Resolving the Spanning Puzzle in Macro-Finance Term Structure Models*

Michael D. Bauer and Glenn D. Rudebusch
Federal Reserve Bank of San Francisco

December 19, 2014

Abstract

Previous macro-finance term structure models (MTSMs) imply that macroeconomic state variables are spanned by (i.e., perfectly correlated with) model-implied bond yields. However, this theoretical implication appears inconsistent with regressions showing that much macroeconomic variation is unspanned and that the unspanned variation helps forecast excess bond returns and future macroeconomic fluctuations. We resolve this contradiction—or "spanning puzzle"—by reconciling spanned MTSMs with the regression evidence, thus salvaging the previous macro-finance literature. Furthermore, we statistically reject "unspanned" MTSMs, which are an alternative resolution of the spanning puzzle, and show that their knife-edge restrictions are economically unimportant for determining term premia.

Keywords: yield curve, term structure models, macro-finance, unspanned macro risks, monetary policy
JEL Classifications: E43, E44, E52

*The views expressed in this paper are those of the authors and do not necessarily reflect those of others in the Federal Reserve System. We thank Martin Andreasen and Mikhail Chernov for helpful comments, and Marcel Priebsch for providing data and code for the replication of his results. Author contact: michael.bauer@sf.frb.org, glenn.rudebusch@sf.frb.org

1 Introduction

A long literature in finance has modeled bond yields using a small set of factors that are linear combinations of bond yields. The resulting "yields-only" models provide a useful reduced-form description of term structure dynamics but offer little insight into the economic forces that drive changes in interest rates. To provide that underlying insight, much research has used affine macro-finance term structure models (MTSMs) to examine the connections between macroeconomic variables and the yield curve. For example, many papers have estimated reduced-form MTSMs with a vector autoregression for the macroeconomic and yield-curve variables coupled with a reduced-form pricing kernel.[1] In addition, by incorporating economic structural relationships, many researchers have developed equilibrium MTSMs for endowment or production economies.[2] Throughout all of this macro-finance term structure research, the short-term interest rate is represented as an affine function of risk factors (i.e., the state variables) that include macroeconomic variables. Accordingly, the assumption of the absence of arbitrage and the usual form of the stochastic discount factor imply that model-implied yields are also affine in these risk factors. This linear mapping from macro factors to yields can, outside of a knife-edge case, be inverted to express the macro factors as a linear combination of yields. Hence, these models imply "invertibility" (Duffee, 2013b) or "spanning," in which information in the macro variables is completely captured by the contemporaneous yield curve.

Because of this macro spanning implication of previous MTSMs, these models—and by association, the entire past literature of macro-finance term structure research—has come under severe criticism. Indeed, Joslin et al. (2014) (henceforth JPS) argue that essentially all previous MTSMs—both reduced-form and equilibrium approaches—impose "counterfactual restrictions on the joint distribution of bond yields and the macroeconomy" (p. 1197). Their critique is straightforward: While previous MTSMs imply that the information in macroeconomic variables should be spanned by the yield curve, there is strong regression-based evidence for unspanned macro information. This regression evidence is of three types. First, simple regressions of macro variables on observed yields show that there is significant "unspanned macroeconomic variation." For example, JPS find that only 15% of the variation in their measure of economic activity is captured in the first three principal components (PCs) of the yield

[1]Some examples of this approach include Ang and Piazzesi (2003), Bernanke et al. (2004), Ang et al. (2008, 2011), Bikbov and Chernov (2010), Joslin et al. (2013b), and Bauer et al. (2014).

[2]Equilibrium finance models of the term structure include Wachter (2006), Piazzesi and Schneider (2007), Buraschi and Jiltsov (2007), Gallmeyer et al. (2007), Bekaert et al. (2009), and Bansal and Shaliastovich (2013). Among many others, Hördahl et al. (2006), Dewachter and Lyrio (2006), Rudebusch and Wu (2008), and Rudebusch and Swanson (2012) consider term structure implications of macroeconomic models with production economies.

curve, rather than the 100% predicted by theoretical macro spanning.[3] Second, JPS point to other regressions that suggest that the macro variation not captured by the yield curve is useful for predicting excess bond and stock returns—i.e., there is evidence for "unspanned macroeconomic risk" in the sense that unspanned macro information appears relevant for pricing risk (see also Cooper and Priestley, 2008; Ludvigson and Ng, 2009). Finally, Duffee (2013a,b) documents that macro information not captured by the yield curve is also useful for predicting future macro variation because the yield curve does not capture the persistence of macro variables as it should in a standard spanned MTSM. This last element is evidence for what we call "unspanned macroeconomic forecasts."

The apparent conflict between the theoretical spanning condition implicit in MTSMs and the tripartite regression evidence on unspanned macro information constitutes what we term the "spanning puzzle." This spanning puzzle casts doubt on the validity of commonly used MTSMs and much previous macro-finance research and raises questions about the future direction of macro-finance term structure analysis. In his comprehensive survey of macro-finance bond pricing, Duffee (2013a) describes the contradiction between theoretical spanning and the contrary regression evidence as an "important conceptual difficulty with macro-finance models" (p. 412). Similarly, Gürkaynak and Wright (2012) see the spanning puzzle as a "thorny issue with the use of macroeconomic variables in affine models" (p. 350). Indeed, one direction forward for macro-finance research that has been advocated by JPS and others, is to discard existing spanned MTSMs and shift to a new class of unspanned MTSMs, in which macro variables are not pricing factors.[4] In unspanned MTSMs, yields load only on yield factors and not on macro factors.

In this paper, we show that spanned MTSMs can be reconciled with the data; thus our evidence salvages the reduced-form and equilibrium MSTMs and results of previous researchers. Given this solution to the spanning puzzle, the alternative unspanned models are not required. In fact, our statistical tests strongly reject the knife-edge restrictions imposed by unspanned models, while we show that these restrictions are economically unimportant for determining term premia.

Our analysis progresses in several broad stages. First, we review the general classes of spanned and unspanned MTSMs. Then, we revisit the evidence for unspanned macro information and consider in turn the three regressions described above for unspanned macro

[3]Similarly, Duffee (2013b) also contends that much macroeconomic variation is not reflected in the yield curve as macro spanning regressions show that for "typical variables included in macro-finance models, the R^2s are on the wrong side of 1/2" (p. 412).

[4]Examples of research with unspanned (or hidden) factors include Duffee (2011), Wright (2011), Chernov and Mueller (2012), and Coroneo et al. (2013).

variation, risk and forecasts. For robustness, we examine results for ten different macroeconomic variables. We distinguish between two different types of macroeconomic variables: those that are directly relevant for determining monetary policy, and those that are not. The former, which we denote as "policy factors," are closely related to the yield curve and display little to no evidence of unspanned macro variation.[5] Other macro variables, "non-policy factors," are variables that monetary policymakers pay much less attention to when setting the current short-term interest rate. These are the variables for which JPS and Duffee (2013b) document low R^2 in regressions on yields, which is not surprising, since they are also widely found to be unimportant in estimated monetary policy rules. We also examine regression evidence that the unspanned macro variation seems to have significant forecasting power for future bond returns and for future macroeconomic variation. For the former regression, we conclude that such results are sensitive to the macro variables included. In contrast, the latter regression is robust to the use of alternative maarcoeconomic variables.

Next, we show that all previous spanned MTSMs appear to be perfectly consistent with all the regression evidence showing the existence of unspanned macro information. Our resolution of the spanning puzzle centers on measurement error. The theoretical macro spanning implication of MTSMs holds for model-implied yields, while the regression evidence is based on measured yields. A significant wedge between the two is created by even very small amounts of measurement error in yields, and this wedge prevents the spanning regressions from properly identifying the presence or absence of spanning. To demonstrate this result, we estimate the three types of spanning regressions using *simulated* data from empirical spanned and unspanned MTSMs. With just a basis point of measurement error, we can replicate the regression evidence from the real-world data—contradicting the views in JPS, Duffee (2013b), and Priebsch (2014). This novel resolution of the spanning puzzle shows that the theoretical spanning condition has little practical relevance and that the regression evidence provides no reason to prefer unspanned models over spanned ones.

Our resolution to the spanning puzzle shows that spanned models do not contradict the regression evidence. As noted above, other researchers, most prominently JPS, have addressed the spanning puzzle by advocating alternative MTSMs with unspanned macro factors. Such models generate unspanned macro risks by assuming a knife-edge restriction that prevents invertibility. We exploit the nesting of unspanned models within spanned models and perform the first likelihood-ratio tests of the restrictions delineating spanned and unspanned MTSMs.

[5] These include measures of economic slack (such as the unemployment rate) and measures of underlying inflation, which are variables the Federal Reserve and other central banks rely on most directly when setting the short-term interest rate and can be identified from estimated monetary policy rules and the communications of monetary policymakers.

In both of the variations we consider, these restrictions are strongly rejected. This statistical evidence shows that models with unspanned macro risks are at variance with the data and thus cannot resolve the spanning puzzle. Still, while we overwhelmingly reject the unspanned model statistically, our results are more nuanced when considered in economic terms. In particular, we find that spanned and unspanned models imply essentially identical term premia. That is, while the cross-sectional fit and possibly other dimensions of an MTSM are affected by the knife-edge restrictions, the restrictions do not appreciably alter estimates of the risk premium—contrary to the claim of JPS. Since unspanned models may be able to reproduce some of the economic features of spanned models but are more parsimoniously parameterized, they may be a useful approximation for certain purposes.

The paper is structured as follows: Section 2 discusses the theoretical spanning implication of conventional MTSMs and the knife-edge restrictions needed to obtain an unspanned model. Section 3 examines the regression evidence for unspanned macro information. Section 4 describes our spanned and unspanned MTSM estimates and resolves the spanning puzzle by showing that the spanned models are consistent with the regression evidence, so the theoretical spanning implications of previous MTSMs are of little practical relevance. In Section 5, we directly test spanned and unspanned models, show that unspanned models are rejected by the data, and investigate the implications for term premia. Section 6 concludes.

2 Spanning in Macro-Finance Term Structure Models

To lay the groundwork for our analysis, we specify two different types of MTSMs. One is the conventional spanned MTSM that has been widely used for over a decade in which macroeconomic risks are spanned by the model-implied yield curve. The other is an unspanned MTSM as introduced by JPS. Both models parsimoniously capture the behavior of interest rates across maturities and over time with a low-dimensional factor structure. We now describe the theoretical structure of these models and discuss the role of macro spanning.

2.1 The conventional spanned MTSM

Our spanned MTSM closely parallels the formulation in Joslin et al. (2013b) and can represent a vast majority of MTSMs used in the literature. Yields are collected in the vector Y_t, which contains rates for J different maturities. The risk factors that determine yields are denoted Z_t. In a macro-finance model, the risk factors Z_t include both macro factors and yield factors. We denote the $\mathcal{M}$ macro factors by M_t. For the yield factors, we are free to choose any specific yields or linear combination of yields. We write W for a $(J \times J)$ full-rank matrix that defines "portfolios" (linear combinations) of yields, $P_t = WY_t$, and we denote by P_t^j and W^j the first j yield portfolios and their weights. We take the first $\mathcal{L}$ linear combination of yields, $P_t^{\mathcal{L}}$, as the yield factors. We use PCs of observed yields, and the corresponding loadings make up the rows of W. Hence, there are $\mathcal{N} = \mathcal{M} + \mathcal{L}$ risk factors, denoted $Z_t = (M_t', P_t^{\mathcal{L}\prime})$, all of which are observable.

All MTSMs have three components: an equation relating the short-term interest rate to the risk factors, a time series model for the risk factors, and a dynamic specification for the risk factors under the risk-neutral pricing measure (or alternatively, for a stochastic discount factor). The first of these takes the one-period interest rate to be affine in all risk factors:

$$r_t = \rho_0 + \rho_M' M_t + \rho_{\mathcal{L}}' P_t^{\mathcal{L}} = \rho_0 + \rho_1' Z_t. \tag{1}$$

In the second, the risk factors are assumed to follow a Gaussian vector autoregression (VAR) under the risk-neutral probability measure:

$$Z_t = \mu^{\mathbb{Q}} + \phi^{\mathbb{Q}} Z_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}}, \quad \varepsilon_t^{\mathbb{Q}} \overset{iid}{\sim} N(0, I_{\mathcal{N}}). \tag{2}$$

Under these assumptions, bond yields are affine in the risk factors,

$$Y_t = A + BZ_t, \tag{3}$$

where the affine loadings A and B are given in Appendix A. The third and final piece of our MTSM is a time series process for Z_t (under the real-world probability measure). We specify a first-order Gaussian VAR[6]:

$$Z_t = \mu + \phi Z_{t-1} + \Sigma \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, I_{\mathcal{N}}). \tag{4}$$

This completes the model specification.

It's easy to see that the above assumptions generally imply that macro variables are spanned by (i.e., perfectly correlated with) the first $\mathcal{N}$ yield portfolios. First premultiply equation (3) with an $(\mathcal{N} \times J)$ matrix to select $\mathcal{N}$ linear combinations of model-implied yields. When using $W^{\mathcal{N}}$ this yields $P_t^{\mathcal{N}} = W^{\mathcal{N}} A + W^{\mathcal{N}} B Z_t$. This equation can, outside of knife-edge

[6] Joslin et al. (2013a) show how these models can be extended to have a higher-order lag structure under the real-world probability measure.

cases, be inverted for Z_t, and in particular the macro factors:

$$M_t = \gamma_0 + \gamma_1 P_t^{\mathcal{N}}. \tag{5}$$

That is, M_t is a deterministic function of $P_t^{\mathcal{N}}$, or equivalently of any other $\mathcal{N}$ linear combinations of yields.

An important caveat to the spanning condition (5) is that it holds for linear combinations of *model-implied* yields, but the same is not true for *observed* yields—the presence of measurement errors naturally breaks macro-spanning. We investigate in Section 4 how much unspanned macro variation can be generated by plausible small amounts of measurement error. A second caveat, related to the number of yield factors in the projection, is discussed in the next subsection.

2.2 The unspanned MTSM

We now turn to the specification of an MTSM that incorporates unspanned macro risks, based on the work of JPS. This model is a special case of the spanned MTSM. For the short rate equation, instead of (1) this model assumes that the short rate depends only on the yield factors, but not on the macro factors:

$$r_t = \rho_0 + \rho_{\mathcal{L}}' P_t^{\mathcal{L}} + 0_{\mathcal{M}}' M_t$$

Furthermore, instead of equation (2) we have a risk-neutral distribution for $P_t^{\mathcal{L}}$ that is independent of the macro factors:

$$P_t^{\mathcal{L}} = \mu^{\mathbb{Q}} + \phi^{\mathbb{Q}} P_{t-1}^{\mathcal{L}} + 0_{\mathcal{L}\times\mathcal{M}} M_t + \Sigma \varepsilon_t^{\mathbb{Q}}, \quad \varepsilon_t^{\mathbb{Q}} \overset{iid}{\sim} N(0, I_{\mathcal{L}}), \tag{6}$$

with the obvious changes in the dimensions of the parameter matrices. In this model, the risk-neutral distribution of the macro factors is not identified, since there are no payoffs that depend on macro factors and hence no pricing implications. Equation (6) implies that the macro factors do not affect risk-neutral expectations of future yield factors:

$$E^{\mathbb{Q}}(P_{t+h}^{\mathcal{L}} | Z_t) = E^{\mathbb{Q}}(P_{t+h}^{\mathcal{L}} | P_t^{\mathcal{L}}), \quad \forall h.$$

As a consequence of these assumptions, yields only depend on the yield factors but not on macro factors. That is, instead of equation (3) with a full-rank loading matrix, we have

$$Y_t = A + B_P P_t^{\mathcal{L}} + 0_{\mathcal{M}\times\mathcal{M}} M_t. \tag{7}$$

Equation (7) clarifies that in unspanned models, there is no direct link from macro factors to contemporaneous yields.

Under the real-world measure, the VAR for Z_t is the same as in the spanned model—see equation (4). The only way that the macro factors enter into the model is that they affect real-world expectations of future interest rates, and hence term premia, but they do not directly affect current interest rates. In other words, we basically have a yields-only model for the cross section of yields, which is augmented by additional predictors for the calculation of expectations and risk premia.

The spanning condition (5) does not hold in these models because we have exactly the knife-edge case that prevents us from inverting model-implied yields to obtain the risk factors.[7] Instead of (5) we have

$$M_t = \gamma_0 + \gamma_P P_t^{\mathcal{L}} + OM_t, \tag{8}$$

(equation (11) of JPS) where OM_t captures the orthogonal macroeconomic variation not captured by a projection on model-implied yields. This shows that unspanned macro risk is built into the model by construction: Because the direct link between macro factors and yields is broken, macro variation is not fully captured by model-implied yields, and the unspanned macro variation has predictive power for future yields and returns. We will investigate below whether this is necessary or desirable for empirical applications of MTSMs.

When comparing equations (5) and (8) it may appear as though spanned models impose a restrictive constraint while unspanned models allow for more flexibility. This, however, is incorrect. In fact, equation (8) also holds in spanned models, since a projection on $\mathcal{L} < \mathcal{N}$ linear combination of yields cannot fully explain the macroeconomic variation and naturally leaves an orthogonal projection residual. Because the risk factors are the same in both models, γ_0 and γ_P are also identical across models, as is OM_t.[8] Importantly, the spanned model is more flexible than, and in fact nests the unspanned model, since it does not impose the knife-edge restrictions.

[7] Formally, the matrix $W^{\mathcal{N}}B$ is singular and cannot be inverted to yield Z_t as a function of $P_t^{\mathcal{N}}$.

[8] JPS claim that conventional, spanned MTSMs impose that OM_t in equation (8) is zero (p. 1205). This would only be true when comparing, say, a $USM(3,2)$ model with an $SM(1,2)$ model. These models have different risk factors and vastly different economic implications and are not properly comparable. An appropriate comparison requires spanned and unspanned models with the same risk factors.

3 Evidence for unspanned macro information

We start our investigation of the spanning puzzle by examining the regressions for unspanned macro variation, unspanned macro risk, and unspanned macro forecasts that have been so influential in JPS, Ludvigson and Ng (2009), Duffee (2013a), (Duffee, 2013b), and others. This varied regression evidence suggests the existence of substantial unspanned macro information. Here, we reexamine each of these three strands of regression evidence using, for robustness, ten different macroeconomic variables.

3.1 Is macro variation spanned by the yield curve?

For our empirical investigation of the extent of unspanned macro information, we consider ten different macroeconomic inflation and economic activity variables. Regarding the inflation series, we use year-over-year growth in the Consumer Price Index (CPI) and the Personal Consumption Expenditure Price Index (PCEPI), in both cases excluding food and energy prices, i.e., core inflation.[9] We also include the INF measure used by JPS, which is based on survey expectations of CPI inflation expectations over the coming year (from the Blue Chip Financial Forecasts) and therefore also captures the underlying core inflation pressures in the economy. Regarding the activity measures, we include measures of the *level* and the *growth* of activity in the U.S. economy, and it is important to distinguish between them. Level measures capture deviations of economic activity from the full-employment or potential level of activity. That is, they measure the degree of slack in the economy. Our preferred measure of slack is the unemployment gap, calculated as the difference between the actual unemployment rate and the estimate of the natural rate of unemployment from the Congressional Budget Office (CBO). As a second measure of slack, we consider the output gap, measured as the difference between the log-level of GDP and the log-level of potential GDP as estimated by the CBO.[10] We consider five measures of growth in economic activity: GRO, the measure used by JPS which corresponds to the three-month moving average of the Chicago Fed's National Activity Index[11]; growth of monthly real GDP, smoothed by using either a three-month moving average (ma3) or year-over-year (yoy) growth rates; growth of industrial production; and growth of

[9]In contrast to core inflation, headline inflation, which includes volatile food and energy prices, is noisy and displays a much weaker link to monetary policy actions and interest rates. A focus on core inflation is consistent with the views and statements of monetary policymakers.

[10]To obtain monthly numbers for GDP, we use monthly estimates from Macroeconomic Advisers starting in 1992, and quarterly GDP data from the Bureau of Economic Analysis (BEA) interpolated to monthly values before 1992.

[11]It is formulated so that negative values indicate below-average economic growth and positive values indicate above-average growth.

nonfarm payroll employment (the last two are measured as three-month moving averages). Our sample period, which coincides with that used by JPS, extends from January 1985 to December 2007.

Level and growth indicators are essentially uncorrelated with each other over the business cycle. For example, just after a recession ends, growth will turn positive and even shift above trend while the level of output and employment remains depressed. Importantly, the empirical monetary policy rules literature has identified level not growth variables as most important to central banks in setting the short-term interest rate.[12] To see this difference, we run policy-rule regressions for the federal funds rate on pairs of macro variables. Each economic activity indicator that we consider is paired with CPI, whereas the inflation measures are each paired with $UGAP$. The first two columns of Table 2 report the R^2 for these policy rules, and the partial R^2 for each macro variable under consideration (that is, its explanatory power of a given macro variables in the policy rule, controlling for the effect of the other variable). The R^2 in the first column can be compared to the R^2 in univariate regressions of the policy rate on CPI, which is 0.51, and on $UGAP$, which is 0.17.

Level indicators (i.e., measures of slack) are important determinants of monetary policy, according to their explanatory power in simple policy rules. When paired with core CPI, they increase R^2 up to about 80%. In contrast, growth measures do not show a close association with the policy rate. They barely increase the R^2 compared to a univariate regression with only inflation. In particular, GRO increases the policy rule R^2 only from 0.51 to 0.53. Those variables that appear to drive monetary policy we term "policy factors." Those variables—notably the growth variables—that display only a weak direct relationship with the policy rate we term "non-policy factors." Note that all three inflation measures are very closely linked to the policy rate, giving an R^2 of about 75-80%, and are therefore included in our set of policy factors. The partial R^2 of our policy factors are all about 60% or higher, whereas the partial R^2 of the non-policy factors are at most 20%. This clearly shows the dichotomy between these two groups of variables.

To measure how much macroeconomic variation is captured by the yield curve, we regress each of the ten macroeconomic variables on the first three PCs of yields. The R^2's from these regressions, displayed in the third column of Table 2, show that most of the variation in each of the policy factors is explained by the yield curve, with R^2's around 60–70%. This is true

[12] Notably, the Taylor rule uses a levels output gap and not a growth rate. More generally, the use of core CPI and the unemployment gap are supported by estimated monetary policy rules and by the statements of monetary policymakers. See, among many others, Taylor (1993), Taylor (1999), Orphanides (2003), Bean (2005), and Rudebusch (2006). The low weight on growth in monetary policy rules can also be shown to be optimal (e.g. Rudebusch, 2002).

for measures of slack as well as for core inflation measures. In contrast, only a small portion of the variation in non-policy factors, including GRO, is captured by yields—the R^2's are all near or below 30%.

To help uncover *how* the yield curve captures macro variation, the last three columns of Table 2 report R^2's for univariate regressions of macro variables on each yield PC separately. As usual, the first three PCs correspond to level, slope, and curvature of the yield curve. Measures of slack are most closely correlated with the slope of the yield curve, while inflation measures are mainly correlated with the level. In contrast, growth measures are correlated most strongly with the curvature. Given that the curvature accounts for only a small and noisy portion of yield curve variation, this correlation with growth variables could be a sign of overfitting and further evidence of the tenuous relationship between yields and growth measures.

To further document these differing correlations, Figure 1 provides an expanded and reinterpreted version of Figure 2 in JPS, which showed that the three yield PCs had weak correlations with GRO. Our Figure 1 shows the second PC (slope), GRO, and $UGAP$, with all three series are standardized to have zero mean and unit variance. This figure illustrates the crucial difference between gap (policy) and growth (non-policy) factors. The strong comovement of $UGAP$ and the slope is very clear—the correlation coefficient is 0.84. On the other hand, GRO is basically uncorrelated with the slope. In other words, policy factors are closely related to, that is essentially spanned by, the yield curve.

What emerges from this evidence is an explanation for the source of spanned and unspanned macro variation that is based on the monetary policy reaction function, which provides a systematic link between certain macroeconomic variables and interest rates.[13] Policy factors—those macro variables that subtantially drive monetary policy—show very little evidence for unspanned macro variation, and are essentially spanned by the yield curve. This is true for measures of resource slack and for core inflation. On the other hand, the non-policy factors—and specifically, growth variables—display a weak relationship with the policy rate and, consequently, also exhibit significant unspanned variation. This reflects the low weight these variables have in directly setting the short-term interest rate by the monetary authority.

[13] This is consistent with the findings in Diebold et al. (2006) and Rudebusch and Wu (2008) in which the central bank adjusts the short rate and the slope of the yield curve in response to cyclical fluctuations in resource utilization, and the level of the yield curve adjusts to changes in inflation expectations and the perceived central bank inflation target.

3.2 Are expected bond returns spanned by the yield curve?

We now turn to the question as to whether macroeconomic variables contain information that is useful for predicting excess bond returns, above and beyond the information contained in the yield curve. That is, does the unspanned macro variation represent unspanned macro risk? Following standard practice—see, for example, Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009)—we run predictive regressions for annual excess bond returns, using the average excess returns for bonds of different maturities (two through ten years in our case) as the dependent variable. The predictors are the first three PCs of the yield curve and each of our macroeconomic variables.

Table 3 shows the results for these regression-based tests for unspanned macro risk. The first column reports the R^2 of the predictive regression. Yield PCs alone predict annual excess returns with an R^2 of 20%, which is the benchmark against which to compare the R^2 reported in the first column of the table. The second column shows t-statistics for testing the hypothesis that the macro variable can be excluded from this regression, using heteroskedasticity- and autocorrelation-consistent (HAC) standard errors. We also calculate the relative root-mean-squared error of the predictions that include the macro variable compared to those that use only yield PCs—values below one indicate improvements of forecast accuracy due to inclusion of macro information. While the t-statistics measure statistical significance, the relative RMSEs measure the economic significance of unspanned macro information for predicting excess bond returns.

Measures of economic slack do not help to predict bond returns in our data.[14] The same holds true for core inflation measures. In general, we find that the policy factors—those variables that display little unspanned variation—do not add significant predictive power beyond the yield factors. This is not surprising because the policy factors contain very little independent information. On the other hand, four out of five non-policy factors display in-sample predictive power that is significant at least at the 10% level. As for our cross-sectional spanning regressions, there appears to be a dichotomy between policy and non-policy factors, though one exception to this pattern is INF, a policy factor which exhibits significant predictive power. However, even for those variables that exhibit statistically significant predictive power, the economic significance is quite modest—the RMSE is improved by only 1-8% by inclusion of a predictor that measures economic growth.

Overall, there is some evidence for predictive power of macro variables for bond returns,

[14] This stands in contrast to the findings of Cooper and Priestley (2008), who document predictive power of the output gap for excess bond returns. They use a similar sample period as Cochrane and Piazzesi (2005), which has been found by Duffee (2013b) to exhibit extraordinary predictability of bond returns.

i.e., for unspanned macro risk, consistent with the findings of Cooper and Priestley (2008) and Ludvigson and Ng (2009). But the evidence appears somewhat weak and variable across different macroeconomic data series. Duffee (2013b) is also skeptical about its robustness, since he finds that it may be sensitive to the sample period used. Also, inference about significant predictive power in bond return regressions is notoriously problematic, because of coefficient bias and well-known issues with HAC standard errors in small samples, which are due to the highly persistent predictors and overlapping (annual) returns. Finally, the evidence on the predictive power of macro variables is in-sample and may not be reflected in out-of-sample forecasting.[15] In any case, in Section 4 we show that this excess returns regression evidence—even if taken at face value—is not able to invalidate the spanned models or really distinguish spanned and unspanned models.

3.3 Are macro forecasts spanned by the yield curve?

The third and final dimension of macro spanning that we investigate concerns forecasts of macro variables. If macro information is spanned by the yield curve, then macroeconomic forecasts made using the information in the yield curve cannot be improved upon by including own lags of macro variables. In particular, the persistence in macro variables would be completely captured by the yield curve. This would hold true even in the case that macro spanning holds but observed macro variables are measured with (serially uncorrelated) measurement errors. For this reason, Duffee (2013b) considers this analysis "more direct evidence that the problem is misspecification" (p. 412). Here we will consider the evidence for what we call "unspanned macro forecasts" in our ten macro variables. In Section 4 we will then investigate whether spanned MTSMs are inconsistent with this evidence, that is, whether empirical findings of unspanned macro forecasts in fact constitute evidence for misspecification of these models.

To study unspanned macro forecasts we take a similar approach as Duffee (2013b). We obtain yields-only macro forecasts by regressing the value of the macro variable in month $t+1$ on three PCs of yields dated at t, and compare these to macro-yields forecast in which the predictors are augmented by the lagged macro variable. In Table 3 we report HAC t-statistics for the null hypothesis that the lagged macro variable can be excluded, i.e., that macro forecasts are spanned, as well as relative RMSE comparing macro-yields to yields-only forecasts. We also report first-order autocorrelations of macro variables to help our understanding of the role of persistence in this context.

[15]Note that *GRO*, the Chicago National Activity Index, was not available to investors in real time, but only became available in 1999.

The evidence for unspanned macro forecasts is strong. Rejections of the spanning hypothesis are both statistically and economically highly significant in our monthly data set. As one would expect, the more persistent variables generally display a larger degree of unspanned persistence, i.e., a larger improvement of forecast accuracy with the inclusion of own macro lags. Our results are qualitatively consistent with Duffee (2013b), but our evidence is stronger because our variables are more persistent—our data are monthly, and our measures of slack and (year-over-year) core inflation are more persistent than Duffee's growth and (quarter-over-quarter) headline inflation measures.

Overall, our empirical investigation has shown that there are three distinctly different dimensions of unspanned macro information. If macro spanning was true for some macro variables in the data-generating process, these dimensions would be equivalent and these variables would exhibit neither unspanned macro variation, nor unspanned risk, nor unspanned forecasts. However, macro spanning does not hold for observed yields and macro variables, and these three different dimensions of the violation of spanning are distinct in the sense that those variables with a large degree of unspanned macro variation are not necessarily the same as those with large unspanned macro risk or unspanned forecasts. In particular, we find that slack and inflation measures display little unspanned variation, which we can rationalize with observations about monetary policy, while at the same time these variables exhibit a significant degree of unspanned forecasts, which is explained by their high persistence. In this sense, we have shed new light on the precise character of the lack of macro spanning in the data. We will now link this evidence to the commonly used models for capturing macro-finance interactions. Importantly, we will show that these models in fact are flexible enough to capture all three dimensions of this evidence, in contrast to claims to the contrary in the macro-finance literature.

4 Do spanned MTSMs contradict the regressions?

In this section, we assess whether MTSMs can reproduce the regression evidence described above. To do so, we first estimate both spanned and unspanned MTSMs from the actual data. Then, we generate simulated data samples from these models and use those samples to estimate the three regressions above to assess the amount of unspanned macro information. Contrary to claims in the literature, we find that spanned MTSMs are completely consistent with the regression evidence on unspanned macro information.[16]

[16] That is, the unspanned macro information regression tests have no power to reject reasonable empirical MTSMs even if those models impose theoretical macro spanning.

4.1 Estimated spanned and unspanned MTSMs

We consider both spanned and unspanned models in our empirical assessment of the spanning puzzle, using the model specifications laid out in Sections 2.1 and 2.2. We denote the spanned models by $SM(\mathcal{L}, \mathcal{M})$, and the un-spanned models by $USM(\mathcal{L}, \mathcal{M})$. Note that each model $USM(\mathcal{L}, \mathcal{M})$ is a nested special case of model $SM(\mathcal{L}, \mathcal{M})$, with knife-edge restrictions imposed that sever the direct link from macro variables to yields. These restrictions will be tested explicitly in Section 5. We focus on models with three yield factors and two macro factors, that is, the $SM(3, 2)$ and $USM(3, 2)$ models. However, all of our results were robust to changes in the number of yield factors employed. In particular, we have estimated models with one or two yield factors, and found our conclusions regarding the implications of macro spanning and of knife-edge unspanned restrictions unchanged.

Our models are estimated with yield data that match JPS in construction and sample and consist of monthly observations from January 1985 to December 2007. The mid-1980s start date avoids mixing different monetary policy regimes (Rudebusch and Wu, 2007) while ending the sample before 2008 avoids the recent zero-lower-bound episode, which is troublesome for affine models (Bauer and Rudebusch, 2013). The yields are unsmoothed zero-coupon Treasury yields, bootstrapped from observed bond prices using the Fama-Bliss methodology.[17] The yield maturities are three and six months, and one through ten years. To show the robustness of our results, we estimate our models using two different sets of macroeconomic series. The first set follows JPS and includes their measure of the macroeconomic growth (GRO), and their survey-based measure of inflation (INF). The second set includes measures of economic activity and inflation that are policy factors and therefore more standard in the context of monetary policy analysis, namely, the unemployment gap ($UGAP$) and core CPI inflation (CPI). As noted above, while INF and CPI are highly correlated, the two cyclical indicators GRO and $UGAP$ are very different.

An additional important aspect for the estimation of any term structure model is the choice of measurement error specification. Because a low-dimensional factor model cannot perfectly match the entire yield curve, it has always been common—even in yields-only models—to include measurement errors to avoid stochastic singularity. We denote the observed yields $Y_t^o = Y_t + e_t$, where the J-vector e_t contains serially uncorrelated Gaussian measurement errors. We assume that the errors on each maturity have equal variance, σ_e^2, so that the likelihood tries equally hard to match yields of all maturities.[18] This same specification is

[17] We thank Anh Le for supplying these data.

[18] Our assumption that the yield factors are observable implies that $W^{\mathcal{L}} e_t = 0$, so that e_t effectively contains $J - \mathcal{L}$ independent errors. This assumption is unimportant for estimation, since assuming unobserved yield factors and using the Kalman filter in the estimation leads to virtually identical parameter estimates.

standard and is used by JPS and many others. The presence of measurement error drives a wedge between model-implied and observed yields, which we will show below to have important implications for the observability of the theoretical macro-spanning condition.[19]

The macroeconomic variables are assumed to be observed without errors. This is the usual assumption in the macro-finance term structure literature. For example, there are no macro measurement errors in the (spanned) TS^n models in Joslin et al. (2013b) and in the (unspanned) models in JPS. Of course, measurement error for the macro variables would also break macro spanning, by creating unspanned macro variation. This would reinforce our resolution to the spanning puzzle. We do not pursue this route because we want to challenge the spanned MTSM as much as possible and investigate whether it can produce unspanned macro information for specifications that are typical in this literature, with only small yield measurement errors.

Finally, estimation is carried out using maximum likelihood. Normalization assumptions are needed to identify the parameters of the model, because otherwise there are invariant rotations that change the parameters but not the observable implications. For the spanned model, we use the canonical form and parameterization of Joslin et al. (2013b). This normalization is based on the idea that one can rotate the risk factors into $P_t^{\mathcal{N}}$, and then apply the canonical form of Joslin et al. (2011). The fundamental parameters of the model are $r_\infty^{\mathbb{Q}}$, the long-run risk-neutral mean of the short rate, $\lambda^{\mathbb{Q}}$, the eigenvalues of $\phi^{\mathbb{Q}}$, the spanning parameters γ_0, γ_1, the VAR parameters μ, ϕ, and Σ, and the standard deviation of the measurement errors, σ_e.[20] For the unspanned model, we use the canonical form of JPS. In this case, the free parameters are $r_\infty^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$, μ, ϕ, Σ, and σ_e. Conveniently, for both spanned and unspanned models μ and ϕ, as well as σ_e can be concentrated out of the likelihood function, meaning that for given values of the other parameters, the optimal values of these parameters can be found analytically. Hence we only need to search numerically for the maximum likelihood over the remaining parameters, of which there are 33 for $SM(3,2)$ and 19 for $USM(3,2)$.

Our model specifications do not impose any overidentifying restrictions, i.e., they are maximally flexible. An alternative is to impose restrictions on risk prices, which typically improves inference about risk premia by making better use of the information in the cross section of interest rates—for an in-depth discussion see Bauer (2014).[21] In their estimation

[19] Some have argued that yield measurement errors may be autocorrelated, including Hamilton and Wu (2011) and Adrian et al. (2012). While this may be the case, assuming autocorrelated errors would of course not change the fact that yield measurement errors introduce unspanned macro variation.

[20] The parameters ρ_0, ρ_1, $\mu^{\mathbb{Q}}$, and $\phi^{\mathbb{Q}}$ are determined by the fundamental parameters according the mapping provided in Appendix A of Joslin et al. (2013b).

[21] The underlying problem is the short length of the interest rate data samples commonly used, coupled with the high persistence of the level of interest rates. This results in small-sample bias in estimates of the time

of an MTSM with unspanned macro risks, JPS impose a number of zero restrictions on risk price parameters, guided in their model choice by information criteria.[22] We have carried out a similar model selection exercise and arrived at very similar restrictions, but we found these restrictions to be immaterial for our results. The reason is that these restrictions mainly affect the VAR dynamics, which are unimportant for our results regarding macro spanning. To allow for an easy comparison across different models—including spanned and unspanned macro-finance models as well as yields-only models—we focus exclusively on maximally flexible models.

In terms of estimation results, we omit individual parameter estimates for the sake of brevity, and instead focus on the cross-sectional fit of the models. Table 1 reports root-mean-squared errors (RMSEs), calculated for selected individual yields as well as across all yields. All four models fit yields well, with fitting errors on average being around five to six basis points.[23] The accurate fit of our models is due to the fact that the three yield factors well capture the variation in the yield curve (Litterman and Scheinkman (1991)). The spanned models achieve a slightly better fit because the macro variables also affect model-implied yields and can capture some additional yield variation. Our spanned models fit the yield curve much better than those in Joslin et al. (2013b) or Joslin et al. (2013a), which allow for only one or two yield factors.

4.2 Simulation design

With parameter estimates for four different MTSMs in hand—$SM(3,2)$ and $USM(3,2)$, each estimated with two different macro data sets (GRO/INF and $UGAP/CPI$)—we simulate 500 macro/yield data sets from each model. Our procedure is to simulate yield and macro factors from the VAR, construct fitted yields using the affine factor loadings, and add *i.i.d.* Gaussian measurement error to obtain simulated yields.[24] The measurement errors have a standard deviation of σ, which is the same across maturities. We consider the effects of changing σ from the maximum likelihood estimates to a small value of 1 basis point and to

series parameters. One approach to address this problem is to take advantage of the information in the cross section of interest rates by selecting plausible restrictions on risk pricing, as in Bauer (2014) and JPS. An alternative approach is to directly adjust for the small-sample bias, as in Bauer et al. (2012).

[22]In addition, they restrict the largest eigenvalue of ϕ to equal the largest eigenvalue of $\phi^{\mathbb{Q}}$.

[23]The two unspanned models achieve the exact same fit to the yield curve because the yield factors are the same and macro variables do not enter into the bond pricing.

[24]We initialize the state vector with zeros and simulate the VAR for 500 periods, which approximates a draw from the unconditional distribution. Then we use the next T observations as the simulated series. Note that in our estimation we assumed observable yield factors, whereas in our simulation setup all yields are measured with error and yield factors are latent. This is inconsequential because assuming latent factors in the estimation gives virtually identical parameter estimates (see also footnote 18).

zero. The simulated data sets have the same length as the actual data, which is $T = 276$ months. In each simulated yield data set, we estimate PCs, which capture the information in the simulated yield curves. Then we run regressions in the simulated macro/yield data that are similar to those that we estimated in Section 3. We now discuss the simulation-based results for each of the three dimensions of unspanned macro information in turn. In Appendix B, we consider large-sample results for these spanning regressions based on model-implied population moments, which provides further detail and intuition for the regression evidence on unspanned macro information.

4.3 Unspanned macro variation in MTSMs

We first focus on the cross-sectional regressions of macroeconomic variables on the PCs of contemporaneous yields. The R^2's of these regressions are shown in Table 4. The top two rows report the results of "macro on yields" regressions using the actual data on macro variables and either three or five yield PCs. These R^2's provide the benchmark against which to compare the model-based simulated R^2 distribution. As noted in Section 3, GRO exhibits much unspanned variation, while $UGAP$ and both inflation measures are largely spanned by the information in yields.

To investigate whether these results provide any information with which to discriminate between spanned and unspanned models, we examine regressions of simulated macro data on PCs of simulated yield data. Let us first consider the results for the spanned model $SM(3, 2)$. This model has five factors, so five (linear combinations of) model-implied yields completely span the macro variation. Hence, when we simulate yields without measurement error (the line with $\sigma = 0bp$) and regress macro variables on five PCs, we find an R^2 of 1 in every simulated data set. In this case, which is implicitly what JPS base their arguments on, the model cannot reproduce any unspanned macro variation and appears inconsistent with the data. However, this conclusion is not justified for an empirical application with plausible measurement error. This measurement error breaks theoretical spanning by driving a wedge between model-implied and observed yields. Adding measurement error with a standard deviation equal to the maximum likelihood estimate from the real-world data ($\sigma = 5.8bp$), the amount of unspanned macro variation in the simulated data closely matches that in the actual data. Using five PCs in the spanning regression, the mean simulated R^2 for GRO is 0.37 with a standard deviation of 0.11 across the 500 simulated data sets. The R^2 of .38 in the actual data is essentially at the center of the simulated distribution. Similar results hold for the INF, $UGAP$, and CPI regressions on five PCs. That is, with plausible measurement error, the spanning regressions provide no evidence against a spanned MTSM. These results hold with just a minuscule measurement error with $\sigma = 1bp$ (with the exception of GRO). Overall then, we find that measurement error easily reconciles spanned models with the regression evidence. In particular, the presence of measurement error breaks the theoretical spanning condition and can quantitatively reproduce the patterns we observe in the data. Importantly, even tiny measurement errors that are smaller in magnitude than usual are able to generate substantial unspanned macro variation.[25]

It's also worth noting that theoretical spanning only holds if the number of PCs in the spanning regression is at least as large as the number of pricing factors. So for $SM(3,2)$ spanning only occurs when using information in five (linear combinations) of the model-implied yields. For regressions using only three PCs of simulated yields, the R^2's are well below 1 and close to values in the actual data even without any measurement error, as shown in the middle panel of Table 4. Evidently, even in a spanned MTSM with three yield factors in addition to the macro factors, which fits the yield curve well, three PCs of yields leave a substantial amount of macro variation unspanned. This lessens the significance of the theoretical spanning condition, which only holds for an information set that includes at least as many linear combinations of yields as the total number of risk factors, which for some models may be quite a large number. Spanned models do not impose that macro variables are spanned by only $\mathcal{L}$ PCs of yields. Only when we include higher-order PCs, i.e., condition on $\mathcal{N}$ linear combinations of yields, does the spanning condition in conventional MTSMs have any testable implications. But higher-order PCs have less explanatory power for the cross section of yields and are measured much more imprecisely.

Turning to the unspanned models, $USM(3,2)$, we ask whether they can match the regression evidence better than the spanned models. In this case we use three PCs of yields, because only the three yield factors enter the affine yield equations—including more PCs does not increase explanatory power of yields but leads to multicollinearity. The results of this exercise show, not surprisingly, that the unspanned models are able to replicate the unspanned macro variation in the data. But note that both spanned and unspanned models have the exact same implications for the relationship between macro variables and $\mathcal{L}$ linear combinations of yields, because equation (8) holds in both types of models. Therefore, the results reported in the top panel of Table 4 are essentially identical to those in the middle panel.

Conventional MTSMs directly link macro variables to yields and therefore imply theoretical spanning. In contrast, our macroeconomic intuition and much empirical evidence tells us that spanning cannot be literally true. The evidence we presented here resolves this puzzle. These

[25] JPS claim that "the spanning property is independent of the issue of errors in measuring either bond yields or macro factors" (p. 1206). We disagree with this statement, based on our evidence that measurement error can reconcile spanned models with evidence on unspanned macro risks.

results closely parallel those when using three PCs of yields that were simulated by our spanned models. There is simply no reason to prefer unspanned models over spanned models based on these findings.

4.4 Unspanned macro risk in MTSMs

We now turn to the models' implications for the predictive power of macro variables for bond returns, i.e., for unspanned macro risk. We regress annual excess bond returns, again averaged over all maturities from two to ten years, on PCs of current yields and macro variables. Table 5 reports (HAC) p-values for the null hypothesis that the two macro variables can be excluded from the predictive regressions, and relative RMSEs that compare the accuracy of macro-yields forecasts to that of yields-only forecasts. The first two rows show the results in the data when we condition on either three or five PCs of yields. The macro variables GRO and INF increase the predictive power substantially, as also reported by JPS, and are highly significant in the predictive regressions. On the other hand, $UGAP$ and CPI increase the predictive accuracy only marginally, and this increase is not statistically significant—with p-values of 0.28 and 0.18, respectively.

Can spanned models reproduce the regression evidence on bond return predictability? Of course, without measurement error (so $\sigma = 0$) and with five PCs used as regressors, the $SM(3,2)$ model results show that with certainty neither macro data set provides any additional predictive power. However, this theoretical spanning result can be overturned either by introducing measurement error or by conditioning on only three yield factors instead of all five factors. In these cases, adding macro variables increases the predictive power of the return regression, as is evident from the relative RMSEs being below one—that is, there is evidence for unspanned macro risk in the simulated data. The intuitive reason is that the PCs of observed yields do not fully capture the information in the yield curve—either due to the noise introduced by measurement error or due to missing higher-order PCs—and as a consequence, macro variables, though theoretically spanned by model-implied yields, contain additional information useful for predictions.

Our MTSMs can easily reproduce the predictive power of $UGAP$ and CPI, which is only modest in the real-world data. Both the p-values and the relative RMSEs are often near or below the values in the data. However, for GRO and INF, neither the spanned nor unspanned models produce a sufficient amount of unspanned macro risk to match the evidence in the data. The distributions of p-values and relative RMSEs across simulated samples indicate that the predictive power of the macro variables is usually neither statistically nor economically significant. There are two possible explanations for why none of the models, including the one used by JPS, can match this specific aspect of return predictability in the data. First, this may reflect the fragility of the regression evidence for the predictive power of GRO and INF in the real-world data. Second, the models' first-order Markov structure for monthly yields and macro variables can only partially capture the predictability of annual excess returns (Cochrane and Piazzesi, 2005).

The important point to take away from this analysis is that unspanned models do not have any advantage over spanned models in capturing the predictability of excess bond returns. In the existing literature, the use of unspanned MTSMs is typically motivated with the regression evidence for the association of macro variables with bond risk premia—see JPS, Coroneo et al. (2013), and Priebsch (2014), among others. However, we clearly show that the evidence on unspanned macro risk does not give any reason to prefer unspanned models over spanned models.

4.5 Unspanned macro forecasts in MTSMs

Finally, we investigate which MTSMs can reproduce the empirical evidence on unspanned macro forecasts, that is, the fact that information in the yield curve does not fully capture the persistence of macroeconomic variables. The metric we focus on is the relative RMSE of (one-month-ahead) macro-yields forecasts vs. yields-only forecasts for each macro variable—values below one indicate the presence and magnitude of unspanned macro forecasts (see also the last column of Table 3). Again, we compare the values obtained for regressions using the real-world data to the distribution of values in regressions using simulated data.

Table 6 shows the results of this analysis. The first two rows report the values for the data, using either three or five PCs of yields to obtain macro forecasts. As was noted in Section 3, the evidence for unspanned macro forecasts is very strong in our data. The results for the simulated data show that both unspanned and spanned MTSMs can match this evidence quite easily. The only case in which data simulated from an MTSM does not exhibit any unspanned macro forecasts is when we use the spanned model, condition on all five PCs, and do not allow for any yield measurement errors. We know that in this case macro spanning holds in the simulated data—we saw this already in Tables 4 and 5. Therefore, lags of macro variables are not needed for macro forecasts, because the yield curve completely captures all relevant information. However, outside of this special case, the MTSMs generate a substantial amount of unspanned macro forecasts, generally sufficient to match the evidence in the real-world data. In almost all specifications, the distribution of the relative RMSEs across simulations is centered near the value in the data, and a sizable fraction of the simulations exhibit at least as much unspanned macro forecasts as in the real-world data.

The regression evidence on unspanned macro forecasts in simulated data is in line with our results regarding unspanned macro variation and unspanned risk: spanned models are able to match this evidence just as well as unspanned models. That is, the regression evidence simply cannot discriminate between these spanned and unspanned MTSMs, and does not provide any grounds for preferring a particular class of models.[26] Importantly, since we have shown spanned models to be consistent with the regression evidence, this resolves the spanning puzzle.

4.6 The role of measurement error

Our simulations have shown that if we condition on $\mathcal{N}$ linear combinations of yields in a spanned model, then measurement error plays a key role in breaking theoretical spanning. But the large amount of unspanned macro variation that can be generated using very small measurement errors appears puzzling. After all, "the measurement error of yields is tiny relative to the variability of yields" (Duffee, 2013b, p. 412).[27] Furthermore, JPS and Joslin et al. (2013b) note that low-order PCs of observed yields are generally very close to PCs of model-implied, "true" yields, not only because measurement error is small, but also because it gets washed out by taking these linear combinations of yields.[28] How does the spanned model generate such a substantial amount of unspanned macro information with tiny measurement errors? In other words, what is the intuition for the measurement-error solution to the spanning puzzle?

The crucial point is that measurement error matters little for low-order PCs, but it substantially affects higher-order PCs. The amount of macro information captured by low-order PCs (say, level, slope, and curvature) is essentially identical in spanned and unspanned models, and largely unaffected by the presence of measurement error. This is clearly evident in the top and middle panels of Tables 4 through 6. The theoretical spanning condition implies that spanned models capture the remaining macro information in the higher-order PCs, for example the fourth and fifth PCs in $SM(3,2)$. These have very different properties from low-

[26]Our results are not driven by small-sample bias. We have investigated the simulation-based evidence in large samples (and using model-implied population moments), and our conclusions remain the same. While return-predictability is well-known to be affected by small-sample bias (Stambaugh, 1999) and this is true in our setting as well, the finding that spanned and unspanned models deliver the same implications for bond risk premia is unaffected by this fact.

[27]Duffee goes on to conclude that "if the five-factor model is correctly specified, the observed macro variables are primarily noise." However, we do not allow for measurement errors on macro variables, so this cannot be the correct explanation.

[28]Specifically, JPS remark that "experience shows that the observed low-order PCs comprising [the yield factors] are virtually identical to their filtered counterparts in models that accommodate errors in all PCs" (pp. 1206).

order PCs. In particular, they are on average smaller in magnitude, and are estimated much less precisely. Importantly, adding (the same linear combinations of) measurement error substantially lowers the information content in the higher-order PCs. This is how, mechanically, the introduction of small measurement error reduces the information content in yield PCs by a large amount. Appendix B shows this intuition mathematically. In other words, only the level and slope (and to some extent the curvature) of the yield curve are large relative to measurement errors, but spanned macro-finance models imply that other, more subtle linear combinations of yields also contain macro information, which do not survive the introduction of even small noise for yields.

There is a close parallel between the role of measurement error for macro spanning and the usual motivation for the inclusion of measurement error in term structure models. A low-dimensional factor structure cannot hold exactly in the data and would therefore be rejected. Nevertheless, we commonly use models with only few factors, augmenting them with measurement error to avoid stochastic singularity. Similarly, theoretical spanning cannot hold exactly in the data and is easily rejected. But again, measurement error comes to the rescue in a similar way, in this case by breaking theoretical macro spanning, just as it breaks stochastic singularity. Measurement error generally plays a crucial role in the yield-curve literature, by reconciling term structure models with the data.

5 Direct tests of spanned and unspanned models

In the previous two sections, we showed that spanned MTSMs could account for the regression evidence on unspanned macro variation and risk; that is, when properly evaluated, there was no spanning puzzle. We now turn to direct likelihood ratio tests comparing spanned and unspanned models. We also contrast their economic implications for risk premia.

5.1 Testing the knife-edge unspanned risk restriction

The distinguishing feature of an unspanned MTSM is that all of its loadings of yields on macro factors are equal to zero. Hence, yields cannot be inverted to infer any of the macro variables. This parallels the models of unspanned volatility proposed by Collin-Dufresne and Goldstein (2002) and others, where yields have zero loadings on volatility factors. In both classes of models, "knife-edge" restrictions are required in order for unspanned factors to exist.[29] Such

[29] Collin-Dufresne and Goldstein (2002) speak of "knife-edge" parameterizations that give rise to unspanned volatility factors, and Duffee (2013a) uses this term in the context of unspanned macro factors. Knife-edge restrictions have the effect that the relevant factor loadings, which are determined by the model's parameters,

macro-finance knife-edge restrictions could hold in principle, as in the case of unspanned volatility.[30] However, unlike unspanned volatility factors, unspanned macro factors have a somewhat peculiar interpretation. For a macro unspanned factor to exist, it must affect, at each maturity, the risk-neutral yield and the risk premium with exactly the same magnitude but with opposite sign so that the yield at that maturity is unchanged, i.e., it has a zero loading on the macro variables. While several authors have noted the presence of knife-edge restrictions for unspanned MTSMs—including Duffee (2011), Gürkaynak and Wright (2012), and Joslin et al. (2013b)—we are the first to actually conduct an empirical hypothesis test of these restrictions.

These tests are based on the straightforward nesting relationship between spanned and unspanned models. As described in Section 2, for any number of yield factors, the $SM(\mathcal{L},\mathcal{M})$ nests the $USM(\mathcal{L},\mathcal{M})$. The risk factors are exactly the same in both models, while the knife-edge restrictions lead to a lower-dimensional parameter space for the unspanned model $USM(\mathcal{L},\mathcal{M})$. Specifically, there are $\mathcal{M}(2+\mathcal{N})$ fewer parameters in an unspanned model relative to the spanned model—there are $\mathcal{M}$ zero restrictions in the short-rate equation, and the $\mathcal{M}(1+\mathcal{N})$ spanning parameters in γ_0 and γ_1 are absent.

For our test of unspanned risk ristrictions, we examine the models $USM(3,2)$ and $SM(3,2)$ (although, again, our results are robust to the number of factors employed). The $USM(3,2)$ is a restricted version of the $SM(3,2)$ with 14 fewer parameters. The log-likelihood values for these models are shown in Table 7 for two different pairs of macro variables (GRO, INF) and $(UGAP, CPI)$. For either macro pair, the spanned models fit the data substantially better than the unspanned ones. Indeed, likelihood-ratio tests, also reported in Table 7, reject the unspanned restrictions with large χ^2-statistics in all cases—two orders of magnitude larger than the 5% critical value. These rejections reflect the fact that inclusion of macro factors in the measurement equations for yields improves the cross-sectional fit.[31] Allowing macro variables to enter as pricing factors lowers the root-mean-squared fitting errors of the models—see Table 1. While these declines are modest, they are highly statistically significant, given that they translate into significant increases in the log-likelihood function.

This evidence casts doubt on the unspanned model proposed by JPS, as the key restrictions on which it relies are rejected by the data. The use of unspanned macro models has typically been motivated and justified only indirectly, on the basis of the regression evidence

end up being exactly zero.

[30] Bikbov and Chernov (2009) conduct an analysis of unspanned stochastic volatility that is in many ways similar to ours.

[31] The time series fit remains essentially unchanged by the knife-edge restrictions, since the VAR parameters μ and ϕ are identical in the spanned and unspanned models (see also JPS, p. 1207, on this point).

for unspanned macro information. In contrast, we provide the first direct statistical evidence of the plausibility of unspanned macro-finance models. Our tests reject these models with high statistical significance.

5.2 The spanning restrictions tested by JPS

On the surface, our results seem to completely contradict the test results in JPS (p. 1214) that purport to show that the spanned MTSM is rejected in favor of the unspanned MTSM. Just as puzzling is the fact that JPS test the spanned model as a restricted version of the unspanned model. This is precisely the reverse direction of the specification nesting that we have described above in which the unspanned model involves knife-edge zero restrictions on the spanned model. Here we elucidate the nature of the restrictions tested by JPS.

JPS start with an unspanned model, and then consider the restricted model, which they denote by $\mathcal{M}_{span}$, in which those columns of ϕ that correspond to the macro factors are set to zero. A likelihood-ratio test strongly rejects these restrictions, with a reported χ^2-statistic of 1,189. What is the interpretation of this result?

The restrictions of $\mathcal{M}_{span}$ do not imply a spanned MTSM, but instead effectively lead to a yields-only model: Only yield factors affect bond prices and only yield factors determine expectations and risk premia.[32] The only difference between $\mathcal{M}_{span}$ and a proper yields-only model is that $\mathcal{M}_{span}$ includes two equations for forecasting macro variables in the VAR, and so its likelihood function includes (large) macro forecast errors resulting from yields-only forecasts. For yields and risk premia, $\mathcal{M}_{span}$ and yields-only models have the exact same observational implications.

The likelihood-ratio test in JPS is not a direct test of macro-spanning in MTSMs, but it is an assessment of whether current yields are sufficient to forecast macro and yield factors. The rejection reflects the fact that in these forecasting regressions, macro variables have predictive power.[33] The test result in JPS should not be viewed as evidence that we should be using unspanned MTSMs. Instead, it is simply evidence of unspanned macro risks in GRO and INF, and exactly parallels the regression-based evidence. We have shown in Section 4 that

[32] The restrictions of $\mathcal{M}_{span}$ do not imply that macro variables are spanned by yields, but only that expectations are spanned. While the former (exact spanning) is a sufficient condition for the latter (expectations spanning), it is not a necessary condition.

[33] The particularly large χ^2-statistic is driven by the fact that the persistence of macro variables is not fully captured by conditioning on current yields (Duffee, 2013b). The zero restrictions in $\mathcal{M}_{span}$ on ϕ zero out *own lags* of macro variables. If instead we only restrict to zero the coefficients on macro variables in the VAR equations for the yield factors, the χ^2-statistics are only 74 for the $USM(3, 2)$ model with GRO and INF (the JPS model), and 18 for the same model with $UGAP$ and CPI. These weaker rejections reflect the predictive power of macro variables for yields, which is modest.

such regression evidence for unspanned macro risk can easily be reproduced by a spanned MTSM.

5.3 Risk premia in spanned and unspanned MTSMs

We now investigate the role of spanned and unspanned macro risks for model-based estimation of term premia. First, we revisit the estimates of JPS, using the same macro data, GRO and INF. Figure 2 shows two-to-three-year forward term premia from models $USM(3,2)$ and $SM(3,2)$, as well as from a three-factor yields-only model.[34] This figure is comparable to Figure 1 of JPS: Our $USM(3,2)$ model corresponds to their $\mathcal{M}_{us}$ model, and our yields-only model corresponds to their $\mathcal{M}_{span}$ model. The comparison shows that our estimated term premia closely resemble those of JPS. This is true although our models are maximally flexible and in this regard differ from those in JPS who impose various overidentifying restrictions (see our discussion in Section 4.1) and although we use a slightly different yield data set.

Figure 2 shows that the spanned and unspanned models imply essentially identical forward term premia—the two lines corresponding to models $SM(3,2)$ and $USM(3,2)$ lie almost exactly on top of each other. Evidently, the knife-edge restrictions of unspanned models do not materially affect estimated term premia. While we have shown above that the rejections of these restrictions are statistically significant, Figure 2 reveals that from the perspective of term premium estimation, these rejections are not economically significant. The same holds for the improvements in cross-sectional fit achieved by spanned models, which are on the order of one basis point or less, and hence are also not economically significant. We cannot rule out that there are other empirical objects of interest for which the knife-edge restrictions have a material impact, but for the estimation of term premia they are inconsequential—spanned and unspanned models give practically identical results.

This finding sharply contrasts with the claim in JPS that unspanned models "accommodate much richer dynamic codependencies among risk premiums and the macroeconomy than in extant MTSMs" (p. 1198). First, unspanned models are in fact restricted versions of extant spanned models. Second, both types of models allow for essentially the same risk premium dynamics—incorporating unspanned macro risks in an MTSM does not change the term premium implications of the model.

The yields-only model (which corresponds to the $\mathcal{M}_{span}$ model in JPS) implies a very different term premium than the macro-finance models. Just like the rejections of the VAR

[34] Our yields-only model is a maximally flexible affine model, with the first three PCs of observed yields as risk factors. We use the normalization of Joslin et al. (2011) and estimate the model using maximum likelihood.

restrictions emphasized by JPS (see Section 5.2), this difference in term premia is simply due to the in-sample predictive power of the unspanned components of GRO and INF. Should we prefer this particular macro-finance term premium over the yields-only term premium? One way to answer this question is to judge the plausibility of the behavior of these risk premia from a macroeconomic perspective. On these grounds, JPS argue in favor of the macro-finance term premia, because they "show a pronounced cyclical pattern with peaks during recessions" (p. 1198). However, these peaks occur early in recessions or even before the beginning of the recessions, while the economy is still expanding briskly and risk aversion and risk compensation are low. A more plausible business cycle pattern for risk premia is to be high in troughs and low at peaks (Cochrane and Piazzesi, 2005). The yields-only term premium is therefore more plausible, since it is low late in the expansion and rises throughout recessions. It peaks near the end of the recession or early in the recovery, when economic prospects are highly uncertain. From a macroeconomic perspective, there are clearly reasons to question the plausibility of the term premium implied by an MTSM with GRO and INF. The fact that these variables display substantial unspanned variation is not necessarily a good reason to include them in an MTSM. Not only does the in-sample predictive power most likely lack robustness, it also leads to quite puzzling behavior of the resulting term premia.[35]

An MTSM with more conventional macroeconomic variables delivers term premia which do not show this puzzling behavior. In Figure 3 we show the forward term premia obtained from MTSMs with $UGAP$ and CPI, together with the yields-only term premium. Again, the implied term premia from the spanned and unspanned models are essentially identical. In this case, they both resemble the term premium from the yields-only model. The reason is that these macro variables are closely tied to monetary policy and to the yield curve and display little to no unspanned macro risks (see Section 3). From a macroeconomic perspective, the term premia in Figure 3 are much more plausible than the macro-finance term premia in Figure 2, given their more reasonable cyclical behavior.

Our evidence can be viewed as a caution against including in MTSMs macro variables that are selected based on high in-sample predictive power for excess returns. Such variables can substantially alter estimated risk premia and can reduce their plausibility.[36]

Our key point here is that spanned and unspanned models imply essentially identical

[35] Bauer et al. (2012) and Bauer et al. (2014) also discuss the countercyclical behavior of term premia estimated from term structure models.

[36] Duffee (2013b) notes that "the spanning requirement [...] reduces significantly the ability of researchers to fish for variables that forecast excess returns." Doing away with the spanning constraint removes the discipline imposed by it. Instead of simply adding variables that are found to have in-sample significance in forecasting regressions, it will be important to document robust and stable predictive power before using any particular series to augment MTSMs.

term premia. While we come to different conclusions in our comparison of spanned and unspanned models than JPS, we view their novel class of unspanned MTSMs as potentially quite useful in applications. Importantly, we have shown that the knife-edge restrictions are rejected on statistical grounds, but leave the economic implications of affine MTSMs essentially unchanged. Hence, there is no grave danger in using unspanned models for, say, analysis of bond risk premia. A benefit of these models is that the number of free parameters is reduced quite substantially, which has practical advantages. In addition to simplifying the estimation and inference, the increased parsimony may also prove useful in forecasting, where more tightly parameterized models typically do better. Overall, unspanned models may be a useful shortcut in practical applications of MTSMs.[37] We emphasize, however, that these models are not needed to match the regression evidence that is usually cited to justify their use, and that they give the same answers as spanned models to questions about risk premia.

6 Conclusion

In this paper, we have resolved the spanning puzzle in macro-finance and salvaged the conventional MTSMs widely used in the literature. First, we show that some key macroeconomic variables—those with a close connection to monetary policy—are effectively spanned by the yield curve. Specifying a spanned MTSM with these "policy factors" appears quite appropriate. We also show that, regardless of the variables used, the theoretical spanning between macro variables and model-implied yields does not have any practical significance in terms of the regression evidence. Specifically, a typical spanned MTSM with small measurement errors on yields does not conflict with the regression evidence on unspanned macro variation and unspanned macro risk. These findings should reassure the many past and present researchers employing conventional spanned models for analyzing macro-finance interactions. We also cast some doubt on the validity of the alternative unspanned MTSMs. Their knife-edge restrictions are strongly rejected in the data. At the same time, for the estimation of term premia in long-term interest rates, spanned and unspanned models appear to deliver similar results. That is, the rejections of the knife-edge restrictions of unspanned models are statistically but not economically significant. One interpretation of our results is that the choice between spanned and unspanned models is less important than the choice of which macro variables should be used to augment the information set for forecasting and inference about risk premia.

[37] An established example of another such usefully constrained model in the literature is the arbitrage-free Nelson-Siegel (AFNS) model of Christensen et al. (2011). Although modest in-sample statistical rejections of the three parameter restrictions associated with the AFNS model are not uncommon, the AFNS model provides notable economic benefits in terms of parsimony, tractability, and intuition.

One could imagine alternative solutions to the spanning puzzle. One is nonlinearity: Some structural and reduced-form MTSMs imply nonlinear mappings from risk factors to bond yields. While this theoretically breaks the (linear) spanning condition, it remains an empirical question how much unspanned macro information such nonlinearities can generate. Another possible solution is regime-switching or parameter instability across subsamples. If macro spanning holds but the parameters in the spanning relation change, then regressions using the full sample would find evidence for unspanned macro information. While our sample period is chosen to minimize the likelihood of possible structural breaks (for example due to changes in policy), we cannot rule out this possibility.[38] These explanations may contribute to the unspanned phenomenon in the data, but our results reconcile spanned models with the regression evidence without adding any assumptions or model features not commonly included in the MTSM literature.

An open question about unspanned MTSMs is whether the severing of the direct link between macro variables and yields has any serious consequences. It appears that since direct effects of macro state variables on asset prices are ruled out, the usefulness of these models for policy analysis may be limited, in particular for studying the effects of monetary policy. However, the indirect link through the interaction of macro and yield factors might be sufficient to answer most questions about macro-yield interactions.[39] More generally, the question is whether there are economic implications of MTSMs other than the one we study here, based on which either spanned or unspanned models seem preferable. Depending on the answer to this question, it may be useful to develop hybrid models with both spanned and unspanned macroeconomic risks. We leave these issues for future research.

References

Adrian, Tobias, Richard K. Crump, and Emanuel Moench (2012) "Pricing the Term Structure with Linear Regressions," Staff Report No. 340, Federal Reserve Bank of New York.

Ang, Andrew, Geert Bekaert, and Min Wei (2008) "The Term Structure of Real Rates and Expected Inflation," *Journal of Finance*, Vol. 63, pp. 797–849.

[38] Song (2014) claims to generate unspanned macro variation with regime-switching in an equilibrium MTSM. However, it remains unclear to what extent the yield measurement errors vs. the regime-switches are the cause of unspanned macro information.

[39] Priebsch (2014) argues that unspanned models can be used to study various issues related to monetary policy.

Ang, Andrew, Jean Boivin, Sen Dong, and Rudy Loo-Kung (2011) "Monetary Policy Shifts and the Term Structure," *Review of Economic Studies*, Vol. 78, pp. 429–457.

Ang, Andrew and Monika Piazzesi (2003) "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, Vol. 50, pp. 745–787.

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

Bauer, Michael D. (2014) "Bayesian Estimation of Dynamic Term Structure Models under Restrictions on Risk Pricing," Working Paper 2011-03, Federal Reserve Bank of San Francisco.

Bauer, Michael D. and Glenn D. Rudebusch (2013) "Monetary Policy Expectations at the Zero Lower Bound," Working Paper 2013-18, Federal Reserve Bank of San Francisco.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu (2012) "Correcting Estimation Bias in Dynamic Term Structure Models," *Journal of Business and Economic Statistics*, Vol. 30, pp. 454–467.

——— (2014) "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Comment," *American Economic Review*, Vol. 104, pp. 1–16.

Bean, Charles (2005) "Commentary: Separating the Business Cycle from Other Economic Fluctuations," in *Federal Reserve Bank of Kansas City Proceedings*, pp. 181–186, Aug.

Bekaert, Geert, Eric Engstrom, and Yuhang Xing (2009) "Risk, Uncertainty, and Asset Prices," *Journal of Financial Economics*, Vol. 91, pp. 59–82.

Bernanke, Ben S., Vincent R. Reinhart, and Brian P. Sack (2004) "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment," *Brookings Papers on Economic Activity*, Vol. 2004, pp. 1–100.

Bikbov, Ruslan and Mikhail Chernov (2009) "Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options," *Management Science*, Vol. 55, pp. 1292–1305.

——— (2010) "No-Arbitrage Macroeconomic Determinants of the Yield Curve," *Journal of Econometrics*, Vol. 159, pp. 166–182.

Buraschi, Andrea and Alexei Jiltsov (2007) "Habit Formation and Macroeconomic Models of the Term Structure of Interest Rates," *The Journal of Finance*, Vol. 62, pp. 3009–3063.

Chernov, Mikhail and Philippe Mueller (2012) "The Term Structure of Inflation Expectations," *Journal of Financial Economics*, Vol. 106, pp. 367–394.

Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch (2011) "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," *Journal of Econometrics*, Vol. 164, pp. 4–20.

Cochrane, John H. and Monika Piazzesi (2005) "Bond Risk Premia," *American Economic Review*, Vol. 95, pp. 138–160.

Collin-Dufresne, Pierre and Robert S Goldstein (2002) "Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility," *The Journal of Finance*, Vol. 57, pp. 1685–1730.

Cooper, Ilan and Richard Priestley (2008) "Time-Varying Risk Premiums and the Output Gap," *Review of Financial Studies*, Vol. 22, pp. 2801–2833.

Coroneo, Laura, Domenico Giannone, and Michle Modugno (2013) "Unspanned Macroeconomic Factors in the Yields Curve," Working Papers ECARES ECARES 2013-07, ULB – Universite Libre de Bruxelles.

Dewachter, Hans and Marco Lyrio (2006) "Macro Factors and the Term Structure of Interest Rates," *Journal of Money, Credit and Banking*, Vol. 38, pp. 119–140.

Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba (2006) "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of econometrics*, Vol. 131, pp. 309–338.

Duffee, Gregory R. (2011) "Information In (and Not In) the Term Structure," *Review of Financial Studies*, Vol. 24, pp. 2895–2934.

——— (2013a) "Bond Pricing and the Macroeconomy," in Milton Harris George M. Constantinides and Rene M. Stulz eds. *Handbook of the Economics of Finance*, Vol. 2, Part B: Elsevier, pp. 907–967.

——— (2013b) "Forecasting Interest Rates," in Graham Elliott and Allan Timmermann eds. *Handbook of Economic Forecasting*, Vol. 2, Part A: Elsevier, pp. 385–426.

Gallmeyer, Michael F., Burton Hollifield, Francisco Palomino, and Stanley E. Zin (2007) "Arbitrage-Free Bond Pricing with Dynamic Macroeconomic Models," Working Paper 13245, National Bureau of Economic Research.

Gürkaynak, Refet S. and Jonathan H. Wright (2012) "Macroeconomics and the Term Structure," *Journal of Economic Literature*, Vol. 50, pp. 331–367.

Hamilton, James D. and Jing Cynthia Wu (2011) "Testable Implications of Affine Term Structure Models," NBER Working Papers 16931, National Bureau of Economic Research, Inc.

Hördahl, Peter, Oreste Tristani, and David Vestin (2006) "A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics," *Journal of Econometrics*, Vol. 131, pp. 405–444.

Joslin, Scott, Anh Le, and Kenneth J. Singleton (2013a) "Gaussian Macro-Finance Term Structure Models with Lags," *Journal of Financial Econometrics*, Vol. 11, pp. 581–609.

——— (2013b) "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs," *Journal of Financial Economics*, Vol. 109, pp. 604–622.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton (2014) "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, Vol. 69, p. 11971233.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu (2011) "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, Vol. 24, pp. 926–970.

Litterman, Robert and J. Scheinkman (1991) "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, Vol. 1, pp. 54–61.

Ludvigson, Sydney C. and Serena Ng (2009) "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, Vol. 22, pp. 5027–5067.

Orphanides, Athanasios (2003) "Historical Monetary Policy Analysis and the Taylor Rule," *Journal of Monetary Economics*, Vol. 50, pp. 983–1022.

Piazzesi, Monika and Martin Schneider (2007) "Equilibrium Yield Curves," in *NBER Macroeconomics Annual 2006, Volume 21*: MIT Press, pp. 389–472.

Priebsch, Marcel (2014) "(Un)Conventional Monetary Policy and the Yield Curve," working paper, Federal Reserve Board, Washington, D.C.

Rudebusch, Glenn D. (2002) "Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty," *The Economic Journal*, Vol. 112, pp. 402–432.

——— (2006) "Monetary Policy Inertia: Fact or Fiction?" *International Journal of Central Banking*, Vol. 2, pp. 85–135.

Rudebusch, Glenn D and Eric T Swanson (2012) "The bond premium in a DSGE Model with Long-Run Real and Nominal Risks," *American Economic Journal: Macroeconomics*, Vol. 4, pp. 105–143.

Rudebusch, Glenn D. and Tao Wu (2007) "Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models," *Journal of Money, Credit and Banking*, Vol. 39, pp. 395–422.

——— (2008) "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," *Economic Journal*, Vol. 118, pp. 906–926.

Song, Dongho (2014) "Bond Market Exposures to Macroeconomic and Monetary Policy Risks," unpublished manuscript, Boston College.

Stambaugh, Robert F. (1999) "Predictive regressions," *Journal of Financial Economics*, Vol. 54, pp. 375–421.

Taylor, John B. (1993) "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 39, pp. 195–214.

——— (1999) "A Historical Analysis of Monetary Policy Rules," in *Monetary Policy Rules*: National Bureau of Economic Research, pp. 319–348.

Wachter, Jessica A. (2006) "A Consumption-Based Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, Vol. 79, pp. 365–399.

Wright, Jonathan H. (2011) "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review*, Vol. 101, pp. 1514–1534.

A Affine bond pricing

Under the assumptions of Section 2, bond prices are exponentially affine functions of the pricing factors:

$$B_t^m = e^{\mathcal{A}_m + \mathcal{B}_m X_t},$$

and the loadings $\mathcal{A}_m = \mathcal{A}_m(\mu^{\mathbb{Q}}, \phi^{\mathbb{Q}}, \delta_0, \delta_1, \Sigma)$ and $\mathcal{B}_m = \mathcal{B}_m(\phi^{\mathbb{Q}}, \delta_1)$ follow the recursions

$$\begin{aligned}
\mathcal{A}_{m+1} &= \mathcal{A}_m + (\mu^{\mathbb{Q}})'\mathcal{B}_m + \frac{1}{2}\mathcal{B}_m'\Sigma\Sigma'\mathcal{B}_m - \delta_0 \\
\mathcal{B}_{m+1} &= (\phi^{\mathbb{Q}})'\mathcal{B}_m - \delta_1
\end{aligned}$$

with starting values $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$. Model-implied yields are determined by $y_t^m = -m^{-1}\log B_t^m = A_m + B_m X_t$, with $A_m = -m^{-1}\mathcal{A}_m$ and $B_m = -m^{-1}\mathcal{B}_m$. Risk-neutral yields, the yields that would prevail if investors were risk-neutral, can be calculated using

$$\tilde{y}_t^m = \tilde{A}_m + \tilde{B}_m X_t, \quad \tilde{A}_m = -m^{-1}\mathcal{A}_m(\mu, \phi, \delta_0, \delta_1, \Sigma), \quad \tilde{B}_m = -m^{-1}\mathcal{B}_m(\phi, \delta_1).$$

Risk-neutral yields reflect policy expectations over the life of the bond, $m^{-1}\sum_{h=0}^{m-1} E_t r_{t+h}$, plus a convexity term. The yield term premium is defined as the difference between actual and risk-neutral yields, $ytp_t^m = y_t^m - \tilde{y}_t^m$.

B Spanning results in large samples

In the paper we use repeated simulations of samples with length similar to our real-world data to investigate the implications of our MTSMs for the three types of spanning regressions. An alternative to such repeated simulations of short samples is to use one very large simulated sample or, equivalently, model-implied population moments. This is what we do in this appendix.

We first define notation and show the expressions we use to calculate population R^2 and relative RMSEs. Given the linear model

$$y_t = \beta' x_t + \epsilon_t,$$

where it is assumed that (i) $\{y_t, x_t\}$ are jointly stationary and ergodic, (ii) all N regressors x_t are predetermined, and (iii) $E(x_t x_t')$ has full rank, the regression R^2 converges in probability to

$$R^2 = \frac{Var(\beta' x_t)}{Var(y_t)} = \frac{\beta' Cov(x_t)\beta}{Var(y_t)} = \frac{Cov(y_t, x_t)Cov^{-1}(x_t)Cov(x_t, y_t)}{Var(y_t)}.$$

In our notation, $Cov(y_t, x_t)$ is a $(1 \times N)$ row vector, $Cov^{-1}(x_t)$ is the inverse of the $(N \times N)$ variance-covariance matrix $Cov(x_t)$, and $Cov(x_t, y_t)$ is an $(N \times 1)$ column vector. Since the mean-squared-error converges to $Var(\epsilon_t) = Var(y_t) - Var(\beta' x_t)$, the relative RMSE of an unrestricted and a restricted model converges to

$$\frac{RMSE_{ur}}{RMSE_r} = \sqrt{\frac{1 - R^2_{ur}}{1 - R^2_r}},$$

where R^2_{ur} and R^2_r are the population R^2 of the unrestricted and restricted models, respectively.

B.1 Unspanned macro variation

In the first type of regressions, macroeconomic variables are regressed on PCs of contemporaneous yields. We denote the macroeconomic variable under consideration as m_t. To emphasize the role of measurement error we write for observed yields

$$\tilde{Y}_t = Y_t + e_t = A + BZ_t + e_t,$$

where we have $Cov(e_t) = \sigma^2 I_J$. Note that for unspanned models, the rows of B corresponding to the macro factors contain only zeros. The loadings for the principal components will be taken as fixed in this analysis, corresponding to PCs in the real-world data. They are given in the matrix W, which is a $(3 \times J)$ or $(5 \times J)$ matrix, depending on how many yield PCs are used as regressors. Hence the regressors are $\tilde{P}_t = W\tilde{Y}$, and we also define the PCs of model-implied yields as $P_t = WY_t$. This notation does not make it explicit how many yield PCs are contained in P_t—we consider the two cases with either $\mathcal{L} = 3$ or $\mathcal{N} = 5$ yield PCs. With this notation, the spanning regressions are

$$m_t = const + \beta' \tilde{P}_t + \epsilon_t.$$

The population R^2 is

$$R^2 = \frac{Cov(m_t, P_t) Cov^{-1}(\tilde{P}_t) Cov(P_t, m_t)}{Var(m_t)},$$

where we have used that fact that $Cov(\tilde{P}_t, m_t) = Cov(P_t, m_t)$. The relevant population moments are

$$\begin{aligned} Cov(\tilde{P}_t) &= WCov(\tilde{Y}_t)W' = WBCov(Z_t)(BW)' + \sigma^2 WW', \\ Cov(P_t, m_t) &= WBCov(Z_t)\iota_m, \\ Var(m_t) &= \iota_m' Cov(Z_t)\iota_m. \end{aligned}$$

where the unconditional covariance matrix of the risk factors $Cov(Z_t)$ is determined by the VAR parameters using $vec(Cov(Z_t)) = (I_{\mathcal{N}^2} - \phi \otimes \phi)^{-1} vec(\Sigma\Sigma')$, and ι_m is a column vector that selects m_t from Z_t, i.e., $m_t = \iota_m' Z_t$.

Table B.1 shows the model-implied R^2 for each of the cases and models that we considered in Section 4.3. The first thing to note is that in the case of 3 PCs, models $USM(3,2)$ and $SM(3,2)$ have essentially identical implications—both generate a very substantial amount of unspanned macro variation, sufficient to fit the R^2 in the data. Second, in the case that only 3 PCs are used, measurement error does not noticeably affect the R^2. Third, confirming our

Table B.1: Unspanned macro variation in MTSMs – large-sample results

		GRO	INF	CPI	$UGAP$
Data	3 PCs	0.279	0.812	0.809	0.725
	5 PCs	0.379	0.864	0.812	0.751
		GRO/INF Model		$UGAP/CPI$ Model	
$USM(3,2)$	3 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.280	0.785	0.761	0.709
	$\sigma = 1bp$	0.293	0.788	0.768	0.712
	$\sigma = 0$	0.293	0.788	0.768	0.712
$SM(3,2)$	3 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.278	0.773	0.753	0.723
	$\sigma = 1bp$	0.288	0.775	0.759	0.726
	$\sigma = 0$	0.288	0.775	0.759	0.726
$SM(3,2)$	5 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.480	0.875	0.759	0.745
	$\sigma = 1bp$	0.769	0.962	0.877	0.939
	$\sigma = 0$	1.000	1.000	1.000	1.000

Unspanned macro variation in large samples, measured by the theoretical R^2 implied by model parameters, for regressions of macro variables on contemporaneous yield PCs. The first two rows show R^2 for the actual historical data for comparison.

simulation results, macro spanning holds for model $SM(3,2)$ only if we use 5 PCs and do not have measurement error. In this case, the regressors are $P_t^{\mathcal{N}}$, which spans m_t, hence the R^2 is 1. Finally, even in the case that theoretical spanning holds, small measurement error with $\sigma = \hat{\sigma}_e^{MLE}$ generates substantial unspanned macro variation, usually enough to match the values in the real-world data. Overall, our large-sample results are quite similar to the small-sample results, which shows that model-implied unspanned macro variation is not a small-sample phenomenon.

B.2 Unspanned macro risk

For the analysis of unspanned macro risk we consider two alternative regressions for excess bond returns: The unrestricted model includes both $\tilde{P}_t$ and the macro variables M_t, whereas the restricted contains only $\tilde{P}_t$ as predictors.

To begin with, we need the model-implied excess returns for a bond with maturity n and a holding period of h months, denoted by $rx_{t,t+h}^{(n)}$.[40] We first consider the expected excess return, for which we can write

$$E_t rx_{t,t+h}^{(n)} = \beta_n' Z_t, \quad \beta_n' = \mathcal{B}_{n-h}' \phi^h - \mathcal{B}_n' + \mathcal{B}_h'.$$

This follows easily from the definition of the expected excess return, which is $E_t rx_{t,t+h}^{(n)} = E_t p_{t+h}^{(n-h)} - p_t^{(n)} - y_t^{(1)}$ ($p_t^{(n)}$ is the log bond price, $\log P_t^{(n)}$), the affine formulas for log bond prices (see Appendix A), and the VAR specification for Z_t. The surprise component of the excess return is

$$rx_{t,t+h}^{(n)} - E_t rx_{t,t+h}^{(n)} = \mathcal{B}_{n-h}' \nu_{t,t+h}, \quad \nu_{t,t+h} = \sum_{i=1}^{h} \phi^{h-i} \Sigma \varepsilon_{t+i},$$

where we defined the VAR forecast errors $\nu_{t,t+h} = Z_{t+h} - E_t Z_{t+h}$. The dependent variable in our regressions is the average excess return across all maturities longer than h periods, which we write as

$$\bar{rx}_{t,t+h} = K^{-1} \sum_n rx_{t,t+h}^{(n)},$$

denoting the number of relevant maturities by K, which is equal to 9 (from 2 to 10 years) in our paper. For the average return we have $\bar{rx}_{t,t+h} = \bar{\beta}' Z_t + \bar{\mathcal{B}}' \nu_{t,t+h}$, where $\bar{\beta}$ denotes the average of β_n and $\bar{\mathcal{B}}$ denotes the average of $\mathcal{B}_{n-h}$ across these K maturities.

For the restricted regression of excess returns on only the yield PCs we have

$$R_r^2 = \frac{Cov(\bar{rx}_{t,t+h}, \tilde{P}_t) Cov^{-1}(\tilde{P}_t) Cov(\tilde{P}_t, \bar{rx}_{t,t+h})}{Var(\bar{rx}_{t,t+h})},$$

[40] To simplify this analysis, we ignore the yield measurement errors that enter observed excess returns. Their effects are negligibly small and unimportant for our results and intuition.

which can be calculated based on $Cov(\tilde{P}_t)$ and the following moments:

$$\begin{aligned} Cov(\bar{rx}_{t,t+h}, \tilde{P}_t) &= Cov(\bar{rx}_{t,t+h}, P_t) = \bar{\beta}' Cov(Z_t)(WB)', \\ Var(\bar{rx}_{t,t+h}) &= \bar{\beta}' Cov(Z_t)\bar{\beta} + \bar{\mathcal{B}}' Cov(\nu_{t,t+h})\bar{\mathcal{B}}, \\ Cov(\nu_{t,t+h}) &= \sum_{i=1}^{h} \phi^{h-i}\Sigma\Sigma'(\phi^{h-i})'. \end{aligned}$$

The first equality is due to the fact that we focus on model-implied returns.

For the unrestricted regression the regressors are $V_t = (\tilde{P}_t', M_t')'$. The population R^2 is

$$R^2_{ur} = \frac{Cov(\bar{rx}_{t,t+h}, V_t)Cov^{-1}(V_t)Cov(V_t, \bar{rx}_{t,t+h})}{Var(\bar{rx}_{t,t+h})},$$

and the additional required population moments are

$$\begin{aligned} Cov(\bar{rx}_{t,t+h}, M_t) &= \bar{\beta}' Cov(Z_t)\iota_M, \\ Cov(\tilde{P}_t, M_t) &= WBCov(Z_t)\iota_M, \text{and} \\ Cov(M) &= \iota_M' Cov(Z_t)\iota_M. \end{aligned}$$

where ι_M is a selection matrix such that $M_t = \iota_M' Z_t$. Note that if spanning holds and M_t and $\tilde{P}_t$ are perfectly correlated, then $Cov(V_t)$ is not invertible. In this case the collinear regressors M_t are dropped and hence we have $R^2_{ur} = R^2_r$.

Table B.2 shows the population R^2 and the relative RMSEs for the return regressions in the data and in population. Similar observations hold as for Table B.1: For 3 PCs, spanned and unspanned models have the same implications, and measurement error has essentially no effect; when using 5 PCs for the spanned model, measurement is needed to break theoretical spanning. What is particular about the results here is that the predictability of excess returns is much smaller in population than in small samples. Clearly there is a sizable small-sample bias, which is due to the lack of strict exogeneity and the high persistence of the regressors (Stambaugh, 1999). Hence, when comparing model implications for unspanned macro risk to real-world data, we need to use short simulated samples, as in Table 5. However, while the absolute magnitude of return predictability is affected by small-sample issues, the qualitative conclusions about unspanned macro risk and unspanned vs. spanned models remain unchanged.

B.3 Unspanned macro forecasts

For investigating unspanned macro forecasts, we compare unrestricted forecasts of m_{t+1} using both $\tilde{P}_t$ and m_t as predictors, and restricted forecasts with only P_t as predictors. For the restricted model we can calculate the R^2 using results given above and

$$\begin{aligned} Cov(\tilde{P}_t, m_{t+1}) &= Cov(P_t, m_{t+1}) = WBCov(Z_t, m_{t+1}) \\ &= WBCov(Z_t, Z_{t+1})\iota_m = WBCov(Z_t)\phi'\iota_m. \end{aligned}$$

Table B.2: Unspanned macro risk in MTSMs – large-sample results

		GRO/INF			$UGAP/CPI$		
		R^2 Y	R^2 Y+M	RMSE	R^2 Y	R^2 Y+M	RMSE
Data	3 PCs	0.277	0.469	0.857	0.277	0.338	0.957
	5 PCs	0.334	0.489	0.876	0.334	0.414	0.938
$USM(3,2)$	3 PCs						
	$\sigma = \hat{\sigma}_e^{MLE}$	0.107	0.200	0.947	0.159	0.177	0.990
	$\sigma = 1bp$	0.108	0.201	0.946	0.160	0.177	0.990
	$\sigma = 0$	0.108	0.201	0.946	0.160	0.177	0.990
$SM(3,2)$	3 PCs						
	$\sigma = \hat{\sigma}_e^{MLE}$	0.103	0.207	0.940	0.160	0.179	0.989
	$\sigma = 1bp$	0.104	0.208	0.940	0.160	0.179	0.989
	$\sigma = 0$	0.104	0.208	0.940	0.161	0.179	0.989
$SM(3,2)$	5 PCs						
	$\sigma = \hat{\sigma}_e^{MLE}$	0.168	0.207	0.976	0.161	0.179	0.989
	$\sigma = 1bp$	0.206	0.208	0.999	0.169	0.179	0.994
	$\sigma = 0$	0.208	0.208	1.000	0.179	0.179	1.000

Unspanned macro risk in large samples, measured by the theoretical R^2 for regressions of one-year excess bond returns on yield PCs ("Y") and on both yield PCs and macro variables ("Y+M"), and by the relative root-mean-squared error (RMSE) of return forecasts with and without macro variables. The first two rows show these metrics for the actual historical data for comparison.

Table B.3: Unspanned macro forecasts in MTSMs – large-sample results

		GRO	INF	CPI	$UGAP$
Data	3 PCs	0.472	0.297	0.293	0.325
	5 PCs	0.499	0.340	0.294	0.340
		GRO/INF Model		$UGAP/CPI$ Model	
$USM(3,2)$	3 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.465	0.306	0.295	0.331
	$\sigma = 1bp$	0.469	0.307	0.299	0.332
	$\sigma = 0$	0.469	0.307	0.299	0.332
$SM(3,2)$	3 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.466	0.307	0.299	0.336
	$\sigma = 1bp$	0.469	0.307	0.302	0.337
	$\sigma = 0$	0.469	0.307	0.302	0.337
$SM(3,2)$	5 PCs				
	$\sigma = \hat{\sigma}_e^{MLE}$	0.527	0.403	0.302	0.349
	$\sigma = 1bp$	0.681	0.632	0.401	0.607
	$\sigma = 0$	1.000	1.000	1.000	1.000

Unspanned macro forecasts in large samples, measured by the relative root-mean-squared error (RMSE) of macro forecasts using yield PCs, with and without inclusion of own macro lags. The first two rows show R^2 for the actual historical data for comparison.

For the unrestricted model we also need

$$Cov(m_t, m_{t+1}) = \iota_m' Cov(Z_t) \phi' \iota_m.$$

Note that as for the unspanned macro risk regressions, spanning leads to perfect multicollinearity in the unrestricted model, hence in that case $R^2_{ur} = R^2_r$.

Table B.3 shows the population results for the model-implied unspanned macro forecasts. As in the case of unspanned macro variation, the large-sample results closely correspond to the small-sample results. Spanned and unspanned models match the data equally well when using 3 yield PCs in the regressions. The results for the spanned model with 5 yield PCs again show the importance of measurement error. In these results it is particularly noteworthy that even very small measurement errors create a substantial amount of unspanned macro forecasts. Even errors with standard deviation of only 1 basis point create a sufficiently big wedge such that in the case of the $CPI/UGAP$ model, inclusion of lagged CPI improves CPI forecasts accuracy by 60%.

B.4 Effects of measurement error on the variance of yield PCs

In Section 4.6 we discuss how the introduction of small measurement error can significantly reduce the information content in yield PCs. To see this more clearly, we now compare the

Table B.4: Variance of model-implied yield PCs

	$Var(\tilde{P}_t^{\mathcal{N}})$	$Var(P_t^{\mathcal{N}})$	$Var(We_t)$
PC 1	247.58	247.54	0.03
PC 2	222.93	222.53	0.40
PC 3	8.87	8.47	0.40
PC 4	0.44	0.04	0.40
PC 5	0.41	0.01	0.40

covariance matrix of the PCs of observed yields and true yields. We focus on the case with $\mathcal{N} = 5$ PCs, since this is the case where measurement error matters. Note that $Cov(\tilde{P}_t) = Cov(P_t) + \sigma^2 WW'$. Since the rows of W are orthogonal, WW' is diagonal and we can focus on the diagonal elements of $Cov(\tilde{P}_t)$ and $Cov(P_t)$. For model $SM(3,2)$ with $UGAP/CPI$, the numerical values of the diagonal elements of these matrices and of $\sigma^2 WW'$ are shown in Table B.4. The absolute magnitudes of these variances are of little importance, since they are determined by our scaling of W.[41] What is important are the relative changes between the first column and the second column. For the lower-order PCs—the level, slope, and curvature—the variances are essentially not affected by measurement error. However, the fourth and fifth PC are tiny, relative to the error variance, and hence they get overwhelmed by the measurement errors. These higher-order PCs "complete" the spanning in the $SM(3,2)$ models, in the sense that they contain all of the information about M_t that is not in the first three PCs. Therefore, their information content is crucial, but it does not survive the introduction of measurement error. This is the intuition for why even small measurement errors can lead to a substantial degree of unspanned macro information.

[41] To construct W we start from the orthonormal eigenvectors of actual yields, scale the loadings for the first PC to add up to unity, and then scale all loadings by 1200 so that they correspond to annualized percentages.

Table 1: Cross-sectional fit of spanned and unspanned MTSMs

Model	Macro Variables	All	3m	6m	1y	2y	3y	5y	7y	10y
$SM(3,2)$	GRO, INF	5.1	5.2	4.8	6.9	3.7	3.9	5.2	6.0	6.3
$SM(3,2)$	$UGAP, CPI$	5.5	6.0	5.0	7.6	4.0	4.0	5.4	6.1	7.3
$USM(3,2)$	GRO, INF	5.7	6.1	4.8	8.4	4.2	4.4	5.6	6.1	7.7
$USM(3,2)$	$UGAP, CPI$	5.7	6.1	4.8	8.4	4.2	4.4	5.6	6.1	7.7

This table presents the root-mean-squared fitting errors for four different MTSMs in annualized basis points across all maturities and for selected individual maturities.

Table 2: Monetary policy rules and unspanned macro variation

	Policy rule		Macro-spanning R^2			
	R^2	partial	joint	level	slope	curvature
Policy factors						
1.) Unemp. gap	0.80	0.59	0.72	0.01	0.67	0.04
2.) Output gap	0.79	0.58	0.57	0.01	0.45	0.10
3.) INF (JPS)	0.75	0.71	0.81	0.74	0.03	0.04
4.) CPI inflation	0.80	0.76	0.81	0.67	0.04	0.10
5.) PCEPI inflation	0.74	0.68	0.77	0.60	0.05	0.12
Non-policy factors						
6.) GRO (JPS)	0.53	0.05	0.28	0.01	0.00	0.27
7.) GDP growth (ma3)	0.52	0.02	0.14	0.01	0.01	0.12
8.) GDP growth (yoy)	0.51	0.01	0.20	0.00	0.00	0.19
9.) IP growth (ma3)	0.60	0.20	0.32	0.14	0.02	0.16
10.) Jobs growth (ma3)	0.61	0.20	0.20	0.04	0.01	0.15

The first two columns report monetary policy rule regressions in which each economic activity measure (variables 1, 2, 6-10) are paired with CPI inflation (which has a univariate $R^2 = 0.51$ in the rule regression) and the inflation measures (variables 3-5) are paired with the unemployment gap (univariate $R^2 = 0.17$). The first column shows the R^2 of these bivariate regressions, and second column shows the partial R^2 for each macro variable. The last four columns document whether yields span macro variables by providing four R^2 for the regression of each macro variable on the three PCs of yields—denoted level, slope, and curvature—jointly and one at a time.

Table 3: Unspanned macro risk and unspanned macro forecasts

	Excess returns			Macro forecasts		
	R^2	t-stat.	RMSE	Autocorr.	t-stat.	RMSE
Policy factors						
1.) Unemp. gap	0.20	0.67	1.00	0.98	53.23	0.34
2.) Output gap	0.20	0.73	1.00	0.95	33.21	0.46
3.) INF (JPS)	0.36	4.14	0.89	0.99	42.85	0.34
4.) CPI inflation	0.26	1.43	0.96	0.99	44.06	0.29
5.) PCEPI inflation	0.23	1.67	0.98	0.98	55.32	0.32
Non-policy factors						
6.) GRO (JPS)	0.25	2.75	0.97	0.91	26.55	0.50
7.) GDP growth (ma3)	0.21	2.18	0.99	0.47	5.04	0.92
8.) GDP growth (yoy)	0.20	0.88	1.00	0.77	14.65	0.71
9.) IP growth (ma3)	0.32	3.81	0.92	0.94	31.59	0.42
10.) Jobs growth (ma3)	0.22	1.72	0.98	0.87	22.77	0.53

The first three columns assess unspanned macro risk via the predictive power of macro variables for one-year excess bond returns. The first column shows the R^2 when a macro variable is included as predictor along with three PCs of yields. When using only the three yield PCs, the R^2 is 0.195. The second column shows the t-statistic for testing the null hypothesis that the macro variable can be excluded, using heteroskedasticity- and autocorrelation-robust standard errors. The third column shows the relative root-mean-squared error (RMSE) of forecasts with and without macroeconomic information—values below one indicate improvement in predictive accuracy. The last three columns document the predictive power of macro variables at time t for their value at $t+1$, conditional on three PCs of the yield curve at time t, i.e., unspanned macro forecasts. The fourth column reports the first-order autocorrelation of the macro variables. The fifth column shows the t-statistics for testing the null hypothesis that macro variables can be excluded from the forecasting regressions. The last column shows the relative RMSE of macro-yield vs. yields-only forecast.

Table 4: Unspanned macro variation in MTSMs

			GRO	INF	CPI	$UGAP$
Data	3 PCs		0.279	0.812	0.809	0.725
Data	5 PCs		0.379	0.864	0.812	0.751
			GRO/INF Model		$UGAP/CPI$ Model	
$USM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.235	0.680	0.569	0.645
			(0.115)	(0.152)	(0.194)	(0.152)
			[0.634]	[0.806]	[0.906]	[0.644]
		$\sigma = 1bp$	0.304	0.708	0.640	0.668
			(0.118)	(0.138)	(0.174)	(0.152)
			[0.416]	[0.732]	[0.856]	[0.572]
		$\sigma = 0$	0.329	0.709	0.657	0.699
			(0.112)	(0.140)	(0.164)	(0.142)
			[0.344]	[0.752]	[0.800]	[0.490]
$SM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.206	0.678	0.559	0.650
			(0.124)	(0.145)	(0.193)	(0.154)
			[0.724]	[0.824]	[0.916]	[0.650]
		$\sigma = 1bp$	0.389	0.710	0.623	0.684
			(0.159)	(0.151)	(0.176)	(0.136)
			[0.246]	[0.720]	[0.856]	[0.554]
		$\sigma = 0$	0.447	0.713	0.647	0.692
			(0.190)	(0.150)	(0.168)	(0.143)
			[0.200]	[0.718]	[0.840]	[0.534]
$SM(3,2)$	5 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.371	0.733	0.621	0.698
			(0.114)	(0.124)	(0.175)	(0.136)
			[0.530]	[0.904]	[0.860]	[0.606]
		$\sigma = 1bp$	0.707	0.863	0.714	0.805
			(0.081)	(0.089)	(0.144)	(0.098)
			[0.002]	[0.360]	[0.706]	[0.254]
		$\sigma = 0$	1.000	1.000	1.000	1.000
			(0.000)	(0.000)	(0.000)	(0.000)
			[0.000]	[0.000]	[0.000]	[0.000]

This table documents unspanned macro variation in data that is simulated from MTSMs, measured by the R^2's from regressions of macro variables on contemporaneous yield PCs. The first two rows use the actual historical data while the rest of the table uses data simulated from MTSMs. Numbers in parentheses are standard deviations across simulations. Numbers in squared brackets are the fractions of the simulated statistics that are below those in the data.

Table 5: Unspanned macro risk in MTSMs

			GRO/INF Model		$UGAP/CPI$ Model	
			p-value	RMSE	p-value	RMSE
Data	3 PCs		0.000	0.857	0.278	0.957
Data	5 PCs		0.003	0.876	0.180	0.938
$USM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.123	0.920	0.228	0.953
			(0.222)	(0.057)	(0.278)	(0.039)
			[0.200]	[0.164]	[0.676]	[0.440]
		$\sigma = 1bp$	0.120	0.915	0.237	0.953
			(0.219)	(0.064)	(0.290)	(0.040)
			[0.236]	[0.204]	[0.680]	[0.434]
		$\sigma = 0$	0.120	0.922	0.236	0.956
			(0.209)	(0.057)	(0.279)	(0.038)
			[0.182]	[0.152]	[0.678]	[0.392]
$SM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.087	0.910	0.239	0.953
			(0.179)	(0.061)	(0.288)	(0.043)
			[0.238]	[0.178]	[0.660]	[0.430]
		$\sigma = 1bp$	0.118	0.920	0.242	0.951
			(0.214)	(0.061)	(0.291)	(0.042)
			[0.208]	[0.154]	[0.652]	[0.472]
		$\sigma = 0$	0.144	0.924	0.248	0.957
			(0.236)	(0.059)	(0.290)	(0.039)
			[0.196]	[0.142]	[0.654]	[0.390]
$SM(3,2)$	5 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.101	0.928	0.239	0.954
			(0.191)	(0.051)	(0.285)	(0.040)
			[0.456]	[0.164]	[0.572]	[0.258]
		$\sigma = 1bp$	0.172	0.969	0.241	0.959
			(0.253)	(0.031)	(0.289)	(0.036)
			[0.278]	[0.022]	[0.588]	[0.228]
		$\sigma = 0$	1.000	1.000	1.000	1.000
			(0.000)	(0.000)	(0.000)	(0.000)
			[0.000]	[0.000]	[0.000]	[0.000]

This table shows the degree to which expected one-hear excess bond returns are spanned by yields, measured using p-values for testing the null hypothesis (using heteroskedasticity- and autocorrelation-robust Wald tests) that macro variables can be excluded from predictive regressions with yield PCs and macro variables, as well as with relative root-mean-squared errors (RMSE) comparing predictions with and without macro variables. Numbers in parentheses are standard deviations across simulations. Numbers in squared brackets are the fractions of the simulated statistics that are below those in the data.

Table 6: Unspanned macro forecasts in MTSMs

			GRO	INF	CPI	$UGAP$
Data	3 PCs		0.472	0.297	0.293	0.325
Data	5 PCs		0.499	0.340	0.294	0.340
			GRO/INF Model		$UGAP/CPI$ Model	
$USM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.460	0.347	0.304	0.356
			(0.055)	(0.054)	(0.051)	(0.057)
			[0.612]	[0.180]	[0.450]	[0.308]
		$\sigma = 1bp$	0.487	0.356	0.329	0.370
			(0.053)	(0.053)	(0.050)	(0.049)
			[0.414]	[0.134]	[0.228]	[0.182]
		$\sigma = 0$	0.495	0.358	0.335	0.374
			(0.049)	(0.050)	(0.046)	(0.051)
			[0.304]	[0.108]	[0.158]	[0.162]
$SM(3,2)$	3 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.457	0.352	0.306	0.357
			(0.060)	(0.057)	(0.051)	(0.058)
			[0.606]	[0.172]	[0.434]	[0.302]
		$\sigma = 1bp$	0.520	0.372	0.326	0.367
			(0.061)	(0.067)	(0.049)	(0.052)
			[0.216]	[0.116]	[0.244]	[0.216]
		$\sigma = 0$	0.549	0.372	0.341	0.375
			(0.086)	(0.073)	(0.059)	(0.061)
			[0.188]	[0.120]	[0.174]	[0.182]
$SM(3,2)$	5 PCs	$\sigma = \hat{\sigma}_e^{MLE}$	0.501	0.379	0.327	0.380
			(0.051)	(0.056)	(0.047)	(0.052)
			[0.484]	[0.248]	[0.240]	[0.218]
		$\sigma = 1bp$	0.660	0.516	0.368	0.454
			(0.051)	(0.076)	(0.045)	(0.054)
			[0.002]	[0.022]	[0.036]	[0.014]
		$\sigma = 0$	1.000	1.000	1.000	1.000
			(0.000)	(0.000)	(0.000)	(0.000)
			[0.000]	[0.000]	[0.000]	[0.000]

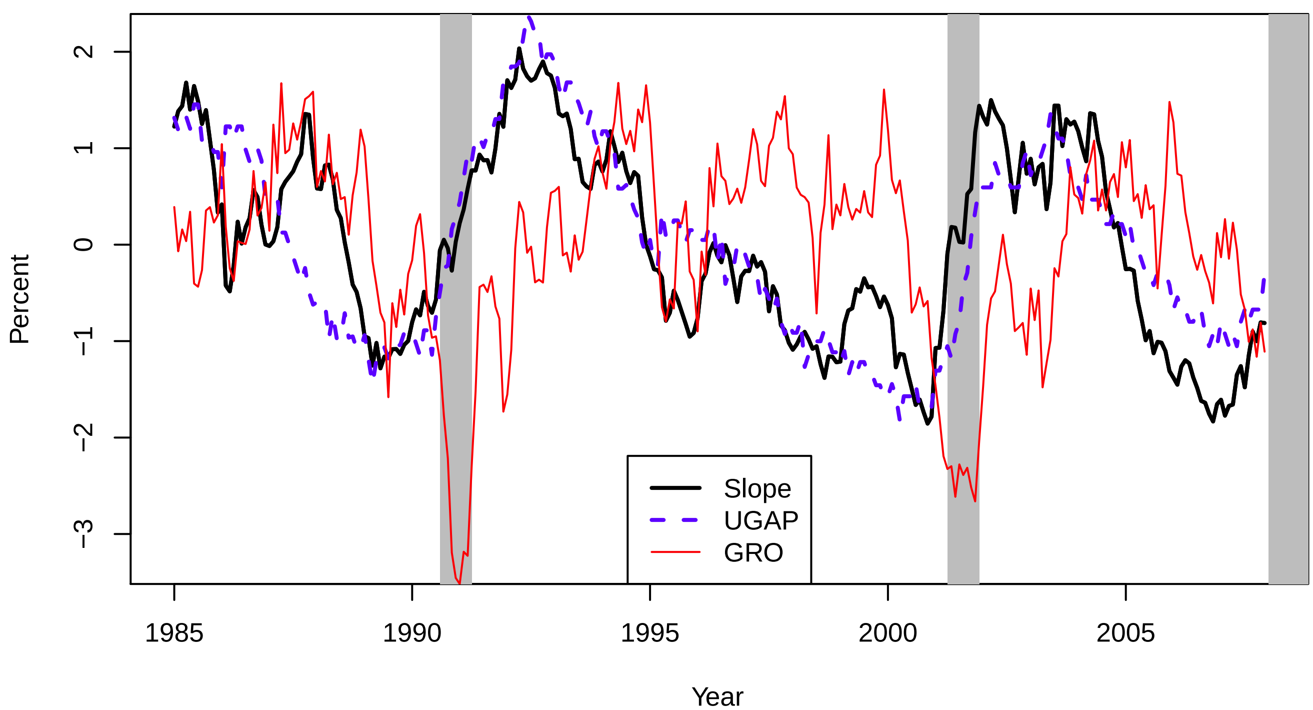
This table shows the degree to which one-step-ahead macro forecasts are spanned by yields, measured by the relative root-mean-squared error of forecasts based on both macro variables and yields to forecasts based only on yields. Numbers in parentheses are standard deviations across simulations. Numbers in squared brackets are the fractions of the simulated statistics that are below those in the data.

Table 7: Likelihood ratio test of unspanned model restriction

	Macro variables	
Statistic	GRO, INF	$UGAP$, CPI
Log-L $SM(3,2)$	22,737	21,300
Log-L $USM(3,2)$	22,439	21,210
$\chi^2(14)$	595	182

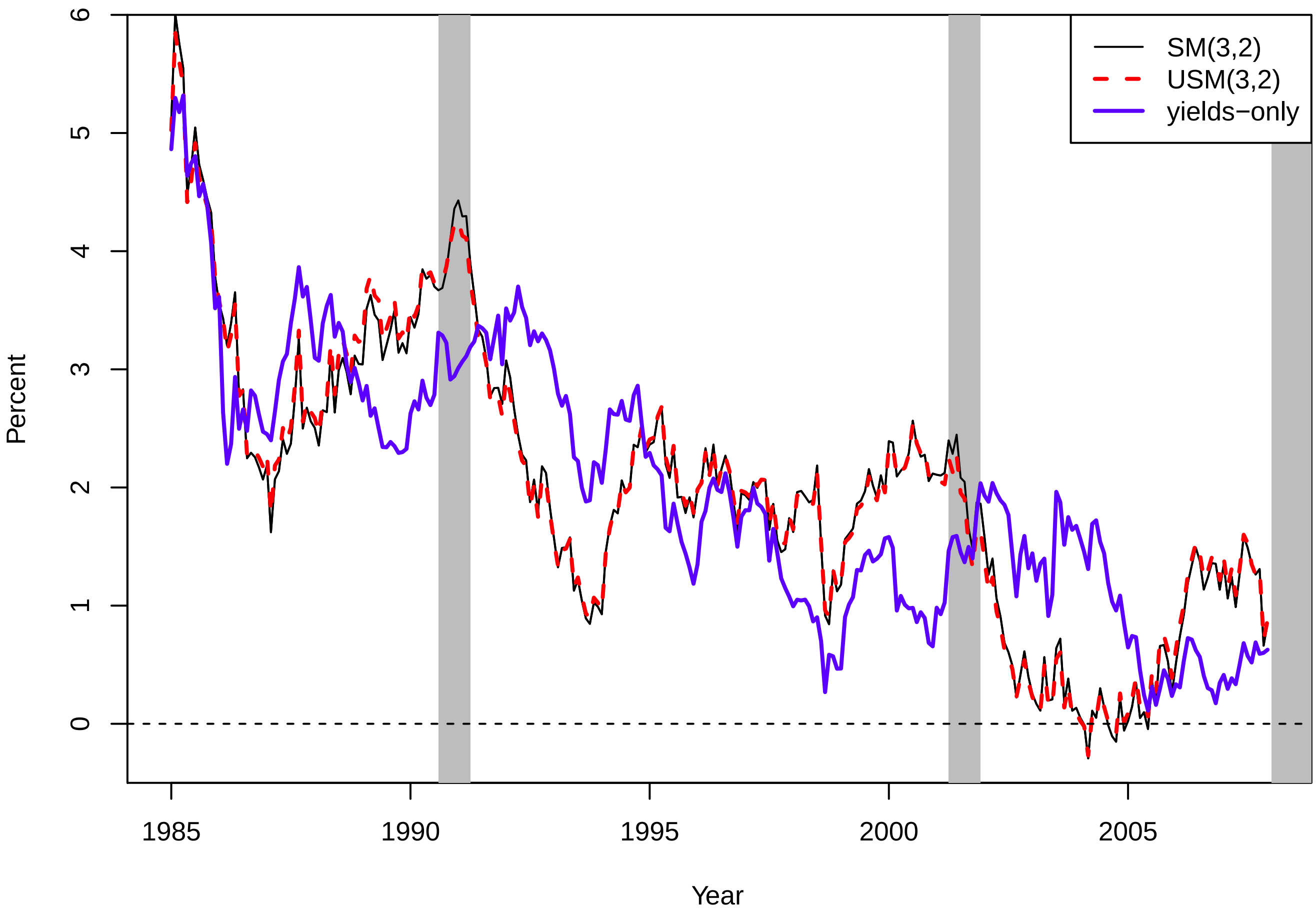
For each pair of macro variables, this table presents the log-likelihood values for spanned and unspanned MTSMs—$SM(3,2)$ and $USM(3,2)$, respectively—and χ^2 statistics for the likelihood-ratio test that $USM(3,2)$ is an acceptable restricted version of $SM(3,2)$. The 5%-critical value for a $\chi^2(14)$-distributed random variable is 6.57.

Figure 1: Slope of the yield curve and macroeconomic variables



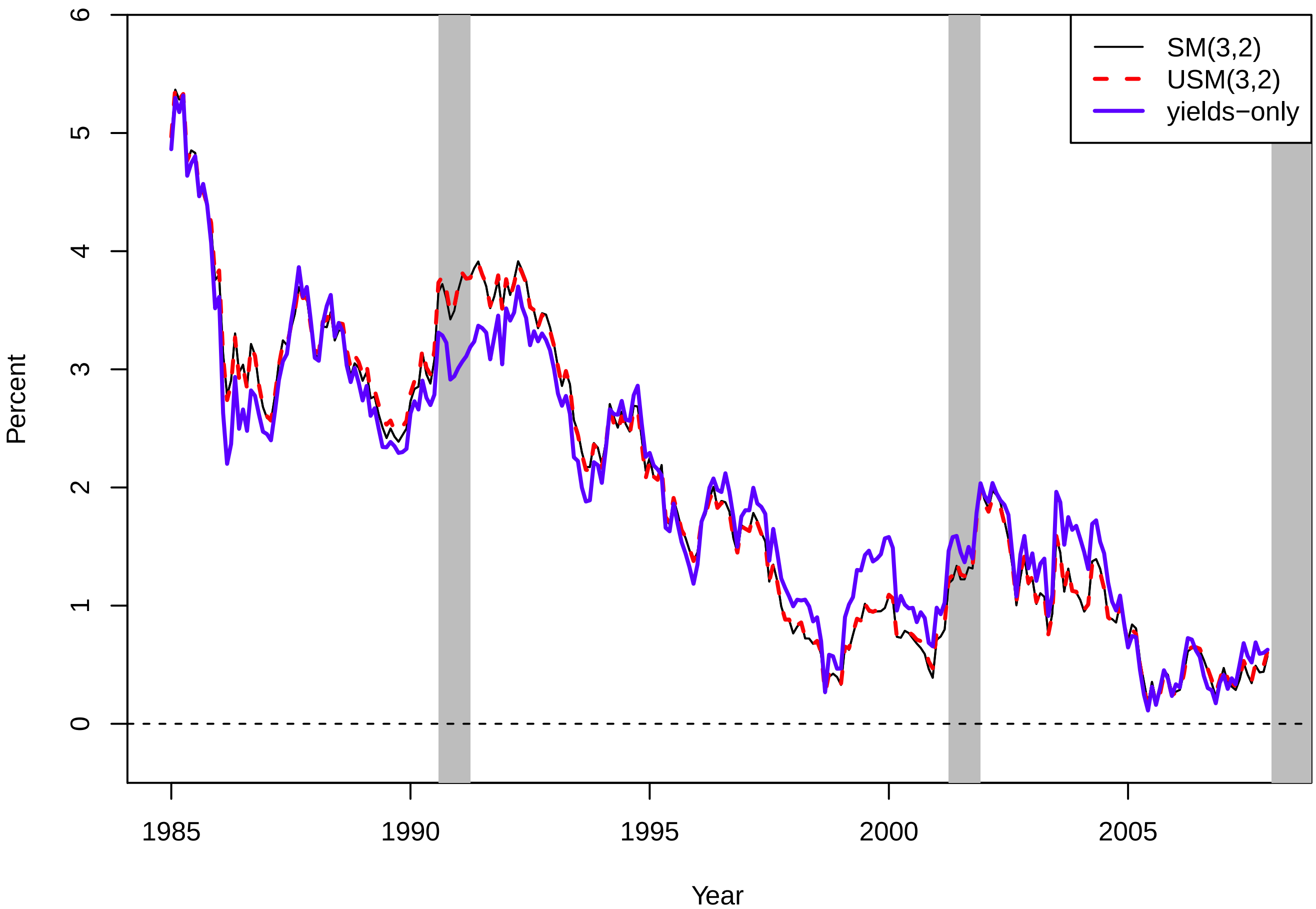
This figure plots the slope of the yield curve, which is measured as the second principal component of yields; $UGAP$, which is the unemployment gap; and GRO, which is the economic growth indicator used by Joslin et al. (2014). All variables are standardized to have mean zero and unit variance.

Figure 2: Term premium estimates from MTSMs with GRO and INF



This figure depicts the two-to-three-year forward term premium estimated from spanned and unspanned macro-finance models—$SM(3,2)$ and $USM(3,2)$, respectively—using GRO and INF macro data, as well as from a three-factor yields-only model.

Figure 3: Term premium estimates from MTSMs with $UGAP$ and CPI



This figure depicts the two-to-three-year forward term premium estimated from spanned and unspanned macro-finance models—$SM(3,2)$ and $USM(3,2)$, respectively—using $UGAP$ and CPI macro data, as well as from a three-factor yields-only model.