov}(u^{(1)}_{t+1},u^{(1)}_{t}) \left( \hat{\lambda}^{(1)}_{x} \right) \quad
\]

\[
- \bar{Cov}(u^{(2)}_{t+1},u^{(2)}_{t}) \left( \hat{\lambda}^{(2)}_{x} \right) \quad
\]

\[
\text{with } \hat{w}_{t+1} = \hat{x}_{t+1} - 1_{\{z_{t+1} \geq c\}} \hat{\lambda}^{(1)}_{0} - (1 - 1_{\{z_{t+1} \geq c\}}) \hat{\lambda}^{(2)}_{0} \hat{x}^{(1)}_{t} - \hat{\lambda}^{(2)}_{x} \hat{x}^{(2)}_{t}. \quad
\]

Note that we adopt a standard degree of freedom correction to the first term in (19) because we estimate 2 \((n_{x} + 1)\) unknown parameters per equation in the model. The asymptotic distribution of \(\theta^{x}_{2}\) for \(T \rightarrow \infty\) follows from Hansen (1982) when applied on the moment conditions in (18). The estimated loadings in the market prices of risk are then given by \(\hat{\lambda}^{(1)}_{0} = \hat{\lambda}^{(i)}_{0}\) and \(\hat{\lambda}^{(2)}_{0} = \hat{\lambda}^{(i)}_{x} - (I - \hat{\Phi})\) for \(i = \{1, 2\}\).

Finally, the standard errors for \(\hat{\lambda}^{(1)}_{0}\) and \(\hat{\lambda}^{(2)}_{0}\) are identical to those for \(\hat{\lambda}^{(i)}_{0}\) and \(\hat{\lambda}^{(i)}_{x}\), respectively. That is, we omit uncertainty about \(\hat{\Phi}\), because this estimator uses \(Tn_{y}\) observations and therefore tends faster to infinity than \(\hat{\theta}^{x}_{2}\) when also \(n_{y} \rightarrow \infty\), as noted in Andreasen and Christensen (2015).

### A.5 Step 2 of the SR Approach: inducing stationarity

If the stability condition for \(x_{t}\) is not satisfied, then \(I - \Phi + \lambda^{(i)}_{x}\) is downscaled by \(\delta_{i}\) for \(i = \{1, 2\}\) if the eigenvalues of \(I - \Phi + \lambda^{(i)}_{x}\) are greater than or equal to one. The values of \(\delta_{1}\) and \(\delta_{2}\) are determined as in Andreasen and Meldrum (2014), i.e. by

\[
(\delta_{1}, \delta_{2}) = \arg \min_{\{\delta_{\text{lower}} \leq \delta_{i} < 1\}^{2}} \sum_{i=1}^{n_{x}} \left( \frac{\sigma^{2}_{t, \text{model}}(\delta_{1}, \delta_{2}) - \sigma^{2}_{t, \text{sample}}}{\sigma^{2}_{t, \text{sample}}} \right)^{2}.
\]

We follow Andreasen and Meldrum (2014) and estimate the unconditional variance of the \(i\)th pricing factor in the sample from \(\hat{x}_{i,t}\) to

\[
\hat{\sigma}^{2}_{i, \text{sample}} = \frac{1}{T - 1} \sum_{t=1}^{T} \left( \hat{x}_{i,t} - \hat{E}[\hat{x}_{i}] \right)^{2} - \frac{1}{T} \sum_{t=1}^{T} \bar{Var}(u_{i,t}),
\]

where \(\hat{E}[\hat{x}_{i,t}] = 1/T \sum_{t=1}^{T} \hat{x}_{i,t}\) and \(\bar{Var}(u_{i,t})\) refers to the estimated variance of \(\hat{x}_{i,t}\). The value of the unconditional variance of \(x_{i,t}\) in the model \(\sigma^{2}_{t, \text{model}}(\delta_{1}, \delta_{2})\) is computed by simulation, using

\[
x_{t+1} = 1_{\{z_{t} \geq c\}} \lambda^{(1)}_{0} + (1 - 1_{\{z_{t} \geq c\}}) \lambda^{(2)}_{0} + (\delta_{1} 1_{\{z_{t} \geq c\}} \{I - \Phi + \lambda^{(1)}_{x}\}) + \delta_{2} (1 - 1_{\{z_{t} \geq c\}}) \{I - \Phi + \lambda^{(2)}_{x}\}) x_{t} + \Sigma x \epsilon_{x,t+1}^{P}
\]

and (12).
Table 1: Statistical tests in ordinary and modified Campbell-Shiller regressions: Fama-Bliss yields

The table reports results from ordinary and modified Campbell-Shiller regressions, where the latter are estimated using $y_{t+m,k-m} - y_{t,k} = \alpha_k^{EXP} + \alpha_k^\Delta 1_{\{z_t<\epsilon\}} + (\beta_k^{EXP} + \beta_k^\Delta 1_{\{z_t<\epsilon\}}) \frac{y_{t,k} - y_{t,m}}{m} + \hat{u}_{t+m,k}$ with $m = 3$ and PMI to identify recessions. The reported intercepts are multiplied by 100. In parentheses are t-statistics obtained from Newey-West standard errors computed with $m$ lags. Significance at the 10 and 5 percent level is denoted by * and **, respectively. The estimation is carried out on monthly data from 1961:6 to 2013:12 using Fama-Bliss (FB) bond yields, except for some long-term yields where the estimation starts somewhat later due to lack of data availability (see Appendix A.1).

<table>
<thead>
<tr>
<th>$k$</th>
<th>No switch</th>
<th>Recessions vs. expansions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_k$</td>
<td>$\beta_k$</td>
</tr>
<tr>
<td>24</td>
<td>0.05</td>
<td>-1.17*</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
<td>(-1.84)</td>
</tr>
<tr>
<td>48</td>
<td>0.08</td>
<td>-2.17**</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(-2.74)</td>
</tr>
<tr>
<td>72</td>
<td>0.08</td>
<td>-2.43**</td>
</tr>
<tr>
<td></td>
<td>(1.18)</td>
<td>(-2.51)</td>
</tr>
<tr>
<td>96</td>
<td>0.07</td>
<td>-2.57**</td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
<td>(-2.35)</td>
</tr>
<tr>
<td>120</td>
<td>0.08</td>
<td>-3.12**</td>
</tr>
<tr>
<td></td>
<td>(1.46)</td>
<td>(-2.83)</td>
</tr>
</tbody>
</table>
Table 2: Statistical tests in ordinary and modified Campbell-Shiller regressions: Gürkaynak-Sack-Wright yields
The table reports results from ordinary and modified Campbell-Shiller regressions, where the latter are estimated using:

\[ y_{t+m,k-m} - y_{t,k} = \alpha_k^{EXP} + \alpha_k^{\Delta 1_{\{z_t<0\}}} + (\beta_k^{EXP} + \beta_k^{\Delta 1_{\{z_t<0\}}}) m \frac{y_{t,k}}{m} (y_{t,k} - y_{t,m}) + \tilde{u}_{t+m,k} \]

with \( m = 3 \) and PMI to identify recessions. The reported intercepts are multiplied by 100. In parentheses are t-statistics obtained from Newey-West standard errors computed with \( m \) lags. Significance at the 10 and 5 percent level is denoted by * and **, respectively. The estimation is carried out on monthly data from 1961:6 to 2013:12 using Gürkaynak-Sack-Wright (GSW) bond yields.

<table>
<thead>
<tr>
<th>( k )</th>
<th>No switch</th>
<th>Recessions vs. expansions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha_k )</td>
<td>( \beta_k )</td>
</tr>
<tr>
<td>24</td>
<td>-0.00</td>
<td>-1.05*</td>
</tr>
<tr>
<td></td>
<td>(-0.01)</td>
<td>(-1.89)</td>
</tr>
<tr>
<td>48</td>
<td>0.03</td>
<td>-1.56**</td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td>(-2.10)</td>
</tr>
<tr>
<td>72</td>
<td>0.05</td>
<td>-2.03**</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(-2.45)</td>
</tr>
<tr>
<td>96</td>
<td>0.06</td>
<td>-2.46**</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(-2.70)</td>
</tr>
<tr>
<td>120</td>
<td>0.07</td>
<td>-2.87**</td>
</tr>
<tr>
<td></td>
<td>(1.59)</td>
<td>(-2.88)</td>
</tr>
</tbody>
</table>
Table 3: Wald tests in modified Campbell-Shiller regressions

The table reports Wald statistics and \( p \)-values in parentheses for the null hypothesis of a joint switch in the intercepts and for a joint switch in the slope coefficients when using an increasing number of maturities. All tests are carried out using Newey-West covariance matrices computed with \( m \) lags and with recessions identified from the PMI. Significance at the 10 and 5 percent level is denoted by * and **, respectively. The tests are implemented on monthly Fama-Bliss bond yields over the period from 1961:06 to 2013:12. As the panel of bond yields is unbalanced, the Lynch and Wachter (2013) adjusted moments estimator is used.

<table>
<thead>
<tr>
<th>( H_0 )</th>
<th>Number of maturities</th>
<th>Selected maturities ( k )</th>
<th>( W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_k = 0 )</td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>57.5** (0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>76.3** (0.00)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6, 12, ..., 114, 120</td>
<td>150.8** (0.00)</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>6, 9, ..., 117, 120</td>
<td>398.2** (0.00)</td>
</tr>
<tr>
<td>( \beta_k = 0 )</td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>19.4** (0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>56.2** (0.00)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6, 12, ..., 114, 120</td>
<td>123.1** (0.00)</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>6, 9, ..., 117, 120</td>
<td>282.5** (0.00)</td>
</tr>
</tbody>
</table>
Table 4: Wald tests in modified Campbell-Shiller regressions: subsample analysis

The table reports Wald statistics and p-values in parentheses for the null hypothesis of a joint switch in the intercepts and for a joint switch in the slope coefficients when using an increasing number of maturities. All tests are carried out using Newey-West covariance matrices computed with m lags and with recessions identified from the PMI. Significance at the 10 and 5 percent level is denoted by * and **, respectively. The tests are implemented on monthly Fama-Bliss bond yields over the period from 1972:01 to 2013:12 (to get a balanced panel of bond yields) and from 1983:1 to 2013:12 (the post-Volcker period).

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>Number of maturities</th>
<th>Selected maturities $k$</th>
<th>1972:1-2013:12 W</th>
<th>1983:1-2013:12 W</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k^\Delta = 0$</td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>41.1**(0.00)</td>
<td>32.0**(0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>63.1**(0.00)</td>
<td>99.6**(0.00)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6, 12, ..., 114, 120</td>
<td>148.8**(0.00)</td>
<td>221.8**(0.00)</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>6, 9, ..., 117, 120</td>
<td>369.9**(0.00)</td>
<td>647.0**(0.00)</td>
</tr>
<tr>
<td>$\beta_k^\Delta = 0$</td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>24.3**(0.00)</td>
<td>24.6**(0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>71.7**(0.00)</td>
<td>80.0**(0.00)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6, 12, ..., 114, 120</td>
<td>156.9**(0.00)</td>
<td>151.9**(0.00)</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>6, 9, ..., 117, 120</td>
<td>284.8**(0.00)</td>
<td>422.6**(0.00)</td>
</tr>
</tbody>
</table>
Table 5: Return predictability by the yield spread: robustness

This table shows the results of regressing the three-month ahead excess returns from the 20-year Ibbotson bond return series on the corresponding 20-year yield spread when accounting for regime switching. This is implemented by estimating $x_{t+m} = \mu^{EXP} + \mu^A_{1(NBER)} + (\theta^{EXP} + \theta^A_{1(NBER)}) x_t + \epsilon_{t+m}$ with $m = 3$. Regimes are identified using NBER recessions. Newey-West $t$-statistics with $m$ lags are provided in parentheses, with significance at the 10 and 5 percent level denoted by * and **, respectively. $T$ is the total number of observations and $T_{REC}$ is the number of recession months.

<table>
<thead>
<tr>
<th>Sample period</th>
<th>$T$ ($T_{REC}$)</th>
<th>No switch</th>
<th>Recessions vs. expansions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\theta$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1926:1-2013:12</td>
<td>1053</td>
<td>$-0.71^{**}$</td>
<td>$0.73^{**}$</td>
</tr>
<tr>
<td>(199)</td>
<td>$(-2.14)$</td>
<td>$(4.04)$</td>
<td></td>
</tr>
<tr>
<td>1926:1-1961:5</td>
<td>422</td>
<td>$-0.08$</td>
<td>$0.36^{*}$</td>
</tr>
<tr>
<td>(116)</td>
<td>$(-0.25)$</td>
<td>$(1.90)$</td>
<td></td>
</tr>
<tr>
<td>1961:6-2013:12</td>
<td>628</td>
<td>$-0.95^{**}$</td>
<td>$0.85^{**}$</td>
</tr>
<tr>
<td>(83)</td>
<td>$(-2.04)$</td>
<td>$(3.74)$</td>
<td></td>
</tr>
</tbody>
</table>
Given the low estimates of $\Phi (1, 1)$, the value of $\alpha$ is unidentified and set to 0 for all models. $t$-statistics are reported in parentheses and computed based on asymptotic standard errors for $\Phi$ that are robust to measurement errors $v_{t,k}$ displaying heteroskedasticity in the time series dimension, cross-sectional correlation, and autocorrelation. We use $w_D = 5$ and $w_T = 10$ in the provided estimator of Andreasen and Christensen (2015). The $t$-statistics for $\Sigma_x$ are computed based on the standard errors described in Appendix A.4, whereas the $t$-statistics related to the parameters for PMI are obtained as in Andreasen and Christensen (2015). Significance at the 10 and 5 percent level is denoted by * and **, respectively.

<table>
<thead>
<tr>
<th>PMI</th>
<th>$M^{ATSM}$</th>
<th>$M^{ATSM}_{\lambda_0}$</th>
<th>$M^{ATSM}_{\lambda_x}$</th>
<th>$M^{ATSM}_{\lambda_0\lambda_x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi (1, 1)$</td>
<td>$1.18 \times 10^{-8}$</td>
<td>$1.54 \times 10^{-8}$</td>
<td>$1.19 \times 10^{-8}$</td>
<td>$1.28 \times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td>$(2.71 \times 10^{-5})$</td>
<td>$(3.52 \times 10^{-5})$</td>
<td>$(1.61 \times 10^{-5})$</td>
<td>$(2.93 \times 10^{-5})$</td>
</tr>
<tr>
<td>$\Phi (2, 2)$</td>
<td>0.0352**</td>
<td>0.0353**</td>
<td>0.0309**</td>
<td>0.0352**</td>
</tr>
<tr>
<td></td>
<td>(7.62)</td>
<td>(7.60)</td>
<td>(5.76)</td>
<td>(7.63)</td>
</tr>
<tr>
<td>$\Phi (3, 3)$</td>
<td>0.0416**</td>
<td>0.0416**</td>
<td>0.0350**</td>
<td>0.0416**</td>
</tr>
<tr>
<td></td>
<td>(9.40)</td>
<td>(9.37)</td>
<td>(7.37)</td>
<td>(9.41)</td>
</tr>
<tr>
<td>$\Sigma_x (1, 1)$</td>
<td>$3.64 \times 10^{-4**}$</td>
<td>$3.64 \times 10^{-4**}$</td>
<td>$5.39 \times 10^{-4**}$</td>
<td>$3.65 \times 10^{-4**}$</td>
</tr>
<tr>
<td></td>
<td>(18.49)</td>
<td>(18.52)</td>
<td>(21.94)</td>
<td>(18.83)</td>
</tr>
<tr>
<td>$\Sigma_x (2, 1)$</td>
<td>$-0.0034**$</td>
<td>$-0.0034**$</td>
<td>$-0.0049**$</td>
<td>$-0.0034**$</td>
</tr>
<tr>
<td></td>
<td>$(-9.38)$</td>
<td>$(-9.31)$</td>
<td>$(-9.18)$</td>
<td>$(-9.44)$</td>
</tr>
<tr>
<td>$\Sigma_x (2, 2)$</td>
<td>0.0042**</td>
<td>0.0042**</td>
<td>0.0068**</td>
<td>0.0041**</td>
</tr>
<tr>
<td></td>
<td>(24.57)</td>
<td>(24.45)</td>
<td>(26.96)</td>
<td>(24.61)</td>
</tr>
<tr>
<td>$\Sigma_x (3, 1)$</td>
<td>0.0031**</td>
<td>0.0031**</td>
<td>0.0045**</td>
<td>0.0031**</td>
</tr>
<tr>
<td></td>
<td>(8.97)</td>
<td>(8.87)</td>
<td>(8.90)</td>
<td>(9.02)</td>
</tr>
<tr>
<td>$\Sigma_x (3, 2)$</td>
<td>$-0.0041**$</td>
<td>$-0.0041**$</td>
<td>$-0.0066**$</td>
<td>$-0.0041**$</td>
</tr>
<tr>
<td></td>
<td>$(-23.98)$</td>
<td>$(-23.82)$</td>
<td>$(-26.50)$</td>
<td>$(-24.09)$</td>
</tr>
<tr>
<td>$\Sigma_x (3, 3)$</td>
<td>$3.63 \times 10^{-4**}$</td>
<td>$3.51 \times 10^{-4**}$</td>
<td>$4.06 \times 10^{-4**}$</td>
<td>$3.40 \times 10^{-4**}$</td>
</tr>
<tr>
<td></td>
<td>(8.97)</td>
<td>(9.58)</td>
<td>(20.48)</td>
<td>(15.19)</td>
</tr>
</tbody>
</table>

$\gamma_0$ | $0.0094**$ | $0.0095**$ | $0.0094**$ | $0.0094**$ |
| | $(2.21)$ | $(2.16)$ | $(2.21)$ | $(2.21)$ |

$\gamma_z$ | $0.9338**$ | $0.9338**$ | $0.9338**$ | $0.9338**$ |
| | $(51.81)$ | $(51.75)$ | $(51.81)$ | $(51.81)$ |

$\gamma_x (1, 1)$ | $-1.4022**$ | $-1.3459**$ | $-1.4023**$ | $-1.4023**$ |
| | $(-2.02)$ | $(-1.99)$ | $(-2.02)$ | $(-2.02)$ |

$\gamma_x (2, 1)$ | $-3.1301**$ | $-2.9007**$ | $-3.1288**$ | $-3.1288**$ |
| | $(-4.61)$ | $(-4.52)$ | $(-4.61)$ | $(-4.61)$ |

$\gamma_x (3, 1)$ | $-3.4533**$ | $-3.1354**$ | $-3.4552**$ | $-3.4552**$ |
| | $(-4.69)$ | $(-4.62)$ | $(-4.69)$ | $(-4.69)$ |

$\Sigma_{zz}$ | $0.0222**$ | $0.0222**$ | $0.0222**$ | $0.0222**$ |
Table 7: Estimation results: the market prices of risk

Estimates of $\lambda_0$ are common across expansions (EXP) and recessions (REC) in $M^\text{ATS}$, while estimates of $\lambda_x$ are common across expansions and recessions in $M^\text{ATS}_{\lambda_x}$. The $t$-statistics in parentheses are computed based on the standard errors described in Appendix A.4. Significance at the 10 and 5 percent level is denoted by * and **, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$M^\text{ATS}$</th>
<th>$M^\text{ATS}_{\lambda_0}$</th>
<th>$M^\text{ATS}_{\lambda_x}$</th>
<th>$M^\text{ATS}_{\lambda_0\lambda_x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intercepts:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 (1,1)$</td>
<td>$2.37 \times 10^{-4**}$</td>
<td>$2.29 \times 10^{-4**}$</td>
<td>$1.64 \times 10^{-4*}$</td>
<td>$0.0010**$</td>
</tr>
<tr>
<td></td>
<td>(4.13)</td>
<td>(3.96)</td>
<td>(1.82)</td>
<td>(8.64)</td>
</tr>
<tr>
<td>$\lambda_0 (2,1)$</td>
<td>$-0.0033**$</td>
<td>$-0.0032**$</td>
<td>$-0.0019$</td>
<td>$-0.0059**$</td>
</tr>
<tr>
<td></td>
<td>(-3.60)</td>
<td>(-3.42)</td>
<td>(-1.39)</td>
<td>(-3.60)</td>
</tr>
<tr>
<td>$\lambda_0 (3,1)$</td>
<td>$0.0032**$</td>
<td>$0.0030**$</td>
<td>$0.0015$</td>
<td>$0.0053**$</td>
</tr>
<tr>
<td></td>
<td>(3.54)</td>
<td>(3.32)</td>
<td>(1.10)</td>
<td>(3.36)</td>
</tr>
<tr>
<td><strong>Slopes:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_x (1,1)$</td>
<td>$-0.0316**$</td>
<td>$-0.0293**$</td>
<td>$-0.1651**$</td>
<td>$-0.0517**$</td>
</tr>
<tr>
<td></td>
<td>(-3.37)</td>
<td>(-3.05)</td>
<td>(-7.87)</td>
<td>(-2.02)</td>
</tr>
<tr>
<td>$\lambda_x (1,2)$</td>
<td>$0.0224**$</td>
<td>$0.0221**$</td>
<td>$0.0217$</td>
<td>$0.0244$</td>
</tr>
<tr>
<td></td>
<td>(2.58)</td>
<td>(2.66)</td>
<td>(1.22)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>$\lambda_x (1,3)$</td>
<td>$0.0188**$</td>
<td>$0.0185**$</td>
<td>$0.0187$</td>
<td>$0.0191$</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(2.90)</td>
<td>(0.97)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>$\lambda_x (2,1)$</td>
<td>$0.5096**$</td>
<td>$0.4666**$</td>
<td>$0.6461**$</td>
<td>$0.8623**$</td>
</tr>
<tr>
<td></td>
<td>(3.14)</td>
<td>(2.70)</td>
<td>(2.31)</td>
<td>(3.59)</td>
</tr>
<tr>
<td>$\lambda_x (2,2)$</td>
<td>$0.0721$</td>
<td>$0.0712$</td>
<td>$-0.0742$</td>
<td>$-0.4898$</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.44)</td>
<td>(-2.6)</td>
<td>(-0.78)</td>
</tr>
<tr>
<td>$\lambda_x (2,3)$</td>
<td>$0.1546$</td>
<td>$0.1602$</td>
<td>$0.1282$</td>
<td>$-0.3804$</td>
</tr>
<tr>
<td></td>
<td>(0.81)</td>
<td>(0.84)</td>
<td>(0.43)</td>
<td>(-0.56)</td>
</tr>
<tr>
<td>$\lambda_x (3,1)$</td>
<td>$-0.4976**$</td>
<td>$-0.4455**$</td>
<td>$-0.6015**$</td>
<td>$-0.9047**$</td>
</tr>
<tr>
<td></td>
<td>(-3.11)</td>
<td>(-2.67)</td>
<td>(-2.17)</td>
<td>(-3.84)</td>
</tr>
<tr>
<td>$\lambda_x (3,2)$</td>
<td>$-0.0710$</td>
<td>$-0.0711$</td>
<td>$-0.0363$</td>
<td>$0.2598$</td>
</tr>
<tr>
<td></td>
<td>(-0.41)</td>
<td>(-0.45)</td>
<td>(-0.13)</td>
<td>(0.43)</td>
</tr>
<tr>
<td>$\lambda_x (3,3)$</td>
<td>$-0.1461$</td>
<td>$0.1528$</td>
<td>$-0.2314$</td>
<td>$0.1416$</td>
</tr>
<tr>
<td></td>
<td>(-0.79)</td>
<td>(-0.83)</td>
<td>(-0.78)</td>
<td>(0.22)</td>
</tr>
</tbody>
</table>


Table 8: Slope coefficients in modified Mincer-Zarnowitz regressions

This table reports the regression slopes of modified Mincer-Zarnowitz regressions of realized three months excess returns on a constant and model-implied expected excess returns when conditioning on the state of the economy using NBER recessions. Excess returns are pooled within bins of two years for these regressions, given a total of eight excess returns per reported maturity range. Thompson (2011) two-way clustered $t$-statistics with three lags are provided in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Expansions</th>
<th>Recessions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maturity range in years:</td>
<td>Maturity range in years:</td>
</tr>
<tr>
<td></td>
<td>0-2</td>
<td>2-4</td>
</tr>
<tr>
<td>$M^{ATSM}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.98**</td>
<td>1.16**</td>
</tr>
<tr>
<td></td>
<td>(4.48)</td>
<td>(4.76)</td>
</tr>
<tr>
<td>$M^{ATSM}_{\lambda_0\lambda_x}$</td>
<td>0.65**</td>
<td>0.76**</td>
</tr>
<tr>
<td></td>
<td>(3.68)</td>
<td>(4.09)</td>
</tr>
<tr>
<td>$M^{SRM}_{\lambda_0\lambda_x}$</td>
<td>0.71**</td>
<td>0.84**</td>
</tr>
<tr>
<td></td>
<td>(3.67)</td>
<td>(4.24)</td>
</tr>
<tr>
<td>$M^{AFNS}_{\lambda_0(2,1)\lambda_x(2,2)}$</td>
<td>0.77**</td>
<td>0.86**</td>
</tr>
<tr>
<td></td>
<td>(3.92)</td>
<td>(4.05)</td>
</tr>
<tr>
<td>$M^{Linear}$</td>
<td>0.94**</td>
<td>1.14**</td>
</tr>
<tr>
<td></td>
<td>(5.54)</td>
<td>(5.36)</td>
</tr>
<tr>
<td>$M^{Spanned}$</td>
<td>0.93**</td>
<td>1.15**</td>
</tr>
<tr>
<td></td>
<td>(5.58)</td>
<td>(5.41)</td>
</tr>
</tbody>
</table>
Table 9: Additional model estimates: risk-neutral coefficients and dynamics of PMI

Given the low estimates of $\Phi(1,1)$, the value of $\alpha$ is unidentified and set to 0 for all models. $t$-statistics for $\Phi$ are provided in parentheses and computed using asymptotic standard errors that are robust to measurement errors $v_{t,k}$ displaying heteroskedasticity in the time series dimension, cross-sectional correlation, and autocorrelation. We use $w_D = 5$ and $w_T = 10$ in the provided estimator of Andreasen and Christensen (2015). The $t$-statistics for $\Sigma_x$ and the parameters for PMI are reported in parentheses and computed using the asymptotic standard errors described in Appendix A.4. For $M^{\text{Spanned}}$, we let $\mu(4,1) = \Phi(4,i) = \Phi(i,4) = 0$ for $i = \{1,2,3\}$ as the fit of bond yields is completely unaffected by these coefficients. Significance at the 10 and 5 percent level is denoted by * and **, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$M_{30\lambda x}$</th>
<th>$M_{30(2,1)\lambda x}(2,2)$</th>
<th>$M^{\text{Linear}}$</th>
<th>$M^{\text{Spanned}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta(4,1)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>0.1283** (6.30)</td>
</tr>
<tr>
<td>$\tilde{\lambda}$</td>
<td>$-$</td>
<td>0.0388** (18.32)</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi(1,1)$</td>
<td>$1.00 \times 10^{-8}$ (2.75 $\times 10^{-5}$)</td>
<td>$1.09 \times 10^{-8}$ (2.45 $\times 10^{-5}$)</td>
<td>$1.11 \times 10^{-8}$ (1.60 $\times 10^{-5}$)</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Phi(2,2)$</td>
<td>0.0351** (5.77)</td>
<td>$-$</td>
<td>0.0352** (7.60)</td>
<td>0.0430</td>
</tr>
<tr>
<td>$\Phi(3,3)$</td>
<td>0.0536** (10.15)</td>
<td>$-$</td>
<td>0.0416** (9.40)</td>
<td>0.0447</td>
</tr>
<tr>
<td>$\Phi(4,4)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>0.0513** (3.00)</td>
</tr>
<tr>
<td>$\Sigma_x(1,1)$</td>
<td>$3.44 \times 10^{-4}$ ** (20.65)</td>
<td>$3.67 \times 10^{-4}$ ** (18.55)</td>
<td>$3.65 \times 10^{-4}$ ** (11.95)</td>
<td>$3.16 \times 10^{-4}$ ** (13.66)</td>
</tr>
<tr>
<td>$\Sigma_x(2,1)$</td>
<td>$-0.0012**$ (9.83)</td>
<td>$-0.0003**$ (10.18)</td>
<td>$-0.0004**$ (19.31)</td>
<td>$-0.0009**$ (4.92)</td>
</tr>
<tr>
<td>$\Sigma_x(2,2)$</td>
<td>$0.0019**$ (25.97)</td>
<td>$3.56 \times 10^{-4}**$ (13.68)</td>
<td>$0.0042**$ (20.72)</td>
<td>$0.0194**$ (19.12)</td>
</tr>
<tr>
<td>$\Sigma_x(3,1)$</td>
<td>$9.68 \times 10^{-4}$ ** (7.72)</td>
<td>$-0.0006**$ (7.21)</td>
<td>$0.0031**$ (6.06)</td>
<td>$0.0088**$ (4.68)</td>
</tr>
<tr>
<td>$\Sigma_x(3,2)$</td>
<td>$-0.0018**$ (24.13)</td>
<td>$7.92 \times 10^{-5}$ ** (1.83)</td>
<td>$-0.0041**$ (20.34)</td>
<td>$-0.0210**$ (18.92)</td>
</tr>
<tr>
<td>$\Sigma_x(3,3)$</td>
<td>$3.69 \times 10^{-4}$ ** (16.88)</td>
<td>$7.05 \times 10^{-4}$ ** (23.73)</td>
<td>$3.53 \times 10^{-4}$ ** (5.71)</td>
<td>$0.0023**$ (21.04)</td>
</tr>
<tr>
<td>$\Sigma_x(4,1)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$6.62 \times 10^{-4}$ (0.73)</td>
<td>$2.71 \times 10^{-4}$ (0.31)</td>
</tr>
<tr>
<td>$\Sigma_x(4,2)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.0040$** (3.30)</td>
<td>$0.0134$** (10.36)</td>
</tr>
<tr>
<td>$\Sigma_x(4,3)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.0051$** (2.69)</td>
<td>$-0.0176$** (19.47)</td>
</tr>
<tr>
<td>$\Sigma_x(4,4)$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.0213$** (21.70)</td>
<td>$0.0027$** (6.00)</td>
</tr>
<tr>
<td>PMI</td>
<td>$\gamma_0$</td>
<td>0.0101** (2.41)</td>
<td>$0.0094$** (2.21)</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_z$</td>
<td>0.9320** (51.66)</td>
<td>0.9338** (51.81)</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_x(1,1)$</td>
<td>$1.4882$** (51.66)</td>
<td>$1.4000$** (51.81)</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_x(2,1)$</td>
<td>$-3.1238$** (2.18)</td>
<td>$-3.2161$** (2.01)</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_x(3,1)$</td>
<td>$-3.8735$** (4.70)</td>
<td>$1.8307$** (4.65)</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$\Sigma_{zz}$</td>
<td>0.0222** (24.94)</td>
<td>0.0221** (24.79)</td>
<td>$-$</td>
</tr>
</tbody>
</table>
Table 10: Additional model estimates: the market prices of risk in the SRM and AFNS models
The t-statistics in parentheses are computed based on the standard errors described in Appendix A.4. Significance at the 10 and 5 percent level is denoted by * and **, respectively.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th><strong>M</strong>^{SRM}_{\lambda_0, \lambda_x}</th>
<th><strong>M</strong>^{AFNS}_{\lambda_0(2,1), \lambda_x(2,2)}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EXP</td>
<td>REC</td>
</tr>
<tr>
<td>\lambda_0 (1, 1)</td>
<td>2.33 × 10^{-4}***</td>
<td>5.84 × 10^{-4}***</td>
</tr>
<tr>
<td>\lambda_0 (2, 1)</td>
<td>-0.0012**</td>
<td>-0.0014</td>
</tr>
<tr>
<td>\lambda_0 (3, 1)</td>
<td>0.0011**</td>
<td>6.30 × 10^{-4}</td>
</tr>
<tr>
<td>\lambda_x (1, 1)</td>
<td>-0.0338**</td>
<td>-0.0728*</td>
</tr>
<tr>
<td>\lambda_x (1, 2)</td>
<td>0.0206**</td>
<td>0.0376</td>
</tr>
<tr>
<td>\lambda_x (1, 3)</td>
<td>0.0156</td>
<td>0.0183</td>
</tr>
<tr>
<td>\lambda_x (2, 1)</td>
<td>0.2008**</td>
<td>0.1186</td>
</tr>
<tr>
<td>\lambda_x (2, 2)</td>
<td>0.0889</td>
<td>-0.2956</td>
</tr>
<tr>
<td>\lambda_x (2, 3)</td>
<td>0.1766**</td>
<td>-0.2380</td>
</tr>
<tr>
<td>\lambda_x (3, 1)</td>
<td>-0.1801**</td>
<td>-0.0880</td>
</tr>
<tr>
<td>\lambda_x (3, 2)</td>
<td>-0.0821</td>
<td>-0.0039</td>
</tr>
<tr>
<td>\lambda_x (3, 3)</td>
<td>-0.1516*</td>
<td>-0.0769</td>
</tr>
</tbody>
</table>
Table 11: Additional model estimates: the market prices of risk in the linear unspanned and spanned macroeconomic risk models

Estimates of $\lambda_0$ and $\lambda_x$ are common across expansions (EXP) and recessions (REC) in $M^{\text{Linear}}$ and $M^{\text{Spanned}}$. The $t$-statistics in parentheses are computed based on the standard errors described in Appendix A.4. Significance at the 10 and 5 percent level is denoted by * and **, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$M^{\text{Linear}}$</th>
<th>$M^{\text{Spanned}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EXP</td>
<td>REC</td>
</tr>
<tr>
<td><strong>Intercepts:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0$ (1, 1)</td>
<td>$2.36 \times 10^{-4**}$</td>
<td>$1.96 \times 10^{-4**}$</td>
</tr>
<tr>
<td></td>
<td>(3.26)</td>
<td>(3.05)</td>
</tr>
<tr>
<td>$\lambda_0$ (2, 1)</td>
<td>$-0.0030**$</td>
<td>$-0.0064$</td>
</tr>
<tr>
<td></td>
<td>(-2.97)</td>
<td>(-1.59)</td>
</tr>
<tr>
<td>$\lambda_0$ (3, 1)</td>
<td>$0.0027**$</td>
<td>$0.0052$</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
<td>(1.20)</td>
</tr>
<tr>
<td>$\lambda_0$ (4, 1)</td>
<td>$0.0094**$</td>
<td>$0.0070$</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(1.49)</td>
</tr>
<tr>
<td><strong>Slopes:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_x$ (1, 1)</td>
<td>$-0.0315**$</td>
<td>$-0.0237**$</td>
</tr>
<tr>
<td></td>
<td>(2.78)</td>
<td>(-2.39)</td>
</tr>
<tr>
<td>$\lambda_x$ (1, 2)</td>
<td>$0.0224**$</td>
<td>$0.0230**$</td>
</tr>
<tr>
<td></td>
<td>(2.29)</td>
<td>(2.70)</td>
</tr>
<tr>
<td>$\lambda_x$ (1, 3)</td>
<td>$0.0188*$</td>
<td>$0.0224**$</td>
</tr>
<tr>
<td></td>
<td>(1.78)</td>
<td>(2.57)</td>
</tr>
<tr>
<td>$\lambda_x$ (1, 4)</td>
<td>$6.44 \times 10^{-6}$</td>
<td>$0.0026**$</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(2.16)</td>
</tr>
<tr>
<td>$\lambda_x$ (2, 1)</td>
<td>$0.4888**$</td>
<td>$1.1312**$</td>
</tr>
<tr>
<td></td>
<td>(3.31)</td>
<td>(2.03)</td>
</tr>
<tr>
<td>$\lambda_x$ (2, 2)</td>
<td>$0.0734$</td>
<td>$-0.8757$</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td>(-1.20)</td>
</tr>
<tr>
<td>$\lambda_x$ (2, 3)</td>
<td>$0.1550$</td>
<td>$-0.8365$</td>
</tr>
<tr>
<td></td>
<td>(0.79)</td>
<td>(-1.13)</td>
</tr>
<tr>
<td>$\lambda_x$ (2, 4)</td>
<td>$-0.0024$</td>
<td>$-0.1060$</td>
</tr>
<tr>
<td></td>
<td>(-0.78)</td>
<td>(-1.04)</td>
</tr>
<tr>
<td>$\lambda_x$ (3, 1)</td>
<td>$-0.4658**$</td>
<td>$-0.9329$</td>
</tr>
<tr>
<td></td>
<td>(-3.21)</td>
<td>(-1.55)</td>
</tr>
<tr>
<td>$\lambda_x$ (3, 2)</td>
<td>$-0.0731$</td>
<td>$1.3262*$</td>
</tr>
<tr>
<td></td>
<td>(-0.41)</td>
<td>(1.73)</td>
</tr>
<tr>
<td>$\lambda_x$ (3, 3)</td>
<td>$-0.1469$</td>
<td>$1.3000*$</td>
</tr>
<tr>
<td></td>
<td>(-0.78)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>$\lambda_x$ (3, 4)</td>
<td>$0.0037$</td>
<td>$0.1738$</td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td>(1.62)</td>
</tr>
<tr>
<td>$\lambda_x$ (4, 1)</td>
<td>$-1.4020**$</td>
<td>$-1.4433**$</td>
</tr>
<tr>
<td></td>
<td>(-2.01)</td>
<td>(-2.01)</td>
</tr>
<tr>
<td>$\lambda_x$ (4, 2)</td>
<td>$-3.1287**$</td>
<td>$-3.4870**$</td>
</tr>
<tr>
<td></td>
<td>(-4.68)</td>
<td>(-4.83)</td>
</tr>
<tr>
<td>$\lambda_x$ (4, 3)</td>
<td>$-3.4530**$</td>
<td>$-3.5736**$</td>
</tr>
<tr>
<td></td>
<td>(-4.87)</td>
<td>(-4.88)</td>
</tr>
<tr>
<td>$\lambda_x$ (4, 4)</td>
<td>$-0.0662**$</td>
<td>$-0.5074**$</td>
</tr>
<tr>
<td></td>
<td>(-4.04)</td>
<td>(-5.17)</td>
</tr>
</tbody>
</table>
Figure 1: Ordinary and modified Campbell-Shiller regressions

The upper part of the figure plots estimation results from ordinary Campbell-Shiller regressions, $y_{t+m,k-m} - y_{t,k} = \alpha_k + \beta_k \frac{m}{k-m} (y_{t,k} - y_{t,m}) + \epsilon_{t+m,k}$. The lower part of the figure plots estimation results from modified Campbell-Shiller regressions, $y_{t+m,k-m} - y_{t,k} = \alpha_k EXP \{z_t \geq c\} + \beta_k \frac{m}{k-m} 1\{z_t \geq c\} (y_{t,k} - y_{t,m}) + \alpha_k REC \{1 - 1\{z_t \geq c\}\} + \beta_k \frac{m}{k-m} 1\{1\{z_t \geq c\}\} (y_{t,k} - y_{t,m}) + \epsilon_{t+m,k}$, where recessions are identified by the PMI. These regressions are estimated for $m = 3$ and $k = 6, 9, 12, ..., 120$, implying a total of 39 regressions, using monthly data from 1961:6 to 2013:12 based on Fama-Bliss (FB) and Gürkaynak-Sack-Wright (GSW) bond yields. The intercepts are multiplied by 100. As the panel of FB bond yields is unbalanced, the Lynch and Wachter (2013) adjusted moments estimator is used for this data set.
Figure 2: Goodness-of-fit in ordinary and modified Campbell-Shiller regressions
The figure reports the goodness-of-fit as measured by the $R^2$ statistic (in percent) for the ordinary and modified Campbell-Shiller regressions. The estimation is carried out on monthly data from 1961:6 to 2013:12 using Fama-Bliss (FB) and Gürkaynak-Sack-Wright (GSW) bond yields, except for some long-term FB bond yields where the estimation starts somewhat later due to lack of data availability (see Appendix A.1). Recessions are identified from the PMI.
Figure 3: Ordinary and modified return regressions

These charts report the slope coefficients in ordinary and modified return regressions where excess bond returns three months ahead are regressed on either the yield spread, the forward spread, or the CP factor (with a constant) using monthly data from 1961:6 to 2013:12. The recessions in the modified return regressions are identified from the PMI. The data are Fama-Bliss (FB) and Gürkaynak, Sack and Wright (2007) (GSW) bond yields. Given the well-known reliance of the CP factor on FB forward rates, the CP factor for the GSW dataset is computed as $\gamma_{GSW}^{t} = (\gamma_{GSW}^{t})' F_t$, where $F_t = [1, y_{t, 12}, f_{t}^{(12, 24)}, f_{t}^{(24, 36)}, f_{t}^{(36, 48)}, f_{t}^{(48, 60)}]'$ are FB forward rates but $\gamma_{GSW}^{t}$ is obtained by regressing GSW returns on FB forward rates. As the panel of FB bond yields is unbalanced, the Lynch and Wachter (2013) adjusted moments estimator is used for this data set.
Figure 4: Robustness analysis: NBER recessions

The top chart reports the PMI re-centered at $c = 44.5$ along with the NBER recessions (shaded). The lower part of the figure plots estimation results from modified Campbell-Shiller regressions, $y_{t+m,k-m} - y_{t,k} = \alpha^\text{EXP}_k (1 - 1_{\text{NBER}}) + \beta^\text{EXP}_k \frac{m}{k-m} (1 - 1_{\text{NBER}}) (y_{t,k} - y_{t,m}) + \alpha^\text{REC}_k 1_{\text{NBER}} + \beta^\text{REC}_k \frac{m}{k-m} 1_{\text{NBER}} (y_{t,k} - y_{t,m}) + \tilde{u}_{t+m,k}$, where recessions are identified by the NBER dummy. These regressions are estimated for $m = 3$ and $k = 6, 9, 12, \ldots, 120$, implying a total of 39 regressions, using monthly data from 1961:6 to 2013:12 based on Fama-Bliss (FB) and Gürkaynak-Sack-Wright (GSW) bond yields. The intercepts are multiplied by 100. As the panel of FB bond yields is unbalanced, the Lynch and Wachter (2013) adjusted moments estimator is used for this data set.
Figure 5: Robustness analysis: the forecast horizon

This figure plots estimation results from modified Campbell-Shiller regressions, 
\[ y_{t+m,k-m} - y_{t,k} = \alpha_k^{EXP} 1_{\{z_t \geq c\}} + \beta_k^{EXP} \frac{m}{k-m} 1_{\{z_t \geq c\}} (y_{t,k} - y_{t,m}) + \alpha_k^{REC} (1 - 1_{\{z_t \geq c\}}) + \beta_k^{REC} \frac{m}{k-m} (1 - 1_{\{z_t \geq c\}}) (y_{t,k} - y_{t,m}) + \epsilon_{t+m,k}, \]
where recessions are identified by the PMI. These regressions are estimated for \( m = \{3, 6, 12\} \) using monthly data from 1961:6 to 2013:12 based on Fama-Bliss (FB) and Gürkaynak-Sack-Wright (GSW) bond yields. The intercepts are multiplied by 100. As the panel of FB bond yields is unbalanced, the Lynch and Wachter (2013) adjusted moments estimator is used for this data set.
Figure 6: Model evaluation: ordinary and modified Campbell-Shiller regressions
The empirical intercepts and slope coefficients in ordinary and modified Campbell-Shiller regressions are computed based on monthly Fama-Bliss bond yields from 1961:6 to 2013:12. The model-implied regression loadings are computed at the estimated parameters for $\mathcal{M}_{T,SM}$ using a simulated sample path of 100,000 observations. The intercepts are multiplied by 100.
Figure 7: Model evaluation: ordinary and modified return regressions

The empirical slope coefficients in ordinary and modified return regressions are computed based on monthly Fama-Bliss bond yields from 1961:6 to 2013:12. The model-implied regression loadings are computed at the estimated parameters for $\mathcal{M}^{ATSM}_{\lambda_0,\lambda_x}$ using a simulated sample path of 100,000 observations.
Figure 8: Excess returns: the Gaussian ATSM vs. regime-dependent ATSM
The figure plots quarterly expected excess returns for the 10-year bond when expressed in annualized percent. Expected returns are computed at the estimated states, i.e. \( \hat{x}_t \left( \hat{\theta}^{step1}, \hat{\Sigma}^{step2} \right) \). With regime switching, expected excess returns are computed by Monte Carlo integration using 10,000 draws. Shaded areas denote NBER recessions.
Figure 9: SRM with regime-switching: ordinary and modified Campbell-Shiller regressions

The empirical intercepts and slope coefficients in ordinary and modified Campbell-Shiller regressions are computed based on monthly Fama-Bliss bond yields from 1961:6 to 2013:12. The model-implied regression loadings are computed at the estimated parameters for $\mathcal{M}_{\lambda_0, \lambda_1}^{\text{SRM}}$ using a simulated sample path of 100,000 observations. The intercepts are multiplied by 100.
Figure 10: AFNS model with regime-switching: ordinary and modified Campbell-Shiller regressions
The empirical intercepts and slope coefficients in ordinary and modified Campbell-Shiller regressions are computed based on monthly Fama-Bliss bond yields from 1961:6 to 2013:12. The model-implied regression loadings are computed at the estimated parameters for $\mathcal{M}_{AFNS}^{\lambda_{0}(2,1)\lambda_{2}(2,2)}$ using a simulated sample path of 100,000 observations. The intercepts are multiplied by 100.
Figure 11: AFNS model with regime-switching: excess returns

We report annualized quarterly excess return for the 10-year bond in \( \mathcal{M}_{AFNS}^{0}(2;1) \) computed by the Monte Carlo method using 10,000 draws. The partial effect of a switch in \( \lambda_{0}(2,1) \) is the difference in excess returns between \( \mathcal{M}_{AFNS}^{0}(2;1) \), where \( \lambda_{0}^{(2)}(2,1) = \lambda_{0}^{(1)}(2,1) \) and \( \mathcal{M}_{AFNS}^{x}(2;1) \), where \( \lambda_{0}^{(2)}(2,1) = \lambda_{0}^{(1)}(2,1) \) and \( \lambda_{x}^{(2)}(2,2) = \lambda_{x}^{(1)}(2,2) \), and similarly for the effect of a switch in \( \lambda_{x}(2,2) \). All returns and interest rates are in percent. Shaded regions denote NBER recessions.
Figure 12: Linear and spanned models: ordinary and modified Campbell-Shiller regressions

The empirical intercepts and slope coefficients in ordinary and modified Campbell-Shiller regressions are computed based on monthly Fama-Bliss bond yields from 1961:6 to 2013:12. The model-implied regression loadings are computed at the estimated parameters for $\mathcal{M}_{\text{Linear}}$ and $\mathcal{M}_{\text{Spanned}}$ using a simulated sample path of 100,000 observations. The intercepts are multiplied by 100.
ONLINE APPENDIX:
BOND MARKET ASYMMETRIES ACROSS RECESSIONS AND EXPANSIONS:
NEW EVIDENCE ON RISK PREMIA

Martin M. Andreasen
Aarhus University and CREATEs

Tom Engsted
Aarhus University and CREATEs

Stig V. Møller
Aarhus University and CREATEs

Magnus Sander
Aarhus University and CREATEs

Not for publication
## Contents

1. Data ................................................................. 129

2. Wald tests .......................................................... 129
   2.1 Lynch-Wachter estimator ...................................... 131
   2.2 Longer horizons ................................................ 135

3. The Campbell-Shiller regressions .................................. 136
   3.1 The link between Campbell-Shiller and return regressions .. 136
   3.2 Interpreting the intercept in the Campbell-Shiller regressions . 136

4. Bias-adjusted estimates ............................................ 138

5. Smooth transition regressions ..................................... 140

6. Out-of-sample evidence ............................................ 142

7. ATSM with a SETAR model ......................................... 145
   7.1 Model description ............................................. 145
   7.2 Pricing of zero-coupon bonds ................................. 146
   7.3 Change of measure ............................................ 149
   7.4 Excess return ............................................... 151
   7.5 The Campbell-Shiller regressions ........................... 153
   7.6 Estimation of the SETAR model with generated regressors ... 159
   7.7 Obtaining estimates for the market price of risk .......... 171
   7.8 Inducing stationarity ....................................... 171
   7.9 Normalization of the ATSM .................................. 174
   7.10 Accounting for the Zero Lower Bound ...................... 175

References ............................................................. 176
1 Data

We use monthly bond yields from an unsmoothed Fama and Bliss (1987) dataset running from 1961:06-2013:12. We limit the analysis to maturities from \( m/12 \) to 10 years with \( m \)-month increments where \( m = 3 \). A similar panel of yields is used in Adrian, Crump and Moench (2013) to obtain their pricing factors. At the very long end of the yield curve, we do not observe all increments and, hence, we interpolate between bond yields of the two nearest maturities.\(^1\) We also note that some long-term bond yields are unavailable from 1961:6, as maturities \( k = \{69, 72\} \) start in 1967:11, \( k = \{75, 78\} \) start in 1969:2, \( k = \{81, 84, \ldots, 114\} \) start in 1971:8, and \( k = \{117, 120\} \) start in 1971:11. We also report results for the Gürkaynak, Sack and Wright (2007) dataset, where bond yields in the \( 7 + \frac{m}{12} \) to 10-year maturity spectrum from 1961:6 to 1971:8 are calculated by extrapolation of their estimated curves.

2 Wald tests

We use Wald tests to examine whether the coefficients in the state-dependent Campbell-Shiller regressions change across expansions and recessions. For convenience, we restate the regression model

\[
y_{t+m,k-m} - y_{t,k} = \alpha_k^{EXP} 1\{z_t \geq c\} + \beta_k^{EXP} \frac{m}{k-m} 1\{z_t \geq c\} (y_{t,k} - y_{t,m}) \\
+ \alpha_k^{REC} (1 - 1\{z_t \geq c\}) + \beta_k^{REC} \frac{m}{k-m} (1 - 1\{z_t \geq c\}) (y_{t,k} - y_{t,m}) \\
+ \tilde{u}_{t+m,k},
\]

where \( 1\{z_t \geq c\} \) is the indicator function with a value of one when \( z_t \geq c \) and zero otherwise (see Section 2 in the paper for additional details). We want to test the following joint hypotheses

\[
H_0: \alpha_k^{REC} = \alpha_k^{EXP}, \ k = 6, 9, 12, \ldots, 120
\]
\[
H_0: \beta_k^{REC} = \beta_k^{EXP}, \ k = 6, 9, 12, \ldots, 120.
\]

We conduct the tests by setting up the system in a GMM framework. We collect the parameters in a vector

\[
\Theta' = [\theta'_6, \theta'_9, \ldots, \theta'_{120}],
\]

\(^1\)In addition, the unsmoothed Fama-Bliss forward rates are known to be somewhat rough and we therefore smooth them using a 12-month equal-weighted window.
where $\theta'_k = \left[ \alpha^{{EXP}_k}, \beta^{{EXP}_k}, \alpha^{{REC}_k}, \beta^{{REC}_k} \right]$. Consider the moment condition

$$E [ f_{t,k} (\theta_k) ] = 0,$$

where

$$f_{t,k} = \tilde{u}_{t+m,k} z_{t,k},$$

and

$$z_{t,k}' = \begin{bmatrix} 1 \{ z_t \geq c \}, & \frac{m}{k-m} 1 \{ z_t \geq c \} (y_{t,k} - y_{t,m}), \\ (1 - 1 \{ z_t \geq c \}), & \frac{m}{k-m} (1 - 1 \{ z_t \geq c \} (y_{t,k} - y_{t,m}) \end{bmatrix}.$$ 

The sample moments for the system of linear regressions are

$$g_T(\Theta)' = \left[ g_{T,6}', g_{T,9}', \ldots, g_{T,120}' \right],$$

where

$$g_{T,k} = \frac{1}{T} \sum_{t=1}^{T} f_{t,k}.$$ 

We know from Hansen (1982) that the asymptotic distribution of the GMM estimator is given by

$$\sqrt{T} \left( \hat{\Theta} - \Theta \right) \sim N \left( 0, \left( D^S \Sigma D \right)^{-1} \right).$$

To estimate the spectral density matrix, we use the Newey-West estimator with $L = m$ lags\(^2\)

$$S_T = C_0 + \sum_{j=1}^{L} \frac{L+1-j}{L+1} \left( C_j + C_j' \right),$$

where $C_j = \frac{1}{T} \sum_{t=1+j}^{T} f_{t,k} f_{t,-j,k}'$. Further, we estimate $D$ using

$$D_T = \text{diag} \left( D_{T,6}, D_{T,9}, \ldots, D_{T,120} \right),$$

where

$$D_{T,k} = \frac{\partial g_{T,k}}{\partial \theta'_k} = -\frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} I_t & tms_I I_t & 0 & 0 \\ tms_I I_t & tms^2_I I_t^2 & (1 - I_t) & tms_I (1 - I_t) \\ 0 & 0 & tms (1 - I_t) & tms^2 (1 - I_t) \end{bmatrix}. \quad \quad \quad (1)$$

\(^2\)Results are robust to using other lag specifications (e.g., $2m$ lags).
where $I_t = 1_{\{z_t \geq c\}}$ and $tms_t = \frac{m}{K-m} (y_{t,k} - y_{t,m})$. We use the following specification to conduct the joint tests

$$H_0: R\Theta = c.$$ 

A Wald test of the above joint hypotheses can then be computed using

$$W = (R\hat{\Theta} - c)' \left( R \left( D_T S_T^{-1} D_T \right)^{-1} R' / T \right)^{-1} (R\hat{\Theta} - c) \sim \chi^2_K.$$ 

For the hypothesis in (2), we have

$$R = ID_K \otimes [1,0,-1,0],$$

where $ID_K$ is the identity matrix with dimension $K$ and $c = 0$. $K$ is the number of maturities we examine and is equal to 39 for the full panel. For the hypothesis in (3), we have

$$R = ID_K \otimes [0,1,0,-1],$$

and $c = 0$.

### 2.1 Lynch-Wachter estimator

Conducting the tests for the full sample period involves making some implementation choices as the panel is unbalanced (this is only relevant for the Fama-Bliss dataset). We use the adjusted moments estimator of Lynch and Wachter (2013), which is an efficient way of exploiting the long sample. For simplicity we partition the moments in two parts: One part where we have yields for full the period from 1961:06-2013:12 and one part where we only have yields starting from 1971:11. The first group contains moments on maturities up to 66 months, while the second group contains those with maturities from 69 to 120 months. In principle, we could make many partitions based on the available observations for each moment, but this will heavily complicate the computations. Consider the moment condition

$$E[f(x_t, \Theta)] = 0,$$

where the moments in $f$ are specified in Section 2 and $x_t$ contains the yields in our sample. We partition the data and moments into two groups. Group 1 is available for the full period (short yields in our case) and group 2 of is only available for the later part of the period (long yields):

$$f(x_t, \Theta) = \left[ f_1(x_{1t}, \Theta_1), f_2(x_{2t}, \Theta_2) \right]' ,$$

where

$$f_1(x_{1t}, \Theta_1)' = \left[ f_t', f_t', ..., f_t', f_t', ..., f_t', ... \right].$$
and
\[ f_2(x_{2t}, \Theta_2) = \left[ f_{t,69}', f_{t,72}', \ldots, f_{t,120}' \right]. \]

Let \( \lambda \) denote the fraction of the sample period for which all data are available (the later part). \( x_{1t} \) is observed from \( t = 1, \ldots, T \), and \( x_{2t} \) is observed from \( t = (1 - \lambda) T + 1, \ldots, T \). It is useful to define the following sums (corresponding to full and partial sample moments):

\[
\begin{align*}
g_{1,T} &= \frac{1}{T} \sum_{t=1}^{T} f_1(x_{1t}, \Theta_1) \\
g_{1,\lambda T} &= \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} f_1(x_{1t}, \Theta_1) \\
g_{2,\lambda T} &= \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} f_2(x_{2t}, \Theta_2).
\end{align*}
\]

Consider the partitioned spectral density
\[
S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},
\]

where \( S_{ij} = \sum_{\tau=-\infty}^{\infty} E[f_{i,0}f'_{j,-\tau}], \quad i, j = 1, 2. \) Let
\[
B_{21} = S_{21}S_{11}^{-1}.
\]

The adjusted-moments (AM) estimator of Lynch and Wachter (2013) solves
\[
\hat{\Theta}^A = \arg \min_\Theta \left[ g_{1,T} (\Theta_1)', g_{2,T}^A (\Theta_2)' \right] (S_T^A)^{-1} \left[ g_{1,T} (\Theta_1) \\ g_{2,T}^A (\Theta_2) \right],
\]

where
\[
g_{2,T}^A (\Theta_2) = g_{2,\lambda T} + \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}),
\]

and \( \hat{B}_{21,\lambda T} \) is based on the spectral density estimated using standard GMM on the short overlapping sample. In our case, we can find analytical solutions. For each maturity in the first group, the sample moment conditions will yield the standard OLS estimator for the regression coefficients in (1). Setting \( g_{2,T}^A (\Theta_2) \) to zero gives

132
the AM estimator for coefficients in the second group:

\[ g_{2,\lambda T} + \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}) = \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} f_{2} \left( x_{t}, \Theta_{2} \right) \]

\[ + \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}) \]

\[ = \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} \left[ f'_{t,69}, f'_{t,72}, \ldots, f'_{t,120} \right] ' \]

\[ + \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}) . \]

The first part corresponds to the standard sample moments for the OLS estimator. The last part is an adjustment term:

\[ \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}) = \hat{B}_{21,\lambda T} \left( \frac{1}{T} \sum_{t=1}^{T} f_{1} \left( x_{1t}, \Theta_{1} \right) - \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} f_{1} \left( x_{1t}, \Theta_{1} \right) \right) \]

\[ = \hat{B}_{21,\lambda T} \left( \frac{1}{T} \sum_{t=1}^{T} \left[ f'_{t,6}, f'_{t,9}, \ldots, f'_{t,66} \right] ' \right) \]

\[ - \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} \left[ f'_{t,6}, f'_{t,9}, \ldots, f'_{t,66} \right] ' \]

\[ = -\hat{B}_{21,\lambda T} \left( \frac{1}{T} \sum_{t=(1-\lambda)T+1}^{T} \left[ f'_{t,6}, f'_{t,9}, \ldots, f'_{t,66} \right] ' \right) , \]

since \( \frac{1}{T} \sum_{t=1}^{T} \left[ f'_{t,6}, f'_{t,9}, \ldots, f'_{t,66} \right] ' = 0 \), because we use OLS on the full sample for the first group. \( \hat{B}_{21,\lambda T} (g_{1,T} - g_{1,\lambda T}) \) is a \( 4K_{2} \times 1 \) vector, where \( K_{2} \) is the number of maturities in the second group. In this adjustment vector, let \( a_{k} \) be a vector of the rows corresponding to maturity \( k \) from the second group, \( k = 69, \ldots, 120 \). Further, let \( z_{t,k} \) be the vector of regressors in the Campbell-Shiller regression for maturity \( k \), see Section 2. Then we can write
\[
\frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} f_{t,k} + a_k = \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} \left( (y_{t+m,k-m} - y_{t,k}) - z_{t,k}' \hat{\Theta}_k^A \right) z_{t,k} + a_k \\
= \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} \left( z_{t,k} (y_{t+m,k-m} - y_{t,k}) - z_{t,k} z_{t,k}' \hat{\Theta}_k^A \right) + a_k \\
= \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} z_{t,k} (y_{t+m,k-m} - y_{t,k}) \\
- \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} z_{t,k} z_{t,k}' \hat{\Theta}_k^A + a_k.
\]

Solving for the AM estimate, \( \hat{\Theta}_k^A \), for maturity \( k \) gives

\[
\hat{\Theta}_k^A = \left( \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} z_{t,k} z_{t,k}' \right)^{-1} \left( \frac{1}{\lambda T} \sum_{t=(1-\lambda)T+1}^{T} z_{t,k} (y_{t+m,k-m} - y_{t,k}) + a_k \right),
\]

which corresponds to OLS with an adjustment term.

The asymptotic distribution is

\[
\sqrt{\lambda T} \left( \hat{\Theta}_k^A - \Theta \right) \sim \mathcal{N} \left( 0, \left( D' (S^A)^{-1} D \right)^{-1} \right),
\]

where

\[
D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix},
\]

with \( D_i = E \left[ \frac{\partial f_i}{\partial \Theta} \right] \), \( i = 1, 2 \), and

\[
S^A = \begin{bmatrix} \lambda S_{11} & \lambda S_{12} \\ \lambda S_{21} & S_{22} - (1 - \lambda) S_{21} S_{11}^{-1} S_{12} \end{bmatrix}.
\]

Details on how we estimate \( S \) and \( D \) are in Section 2. We refer to Lynch and Wachter (2013) for further details.
### 2.2 Longer horizons

We present Wald tests of (2) and (3) using \( m = 6 \) and \( m = 12 \) in (1). We confirm that Campbell-Shiller intercept and slope coefficients switch significantly across recessions and expansions even when using longer horizons.

**Table 1: Wald tests in modified Campbell-Shiller regressions: longer horizons**

The table reports Wald statistics and \( p \)-values in parentheses for the null hypothesis of a joint switch in the intercepts and for a joint switch in the slope coefficients when using an increasing number of maturities for two different horizons. All tests are carried out using Newey-West covariance matrices computed with \( m \) lags and with recessions identified from the PMI. Significance at the 10 and 5 percent level is denoted by * and **, respectively. The data run from 1961:06 to 2013:12 and we use the adjusted moments estimator.

<table>
<thead>
<tr>
<th>( H_0 )</th>
<th>Number of bond yields</th>
<th>Selected maturities</th>
<th>( W )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha_k^\Delta = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m = 6 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>24.3** (0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>29.5** (0.00)</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>12, 18, ..., 114, 120</td>
<td>139.9** (0.00)</td>
</tr>
<tr>
<td></td>
<td>( \beta_k^\Delta = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m = 6 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>18.9** (0.00)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12, 24, ..., 108, 120</td>
<td>36.6** (0.00)</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>12, 18, ..., 114, 120</td>
<td>92.4** (0.00)</td>
</tr>
<tr>
<td></td>
<td>( m = 12 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \alpha_k^\Delta = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>35.5** (0.00)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>24, 36, ..., 108, 120</td>
<td>51.4** (0.00)</td>
</tr>
<tr>
<td></td>
<td>( \beta_k^\Delta = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m = 12 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24, 48, 72, 96, 120</td>
<td>21.8** (0.00)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>24, 36, ..., 108, 120</td>
<td>46.6** (0.00)</td>
</tr>
</tbody>
</table>

135
The Campbell-Shiller regressions

3.1 The link between Campbell-Shiller and return regressions

The $m$-period holding period return on a $k$-period bond is

\[ hpr_{t+m,k} = -\frac{k-m}{12}y_{t+m,k-m} + \frac{k}{12}y_{t,k}. \]

Now subtract $\frac{m}{12}y_{t,m}$ on both sides to obtain

\[ hpr_{t+m,k} - \frac{m}{12}y_{t,m} = -\frac{k-m}{12}y_{t+m,k-m} + \frac{k}{12}y_{t,k} - \frac{m}{12}y_{t,m}. \]

Then add and subtract $\frac{m}{12}y_{t,k}$ on the right hand side

\[ hpr_{t+m,k} - \frac{m}{12}y_{t,m} = -\frac{k-m}{12}(y_{t+m,k-m} - y_{t,k}) + \frac{m}{12}(y_{t,k} - y_{t,m}). \]

Multiplying both sides by $\frac{12}{k-m}$ and rearranging, we get

\[ \frac{12}{k-m} \left( hpr_{t+m,k} - \frac{m}{12}y_{t,m} \right) = \frac{m}{k-m} (y_{t,k} - y_{t,m}) - (y_{t+m,k-m} - y_{t,k}). \] \( (4) \)

Regressing each of the three terms in (4) on a constant and $\frac{m}{k-m} (y_{t,k} - y_{t,m})$, we see that the coefficients in the Campbell-Shiller and return regression are linked through the identity

\[ \frac{12}{m} \theta_k = 1 - \beta_k. \]

The same result holds for switching regressions. To see why, instead regress all terms in (4) on a constant and $\frac{m}{k-m} (y_{t,k} - y_{t,m})$ both interacted with $1_{\{z_t \geq c\}}$ and $(1 - 1_{\{z_t \geq c\}})$.

3.2 Interpreting the intercept in the Campbell-Shiller regressions

The expectations hypothesis with constant term premia is given by

\[ y_{t,k} = \frac{1}{k} \sum_{i=0}^{k-1} E_t [r_{t+i}] + TP_k, \] \( (5) \)
where $TP_k$ denotes a maturity-specific term premium. The one-period holding period return is given by

$$hpr_{t,k} = E_t \left[ \log \left( \frac{P_{t+1,k-1}}{P_{t,k}} \right) \right]$$

$$= E_t [\log P_{t+1,k-1} - \log P_{t,k}]$$

$$= - (k - 1) E_t [y_{t+1,k-1}] + k \times y_{t,k}$$

Now, under the expectations hypothesis in (5), we have

$$hpr_{t,k} = - (k - 1) E_t \left[ \frac{1}{k - 1} \sum_{i=0}^{k-2} E_{t+1} [r_{t+1+i}] + TP_{k-1} \right]$$

$$+ k \times \left[ \frac{1}{k} \sum_{i=0}^{k-1} E_t [r_{t+i}] + TP_k \right]$$

$$= - (E_t [r_{t+1}] + E_t [r_{t+2}] + \ldots + E_t [r_{t+k-1}])$$

$$+ (r_t + E_t [r_{t+1}] + E_t [r_{t+2}] + \ldots + E_t [r_{t+k-1}])$$

$$- (k - 1) TP_{k-1} + k \times TP_k$$

$$= r_t - (k - 1) TP_{k-1} + k \times TP_k$$

Hence,

$$- (k - 1) E_t [y_{t+1,k-1}] + k \times y_{t,k} = r_t - (k - 1) TP_{k-1} + k \times TP_k$$

\[ \Downarrow \]

$$- (k - 1) E_t [y_{t+1,k-1}] + (k - 1) \times y_{t,k} + y_{t,k} = r_t - (k - 1) TP_{k-1} + k \times TP_k$$

\[ \Downarrow \]

$$- (k - 1) (E_t [y_{t+1,k-1}] - y_{t,k}) = - (y_{t,k} - r_t) - (k - 1) TP_{k-1} + k \times TP_k$$

\[ \Downarrow \]

$$E_t [y_{t+1,k-1}] - y_{t,k} = \frac{1}{(k - 1)} (y_{t,k} - r_t) + TP_{k-1} - \frac{k}{k - 1} TP_k$$

The unconditional expectation of yields under the expectations hypothesis in (5) is

$$E [y_{t,k}] = \frac{1}{k} \sum_{i=0}^{k-1} E [E_t [r_{t+i}]] + TP_k$$

$$= E [r_t] + TP_k$$

Given that $E [y_{t,k}]$ is increasing in $k$, we have that $TP_k$ is increasing in $k$. We clearly also have that $\frac{k}{k - 1} > 1$ for $k > 1$, meaning that $\frac{k}{k - 1} TP_k > TP_{k-1}$. As a result,
$TP_{k-1} - \frac{k}{k-1} TP_k$ is negative for an upward sloping yield curve. Comparing this to the standard Campbell-Shiller regression (for $m = 1$

$$y_{t+1,k-1} - y_{t,k} = \alpha_k + \frac{\beta_k}{k-1} (y_{t,k} - r_t) + u_{t+1,k},$$

we thus conclude that (5) implies a negative estimate of $\alpha_k$.

### 4 Bias-adjusted estimates

We use a stationary bootstrap procedure to assess the degree of bias in the coefficients from the Campbell-Shiller regressions. The bootstrap procedure resamples the data in blocks of consecutive observations of the left and right hand side of the Campbell-Shiller regressions. We determine the optimal average block size using the Politis and White (2004) approach. When drawing the blocks, we make sure to preserve cross-sectional correlations by resampling from the same time points for all maturities. For each of the 100,000 bootstrap samples, we estimate the coefficients in the Campbell-Shiller regressions. We compute the bias as the difference between the empirical OLS estimate and the average of the simulated coefficients from the 100,000 bootstrap samples. In Figure 1, we plot bias-adjusted and unadjusted estimates from the Campbell-Shiller regressions. The figure illustrates that the bias is quite small. This resembles the finding of Dai and Singleton (2002). Bekaert, Hodrick and Marshall (1997) show that the failure of the expectations hypothesis cannot be explained by small-sample biases. In fact, they show that the slope coefficient in the Campbell-Shiller regression is biased towards accepting the expectations hypothesis.\(^3\)

\(^3\)In contrast to our setup, they simulate under the null that the expectations hypothesis is true.
Figure 1: Plot of bias-adjusted coefficients
The bias is calculated by subtracting the average of the simulated coefficients from the bootstrap procedure from the empirical OLS estimates. The sample period is 1972:1-2013:12.
5 Smooth transition regressions

By using an NBER or PMI-based dummy to classify the state of the economy, we implicitly assume that the best way to model predictability across recessions and expansions is a binary zero-one approach. To motivate this choice, we now estimate Smooth Transition Regressions with Logistic transition functions (LSTR models). The two-state model is

\[
y_{t+m,k-m} - y_{t,k} = \left[ \alpha^{(1)}_{k} + \beta^{(1)}_{k} tms_{t} \right] G(s_{t}, \gamma, c) \\
+ \left[ \alpha^{(2)}_{k} + \beta^{(2)}_{k} tms_{t} \right] (1 - G(s_{t}, \gamma, c)) + \eta_{t+1},
\]

where \( tms_{t} = \frac{m}{k-m} (y_{t,k} - y_{t,m}) \), and \( G(s_{t}, \gamma, c) \) is the first-order logistic function

\[
G(s_{t}, \gamma, c) = \left( 1 + \exp \left\{-\gamma (s_{t} - c)\right\} \right)^{-1}, \quad \gamma > 0.
\]

The parameter \( \gamma \) determines the speed of transition between the two states, \( s_{t} \) is the transition variable, and \( c \) is the location parameter (transition threshold) which determines when the transition happens. For moderately low values of \( \gamma \), the transition between the two states will be smooth, but as \( \gamma \to \infty \) the LSTR model approaches a threshold model in which the switch is instant around \( c \). The dummy regression in (1) is equivalent to this limit as for very high values of \( \gamma \), the transition function \( G(s_{t}, \gamma, c) \) will be a dummy variable with the value one for \( s_{t} > c \) and zero for \( s_{t} < c \).

The transition function is common across maturities. We thus estimate all \( K \) regressions as a system (we consider \( K = 39 \) maturities). We use a NLS approach to estimate the model. Given \( \gamma \) and \( c \), this can be done line by line using OLS and accordingly the nonlinear optimization only has to be carried out for these transition parameters. See, e.g., Teräsvirta and Yang (2014).

It is well known that estimating \( \gamma \) is difficult when the true parameter value is large, see, e.g., Teräsvirta (1994). It both leads to numerical problems when minimizing the objective function as well as a large standard error on the estimate. To enhance the performance of the numerical optimizer, we make two adjustments. First, the \( \gamma \) parameter generally depends on the scale of \( s_{t} \) and to make it scale-free, we follow Teräsvirta (2004) and divide it by the standard deviation of \( s_{t} \). Further, following Schleer (2015), we reparameterize \( \gamma \) using \( \gamma = \exp (\delta) \). This will make the search for reasonable starting values easier as an equidistant grid in \( \delta \) will translate into a grid for \( \gamma \) that is dense in the beginning and then it becomes increasingly coarser.

We now analyze the LSTR model in (6) with the PMI as transition variable using both Fama and Bliss (1987) and Gürkaynak, Sack and Wright (2007) yields. We use the threshold 44.48 from Berge and Jordà (2011) and start in 1972:1 to ensure that the panel is balanced. For both datasets, we estimate \( \delta \) to be 5.0. Figure 2 illustrates how this high value of \( \delta \) makes the transition between regimes practically
instant.

Figure 3 plots estimates of $\alpha$ and $\beta$. As in the simple switching regression in (1), the coefficients flip sign across regimes. In the expansion state, the slope is increasingly negative as a function of maturity. By contrast, in the recession state, the slope is increasingly positive as a function of maturity. Hence, the LSTR model confirms the contrast between expansions and recessions in predictability patterns on the bond market and that the transition between states of the economy happens basically instantly as captured by a dummy variable rather than smoothly.

Figure 2: Instant switch

![Graph showing an instant switch](image-url)
6 Out-of-sample evidence

We run simple forecasting regressions

\[ x_{hpr_{t+m,k}} = \mu_k + \theta_k x_t + e_{t+m,k}, \]  

where \( x_{hpr_{t+m,k}} \) is the log excess return on a \( k \)-period bond from time \( t \) to \( t + m \) and \( x_t \) is either the yield spread, forward spread, or the Cochrane-Piazzesi factor. As in the article, we set \( m = 3 \). To analyze the out-of-sample performance of the models, we use an expanding (recursive) estimation window. We reserve the first \( q \) observations for the initial estimates that generate the first forecast. The first prediction becomes

\[ \hat{x}_{hpr_{q+m,k}} = \hat{\mu}_{k,q} + \hat{\theta}_{k,q} x_q, \]  

where \( \hat{\mu}_{k,q} \) and \( \hat{\theta}_{k,q} \) are the estimates of \( \mu_k \) and \( \theta_k \) in (8) from the regression of \( \{x_{hpr_{t,k}}\}_{t=1+m}^q \) on a constant and \( \{x_t\}_{t=1}^{q-m} \). The next forecast is then generated from the regression of \( \{x_{hpr_{t,k}}\}_{t=1+m}^{q+1} \) on a constant and \( \{x_t\}_{t=1}^{q-m+1} \), and so on. As in the above sections, we set \( m = 3 \). Similar to e.g. Rapach, Strauss and Zhou (2010) and Rapach and Zhou (2013), we divide the forecast errors between
expansions and recessions based on the state of the economy at the time of the forecast. We consider both the PMI and NBER recession indicator.

Following the existing literature on return predictability, we compare the forecasts of the regression models with the historical average return. For the out-of-sample study, the benchmark is recomputed each period (i.e. estimated using an expanding window). We evaluate the performance of the models using the Campbell and Thompson (2008) out-of-sample $R^2$ statistic ($R^2_{OoS}$). To take sampling error into account, we conduct a formal statistical test of predictability. We test the null that $R^2_{OoS} \leq 0$ (no predictability) against the alternative that $R^2_{OoS} > 0$ using the Clark and West (2007) statistic for which the standard normal distribution is a good approximation asymptotically when comparing forecasts from nested models. We use Newey-West standard errors with $m$ lags to take the overlap in forecast errors into account.

Table 2 shows results for an out-of-sample period running from 1980:1 to 2013:12. In recessions the $R^2_{OoS}$ is typically close to zero or even negative and the Clark-West test does not reject the null of no return predictability. However, in expansions the $R^2_{OoS}$ is generally positive and the Clark-West test conclusively rejects the null of no return predictability. Hence, bond returns appear unpredictable during recessions but are strongly predictable during expansions. We find this result using the most classical bond return predictors (the yield spread, forward spread, and the CP factor). Further, the results are robust across a wide range of bond maturities.
The table shows results from forecasting excess bond returns using either the yield spread, the forward spread, or the CP factor. We use Fama-Bliss (1987) data to compute excess bond returns and consider maturities from 2-10 years. The forecast horizon is three months. We report the Campbell-Thompson (2008) out-of-sample $R^2$ statistic in percent. In parentheses are $p$-values from the Clark and West (2007) statistic. We take autocorrelation into account by using the Newey-West estimator. The evaluation period is from 1980:1 to 2013:12. We report separate results for expansions and recessions. We use either PMI (Panel A) or NBER (Panel B) recession dates to define the state of the economy.

<table>
<thead>
<tr>
<th>k</th>
<th>Yield spread</th>
<th></th>
<th>Forward spread</th>
<th></th>
<th>CP factor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_{REC}$</td>
<td>$R^2_{EXP}$</td>
<td>$R^2_{REC}$</td>
<td>$R^2_{EXP}$</td>
<td>$R^2_{REC}$</td>
<td>$R^2_{EXP}$</td>
</tr>
<tr>
<td>Panel A: PMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-0.77 (0.42)</td>
<td>6.90** (0.00)</td>
<td>-1.82 (0.47)</td>
<td>6.02** (0.00)</td>
<td>-4.03 (0.69)</td>
<td>3.40** (0.02)</td>
</tr>
<tr>
<td>48</td>
<td>0.57 (0.34)</td>
<td>9.20** (0.00)</td>
<td>-6.40 (0.74)</td>
<td>4.60** (0.01)</td>
<td>3.50 (0.21)</td>
<td>7.28** (0.00)</td>
</tr>
<tr>
<td>72</td>
<td>-3.71 (0.51)</td>
<td>7.22** (0.00)</td>
<td>-4.01 (0.57)</td>
<td>5.28** (0.00)</td>
<td>-0.78 (0.40)</td>
<td>3.79** (0.01)</td>
</tr>
<tr>
<td>96</td>
<td>-5.46 (0.60)</td>
<td>6.19** (0.00)</td>
<td>-2.77 (0.43)</td>
<td>4.54** (0.00)</td>
<td>0.17 (0.36)</td>
<td>3.36** (0.01)</td>
</tr>
<tr>
<td>120</td>
<td>-1.48 (0.41)</td>
<td>6.37** (0.00)</td>
<td>-1.43 (0.39)</td>
<td>-0.86 (0.40)</td>
<td>-0.02 (0.36)</td>
<td>2.72** (0.02)</td>
</tr>
<tr>
<td>Panel B: NBER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-1.15 (0.46)</td>
<td>9.01** (0.00)</td>
<td>-4.39 (0.60)</td>
<td>9.79** (0.00)</td>
<td>-5.16 (0.75)</td>
<td>6.00** (0.00)</td>
</tr>
<tr>
<td>48</td>
<td>-2.86 (0.50)</td>
<td>12.47** (0.00)</td>
<td>-8.46 (0.80)</td>
<td>7.49** (0.00)</td>
<td>-1.20 (0.44)</td>
<td>10.46** (0.00)</td>
</tr>
<tr>
<td>72</td>
<td>-6.28 (0.64)</td>
<td>9.95** (0.00)</td>
<td>-4.29 (0.64)</td>
<td>6.70** (0.00)</td>
<td>-3.68 (0.60)</td>
<td>5.81** (0.00)</td>
</tr>
<tr>
<td>96</td>
<td>-8.44 (0.72)</td>
<td>8.67** (0.00)</td>
<td>-6.88 (0.63)</td>
<td>7.03** (0.00)</td>
<td>-2.77 (0.58)</td>
<td>4.93** (0.00)</td>
</tr>
<tr>
<td>120</td>
<td>-4.31 (0.55)</td>
<td>8.04** (0.00)</td>
<td>-4.65 (0.59)</td>
<td>0.36 (0.20)</td>
<td>-1.54 (0.52)</td>
<td>3.50** (0.01)</td>
</tr>
</tbody>
</table>
This section presents an ATSM where the \( P \) dynamics for the pricing factors evolve as a Self Exciting Threshold vector Autoregression (SETAR) model with constant volatility. The \( Q \) dynamics for the factors are assumed to evolve as a standard VAR model, implying that the pricing of bonds is standard. We first present the model in Section 7.1. The expression for zero-coupon bond prices and hence the yield curve is derived in Section 7.2. The change of measure is derived in Section 7.3. The expression for excess returns is derived in Section 7.4. Campbell-Shiller regressions are considered in Section 7.5. Estimation of the SETAR model within the SR approach is discussed in Section 7.6. Section 7.7 describes how to obtain estimates for the market prices of risk. We describe the stationarity conditions for the SETAR model in Section 7.8 and present a data-driven calibration procedure to induce stationarity. Section 7.9 presents the normalization restriction for the model. We finally consider the zero lower bound in Section 7.10.

### 7.1 Model description

The short-term risk-free interest rate \( r_t \) is assumed to be an affine function of the pricing factors \( \mathbf{x}_t \) with dimension \( n_x \times 1 \), i.e.

\[
r_t = \alpha + \beta' \mathbf{x}_t, \tag{10}
\]

where \( \alpha \) is a scalar and \( \beta \) is an \( n_x \times 1 \) vector. The pricing factors is assumed to follow a VAR(1) process under the risk-neutral measure \( \mathbb{Q} \), i.e.

\[
\mathbf{x}_{t+1} = (\mathbf{I} - \Phi) \mathbf{x}_t + \Phi \mathbf{\mu} + \Sigma_x \epsilon_x^{Q,t+1}, \tag{11}
\]

where \( \epsilon_{x,t+1} \sim \mathcal{NID}(0, \mathbf{I}) \). Here, \( \Phi \) is an \( n_x \times n_x \) matrix, \( \mathbf{\mu} \) is an \( n_x \times 1 \) vector, and \( \Sigma_x \) is an \( n_x \times n_x \) matrix. The dynamics of the pricing factors under the physical measure \( \mathbb{P} \) is given by a SETAR model of the form

\[
\begin{bmatrix}
\mathbf{x}_{t+1} \\
z_{t+1}
\end{bmatrix} =
\begin{bmatrix}
1_{[z_t \geq c]} (\mathbf{h}_0^{(1)} + \mathbf{h}_x^{(1)} \mathbf{x}_t) + (1 - 1_{[z_t \geq c]}) (\mathbf{h}_0^{(2)} + \mathbf{h}_x^{(2)} \mathbf{x}_t) \\
\gamma_0 + \gamma_z z_t + \gamma_x x_t
\end{bmatrix}
\begin{bmatrix}
\Sigma_x \\
\Sigma_{zz}
\end{bmatrix}
\begin{bmatrix}
\epsilon_{x,t+1} \\
\epsilon_{z,t+1}
\end{bmatrix}, \tag{12}
\]

where the superscript on \( \mathbf{h}_0 \) of dimension \( n_x \times 1 \) and \( \mathbf{h}_x \) of dimension \( n_x \times n_x \) indicates the given regime as determined by the switching variable \( z_t \). We will consider the case where \( z_t \) is some economic indicator to identify the two regimes "expansions" and "recessions", implying that the dynamics of the pricing factors \( \mathbf{x}_t \) switch between expansions and recessions. The dynamics of \( z_t \) depend on its own lag \( \gamma_z \) and lagged values of the pricing factors to allow the yield curve to affect economic activity (i.e. a feedback effect). Hence, \( \gamma_x \) has dimension \( n_x \times 1 \).
If $\gamma_x = \beta \times \tau$, where $\tau$ is some scalar, then it is only the policy rate which affects economic activity (i.e. the short-rate is a sufficient statistic). If $\gamma_x = 0$ and $\Sigma_{zx} = 0$, the SETAR model reduces to a threshold VAR model, as the switching variable is exogenous for the pricing factors. Note that the ordering of the variables in (12) implies that innovations to the pricing factors $\varepsilon_{x,t+1}^2$ also affects $z_{t+1}$ if $\Sigma_{zx} \neq 0$. That is, a disturbance in the bond market will directly affect economic activity, which is a strong assumption. We will therefore require that $\Sigma_{zx} = 0$ to eliminate direct effects which seem contrary to the sluggish behavior of economic activity.

We also note that this model features unspanned macro risk because the economic variable $z_t$ does not enter in the short rate and the risk-neutral distribution, and the model is therefore related to the recent work of Joslin, Priebsch and Singleton (2014).

### 7.2 Pricing of zero-coupon bonds

No arbitrage implies that the price of a $k$-period zero-coupon bond at time $t$, denoted $P_{t,k}$, is equal to the expected price of a $k-1$-period bond at time $t+1$ discounted by the risk-free rate, i.e.

$$P_{t,k} = E_t^Q \left[ \exp \left[ -r_t \right] P_{t+1,k-1} \right], \quad (14)$$

where $E_t^Q$ denotes the conditional expectation under the risk-neutral probability measure. Within this setting bond prices are exponentially affine in the pricing factors, i.e.

$$P_{t,k} = \exp \left\{ A_k + B_k' x_t \right\}, \quad (15)$$

and the yield curve (continuously compounded) is therefore given by

$$y_{t,k} = -\frac{1}{k} A_k - \frac{1}{k} B_k' x_t.$$

To show this, we substitute equation (15) into equation (14) and obtain

$$\exp \left\{ A_k + B_k' x_t \right\} = E_t^Q \left[ \exp \left\{ -r_t \right\} \exp \left\{ A_{k-1} + B_{k-1}' x_{t+1} \right\} \right].$$
Taking logs and substituting in equations (10) and (11) we get

\[ A_k + B'_k x_t = \log E_t^Q \left[ \exp \left\{ -\alpha - \beta' x_t \right\} \right] \]
\[ \times \exp \left\{ A_{k-1} + B'_{k-1} \left[ (I - \Phi) x_t + \Phi \mu + \Sigma_x \epsilon_{x,t+1}^Q \right] \right\} \]
\[ = -\alpha - \beta' x_t + A_{k-1} + B'_{k-1} \left[ (I - \Phi) x_t + \Phi \mu \right] + \log E_t^Q \left[ \exp \left\{ B'_{k-1} \Sigma_x \epsilon_{x,t+1}^Q \right\} \right] \]
\[ = -\alpha - \beta' x_t + A_{k-1} + B'_{k-1} \left[ (I - \Phi) x_t + \Phi \mu \right] + \log \left( \exp \left\{ \frac{1}{2} B'_{k-1} \Sigma_x \Sigma_x' B_{k-1} \right\} \right) \]
\[ = -\alpha - \beta' x_t + A_{k-1} + B'_{k-1} \left( (I - \Phi) x_t + B_{k-1}' \Phi \mu \right) + \frac{1}{2} B'_{k-1} \Sigma_x \Sigma_x' B_{k-1}. \]

Matching coefficients gives

\[ A_k = -\alpha + A_{k-1} + B'_{k-1} \Phi \mu + \frac{1}{2} B'_{k-1} \Sigma_x \Sigma_x' B_{k-1} \quad (16) \]
\[ B'_k = -\beta' + B'_{k-1} (I - \Phi). \quad (17) \]

These recursions are started at \( A_0 = 0 \) and \( B_0 = 0 \) to ensure that \( A_1 = -\alpha \) and \( B'_1 = -\beta' \) as desired, i.e.

\[ \log P_{t,1} = \log E_t^Q \{ \exp [-r_t] \} = -r_t = -\alpha - \beta' x_t. \]

It is instructive to observe the following. First, note that

\[ B'_1 = -\beta' \]
\[ B'_2 = -\beta' + B'_1 (I - \Phi) = -\beta' - \beta' (I - \Phi) \]
\[ B'_3 = -\beta' + B'_2 (I - \Phi) = -\beta' - \beta' (I - \Phi) - \beta' (I - \Phi)^2. \]

Hence, in general

\[ B'_k = -\beta' \sum_{i=0}^{k-1} (I - \Phi)^i \]
\[ = -\beta' \Phi^{-1} \left( I - (I - \Phi)^k \right), \]

because \( S \equiv \sum_{i=0}^{k-1} (I - \Phi)^i \), so \( S - S (I - \Phi) = I - (I - \Phi)^k \), and therefore

\[ S = \Phi^{-1} \left( I - (I - \Phi)^k \right). \]
Secondly,

\[ A_1 = -\alpha \]
\[ A_2 = -\alpha + A_1 + B'_1 \Phi \mu + \frac{1}{2} B'_1 \Sigma_x \Sigma'_x B_1 \]
\[ = -2\alpha - \beta' \Phi \mu + \frac{1}{2} \beta' \Sigma_x \Sigma'_x \beta. \]

and

\[ A_3 = -\alpha + A_2 + B'_2 \Phi \mu + \frac{1}{2} B'_2 \Sigma_x \Sigma'_x B_2 \]
\[ = -\alpha - 2\alpha - \beta' \Phi \mu + \frac{1}{2} \beta' \Sigma_x \Sigma'_x \beta - \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^2 \right) \beta \]
\[ = -3\alpha - \beta' \left( I + \Phi^{-1} \left( I - (I - \Phi)^2 \right) \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Sigma_x \Sigma'_x \beta + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^2 \right) \beta \]
\[ = -3\alpha - \beta' \Phi^{-1} \left( I - (I - \Phi)^1 + \left( I - (I - \Phi)^2 \right) \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Sigma_x \Sigma'_x \beta + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^2 \right) \beta. \]

Next,

\[ A_4 = -\alpha + A_3 + B'_3 \Phi \mu + \frac{1}{2} B'_3 \Sigma_x \Sigma'_x B_3 \]
\[ = -\alpha - 3\alpha - \beta' \Phi^{-1} \left( I - (I - \Phi)^1 + \left( I - (I - \Phi)^2 \right) \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Sigma_x \Sigma'_x \beta + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^2 \right) \beta \]
\[ - \beta' \Phi^{-1} \left( I - (I - \Phi)^3 \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^3 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^3 \right) \beta \]
\[ = -4\alpha - \beta' \Phi^{-1} \left( I - (I - \Phi)^1 + \left( I - (I - \Phi)^2 \right) + \left( I - (I - \Phi)^3 \right) \right) \Phi \mu \]
\[ + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^1 \right) \Sigma_x \Sigma'_x \left( \Phi^{-1} \left( I - (I - \Phi)^1 \right) \right) \beta \]
\[ + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^2 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^2 \right) \beta \]
\[ + \frac{1}{2} \beta' \Phi^{-1} \left( I - (I - \Phi)^3 \right) \Sigma_x \Sigma'_x \Phi^{-1} \left( I - (I - \Phi)^3 \right) \beta. \]
Thus, in general we have

\[
A_k = -k\alpha - \beta' \sum_{i=1}^{k-1} \Phi^{-1}(I - (I - \Phi)^i) \Phi \mu \\
+ \frac{1}{2} \beta' \sum_{i=1}^{k-1} \Phi^{-1}(I - (I - \Phi)^i) \Sigma_x \Sigma_x'(\Phi^{-1}(I - (I - \Phi)^{k-i}))' \beta.
\]

### 7.3 Change of measure

We next specify the market price of risk to obtain the dynamics under the physical probability measure \(P\), or equivalently, we derive the market prices of risk which implies the SETAR model presented above. We assume a Radon-Nikodym derivative of the form

\[
\frac{dP}{dQ} = \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t + \epsilon_{x,t+1}' \lambda_t \right\},
\]

where \(\lambda_t\) is an \(n_x \times 1\) vector for the market price of risk in period \(t\). Since \(\epsilon_{x,t+1} \sim \mathcal{N}(0, I)\) under the risk-neutral probability measure, its density under this measure is given by

\[
Q(\epsilon_{x,t+1}) = \frac{1}{(2\pi)^{n_x/2}} \exp \left\{ -\frac{1}{2} \epsilon_{x,t+1}' \epsilon_{x,t+1} \right\}.
\]

Hence, we have that the physical probability distribution is given by

\[
\mathbb{P}(\epsilon_{x,t+1}) = Q(\epsilon_{t+1}) \frac{dP}{dQ} = \frac{1}{(2\pi)^{n_x/2}} \exp \left\{ -\frac{1}{2} \epsilon_{x,t+1}' \epsilon_{x,t+1} \right\} \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t + \epsilon_{x,t+1}' \lambda_t \right\}
\]

\[
= \frac{1}{(2\pi)^{n_x/2}} \exp \left\{ -\frac{1}{2} \epsilon_{x,t+1}' \epsilon_{x,t+1} - \frac{1}{2} \lambda_t' \lambda_t + \epsilon_{x,t+1}' \lambda_t \right\}
\]

\[
= \frac{1}{(2\pi)^{n_x/2}} \exp \left\{ -\frac{1}{2} (\epsilon_{x,t+1}' \epsilon_{x,t+1} + \lambda_t' \lambda_t - 2\epsilon_{x,t+1}' \lambda_t) \right\}
\]

\[
= \frac{1}{(2\pi)^{n_x/2}} \exp \left\{ -\frac{1}{2} (\epsilon_{x,t+1}' \epsilon_{x,t+1} - \lambda_t)' (\epsilon_{x,t+1} - \lambda_t) \right\}.
\]

Thus,

\[
\epsilon_{x,t+1}^P = \epsilon_{x,t+1}^Q - \lambda_t
\]

\[
\Downarrow
\]

\[
\epsilon_{x,t+1}^Q = \epsilon_{x,t+1}^P + \lambda_t.
\]

149
Hence, the general dynamics for the pricing factors under the physical probability measure is

\[ x_{t+1} = \Phi \mu + (I - \Phi) x_t + \Sigma_x \epsilon^Q_{x,t+1} \]

\[ = \Phi \mu + (I - \Phi) x_t + \Sigma_x (\epsilon^p_{x,t+1} + \lambda_t) \]

\[ = \Phi \mu + (I - \Phi) x_t + \Sigma_x \lambda_t + \Sigma_x \epsilon^p_{x,t+1}. \]

Thus, to have a SETAR model for the pricing factors is equivalent to postulating the following specification for the market price of risk

\[ \lambda_t \equiv 1_{[z_t \geq c]} \sum^{-1}_x \left( \lambda_0^{(1)} + \lambda_x^{(1)} x_t \right) + (1 - 1_{[z_t \geq c]}) \left( \sum^{-1}_x \left( \lambda_0^{(2)} + \lambda_x^{(2)} x_t \right) \right), \]

as it implies

\[ x_{t+1} = \Phi \mu + (I - \Phi) x_t + \Sigma_x \lambda_t + \Sigma_x \epsilon^p_{x,t+1} \]

\[ = \Phi \mu + (I - \Phi) x_t + \left( 1_{[z_t \geq c]} \left( \lambda_0^{(1)} + \lambda_x^{(1)} x_t \right) + (1 - 1_{[z_t \geq c]}) \left( \lambda_0^{(2)} + \lambda_x^{(2)} x_t \right) \right) \]

\[ + \Sigma_x \epsilon^p_{x,t+1} \]

\[ = \Phi \mu + 1_{[z_t \geq c]} \lambda_0^{(1)} + (1 - 1_{[z_t \geq c]}) \lambda_0^{(2)} \]

\[ + \left( (I - \Phi) + 1_{[z_t \geq c]} \lambda_x^{(1)} + (1 - 1_{[z_t \geq c]}) \lambda_x^{(2)} \right) x_t + \Sigma_x \epsilon^p_{x,t+1}. \]

Hence, we have

\[ h_0^{(1)} = \Phi \mu + \lambda_0^{(1)} \quad \text{and} \quad h_0^{(2)} = \Phi \mu + \lambda_0^{(2)}, \]

and

\[ h_x^{(1)} = I - \Phi + \lambda_x^{(1)} \quad \text{and} \quad h_x^{(2)} = I - \Phi + \lambda_x^{(2)}. \]

Note finally that \( M_{t+1} \equiv \frac{dP}{dQ} e^{-r_t} = \exp \left\{ -r_t - \frac{1}{2} \lambda'_t \lambda_t - \epsilon'_{x,t+1} \lambda_t \right\} \) is the stochastic discount factor (or the ratio of state-price deflators) as

\[ E_t^P \left[ M_{t+1} e^{r_t} \right] = 1 \]

\[ \downarrow \]

\[ E_t^P \left[ \frac{dP}{dQ} e^{-r_t} e^{r_t} \right] = 1 \]

\[ \downarrow \]

\[ E_t^P \left[ \frac{dP}{dQ} \right] = 1 \]

\[ \downarrow \]

\[ E_t^P \left[ \exp \left\{ -\frac{1}{2} \lambda'_t \lambda_t - \epsilon'_{x,t+1} \lambda_t \right\} \right] = 1 \]

150
\[ \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t \right\} E_t^p \left[ \exp \left\{ -\lambda_t' \epsilon_{x,t+1} \right\} \right] = 1 \]

\[ \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t \right\} \exp \left\{ \frac{1}{2} \text{Var}_t^p (\lambda_t' \epsilon_{x,t+1}) \right\} = 1 \]

\[ \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t \right\} \exp \left\{ \frac{1}{2} E_t^p \left[ (\lambda_t' \epsilon_{x,t+1}) (\lambda_t' \epsilon_{x,t+1})' \right] \right\} = 1 \]

\[ \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t \right\} \exp \left\{ \frac{1}{2} E_t^p \left[ (\lambda_t' \epsilon_{x,t+1}) (\epsilon_{x,t+1} \lambda_t) \right] \right\} = 1 \]

\[ \exp \left\{ -\frac{1}{2} \lambda_t' \lambda_t \right\} \exp \left\{ \frac{1}{2} \lambda_t' \lambda_t \right\} = 1. \]

### 7.4 Excess return

The ex ante holding period return on a given bond in this model is given by

\[ hpr_{t,k} = E_t \left[ \log \left( \frac{P_{t+1,k-1}}{P_{t,k}} \right) \right] \]

\[ = E_t \left[ \log P_{t+1,k-1} - \log P_{t,k} \right] \]

\[ = E_t \left[ A_{k-1} + B_{k-1}' x_{t+1} - A_k - B_k' x_t \right] \]

\[ = A_{k-1} - A_k - B_k' x_t + B_{k-1}' E_t [x_{t+1}] \]

\[ = A_{k-1} - A_k - B_k' x_t + B_{k-1}' \left( 1_{[z_t \geq c]} \left( h_0^{(1)} + h_0^{(1)} x_t \right) + (1 - 1_{[z_t \geq c]}) \left( h_0^{(2)} + h_0^{(2)} x_t \right) \right). \]
Next, recall that $A_k = -\alpha + A_{k-1} + B'_{k-1} \Phi \mu + \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$ and $B'_k = -\beta' + B'_{k-1} (I - \Phi)$. Hence,

$$hpr_{t,k} = A_{k-1} - \left(-\alpha + A_{k-1} + B'_{k-1} \Phi \mu + \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}\right)$$

$$- B'_k x_t$$

$$+ B'_{k-1} \left(1_{[z_t \geq c]} \left(h^{(1)}_0 + h^{(1)}_x x_t\right) + (1 - 1_{[z_t \geq c]}) \left(h^{(2)}_0 + h^{(2)}_x x_t\right)\right)\)$$

$$= \alpha - B'_{k-1} \Phi \mu - \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$$

$$- (-\beta' + B'_{k-1} (I - \Phi)) x_t$$

$$+ B'_{k-1} \left(1_{[z_t \geq c]} \left(h^{(1)}_0 + h^{(1)}_x x_t\right) + (1 - 1_{[z_t \geq c]}) \left(h^{(2)}_0 + h^{(2)}_x x_t\right)\right)\)$$

$$= \alpha + \beta' x_t - B'_{k-1} \Phi \mu - \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$$

$$- B'_k (I - \Phi) x_t$$

$$+ B'_{k-1} \left(1_{[z_t \geq c]} \left(h^{(1)}_0 + h^{(1)}_x x_t\right) + (1 - 1_{[z_t \geq c]}) \left(h^{(2)}_0 + h^{(2)}_x x_t\right)\right)\)$$.

Now note that $(I - \Phi) = 1_{[z_t \geq c]} (I - \Phi) + (1 - 1_{[z_t \geq c]}) (I - \Phi)$ and $B'_{k-1} \Phi \mu = 1_{[z_t \geq c]} B'_{k-1} \Phi \mu + (1 - 1_{[z_t \geq c]}) B'_{k-1} \Phi \mu$ by definition. Hence, we get

$$hpr_{t,k} = \alpha + \beta' x_t - \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$$

$$+ B'_{k-1} 1_{[z_t \geq c]} \left(h^{(1)}_0 - \Phi \mu + \left(h^{(1)}_x - (I - \Phi) x_t\right)\right)$$

$$+ B'_{k-1} \left(1 - 1_{[z_t \geq c]}) \left(h^{(2)}_0 - \Phi \mu + \left(h^{(2)}_x - (I - \Phi) x_t\right)\right)\right)\)$$

$$= r_t - \frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$$

$$+ B'_{k-1} 1_{[z_t \geq c]} \left(\lambda^{(1)}_0 + \lambda^{(1)}_x x_t\right)$$

$$+ B'_{k-1} \left(1 - 1_{[z_t \geq c]}) \left(\lambda^{(2)}_0 + \lambda^{(2)}_x x_t\right)\right)\)$$.

Note also that the excess holding period return therefore is given by

$$xhpr_{t,k} \equiv hpr_{t,k} - r_t$$

$$= -\frac{1}{2} B'_{k-1} \Sigma x \Sigma'_x B_{k-1}$$

$$+ B'_{k-1} 1_{[z_t \geq c]} \left(\lambda^{(1)}_0 + \lambda^{(1)}_x x_t\right)$$

$$+ B'_{k-1} \left(1 - 1_{[z_t \geq c]}) \left(\lambda^{(2)}_0 + \lambda^{(2)}_x x_t\right)\right)\) .
7.5 The Campbell-Shiller regressions

The Campbell-Shiller regression reads

\[ y_{t+1,k-1} - y_{t,k} = \alpha_k + \frac{\beta_k}{k-1} \left(y_{t,k} - r_t\right) + u_{t+1,k}. \]

To obtain closed-form solutions for the regression loadings in our DTSM, we ignore regime-switching in this section and consider the $P$ process

\[ x_{t+1} = h_0 + h_x x_t + \Sigma_x \epsilon_{x,t+1}^p, \]

or equivalently

\[ x_{t+1} = (\Phi \mu + \lambda_0) + (I - \Phi + \lambda_x) x_t + \Sigma_x \epsilon_{x,t+1}^p, \]

where the first two unconditional moments are

\[ E[x_t] = (I - h_x)^{-1} h_0, \]

and

\[ vec(Var[x_t]) = (I - h_x \otimes h_x)^{-1} vec(\Sigma_x \Sigma_x'). \]

The variable on the left hand side is

\[ y_{t+1,k-1} - y_{t,k} = \left( -\frac{1}{k-1} A_{k-1} - \frac{1}{k-1} B'_{k-1} x_{t+1} \right) - \left( -\frac{1}{k} A_k - \frac{1}{k} B'_{k} x_t \right) \]

\[ = -\frac{1}{k-1} A_{k-1} - \frac{1}{k-1} B'_{k-1} x_{t+1} + \frac{1}{k} A_k + \frac{1}{k} B'_{k} x_t. \]

\[ = -\frac{1}{k-1} A_{k-1} - \frac{1}{k-1} B'_{k-1} (h_0 + h_x x_t) \]

\[ - \frac{1}{k-1} B'_{k-1} \Sigma_x \epsilon_{x,t+1}^p + \frac{1}{k} A_k + \frac{1}{k} B'_{k} x_t. \]

\[ = -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} h_0 \]

\[ + \left[ \frac{1}{k} B'_k - \frac{1}{k-1} B'_{k-1} h_x \right] x_t - \frac{1}{k-1} B'_{k-1} \Sigma_x \epsilon_{x,t+1}^p \]

\[ = \omega_{k,0} + \omega'_{k,x} x_t - \frac{1}{k-1} B'_{k-1} \Sigma_x \epsilon_{x,t+1}^p, \]

where we have defined

\[ \omega_{k,0} \equiv -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} h_0 \]

\[ \omega'_{k,x} \equiv \frac{1}{k} B'_k - \frac{1}{k-1} B'_{k-1} h_x. \]
The slope is given by

\[
\frac{1}{k - 1} (y_{t,k} - r_t) = \frac{1}{k - 1} \left( -\frac{1}{k^2} A_k - \frac{1}{k^2} B_k x_t - \alpha - \beta' x_t \right)
\]

\[
= \frac{1}{k - 1} \left( -\frac{1}{k} A_k - \alpha \right) - \frac{1}{k - 1} \left( \frac{1}{k} B_k' + \beta' \right) x_t
\]

\[
= \gamma_{k,0} + \gamma_{k,x}' x_t,
\]

where we have defined

\[
\gamma_{k,0} \equiv \frac{1}{k - 1} \left( -\frac{1}{k} A_k - \alpha \right)
\]

\[
\gamma_{k,x}' \equiv -\frac{1}{k - 1} \left( \frac{1}{k} B_k' + \beta' \right).
\]

Next, recall the OLS solution for \( y_t = \alpha + \beta x_t + \varepsilon_t \) which reads (in the population)

\[
\hat{\alpha} = E[y_t] - \hat{\beta} E[x_t]
\]

\[
\hat{\beta} = \frac{E[(x_t - E[x_t])(y_t - E[y_t])]}{E[(x_t - E[x_t])^2]}.
\]
Applied to our problem we thus have that

\[
\hat{\beta}_k = \frac{E \left[ \left( \gamma_{k,0} + \gamma_{k,x} x_t - E \left[ \gamma_{k,0} + \gamma_{k,x} x_t \right] \right) \times \left( \omega_{k,0} + \omega_{k,x} x_t - \frac{1}{k-1} B_{k-1}^\prime \Sigma_x \epsilon_{x,t+1}^p - E \left[ \omega_{k,0} + \omega_{k,x} x_t \right] \right) \right]}{E \left[ \left( \gamma_{k,0} + \gamma_{k,x} x_t - E \left[ \gamma_{k,0} + \gamma_{k,x} x_t \right] \right)^2 \right]} \times E \left[ \gamma'_{k,x} (x_t - E [x_t]) \left( \omega'_{k,x} (x_t - E [x_t]) - \frac{1}{k-1} B_{k-1}^\prime \Sigma_x \epsilon_{x,t+1}^p \right) \right] \]
\]
\[
= E \left[ \gamma'_{k,x} (x_t - E [x_t]) \left( \omega'_{k,x} (x_t - E [x_t]) \right) \right] \times \frac{E \left[ \left( \gamma_{k,x} (x_t - E [x_t]) \right)^2 \right]}{E \left[ \left( \gamma'_{k,x} (x_t - E [x_t]) \right)^2 \right]} \times \frac{E \left[ \gamma'_{k,x} Var [x_t] \omega_{k,x} \gamma_{k,x} \right]}{E \left[ \gamma'_{k,x} Var [x_t] \gamma_{k,x} \right]}
\]
\[
= \frac{\gamma'_{k,x} Var [x_t] \left( \frac{1}{k} B_k - \frac{1}{k-1} h_x B_{k-1} \right)}{\gamma'_{k,x} Var [x_t] \gamma_{k,x} \gamma_{k,x}} \times \frac{E \left[ \gamma_{k,x} Var [x_t] \left( \frac{1}{k} B_k - \frac{1}{k-1} \left( (I - \Phi) + \lambda_x \right) b_{k-1} \right) \right]}{E \left[ \gamma_{k,x} Var [x_t] \gamma_{k,x} \right]}
\]
\[
= \frac{\gamma'_{k,x} Var [x_t] \left( \frac{1}{k} B_k - \frac{1}{k-1} \left( (I - \Phi) \right) b_{k-1} \right)}{\gamma'_{k,x} Var [x_t] \gamma_{k,x} \gamma_{k,x}} - \frac{1}{k-1} \frac{\gamma'_{k,x} Var [x_t] \left( \lambda_x \right) b_{k-1} \gamma_{k,x}}{\gamma'_{k,x} Var [x_t] \gamma_{k,x}}
\]

as \( \lambda_x = h_x - (I - \Phi) \). Thus, if \( \lambda_x \) displays regime-switching, then so will \( \hat{\beta}_k \).
As for the intercept

\[ \dot{\alpha}_k = E [\omega_{k,0} + \omega'_{k,x} x_t] - \beta_k E [\gamma_{k,0} + \gamma'_{k,x} x_t] \]

\[ = \omega_{k,0} - \beta_k \gamma_{k,0} + \left( \omega'_{k,x} - \beta_k \gamma'_{k,x} \right) E [x_t] \]

\[ = -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} h_0 - \beta_k \gamma_{k,0} \]

\[ + \left( \frac{1}{k} B_k - \frac{1}{k-1} B_{k-1} h_x - \beta_k \gamma'_{k,x} \right) E [x_t] \]

\[ = -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \left( \lambda_0 + \Phi \mu \right) - \beta_k \gamma_{k,0} \]

\[ + \left( \frac{1}{k} B_k - \frac{1}{k-1} B_{k-1} \left( I - \Phi \right) - \beta_k \gamma'_{k,x} \right) E [x_t] \]

\[ = -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu \]

\[ + \left( \frac{1}{k} B_k - \frac{1}{k-1} B_{k-1} \left( I - \Phi \right) - \beta_k \gamma'_{k,x} \right) E [x_t] \]

\[ - \beta_k \gamma_{k,0} - \frac{1}{k-1} B'_{k-1} \lambda_0 - \frac{1}{k-1} B'_{k-1} \lambda_x E [x_t] \]

\[ = \Pi_k - \beta_k \gamma'_{k,x} E [x_t] - \beta_k \gamma_{k,0} - \frac{1}{k-1} B'_{k-1} \lambda_0 - \frac{1}{k-1} B'_{k-1} \lambda_x E [x_t] \]

\[ = \Pi_k - \beta_k \left( \gamma'_{k,x} E [x_t] + \gamma_{k,0} \right) - \frac{1}{k-1} B'_{k-1} \lambda_0 - \frac{1}{k-1} B'_{k-1} \lambda_x E [x_t], \]

as \( \lambda_0 = h_0 - \Phi \mu \) and

\[ \Pi_k \equiv -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu + \left( \frac{1}{k} B_k - \frac{1}{k-1} B'_{k-1} \left( I - \Phi \right) \right) E [x_t]. \]

Thus, regime-switching in \( \lambda_0 \) and \( \lambda_x \) will generate regime-switching in \( \dot{\alpha}_k \). Note also that

\[ \gamma'_{k,x} \equiv -\frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{ and } E \left[ \frac{1}{k-1} (y_{t,k} - r_t) \right] = E \left[ \gamma_{k,0} + \gamma'_{k,x} x_t \right], \]

which imply that

\[ \beta_k = \frac{\gamma'_{k,x} \text{Var}[x_t]}{\gamma_{k,x} \text{Var}[x_t]} \left( \frac{1}{k} B_k - \frac{1}{k-1} \left( I - \Phi \right)' B_{k-1} \right) \]

\[ = \frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k - \frac{1}{k-1} \left( I - \Phi \right)' B_{k-1} \right) \]

\[ - \frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k + \beta \right) \]

\[ - \frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k + \beta \right) \]

\[ = -\left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k - \frac{1}{k-1} \left( I - \Phi \right)' B_{k-1} \right) \]

\[ + \frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k + \beta \right) \]

\[ + \frac{1}{k-1} \left( \frac{1}{k} B'_k + \beta' \right) \text{Var}[x_t] \left( \frac{1}{k} B_k + \beta \right) \]
Further, we have that
\[ k_1 = 1 + (\frac{1}{k} B_k + \beta') \] due to \( B_k = -\beta + (I - \Phi)' B_{k-1} \iff (I - \Phi)' B_{k-1} = \beta + B_k \)

\[ = - \left( \frac{1}{k} B_k' + \beta' \right) \text{Var}[x_i] \left( 1 - \frac{1}{k} B_{k-1} \right) + \left( \frac{1}{k} B_k' + \beta' \right) \text{Var}[x_i] \left( \frac{1}{k} B_{k-1} \right) \]

Further, we have that
\[
\hat{\alpha}_k = \Pi_k - \hat{\beta}_k E \left[ \frac{y_{t,k} - t_t}{k - 1} \right] = \Pi_k - \hat{\beta}_k E \left[ \frac{y_{t,k} - t_t}{k - 1} \right] - \frac{1}{k - 1} B_k' \lambda_0 - \frac{1}{k - 1} B_k' \lambda_x E [x_i] 
\]

Note also that
\[
\Pi_k = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( \frac{1}{k} B_k' - \frac{1}{k-1} B_k' (I - \Phi) \right) E [x_i]
\]

\[ = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( \frac{1}{k} B_k' - \frac{1}{k-1} (B_k' + \beta') \right) E [x_i] \]

due to \( B_k = -\beta + (I - \Phi)' B_{k-1} \iff B_k' + \beta' = B_{k-1}' (I - \Phi) \)

\[ = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( B_k' \left( \frac{k-1-\beta}{k(k-1)} \right) - \frac{1}{k-1} \beta' \right) E [x_i] 
\]

\[ = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( B_k' \left( \frac{k-1-k}{k(k-1)} \right) - \frac{1}{k-1} \beta' \right) E [x_i] 
\]

\[ = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( B_k' \left( \frac{k-1-k}{k(k-1)} \right) - \frac{1}{k-1} \beta' \right) E [x_i] 
\]

\[ = - \frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B_k' \mu + \left( B_k' \left( \frac{1}{k(k-1)} \right) \right) E [x_i] 
\]

157
Inserting into $\hat{\alpha}_k$ gives

$$\hat{\alpha}_k = \Pi_k - \hat{\beta}_k \left[ \frac{y_{t,k} - r_t}{k-1} \right] - \frac{1}{k-1} B'_{k-1} (\lambda_0 + \lambda_x E [x_t])$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu - \frac{1}{k-1} \left( \frac{1}{k} B'_{k} + \beta' \right) E [x_t]$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu$$

$$- \hat{\beta}_k \left[ \frac{y_{t,k} - r_t}{k-1} \right] - \frac{1}{k-1} \left( B'_{k-1} (\lambda_0 + \lambda_x E [x_t]) + \left( \frac{1}{k} B'_{k} + \beta' \right) E [x_t] \right).$$

To check the formulas, then note that $\hat{\beta}_k$ that under the expectation hypothesis of $\lambda_x = 0$ we should have $\hat{\beta}_k = 1$, which we clearly have. Note also for $\lambda_x = 0$ and $\lambda_0 = 0$, we have

$$\hat{\alpha}_k = \Pi_k - \hat{\beta}_k \left( \gamma'_{k,x} E [x_t] + \gamma_{k,0} \right) - \frac{1}{k-1} B'_{k-1} \lambda_0 - \frac{1}{k-1} B'_{k-1} \lambda_x E [x_t]$$

$$= \Pi_k - \left( \gamma'_{k,x} E [x_t] + \gamma_{k,0} \right)$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu + \left( \frac{1}{k} B'_{k} - \frac{1}{k-1} B'_{k-1} (I - \Phi) \right) E [x_t]$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu + \left( \frac{1}{k} B'_{k} - \frac{1}{k-1} B'_{k-1} (I - \Phi) \right) E [x_t]$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu$$

$$- \hat{\beta}_k \left( \gamma'_{k,x} E [x_t] + \gamma_{k,0} \right)$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu - \gamma_{k,0}$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu - \frac{1}{k-1} \left( -\frac{1}{k} A_k - \alpha \right)$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu + \frac{1}{k-1} \left( \frac{1}{k} A_k + \alpha \right)$$

$$= -\frac{1}{k-1} A_{k-1} + \frac{1}{k} \left( 1 + \frac{1}{k-1} \right) A_k - \frac{1}{k-1} B'_{k-1} \Phi \mu + \frac{1}{k-1} \alpha$$
\[ = - \frac{1}{k-1} A_k + \frac{1}{k} \left( \frac{k-1+1}{k-1} \right) A_k - \frac{1}{k-1} B_{k-1}' \Phi \mu + \frac{1}{k-1} \alpha \]

\[ = \frac{1}{k-1} \left( B_{k-1}' \Phi \mu + \frac{1}{2} B_{k-1}' \Sigma_{\delta} \Sigma_{\delta}' B_{k-1} \right) - \frac{1}{k-1} B_{k-1}' \Phi \mu, \]

because \( A_k = -\alpha + A_{k-1} + B_{k-1}' \Phi \mu + \frac{1}{2} B_{k-1}' \Sigma_{\delta} \Sigma_{\delta}' B_{k-1} \)

\[ = \frac{1}{k-1} \frac{1}{2} B_{k-1}' \Sigma_{\delta} \Sigma_{\delta}' B_{k-1}. \]

that is, we are left with the convexity adjustment which is typically very small.

### 7.6 Estimation of the SETAR model with generated regressors

This section describes how to estimate the SETAR model by GMM within the SR approach, i.e. when accounting for measurement errors in the estimated pricing factors. We let \( 1_{[z_t \geq c]} \) denote the indicator variable which is one if \( z_t \geq c \), otherwise the variable is zero. We then consider the SETAR model

\[
\begin{align*}
x_{t+1} &= 1_{[z_t \geq c]} \left( h_0^{(1)} + h_\delta^{(1)} x_t \right) + (1 - 1_{[z_t \geq c]}) \left( h_0^{(2)} + h_\delta^{(2)} x_t \right) + \Sigma_\delta \epsilon_{t+1} \\
&= 1_{[z_t \geq c]} h_0^{(1)} + h_\delta^{(1)} 1_{[z_t \geq c]} x_t \\
&\quad + (1 - 1_{[z_t \geq c]}) h_0^{(2)} + h_\delta^{(2)} (1 - 1_{[z_t \geq c]}) x_t + \Sigma_\delta \epsilon_{t+1} \\
&= 1_{[z_t \geq c]} h_0^{(1)} + (1 - 1_{[z_t \geq c]}) h_0^{(2)} \\
&\quad + h_\delta^{(1)} x_t + h_\delta^{(2)} x_t + w_{x,t+1},
\end{align*}
\]

and

\[ z_{t+1} = \gamma_0 + \gamma_\delta z_t + \gamma_\delta x_t + \Sigma_\delta \epsilon_{z,t+1}, \]

where

\[
\begin{align*}
w_{x,t+1} &\equiv \Sigma_\delta \epsilon_{t+1} \\
x_t^{(1)} &\equiv 1_{[z_t \geq c]} x_t \\
x_t^{(2)} &\equiv (1 - 1_{[z_t \geq c]}) x_t.
\end{align*}
\]

Moreover, \( \epsilon_{t+1} \sim NTD(0, I) \), implying that \( \text{Var}(w_{x,t}) = \Sigma_\delta \Sigma_\delta' \). In relation to the general notation for the SR approach in Andreasen and Christensen (2015), we have

\[ \theta_{22} \equiv \left[ h_0^{(1)}, h_0^{(2)}, h_\delta^{(1)}, h_\delta^{(2)}, \gamma_\delta, \gamma_\delta \right] \quad \text{and} \quad \theta_{12} \equiv \left[ \text{vech}(\text{Var}(w_{x,t})) \right]. \]

However, this SETAR model cannot be estimated using standard estimators because \( x_t \) is unobserved. The feasible model for estimation is given by replacing \( x_t \) with \( \hat{x}_t \).

That is, we consider the following model for the estimation

\[ \hat{x}_{t+1} = 1_{[z_t \geq c]} h_0^{(1)} + (1 - 1_{[z_t \geq c]}) h_0^{(2)} + h_\delta^{(1)} \hat{x}_t + h_\delta^{(2)} \hat{x}_t + \hat{w}_{x,t+1}, \quad (20) \]
\[ z_{t+1} = \gamma_0 + \gamma_z z_t + \gamma'_x \hat{x}_t + \hat{w}_{z,t+1}, \]  

(21)

where \( \hat{w}_{x,t+1} \equiv \Sigma_x \hat{e}_{t+1} \) and \( \hat{w}_{z,t+1} \equiv \Sigma_{zz} \hat{e}_{z,t+1}^p \). The "hat" on the innovations denotes the residuals when using the true coefficients but the estimated pricing factors with measurement errors. From the SR approach we further have

\[ \hat{x}_t = x_t + u_t \quad u_t \sim \mathcal{NID}(0, \text{Var}(\hat{x}_t)) \]  

for a sufficiently large cross-sectional panel of bond prices or bond yields. In (20) we use the notation

\[ \hat{x}_t^{(1)} = 1_{|z_t| \geq c} \hat{x}_t \]
\[ \hat{x}_t^{(2)} = (1 - 1_{|z_t| \geq c}) \hat{x}_t, \]

to denote the estimated factor value in regime 1 and 2, respectively. A similar notation is used for the estimation error in the pricing factors, i.e.

\[ u_t^{(1)} = 1_{|z_t| \geq c} u_t \]
\[ u_t^{(2)} = (1 - 1_{|z_t| \geq c}) u_t. \]

We then observe that

\[ \hat{w}_{x,t+1} = \hat{x}_{t+1} - 1_{|z_t| \geq c} h_0^{(1)} - (1 - 1_{|z_t| \geq c}) h_0^{(2)} - h_x^{(1)} \hat{x}_t^{(1)} + h_x^{(2)} \hat{x}_t^{(2)} \]
\[ = x_{t+1} + u_{t+1} - 1_{|z_t| \geq c} h_0^{(1)} - (1 - 1_{|z_t| \geq c}) h_0^{(2)} \]
\[ - h_x^{(1)} (x_t^{(1)} + u_t^{(1)}) - h_x^{(2)} (x_t^{(2)} + u_t^{(2)}) \]
\[ = x_{t+1} - 1_{|z_t| \geq c} h_0^{(1)} - (1 - 1_{|z_t| \geq c}) h_0^{(2)} \]
\[ - h_x^{(1)} x_t^{(1)} - h_x^{(2)} x_t^{(2)} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \]
\[ = w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)}, \]

and

\[ \hat{w}_{z,t+1} = z_{t+1} - \gamma_0 - \gamma_z z_t - \gamma'_x \hat{x}_t \]
\[ = z_{t+1} - \gamma_0 - \gamma_z z_t - \gamma'_x (x_t + u_t) \]
\[ = z_{t+1} - \gamma_0 - \gamma_z z_t - \gamma'_x x_t - \gamma'_x u_t \]
\[ = w_{z,t+1} - \gamma'_x u_t, \]

which will be useful below. At this point we settle with estimating \( \theta_{22} \) and \( \theta_{12} \), implying that we do not provide an estimator for the threshold value \( c \) for switching between the two regimes. In all of the derivations below, we assume that \( z_t \) is measured without error.
7.6.1 Moment conditions for estimating (20)

To estimate the coefficients in the process for \( x_t \), we consider the following moments:

\[
E \left[ \dot{w}_{x,t+1}1_{[z_t \geq c]} \right] = E \left[ \left( w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \right)^1_{[z_t \geq c]} \right] = 0,
\]

because \( E [w_{x,t+1}] = 0 \) and \( E [u_t] = E [u_{t+1}] = 0 \),

and

\[
E \left[ \dot{w}_{x,t+1} \left( 1 - 1_{[z_t \geq c]} \right) \right] = E \left[ \left( w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \right) \left( 1 - 1_{[z_t \geq c]} \right) \right] = 0.
\]

Next:

\[
\begin{align*}
E \left[ \dot{w}_{x,t+1} \left( \dot{x}_t^{(1)} \right)^\prime \right] &= E \left[ \left( w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \right) \left( x_t^{(1)} + u_t^{(1)} \right) \right]^\prime \\
&= E \left[ \left( w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \right) \left( x_t^{(1)} \right)^\prime \\
&+ \left( w_{x,t+1} + u_{t+1} - h_x^{(1)} u_t^{(1)} - h_x^{(2)} u_t^{(2)} \right) \left( u_t^{(1)} \right) \right]^\prime \\
&= E \left[ w_{x,t+1} \left( x_t^{(1)} \right)^\prime + u_{t+1} \left( x_t^{(1)} \right)^\prime - h_x^{(1)} u_t^{(1)} \left( x_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( x_t^{(1)} \right)^\prime \\
&+ w_{x,t+1} \left( u_t^{(1)} \right)^\prime + u_{t+1} \left( u_t^{(1)} \right)^\prime - h_x^{(1)} u_t^{(1)} \left( u_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( u_t^{(1)} \right) \right]^\prime \\
&= E \left[ 0 + u_{t+1} \left( x_t^{(1)} \right)^\prime - h_x^{(1)} u_t^{(1)} \left( x_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( x_t^{(1)} \right)^\prime \\
&+ w_{x,t+1} \left( u_t^{(1)} \right)^\prime + u_{t+1} \left( u_t^{(1)} \right)^\prime - h_x^{(1)} u_t^{(1)} \left( u_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( u_t^{(1)} \right)^\prime \right]^\prime,
\end{align*}
\]

since \( w_{x,t+1} \) is iid. and thus independent of \( x_t^{(1)} \), \( E \left[ w_{x,t+1} \left( x_t^{(1)} \right)^\prime \right] = 0 \)

\[
\begin{align*}
&= E \left[ 0 - h_x^{(1)} u_t^{(1)} \left( x_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( x_t^{(1)} \right)^\prime \\
&+ w_{x,t+1} \left( u_t^{(1)} \right)^\prime + u_{t+1} \left( u_t^{(1)} \right)^\prime - h_x^{(1)} u_t^{(1)} \left( u_t^{(1)} \right)^\prime - h_x^{(2)} u_t^{(2)} \left( u_t^{(1)} \right)^\prime \right]^\prime.
\end{align*}
\]
\( u_{t+1} \) is a function of \( v_{t+1} \) (the measurement errors in bond yields) and \( x_{t}^{(1)} \) is a function of \( \{ w_{x,i} \}_{i=1}^{t} \),

which imply no correlation between \( u_{t+1} \) and \( x_{t}^{(1)} \) due to no correlation between \( v_{t} \) and \( w_{x,t} \) at all leads and lags.

Hence \( E \left[ u_{t+1} \left( x_{t}^{(1)} \right) \right] = 0 \)

\[ = E \left[ 0 - 0 + w_{x,t+1} \left( u_{t}^{(1)} \right) - h_{x}^{(1)} u_{t}^{(1)} \left( u_{t}^{(1)} \right) - h_{x}^{(2)} u_{t}^{(2)} \left( u_{t}^{(1)} \right) \right] , \]

across both regimes we have \( u_{t} \) is a function of \( v_{t} \) and \( x_{t} \) is a function of \( \{ w_{x,i} \}_{i=1}^{t} \),

which imply no correlation between \( u_{t+1} \) and \( x_{t}^{(1)} \) due to no correlation between \( v_{t} \) and \( w_{x,t} \) at all leads and lags.

Hence \( E \left[ u_{t}^{(1)} \left( x_{t}^{(1)} \right) \right] = 0 \) and \( E \left[ u_{t}^{(2)} \left( x_{t}^{(1)} \right) \right] = 0 \)

\[ = E \left[ 0 + u_{t+1} \left( u_{t}^{(1)} \right) - h_{x}^{(1)} u_{t}^{(1)} \left( u_{t}^{(1)} \right) - h_{x}^{(2)} u_{t}^{(2)} \left( u_{t}^{(1)} \right) \right] . \]

because \( u_{t}^{(1)} \) is a function of \( v_{t} \), and \( v_{t} \) and \( w_{x,t} \) are uncorrelated at all leads and lags.

Hence \( E \left[ w_{x,t+1} \left( u_{t}^{(1)} \right) \right] = 0 \)

\[ = E \left[ \left( 1[z_{t} \geq c] + (1 - 1[z_{t} \geq c]) \right) u_{t+1} \left( u_{t}^{(1)} \right) - h_{x}^{(1)} u_{t}^{(1)} \left( u_{t}^{(1)} \right) - h_{x}^{(2)} u_{t}^{(2)} \left( u_{t}^{(1)} \right) \right] \]

\[ = E \left[ \left( u_{t}^{(1)} + u_{t+1}^{(2)} \right) \left( u_{t}^{(1)} \right) - h_{x}^{(1)} u_{t}^{(1)} \left( u_{t}^{(1)} \right) - h_{x}^{(2)} u_{t}^{(2)} \left( u_{t}^{(1)} \right) \right] \]

\[ = E \left[ u_{t+1}^{(2)} \left( u_{t}^{(1)} \right) - h_{x}^{(1)} u_{t}^{(1)} \left( u_{t}^{(1)} \right) - 0 \right] , \]

given that we are either in regime one or two, \( E \left[ u_{t+1}^{(2)} \left( u_{t}^{(1)} \right) \right] = 0 \) and \( E \left[ u_{t}^{(2)} \left( u_{t}^{(1)} \right) \right] = 0 \)
Next, consider the moments
\[
E \left[ \hat{w}_{t+1} \left( \bar{x}^{(2)}_t \right)' \right] = Cov \left( \mathbf{u}^{(2)}_{t+1}, \mathbf{u}^{(2)}_t \right) - h^{(2)}_x Var \left( \mathbf{u}^{(2)}_t \right),
\]
where the result follows from similar arguments as above.

To estimate the conditional variances, we consider
\[
Var \left( \hat{w}_{x,t+1} \right) = Var \left( w_{x,t+1} + \left( u_{t+1} - h^{(1)}_x u^{(1)}_t - h^{(2)}_x u^{(2)}_t \right) \right)
\]
\[
= Var \left( w_{x,t+1} \right) + Var \left( u_{t+1} - h^{(1)}_x u^{(1)}_t - h^{(2)}_x u^{(2)}_t \right),
\]
since \( Cov \left( w_{x,t+1}, u_{t+1} \right), Cov \left( w_{x,t+1}, h^{(1)}_x u^{(1)}_t \right), \) and \( Cov \left( w_{x,t+1}, h^{(2)}_x u^{(2)}_t \right) \) are zero
\[
= Var \left( w_{x,t+1} \right) + Var \left( u^{(1)}_{t+1} + u^{(2)}_{t+1} - h^{(1)}_x u^{(1)}_t - h^{(2)}_x u^{(2)}_t \right)
\]
\[
= Var \left( w_{x,t+1} \right) + Var \left( u^{(1)}_{t+1} + u^{(2)}_{t+1} \right)
+ h^{(1)}_x Var \left( u^{(1)}_t \right) \left( h^{(1)}_x \right)' + h^{(2)}_x Var \left( u^{(2)}_t \right) \left( h^{(2)}_x \right)'
+ Cov \left( u^{(1)}_{t+1}, -h^{(1)}_x u^{(1)}_t \right) + Cov \left( -h^{(1)}_x u^{(1)}_t, u^{(1)}_{t+1} \right)
+ Cov \left( u^{(2)}_{t+1}, -h^{(2)}_x u^{(2)}_t \right) + Cov \left( -h^{(2)}_x u^{(2)}_t, u^{(2)}_{t+1} \right),
\]
because all remaining co-variances are zero, given that we either are in regime 1 or 2.
\[
= Var \left( w_{x,t+1} \right) + Var \left( u^{(1)}_{t+1} \right) + h^{(1)}_x Var \left( u^{(1)}_t \right) \left( h^{(1)}_x \right)'
+ h^{(2)}_x Var \left( u^{(2)}_t \right) \left( h^{(2)}_x \right)'
- Cov \left( u^{(1)}_{t+1}, u^{(1)}_t \right) \left( h^{(1)}_x \right)'
- h^{(1)}_x Cov \left( u^{(1)}_t, u^{(1)}_{t+1} \right)
- Cov \left( u^{(2)}_{t+1}, u^{(2)}_t \right) \left( h^{(2)}_x \right)'
- h^{(2)}_x Cov \left( u^{(2)}_t, u^{(2)}_{t+1} \right)
= Var \left( w_{x,t+1} \right) + \Omega_{t+1},
\]
where

\[ \Omega_{t+1} \equiv \text{Var}(u_{t+1}) + h_x^{(1)} \text{Var}(u_t^{(1)}) (h_x^{(1)})' + h_x^{(2)} \text{Var}(u_t^{(2)}) (h_x^{(2)})' \]

\[ - \text{Cov}(u_{t+1}^{(1)}, u_t^{(1)}) (h_x^{(1)})' - h_x^{(1)} \text{Cov}(u_t^{(1)}, u_{t+1}^{(1)}) \]

\[ - \text{Cov}(u_{t+1}^{(2)}, u_t^{(2)}) (h_x^{(2)})' - h_x^{(2)} \text{Cov}(u_t^{(2)}, u_{t+1}^{(2)}). \]

Thus in general, the moment conditions are

\[ \begin{bmatrix}
E [\bar{w}_{x,t+1} 1_{[z_t \geq c]}] \\
E [\bar{w}_{x,t+1} (1 - 1_{[z_t \geq c]})] \\
E [\bar{w}_{x,t+1} (\hat{x}_t^{(1)})'] \\
E [\bar{w}_{x,t+1} (\hat{x}_t^{(2)})'] \\
\text{vech} (\text{Var}(\bar{w}_{x,t+1}))
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
\text{Cov}(u_t^{(1)}, u_t^{(1)}) - h_x^{(1)} \text{Var}(u_t^{(1)}) \\
\text{Cov}(u_t^{(2)}, u_t^{(2)}) - h_x^{(2)} \text{Var}(u_t^{(2)}) \\
\text{vech} (\text{Var}(w_{x,t+1} + \Omega_{t+1}))
\end{bmatrix}. \]

The population moments are estimated by empirical moments, meaning that the moment conditions are

\[ q_{x,T} = \frac{1}{T-1} \sum_{t=1}^{T-1} q_{x,t}(\theta_2) = 0, \]

where

\[ q_{x,t}(\theta_2) =
\begin{bmatrix}
\bar{w}_{x,t+1} 1_{[z_t \geq c]} \\
\bar{w}_{x,t+1} (1 - 1_{[z_t \geq c]}) \\
\text{vec} (\bar{w}_{x,t+1} \hat{x}_t' - \text{Cov}(u_t^{(1)}, u_t^{(1)}) + h_x^{(1)} \text{Var}(u_t^{(1)})) \\
\text{vec} (\bar{w}_{x,t+1} (\hat{x}_t^{(2)})' - \text{Cov}(u_t^{(2)}, u_t^{(2)}) + h_x^{(2)} \text{Var}(u_t^{(2)})) \\
\text{vech} (\bar{w}_{x,t+1} \bar{w}_t' - \text{Var}(w_{x,t+1} - \Omega_{t+1}))
\end{bmatrix}. \]

Note that these moment conditions require knowledge about \( \text{Var}(u_t^{(1)}), \text{Var}(u_t^{(2)}), \text{Cov}(u_{t+1}^{(1)}, u_t^{(1)}), \text{Cov}(u_t^{(1)}, u_{t+1}^{(1)}), \text{Cov}(u_{t+1}^{(2)}, u_t^{(2)}), \text{Cov}(u_t^{(2)}, u_{t+1}^{(2)}). \) Consistent estimates of these moments follow from the first step of the SR approach. A key observation is that the system implied by \( \frac{1}{T-1} \sum_{t=1}^{T-1} q_{x,t}(\theta_2) = 0 \) has a closed-form solution.
7.6.2 Efficient implementation when estimating (20)

We start by re-writing the system as follows:

\[
\dot{x}_{t+1} = 1_{[z_t \geq c]} h_0^{(1)} + (1 - 1_{[z_t \geq c]}) h_0^{(2)} + h_x^{(1)} \dot{x}_t^{(1)} + h_x^{(2)} \dot{x}_t^{(2)} + \dot{w}_{x,t+1}
\]

\[
\dot{x}_{t+1} = \begin{bmatrix}
    h_0^{(1)} & h_0^{(2)} & h_x^{(1)} & h_x^{(2)}
\end{bmatrix}
\begin{bmatrix}
    1_{[z_t \geq c]} \\
    1 - 1_{[z_t \geq c]} \\
    \dot{x}_t^{(1)} \\
    \dot{x}_t^{(2)}
\end{bmatrix}
+ \dot{w}_{x,t+1}
\]

\[
\dot{x}_{t+1} = h_a a_t + \dot{w}_{x,t+1}.
\]

Moreover, the measurement errors to this transformed system read

\[
u_{a,t} = \begin{bmatrix}
    0 \\
    0 \\
    u_t^{(1)} \\
    u_t^{(2)}
\end{bmatrix},
\]

meaning that

\[
\text{Var} (u_{a,t}) = \begin{bmatrix}
    0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 \\
    0 & 0 & \text{Var} (u_t^{(1)}) & 0 \\
    0 & 0 & 0 & \text{Var} (u_t^{(2)})
\end{bmatrix}.
\]

We further define the matrix

\[
A_{t+1} = \begin{bmatrix}
    0 & 0 & \text{Cov} (u_{t+1}^{(1)}, u_t^{(1)}) & \text{Cov} (u_{t+1}^{(2)}, u_t^{(1)}) & \text{Cov} (u_{t+1}^{(1)}, u_t^{(2)}) & \text{Cov} (u_{t+1}^{(2)}, u_t^{(2)})
\end{bmatrix}.
\]

We then note that

\[
\frac{1}{T-t} \sum_{t=1}^{T-1} [\dot{w}_{x,t+1} a_t'] = \]

\[
= \frac{1}{T-t} \sum_{t=1}^{T-1} \left[ \dot{w}_{x,t+1} \begin{bmatrix}
    1_{[z_t \geq c]}' \\
    (1 - 1_{[z_t \geq c]})'
\end{bmatrix} \begin{bmatrix}
    \dot{x}_t^{(1)}' \\
    \dot{x}_t^{(2)}'
\end{bmatrix}
+ \dot{w}_{x,t+1} \begin{bmatrix}
    (1 - 1_{[z_t \geq c]})' \\
    1_{[z_t \geq c]}'
\end{bmatrix} \begin{bmatrix}
    \dot{x}_t^{(1)}' \\
    \dot{x}_t^{(2)}'
\end{bmatrix}
\right]
\]

\[
= \frac{1}{T-t} \sum_{t=1}^{T-1} \left[ \dot{w}_{x,t+1} 1_{[z_t \geq c]}' \dot{w}_{x,t+1} (1 - 1_{[z_t \geq c]})' \dot{w}_{x,t+1} (\dot{x}_t^{(1)})' \dot{w}_{x,t+1} (\dot{x}_t^{(2)})' \right].
\]
Moreover, 
\[ A_{t+1} - h_a Var(u_{a,t}) = \]

\[
= \begin{bmatrix} 0 & 0 & Cov(u_{t+1}^{(1)}, u_t^{(1)}) & Cov(u_{t+1}^{(2)}, u_t^{(2)}) \\
- \begin{bmatrix} h_0^{(1)} & h_0^{(2)} & h_x^{(1)} & h_x^{(2)} \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & Var(u_t^{(1)}) & 0 \\
0 & 0 & 0 & Var(u_t^{(2)}) \end{bmatrix} \\
= \begin{bmatrix} 0 & 0 & Cov(u_{t+1}^{(1)}, u_t^{(1)}) & Cov(u_{t+1}^{(2)}, u_t^{(2)}) \\
- \begin{bmatrix} 0 & 0 \\
h_x^{(1)} \text{Var}(u_t^{(1)}) \\
h_x^{(2)} \text{Var}(u_t^{(2)}) \end{bmatrix} \\
= \begin{bmatrix} 0 & 0 & Cov(u_{t+1}^{(1)}, u_t^{(1)}) - h_x^{(1)} \text{Var}(u_t^{(1)}) & Cov(u_{t+1}^{(2)}, u_t^{(2)}) - h_x^{(2)} \text{Var}(u_t^{(2)}) \end{bmatrix} \].

Thus, the first four sets of moment conditions may be expressed as
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{w}_{x,t+1} a'_t = \frac{1}{T-1} \sum_{t=1}^{T-1} (A_{t+1} - h_a \text{Var}(u_{a,t})) .
\]

But we then note that
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} a'_t = \frac{1}{T-1} \sum_{t=1}^{T-1} (\hat{x}_{t+1} - h_a a_t) a'_t = \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} a'_t - h_a \frac{1}{T-1} \sum_{t=1}^{T-1} a_t a'_t .
\]

Hence, we have:
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} a'_t - h_a \frac{1}{T-1} \sum_{t=1}^{T-1} a_t a'_t = \frac{1}{T-1} \sum_{t=1}^{T-1} A_{t+1} - \frac{1}{T-1} \sum_{t=1}^{T-1} h_a \text{Var}(u_{a,t}) .
\]

\[ \]
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} a'_t - \frac{1}{T-1} \sum_{t=1}^{T-1} A_{t+1} = h_a \frac{1}{T-1} \sum_{t=1}^{T-1} a_t a'_t - h_a \frac{1}{T-1} \sum_{t=1}^{T-1} \text{Var}(u_{a,t}) .
\]

166
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} \hat{a}_t' - \frac{1}{T-1} \sum_{t=1}^{T-1} A_{t+1} = h_a \left( \frac{1}{T-1} \sum_{t=1}^{T-1} a_t \hat{a}_t' - \frac{1}{T-1} \sum_{t=1}^{T-1} \text{Var}(u_{a,t}) \right)
\]

\[
h_a = \left( \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{x}_{t+1} \hat{a}_t' - \frac{1}{T-1} \sum_{t=1}^{T-1} A_{t+1} \right) \times \left( \frac{1}{T-1} \sum_{t=1}^{T-1} a_t \hat{a}_t' - \frac{1}{T-1} \sum_{t=1}^{T-1} \text{Var}(u_{a,t}) \right)^{-1}.
\]

Using the consistent estimates of \(A_{t+1}\) and \(\text{Var}(u_{a,t})\) from the first step in the SR approach, we get

\[
\hat{h}_a = \left( \sum_{t=1}^{T-1} \tilde{x}_{t+1} \tilde{a}_t' - \sum_{t=1}^{T-1} \tilde{A}_{t+1} \right) \left( \sum_{t=1}^{T-1} a_t \hat{a}_t' - \sum_{t=1}^{T-1} \text{Var}(u_{a,t}) \right)^{-1},
\]

where

\[
\tilde{A}_{t+1} = \begin{bmatrix} 0 & 0 & \text{Cov}(u_{t+1}^{(1)}, u_t^{(1)}) & \text{Cov}(u_{t+1}^{(2)}, u_t^{(2)}) \end{bmatrix}.
\]

The solution to the last moment condition is given by

\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \tilde{w}_{x,t+1} \tilde{w}_{x,t+1}' = \frac{1}{T-1} \sum_{t=1}^{T-1} \text{Var}(w_{x,t+1}) + \frac{1}{T-1} \sum_{t=1}^{T-1} \Omega_{t+1}
\]

\[
\text{Var}(w_{x,t+1}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \left( \tilde{w}_{x,t+1} \tilde{w}_{x,t+1}' - \Omega_{t+1} \right),
\]

given the homoskedasticity of \(\text{Var}(w_{x,t+1})\). Hence we have

\[
\sqrt{\text{Var}}(w_{x,t+1}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \left( \tilde{w}_{x,t+1} \tilde{w}_{x,t+1}' - \Omega_{t+1} \right),
\]
where
\[
\tilde{\Omega}_{t+1} = \tilde{\text{Var}}(\mathbf{u}_{t+1}) + \tilde{\text{Var}}^{(1)}(\mathbf{u}_t) \quad (\hat{\mathbf{h}}_x^{(1)})' + \tilde{\text{Var}}^{(2)}(\mathbf{u}_t) \quad (\hat{\mathbf{h}}_x^{(2)})' \\
- \tilde{\text{Cov}}^{(1)}(\mathbf{u}_{t+1}, \mathbf{u}_t) \quad (\hat{\mathbf{h}}_x^{(1)})' - \hat{\mathbf{h}}_x^{(1)} \tilde{\text{Cov}}^{(1)}(\mathbf{u}_t, \mathbf{u}_{t+1}) \\
- \tilde{\text{Cov}}^{(2)}(\mathbf{u}_{t+1}, \mathbf{u}_t) \quad (\hat{\mathbf{h}}_x^{(2)})' - \hat{\mathbf{h}}_x^{(2)} \tilde{\text{Cov}}^{(2)}(\mathbf{u}_t, \mathbf{u}_{t+1}),
\]

with \( \tilde{\mathbf{w}}_{x,t+1} \equiv \hat{\mathbf{x}}_{t+1} - 1_{\{z_t \geq c\}} \hat{\mathbf{h}}_0^{(1)} - (1 - 1_{\{z_t \geq c\}}) \hat{\mathbf{h}}_0^{(2)} - \hat{\mathbf{x}}_t^{(1)} - \hat{\mathbf{h}}_x^{(2)} \hat{\mathbf{x}}_t^{(2)}. \) Note that we may alternatively adopt the standard degree of freedom correction and apply
\[
\tilde{\text{Var}}(\mathbf{w}_{x,t+1}) = \frac{1}{T - 1 - 2(n_x + 1)} \sum_{t=1}^{T-1} \tilde{\mathbf{w}}_{x,t+1} \tilde{\mathbf{w}}_{x,t+1}' - \frac{1}{T - 1} \sum_{t=1}^{T-1} \tilde{\Omega}_{t+1},
\]
because we estimate \( 2(n_x + 1) \) unknown parameters per equation in the model.

### 7.6.3 Moment conditions for estimating (21)

To estimate the parameters in the process for \( z_t \) we recall that \( w_{z,t+1} \equiv \sum_{z \in z} \mathbf{e}^{(z)}_{z,t+1} \) and consider the moments

- \( E[\hat{w}_{z,t+1}] = E[w_{z,t+1} - \gamma_x \mathbf{u}_t] = 0 \) because \( \mathbf{e}^{(z)}_{z,t+1} \) is iid.
- \( E[\hat{w}_{z,t+1} z_t] = E[(w_{z,t+1} - \gamma_x \mathbf{u}_t) z_t] = 0 \) because \( \mathbf{e}^{(z)}_{z,t+1} \) is iid. and the absence of correlation between \( \mathbf{u}_t \) and \( \mathbf{e}^{(z)}_{z,t} \).
- \( E[\hat{w}_{z,t+1} \hat{\mathbf{x}}_t] = E[(w_{z,t+1} - \gamma_x \mathbf{u}_t) (x_t + \mathbf{u}_t)'] \\
  = E[-\gamma_x \mathbf{u}_t (x_t + \mathbf{u}_t)'] = -\gamma_x \text{Var} (\mathbf{u}_t). \)
- \( \text{Var}(\hat{w}_{z,t+1}) = \text{Var}(w_{z,t+1} - \gamma_x \mathbf{u}_t) = \text{Var}(w_{z,t+1}) + \gamma_x \text{Var} (\mathbf{u}_t) \gamma_x. \)

The population moments are estimated by empirical moments, meaning that the moment conditions are
\[
\mathbf{q}_{z,T} \equiv \frac{1}{T - 1} \sum_{t=1}^{T-1} \mathbf{q}_{z,t} = 0,
\]
where
\[
\mathbf{q}_{z,t} \equiv \begin{bmatrix} \hat{w}_{z,t+1} \\ \hat{w}_{z,t+1} z_t \\ \text{vec}(\hat{w}_{z,t+1} \hat{\mathbf{x}}_t + \gamma_x \text{Var} (\mathbf{u}_t)) \\ \hat{w}_{z,t+1} \hat{\mathbf{w}}'_{z,t+1} - \text{Var}(w_{z,t+1}) - \gamma_x \text{Var} (\mathbf{u}_t) \gamma_x \end{bmatrix}.
\]
These moment conditions require knowledge of \( \text{Var} (\mathbf{u}_t) \) which we estimate in the first step of the SR approach. A key observation is that the system implied by
\[
\frac{1}{T - 1} \sum_{t=1}^{T-1} \mathbf{q}_{z,t} (\theta_2) = 0
\]
also has a closed-form solution.
7.6.4 Efficient implementation when estimating (21)

We start by re-writing the equation as follows

\[ z_{t+1} = \gamma_0 + \gamma_z z_t + \gamma'_x \hat{x}_t + \hat{w}_{z,t+1} \]

\[ \downarrow \]

\[ z_{t+1} = \begin{bmatrix} \gamma_0 & \gamma_z & \gamma'_x \\ b_b & z_t & \hat{x}_t \\ b_t & \end{bmatrix} + \hat{w}_{z,t+1} \]

\[ \downarrow \]

\[ z_{t+1} = h_b b_t + \hat{w}_{z,t+1}. \]

Moreover, the measurement errors to this transformed equation read

\[ u_{b,t} \equiv \begin{bmatrix} 0 \\ 0 \\ u_t \end{bmatrix}, \]

meaning that

\[ Var(u_{b,t}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Var(u_t) \end{bmatrix}. \]

We then note that

\[ \frac{1}{T-1} \sum_{t=1}^{T-1} [\hat{w}_{z,t+1} b'_t] = \frac{1}{T-1} \sum_{t=1}^{T-1} [\hat{w}_{z,t+1} \begin{bmatrix} 1 & z_t & \hat{x}'_t \end{bmatrix}] \]

\[ = \frac{1}{T-1} \sum_{t=1}^{T-1} \begin{bmatrix} \hat{w}_{z,t+1} & \hat{w}_{z,t+1} z_t & \hat{w}_{z,t+1} \hat{x}'_t \end{bmatrix}. \]

Moreover,

\[ -h_b Var(u_{b,t}) = -\begin{bmatrix} \gamma_0 & \gamma_z & \gamma'_x \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Var(u_t) \end{bmatrix} \]

\[ = \begin{bmatrix} 0 & 0 & -\gamma'_x Var(u_t) \end{bmatrix}. \]

Thus, the first three sets of moment conditions may be expressed as

\[ \frac{1}{T-1} \sum_{t=1}^{T-1} [\hat{w}_{z,t+1} b'_t] = -\frac{1}{T-1} \sum_{t=1}^{T-1} h_b Var(u_{b,t}). \]
But we then note that
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} [\hat{w}_{z,t+1} b'_t] = \frac{1}{T-1} \sum_{t=1}^{T-1} (z_{t+1} - h_b b_t) b'_t = \frac{1}{T-1} \sum_{t=1}^{T-1} z_{t+1} b'_t - h_b \frac{1}{T-1} \sum_{t=1}^{T-1} b_t b'_t.
\]
Hence, we have:
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} z_{t+1} b'_t - h_b \frac{1}{T-1} \sum_{t=1}^{T-1} b_t b'_t = -\frac{1}{T-1} \sum_{t=1}^{T-1} h_b \text{Var}(u_{b,t})
\]
which is to be expected given the absence of measurement errors in \(z_t\). Using the consistent estimates of \(\text{Var}(u_{b,t})\) from the first step in the SR approach, we get
\[
h_b = \left( \frac{1}{T-1} \sum_{t=1}^{T-1} z_{t+1} b'_t \right) \left( \frac{1}{T-1} \sum_{t=1}^{T-1} (b_t b'_t - \text{Var}(u_{b,t})) \right)^{-1}
\]
and
\[
h_b = \left( \sum_{t=1}^{T-1} z_{t+1} b'_t \right) \left( \sum_{t=1}^{T-1} (b_t b'_t - \text{Var}(u_{b,t})) \right)^{-1},
\]
which is to be expected given the absence of measurement errors in \(z_t\). Using the consistent estimates of \(\text{Var}(u_{b,t})\) from the first step in the SR approach, we get
\[
\hat{h}_b = \left( \sum_{t=1}^{T-1} z_{t+1} b'_t \right) \left( \sum_{t=1}^{T-1} (b_t b'_t - \text{Var}(u_{b,t})) \right)^{-1}.
\]
The solution to the last moment condition is given by
\[
\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{w}_{z,t+1} \hat{w}'_{z,t+1} = \frac{1}{T-1} \sum_{t=1}^{T-1} \text{Var}(w_{z,t+1}) + \frac{1}{T-1} \sum_{t=1}^{T-1} \gamma_x \text{Var}(u_t) \gamma_x
\]
where
\[
\text{Var}(w_{z,t+1}) = \frac{1}{T-1} \sum_{t=1}^{T-1} (\hat{w}_{z,t+1} \hat{w}'_{z,t+1} - \gamma_x \text{Var}(u_t) \gamma_x),
\]
given the homoskedasticity of \(\text{Var}(w_{z,t+1})\). We then use the estimator
\[
\tilde{\text{Var}}(w_{z,t+1}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \left( (\hat{w}_{z,t+1}) (\hat{w}_{z,t+1})' - \gamma_x \text{Var}(u_t) \gamma_x \right),
\]
where \( \hat{w}_{z,t+1} = z_{t+1} - \hat{\gamma}_0 - \hat{\gamma}_x z_t - \hat{\gamma}_x^\prime \hat{x}_t \), or when using a degree of freedom adjustment
\[
\text{Var} (w_{z,t+1}) = \frac{1}{T-1-(n_x+2)} \sum_{t=1}^{T-1} \left( \hat{w}_{z,t+1} \right)^\prime \left( \hat{w}_{z,t+1} \right) - \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{\gamma}_x \text{Var} (u_t) \hat{\gamma}_x.
\]

### 7.7 Obtaining estimates for the market price of risk

We know from subsection 7.3 that
\[
h_0^{(i)} = \Phi \mu + \lambda_0^{(i)},
\]
and
\[
h_x^{(i)} = (I - \Phi) + \lambda_x^{(i)},
\]
for \( i = \{1,2\} \). Hence, we directly have
\[
\hat{\lambda}_0^{(i)} = \hat{h}_0^{(i)} - \hat{\Phi} \hat{\mu},
\]
and
\[
\hat{\lambda}_x^{(i)} = \hat{h}_x^{(i)} - (I - \hat{\Phi}) .
\]
The standard errors are computed using simulation by drawing from the asymptotical normal distribution of the \( \mathbb{P} \)-parameters. Note that it is ok to ignore uncertainty about the risk-neutral parameters because when \( T \to \infty \), then \( NT \) goes faster to infinity as \( N \) also is assumed to go to infinity.

### 7.8 Inducing stationarity

A key requirement in dynamic term structure models is that the factor dynamics is stationary to induce stationarity of bond yields. We first derive the stationarity condition and then present a data-driven method to induce stationarity.

To derive the stationarity condition in the SETAR model, we simply note that
\[
x_{t+1} = 1_{[z_t \geq c]} h_0^{(1)} + 1_{[z_t \geq c]} h_x^{(1)} x_t
\]
\[+ (1 - 1_{[z_t \geq c]}) h_0^{(2)} + (1 - 1_{[z_t \geq c]}) h_x^{(2)} x_t + \Sigma_x \epsilon_{t+1}
\]
\[= 1_{[z_t \geq c]} h_0^{(1)} + (1 - 1_{[z_t \geq c]}) h_0^{(2)}
\]
\[+ (h_x^{(1)} 1_{[z_t \geq c]} + h_x^{(2)} (1 - 1_{[z_t \geq c]}) )x_t + \Sigma_x \epsilon_{t+1}
\]
\[= h_0^{(1,2)} + h_x^{(1,2)} x_t + \Sigma_x \epsilon_{t+1},
\]
where
\[
h_x^{(1,2)} \equiv h_x^{(1)} 1_{[z_t \geq c]} + h_x^{(2)} (1 - 1_{[z_t \geq c]}).
\]
and

\[ h_0^{(1,2)} \equiv 1_{\{z_t \geq c\}} h_0^{(1)} + (1 - 1_{\{z_t \geq c\}}) h_0^{(2)}. \]

Thus, a sufficient condition for stability - and hence stationarity given \( \epsilon_{t+1} \sim \mathcal{NID}(0, I) \) - is that all eigenvalues of \( h_x^{(1,2)} \) are inside the unit circle for all values of \( z_t \). The latter is satisfied if all eigenvalues of \( h_x^{(1)} \) and \( h_x^{(2)} \) are inside the unit circle. Note that this stability condition can also be inferred from Theorem 6.12 in Pötscher and Prucha (1997), page 70-71. To realize this, recall that this theorem shows that the general process \( x_{t+1} = h(x_t, \epsilon_{t+1}) \) is stable if only the system is contracting when it is iterated sufficiently many time periods ahead. That is, iterate \( x_{t+1} = h(x_t, \epsilon_{t+1}) \) forward by \( k \) time periods to obtain

\[ x_{t+k} = h^k \left( x_t, \{\epsilon_{t+j}\}_{j=1}^k \right), \]

then only the function \( h^k \) needs to be contracting for \( x_{t+1} = h(x_t, \epsilon_{t+1}) \) to be stable. In our case, we thus have

\[
x_{t+2} = h_0^{(1,2)} + h_x^{(1,2)} x_{t+1} + \Sigma_x \epsilon_{t+2} \\
= h_0^{(1,2)} + h_x^{(1,2)} \left( h_0^{(1,2)} + \Sigma_x \epsilon_t + h_x^{(1,2)} x_t + \Sigma_x \epsilon_{t+1} \right) + \Sigma_x \epsilon_{t+2} \\
= h_0^{(1,2)} + h_x^{(1,2)} h_0^{(1,2)} + \left( h_x^{(1,2)} \right)^2 x_t + h_x^{(1,2)} \Sigma_x \epsilon_{t+1} + \Sigma_x \epsilon_{t+2},
\]

and in general

\[
x_{t+k} = \sum_{i=0}^{k-1} \left( h_x^{(1,2)} \right)^i h_0^{(1,2)} + \left( h_x^{(1,2)} \right)^k x_t + \sum_{i=1}^k \left( h_x^{(1,2)} \right)^{k-i} \Sigma_x \epsilon_{t+i}.
\]

Hence, if all eigenvalues of \( h_x^{(1,2)} \) are inside the unit circle, then \( \left( h_x^{(1,2)} \right)^k \) converges to zero as \( k \rightarrow \infty \) and hence for sufficiently large \( k \) \( \left| \left( h_x^{(1,2)} \right)^k \right| < 1 \), i.e. the function \( h^k \left( x_t, \{\epsilon_{t+j}\}_{j=1}^k \right) \) is contracting.\(^4\)

Theorem 6.12 in Pötscher and Prucha (1997), page 70-71, also states that if i) \( \|\epsilon_t\|_r < \infty \) for \( t \in \mathbb{N} \), ii) \( \|x_0\|_r < \infty \), and iii) \( h_x^{(1)} \) and \( h_x^{(2)} \) have finite elements, then \( \|x_t\|_r < \infty \) for \( t \in \mathbb{N} \). Here, we define \( \|\epsilon_t\|_r = \left[ \int |\epsilon_t|^r dP \right]^{1/r} \) where \( |\cdot| \) is the Euclidean norm. The third condition relates to the fact that the transition function needs to be Lipschitz, i.e. if for all \( x \) and \( y \) it holds that \( |h(x) - h(y)| \leq L \|x - y\|_2 \). But this condition clearly holds for our piece-wise linear transition function, provided \( h_x^{(1)} \) and \( h_x^{(2)} \) have finite elements. Given that we assume \( \epsilon_t \sim \mathcal{NID}(0, I) \), we only need to assume that the initial value for our system i.e. \( x_0 \), has

\(^4\)Note that here \( |A| \) denotes the norm given by the square root of the largest eigenvalue of the matrix product \( AA \).
finite \( r \)th moment, i.e. \( \|x_0\|_r < \infty \), in order for \( x_t \) to have finite \( r \)th unconditional moment, i.e. \( \|x_t\|_r < \infty \) for \( t \in \mathbb{N} \).

Given that \( x_t \) is stable, we must also require that \( |\gamma_z| < 1 \) for stability of \( z_t \). But note that stability of bond yields actually holds even if \( z_t \) is unstable.

Another way to derive the stability condition for our system is to note that

\[
\begin{bmatrix}
  x_{t+1} \\
  z_{t+1}
\end{bmatrix} = \begin{bmatrix}
  h_0^{(1,2)} + h_x^{(1,2)} x_t \\
  \gamma_0 + \gamma_z z_t + \gamma_x x_t
\end{bmatrix} + \begin{bmatrix}
  \Sigma_x & 0 \\
  \Sigma_{xz} & \Sigma_{zz}
\end{bmatrix} \begin{bmatrix}
  0 \\
  0
\end{bmatrix} + \begin{bmatrix}
  \epsilon^{(p)}_{x,t+1} \\
  \epsilon^{(p)}_{z,t+1}
\end{bmatrix}.
\]

To evaluate the stability of this system, note that the eigenvalues \( \lambda \) are given by:

\[
\begin{aligned}
\left| \begin{bmatrix}
  h_x^{(1,2)} & 0 \\
  \gamma_x & \gamma_z
\end{bmatrix} - \lambda I \right| &= 0 \\
\Leftrightarrow \left| \begin{bmatrix}
  h_x^{(1,2)} - \lambda I & 0 \\
  \gamma_x & \gamma_z - \lambda
\end{bmatrix} \right| &= 0 \\
\Leftrightarrow \left| h_x^{(1,2)} - \lambda I \right| \times |\gamma_z - \lambda| &= 0.
\end{aligned}
\]

Thus, we have the eigenvalues \( \gamma_z \) and \( \text{eig}(h_x^{(1,2)}) \).

If the stability condition is not satisfied, then we suggest to downscale \( h_x^{(1)} \) by \( \delta_1 \) if \( \left| \text{eig} \left( h_x^{(1)} \right) \right| \geq 1 \) and \( h_x^{(2)} \) by \( \delta_2 \) if \( \left| \text{eig} \left( h_x^{(2)} \right) \right| \geq 1 \). We determine the values of \( \delta_1 \) and \( \delta_2 \) as in Andreasen and Meldrum (2014), i.e. by

\[
(\delta_1, \delta_2) = \arg \min_{\{\delta_{\text{lower}} \leq \delta_{i} < 1\}^2} \sum_{i=1}^{n_x} \left( \frac{\sigma^2_{i,\text{model}} (\delta_1, \delta_2) - \sigma^2_{i,\text{sample}}}{\sigma^2_{i,\text{sample}}} \right)^2.
\]

We follow Andreasen and Meldrum (2014) and estimate the unconditional variance of the \( i \)th pricing factor in the sample using

\[
\hat{\sigma}^2_{i,\text{sample}} = \frac{1}{T - 1} \sum_{t=1}^{T} (\hat{x}_{i,t} - \hat{\mu} [\hat{x}_i])^2 - \frac{1}{T} \sum_{t=1}^{T} \text{Var} \left( u_{i,t} \right),
\]

where \( \hat{\mu} [\hat{x}_i] = 1/T \sum_{t=1}^{T} \hat{x}_{i,t} \) and \( \text{Var} \left( u_{i,t} \right) \) is the estimated variance of \( \hat{x}_{i,t} \). The value of the unconditional variance of \( x_{i,t} \) in the model are computed by simulation using

\[
x_{t+1} = 1_{[x_{t,\geq 0}]} h_0^{(1)} + 1_{[x_{t,\geq 0}]} \delta_1 h_x^{(1)} x_t + (1 - 1_{[x_{t,\geq 0}]} h_0^{(2)} + (1 - 1_{[x_{t,\geq 0}]} \delta_2 h_x^{(2)} x_t + \Sigma_x \epsilon^{(p)}_{x,t+1},
\]

173
and

\[ z_{t+1} = \gamma_0 + \gamma_z z_t + \gamma_x' x_t + \Sigma_{xz} e^p_{z,t+1}. \]

### 7.9 Normalization of the ATSM

Given the affine specification for the market price of risk, we adopt the standard normalisation by letting:

1. \( \beta = 1 \)
2. \( \alpha \) is a free parameter
3. \( \Sigma \) is lower triangular
4. \( \Phi \) is diagonal
5. \( \mu = 0 \)
6. \( \lambda_x \) is a free matrix
7. \( \lambda_0 \) are free parameters
8. \( \Sigma_x \) is positive definite

Although these restrictions, in principle, are sufficient to identify the model, we further require that:

9. The diagonal elements (i.e. eigenvalues) of \( \Phi \) are increasing. I.e. \( \phi_{11} \leq \phi_{22} \leq \phi_{33} \) in a three-factor model.

To motivate this additional restriction, we first note that yields in the ATSM is given by

\[ y_{t,k} = -\frac{A_k}{k} - \frac{B_k'}{k} x_t, \]

for \( k = 1, 2, \ldots K \). Let

\[
A \equiv \begin{bmatrix}
-\frac{A_1}{2} \\
-\frac{A_2}{2} \\
\vdots \\
-\frac{A_N}{N}
\end{bmatrix},
\]

\[
B \equiv \begin{bmatrix}
-\frac{B_k'}{2} \\
-\frac{B_k'}{2} \\
\vdots \\
-\frac{B_k'}{K}
\end{bmatrix},
\]
and

\[ Y_t \equiv \begin{bmatrix} y_{t,1} \\ y_{t,2} \\ \vdots \\ y_{t,K} \end{bmatrix}. \]

The OLS regression to obtain the latent factors is then given by

\[ Y_t = A + Bx_t + \epsilon_t. \]

We observe from the solution to the ATSM that \((\mu = 0\) with our normalization\)

\[ A_k = -\alpha + A_{k-1} + \frac{1}{2} B'_{k-1} \Sigma_x \Sigma'_x B_{k-1} \approx -\alpha + A_{k-1}, \]

because \(\Sigma_x \Sigma'_x\) is very small, whereas

\[ B'_k = -1 + B'_{k-1} (I - \Phi). \]

We first note that the pricing factors in the regression filter are identified even when eigenvalues under \(Q\) are identical due to Proposition 1 in Joslin, Singleton and Zhu (2011). Secondly, given that \(\Sigma_x\) is badly identified from the cross-section dimension, the ordering of the factors is therefore also badly identified. A similar finding is reported in Ait-Sahalia and Kimmel (2010). Note also that the standard identification assumptions 1-8 only identify the order of the factors and hence the model if \(\Sigma_x\) is non-diagonal. Hence, if \(\Sigma_x\) has small non-diagonal values, the identification may be very poor. To eliminate this identification issue we therefore require that the eigenvalues of \(\Phi\) are increasing.

### 7.10 Accounting for the Zero Lower Bound

We can account for the zero lower bound (ZLB) by considering the policy rate

\[ r_t = \max \{ 0, s(x_t) \}, \]

and

\[ s(x_t) = \alpha + \beta' x_t. \]

Given that the factor dynamics is conditional normal under \(Q\), even with regime-switching in the market price of risk, we can price zero-coupon bonds in a standard way, i.e. as outlined by Priebsch (2013). The factor dynamics with regime-switching is then estimated as outlined above.
References


PhD dissertations since 1 July 2011

2011-4  Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics
2011-5  Christian Bach: The Game of Risk
2011-6  Stefan Holst Bache: Quantile Regression: Three Econometric Studies
2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management
2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization
2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders
2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade
2012-1  Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting
2012-2  Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping
2012-3  Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing
2012-4  Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing
2012-5  Lasse Frisgaard Gunnersen: Income Redistribution Policies
2012-6  Miriam Wüst: Essays on early investments in child health
2012-7  Yukai Yang: Modelling Nonlinear Vector Economic Time Series
2012-9  Henrik Nørholm: Structured Retail Products and Return Predictability
2012-10 Signe Frederiksen: Empirical Essays on Placements in Outside Home Care
2012-11 Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing
Jens Riis Andersen: Option Games under Incomplete Information
Margit Malmose: The Role of Management Accounting in New Public Management Reforms: Implications in a Socio-Political Health Care Context
Laurent Callot: Large Panels and High-dimensional VAR
Christian Rix-Nielsen: Strategic Investment
Kenneth Lykke Sørensen: Essays on Wage Determination
Tue Rauff Lind Christensen: Network Design Problems with Piecewise Linear Cost Functions
Dominyka Sakalauskaite: A Challenge for Experts: Auditors, Forensic Specialists and the Detection of Fraud
Rune Bysted: Essays on Innovative Work Behavior
Mikkel Nørlem Hermansen: Longer Human Lifespan and the Retirement Decision
Jannie H.G. Kristoffersen: Empirical Essays on Economics of Education
Mark Strøm Kristoffersen: Essays on Economic Policies over the Business Cycle
Philipp Meinen: Essays on Firms in International Trade
Cédric Gorinas: Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective
Ina Charlotte Jäkel: Product Quality, Trade Policy, and Voter Preferences: Essays on International Trade
Anna Gerstrøm: World Disruption - How Bankers Reconstruct the Financial Crisis: Essays on Interpretation
Paola Andrea Barrientos Quiroga: Essays on Development Economics
Peter Bodnar: Essays on Warehouse Operations
Rune Vammen Lesner: Essays on Determinants of Inequality
Peter Arendorf Bache: Firms and International Trade
Anders Laugesen: On Complementarities, Heterogeneous Firms, and International Trade
<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-17</td>
<td>Anders Bruun Jonassen</td>
<td>Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance</td>
</tr>
<tr>
<td>2014-1</td>
<td>David Sloth Pedersen</td>
<td>A Journey into the Dark Arts of Quantitative Finance</td>
</tr>
<tr>
<td>2014-2</td>
<td>Martin Schultz-Nielsen</td>
<td>Optimal Corporate Investments and Capital Structure</td>
</tr>
<tr>
<td>2014-3</td>
<td>Lukas Bach</td>
<td>Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods</td>
</tr>
<tr>
<td>2014-4</td>
<td>Tanja Groth</td>
<td>Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomy analysis</td>
</tr>
<tr>
<td>2014-5</td>
<td>Niels Strange Hansen</td>
<td>Forecasting Based on Unobserved Variables</td>
</tr>
<tr>
<td>2014-6</td>
<td>Ritwik Banerjee</td>
<td>Economics of Misbehavior</td>
</tr>
<tr>
<td>2014-7</td>
<td>Christina Annette Gravert</td>
<td>Giving and Taking – Essays in Experimental Economics</td>
</tr>
<tr>
<td>2014-8</td>
<td>Astrid Hanghøj</td>
<td>Papers in purchasing and supply management: A capability-based perspective</td>
</tr>
<tr>
<td>2014-9</td>
<td>Nima Nonejad</td>
<td>Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics</td>
</tr>
<tr>
<td>2014-10</td>
<td>Tine L. Mundbjerg Eriksen</td>
<td>Essays on Bullying: an Economist’s Perspective</td>
</tr>
<tr>
<td>2014-11</td>
<td>Sashka Dimova</td>
<td>Essays on Job Search Assistance</td>
</tr>
<tr>
<td>2014-12</td>
<td>Rasmus Tangsgaard Varneskov</td>
<td>Econometric Analysis of Volatility in Financial Additive Noise Models</td>
</tr>
<tr>
<td>2015-1</td>
<td>Anne Floor Brix</td>
<td>Estimation of Continuous Time Models Driven by Lévy Processes</td>
</tr>
<tr>
<td>2015-2</td>
<td>Kasper Vinther Olesen</td>
<td>Realizing Conditional Distributions and Coherence Across Financial Asset Classes</td>
</tr>
<tr>
<td>2015-3</td>
<td>Manuel Sebastian Lukas</td>
<td>Estimation and Model Specification for Econometric Forecasting</td>
</tr>
<tr>
<td>2015-4</td>
<td>Sofie Theilade Nyland Brodersen</td>
<td>Essays on Job Search Assistance and Labor Market Outcomes</td>
</tr>
<tr>
<td>2015-5</td>
<td>Jesper Nydam Wulff</td>
<td>Empirical Research in Foreign Market Entry Mode</td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2015-6</td>
<td>Sanni Nørgaard Breining: The Sibling Relationship Dynamics and Spillovers</td>
<td>Sanni Nørgaard Breining</td>
</tr>
<tr>
<td>2015-7</td>
<td>Marie Herly: Empirical Studies of Earnings Quality</td>
<td>Marie Herly</td>
</tr>
<tr>
<td>2015-8</td>
<td>Stine Ludvig Bech: The Relationship between Caseworkers and Unemployed Workers</td>
<td>Stine Ludvig Bech</td>
</tr>
<tr>
<td>2015-9</td>
<td>Kaleb Girma Abreha: Empirical Essays on Heterogeneous Firms and International Trade</td>
<td>Kaleb Girma Abreha</td>
</tr>
<tr>
<td>2015-10</td>
<td>Jeanne Andersen: Modelling and Optimisation of Renewable Energy Systems</td>
<td>Jeanne Andersen</td>
</tr>
<tr>
<td>2015-11</td>
<td>Rasmus Landersø: Essays in the Economics of Crime</td>
<td>Rasmus Landersø</td>
</tr>
<tr>
<td>2015-12</td>
<td>Juan Carlos Parra-Alvarez: Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)</td>
<td>Juan Carlos Parra-Alvarez</td>
</tr>
<tr>
<td>2015-13</td>
<td>Sakshi Girdhar: The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis</td>
<td>Sakshi Girdhar</td>
</tr>
<tr>
<td>2015-14</td>
<td>Wenjing Wang: Corporate Innovation, R&amp;D Personnel and External Knowledge Utilization</td>
<td>Wenjing Wang</td>
</tr>
<tr>
<td>2015-16</td>
<td>Jonas Maibom: Structural and Empirical Analysis of the Labour Market</td>
<td>Jonas Maibom</td>
</tr>
<tr>
<td>2015-18</td>
<td>Orimar Sauri: Lévy Semistationary Models with Applications in Energy Markets</td>
<td>Orimar Sauri</td>
</tr>
<tr>
<td>2015-20</td>
<td>Jonas Nygaard Eriksen: Business Cycles and Expected Returns</td>
<td>Jonas Nygaard Eriksen</td>
</tr>
<tr>
<td>2016-1</td>
<td>Silvia Migali: Essays on International Migration: Institutions, Skill Recognition, and the Welfare State</td>
<td>Silvia Migali</td>
</tr>
<tr>
<td>2016-2</td>
<td>Lorenzo Boldrini: Essays on Forecasting with Linear State-Space Systems</td>
<td>Lorenzo Boldrini</td>
</tr>
<tr>
<td>2016-4</td>
<td>Camilla Pisani: Volatility and Correlation in Financial Markets: Theoretical Developments and Numerical Analysis</td>
<td>Camilla Pisani</td>
</tr>
</tbody>
</table>
2016-5 Anders Kronborg: Methods and Applications to DSGE Models
2016-6 Morten Visby Krægpøth: Empirical Studies in Economics of Education
2016-7 Anne Odile Peschel: Essays on Implicit Information Processing at the Point of Sale: Evidence from Experiments and Scanner Data Analysis
2016-8 Girum Dagnachew Abate: Essays in Spatial Econometrics
2016-9 Kai Rehwald: Essays in Public Policy Evaluation
2016-10 Reza Pourmoayed: Optimization Methods in a Stochastic Production Environment
2016-11 Sune Lauth Gadegaard: Discrete Location Problems – Theory, Algorithms, and Extensions to Multiple Objectives
2016-12 Lisbeth Palmhøj Nielsen: Empirical Essays on Child Achievement, Maternal Employment, Parental Leave, and Geographic Mobility
2016-13 Louise Voldby Beuchert-Pedersen: School Resources and Student Achievement: Evidence From Social and Natural Experiments
2016-14 Mette Trier Damgaard: Essays in Applied Behavioral Economics
2016-15 Andrea Barletta: Consistent Modeling and Efficient Pricing of Volatility Derivatives
2016-16 Thorvardur Tjörvi Ólafsson: Macrofinancial Linkages and Crises in Small Open Economies
2016-17 Carlos Vladimir Rodríguez Caballero: On Factor Analysis with Long-Range Dependence
2016-18 J. Eduardo Vera-Valdés: Essays in Long Memory
2016-19 Magnus Sander: Returns, Dividends, and Optimal Portfolios