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

Michael D. Bauer and Glenn D. Rudebusch
Federal Reserve Bank of San Francisco
September 15, 2015

Abstract

Previous macro-finance term structure models appear incompatible with regressions that show that much macroeconomic variation is not spanned by bond yields and that this unspanned macro variation helps forecast excess bond returns and future macroeconomic fluctuations. This contradiction—or “spanning puzzle”—has prompted calls to reject those previous spanned macro-finance models in favor of new unspanned models. Instead, we provide simulation-based evidence that statistically reconciles the spanned models with the unspanned macro regression results. Hence, our paper salvages the type of models that have been widely used in the previous macro-finance term structure literature. Furthermore, we provide a new statistical rejection of unspanned models and show that their knife-edge restrictions are economically unimportant for term premia.

Keywords: yield curve, term structure models, macro-finance, unspanned macro risk, 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, Mikhail Chernov, Greg Duffee, Jim Hamilton, and Anh Le 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 such macro spanning is a feature of essentially all 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 much regression-based evidence that suggests the presence of 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

---

\(^1\)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.
principal components (PCs) of the yield curve, rather than the 100% predicted by theoretical macro spanning (also see Duffee, 2013b). 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 (also see 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 past empirical MTSMs and the tripartite regression evidence of unspanned macro information constitutes what we term the “spanning puzzle.” The spanning puzzle casts doubt on the validity of essentially all previous macro-finance models used in the literature. It also raises questions about the future direction of macro-finance research and is a major road-block to further work in this field. 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). One direction forward for macro-finance research that has been advocated by JPS and others, is to discard existing spanned MTSMs and instead shift to a new class of unspanned MTSMs. In these models, the assumption of knife-edge restrictions in an otherwise standard MTSM severs the direct link from macro factors to yields.\(^3\) Accordingly, the macro variables are unspanned by construction, as they do not directly determine bond pricing and yields, and yields cannot be inverted for macro factors. If important factors were indeed unspanned macro variables, that would require development of a completely new class of structural economic models to connect bond yields to the economy.\(^4\)

In this paper, we resolve the spanning puzzle through theoretical and empirical comparisons of spanned and unspanned models, which provide strong support for the former models. Our paper therefore salvages the many existing macro-finance models—both equilibrium and reduced-form no-arbitrage models—and their results that JPS and others have recently called

\(^3\)Examples of research using reduced-form unspanned models include Wright (2011), Chernov and Mueller (2012), Priebsch (2014), and Coroneo et al. (2015).

\(^4\)It would also require a complete rethinking of the monetary policy transmission mechanism; for example, as JPS note: “Our results suggest that a monetary authority may affect the output gap and inflation through channels that leave bond yields unaffected, by having a simultaneous [and offsetting] effect on expectations about the future short rates and risk premiums” (p. 1224).
into question. First, we clarify how unspanned models are nested by spanned models, estimate plausible empirical versions of both classes of models, and perform direct likelihood-ratio tests of the knife-edge restrictions required for unspanned models. Our tests strongly reject these restrictions and hence the unspanned models, both for the macro data used by JPS as well as for an alternative data set with more common measures of economic activity and inflation.5 We also test the knife-edge unspanned macro restrictions in reduced-form yield-curve models and find that they are strongly rejected, independent of how many yield factors and which macro data are used.

Our second contribution is to show that plausible spanned MTSMs estimated from the data are in fact not contradicted statistically by the regression evidence on unspanned macro information (as claimed in JPS). Our approach is as follows. Using our empirical spanned MTSMs, which are representative of a broad class of models used in macro-finance research, we generate artificial samples of yields and macro data. Using these simulated data, we estimate regressions that are commonly used to assess the extent of unspanned macro information. This provides the empirical distributions of the various regression test statistics under the null of a spanned MTSM. Comparing the regression statistics obtained from the actual data to these distributions allows us to statistically assess whether the regression results could have plausibly been generated from the spanned model. We find that the regression evidence of unspanned macro information in the real-world data is completely consistent with spanned macro-finance models. In particular, the simulation-based, small-sample $p$-values of the regression statistics that appear to indicate the presence of unspanned macro information are in fact far above conventional significance levels. Our results reconcile the spanned models with the data and demonstrate that the regression evidence of JPS and others provides no empirical reason to reject these models.

How is it that spanned macro-finance models can generate regression results that are consistent with the regression evidence suggesting the presence of unspanned macro information? We provide two reasons. First, for a given spanned model with, say, $N$ risk factors, macro spanning implies that macro variables will be spanned by $N$ linear combinations of yields. However, the macro variables will not be spanned by less than $N$ linear combinations, so the validity of the regression evidence for unspanned macro information depends on using a sufficient number of linear combinations of yields.6 Second, even after incorporating the correct

---

5JPS also claim to conduct a test distinguishing between spanned and unspanned models, but we argue that their test is not useful in this regard, as it effectively compares an unspanned model to a yields-only model.

6For example, consider a spanned MTSM with three yield factors and two macro factors, for a total of five state variables (or risk factors). Projections of macro variables or excess bond returns on three yield factors—say, the level, slope, and curvature of the yield curve—will not provide valid tests of spanning.
number of factors, the regressions are only guaranteed to properly reject a spanned model if that model fits the data exactly—that is, with no measurement error. However, measurement error—a catch-all for model misspecification, data imperfections, or other noise—is a necessary feature in all empirical yield-curve models. Indeed, the addition of measurement error is a critical requirement to reconcile the $\mathcal{N}$-factor models with real-world data that has more than $\mathcal{N}$ variables and never follows an exact factor structure. This same measurement error also resolves the spanning puzzle and reconciles MTSMs with the regression evidence. Of course, adding large amounts of noise can render any two statistical models indistinguishable. But we show that incorporating just the usual, minuscule yield measurement errors in empirical MTSMs is sufficient to generate the appearance of unspanned macro information in the data. Indeed, the wedge created by measurement error with a standard deviation of just one basis point is enough to prevent the spanning regressions from properly identifying the presence or absence of spanning in MTSMs. Our finding that plausible, small measurement errors is sufficient to generate the empirically observed patterns of unspanned macro information stands in marked contrast to claims in the literature, including JPS, Duffee (2013b), and Priebsch (2014).\footnote{For example, JPS claim that “the spanning property is independent of the issue of errors in measuring either bond yields or macro factors” (p. 1206). This claim ignores the fact that assessing the relevance of their statistical evidence depends crucially on how the model fits the data, including the associated measurement error.} Our work shares some features with Duffee (2011) and Cieslak and Povala (2015), who also consider the wedge created by measurement error between information in true and observed factors, but these studies neither investigate the common macro-finance models nor attempt to understand or resolve the spanning puzzle.

Our results show that the regression evidence of JPS and others provides no statistical basis for preferring either unspanned or spanned models. However, one of the main uses of MTSMs has been to estimate risk premia in long-term interest rates and bond returns. Contrary to the claims of JPS, we find that the unspanned knife-edge restrictions are in fact unimportant for estimating such premia. That is, while the rejections of these restrictions are statistically significant, they are not economically significant for this purpose, as spanned and unspanned models imply essentially identical term premia.\footnote{We come to a different conclusion in this regard from JPS because, as described in detail below, they in fact compare an unspanned MTSM to a yields-only model rather than a spanned model.} Because unspanned models may be able to reproduce some economic features of spanned models with a more parsimonious parameterization, they may be a useful approximation for certain purposes.

Finally, our paper also provides new evidence that helps to elucidate, in economic terms, the spanning regression results. We broadly classify two types of macroeconomic variables: those 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 because bond prices are crucially determined by expectations and risk assessments about the short-term policy interest rate set by the central bank.\textsuperscript{9} These policy factors display little if any evidence of unspanned macro variation. Other macro variables, “non-policy factors,” are variables that monetary policymakers pay much less attention to when setting the current short-term interest rate. The non-policy factors 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. These results provide insight about unspanned macro variation based on the conduct of monetary policy, which is a key link between macro variables and the yield curve.

The paper is structured as follows: Section 2 discusses the conventional, spanned macro-finance models, the spanning puzzle, and the solution proposed by JPS which involves knife-edge restrictions needed to obtain an unspanned model. We test these restrictions and show that unspanned models are rejected by the data. In Sections 3, 4, and 5 we assess whether the regression evidence on unspanned macro variation, unspanned macro risk, and unspanned macro forecasts, respectively, can be reconciled with spanned MTSMs, using simulated data from estimated models. In Section 6, we investigate the implications of macro-spanning for term premia. Section 7 concludes.

\section{Spanning in macro-finance term structure models}

To lay the groundwork for our analysis, we first discuss the apparent conflict between conventional macro-finance models and the regression evidence for unspanned macroeconomic information. We also describe a new class of MTSMs recently proposed by JPS, which imposes knife-edge restrictions on the standard model in order to avoid theoretical macro-spanning. We estimate both spanned and unspanned models, and describe the simulation setup that we will use to study empirically the consequences of macro spanning.

\subsection{The conventional macro-finance model}

Especially during the past decade, many studies have used a variety of different MTSMs—both reduced form and equilibrium or structural models—to examine the dynamic interactions

\textsuperscript{9}These macro variables include measures of economic slack (such as the unemployment rate) and measures of underlying inflation, which are the most relevant variables for setting the short-term interest rate as identified from estimated monetary policy rules and the communications of monetary policymakers.
among macroeconomic variables and interest rates of various maturities. Essentially all of these models imply that macroeconomic risks are spanned by the yield curve.

The model described here is representative of a broad class of MTSMs, including equilibrium finance models and macroeconomic models. Our specification closely parallels the formulation in Joslin et al. (2013b). Yields are collected in the vector \( Y_t \), which contains rates for \( J \) different maturities. The risk factors that determine yields are denoted \( Z_t \) and include both yield factors and macro 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 \( P^j_t \) and \( W^j \) the first \( j \) yield portfolios and their weights. We take the first \( \mathcal{L} \) linear combination of yields, \( P^\mathcal{L}_t \), as the yield factors. We use PCs of observed yields, and the corresponding loadings make up the rows of \( W \). Hence, there are \( \mathcal{N} = \mathcal{L} + \mathcal{M} \) risk factors, denoted \( Z_t = (P^\mathcal{L}_t', M_t')' \), all of which are observable.

All no-arbitrage term structure models 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 one-period interest rate is affine in the risk factors:

\[
    r_t = \rho_0 + \rho'_1 Z_t = \rho_0 + \rho'_P P^\mathcal{L}_t + \rho'_M M_t. \tag{1}
\]

The risk factors are assumed to follow a Gaussian vector autoregression (VAR) under the risk-neutral probability measure \( Q \):

\[
    Z_t = \mu^Q + \phi^Q Z_{t-1} + \Sigma^Q \varepsilon^Q_t, \quad \varepsilon^Q_t \overset{iid}{\sim} N(0, I_N). \tag{2}
\]

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

\[
    Y_t = A + BZ_t = A + B_P P^\mathcal{L}_t + B_M M_t, \tag{3}
\]

where the affine loadings \( A \) and \( B \) are given in Appendix A. The time series model for \( Z_t \) (under the real-world probability measure) is a first-order Gaussian VAR:

\[
    Z_t = \mu + \phi Z_{t-1} + \Sigma \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, I_N). \tag{4}
\]
2.2 The spanning puzzle

The model assumptions described above, which are representative of essentially all previous macro-finance models, generally imply that the macro variables are spanned by (i.e., perfectly correlated with) the first \( N \) yield portfolios. To see this, premultiply equation (3) with an \((N \times J)\) matrix, \( W^N \), to select \( N \) linear combinations of model-implied yields, \( P^N_t = W^N A + W^N BZ_t \). This equation can, outside of knife-edge cases, be inverted for \( Z_t \), and in particular for the macro factors:

\[
M_t = \gamma_0 + \gamma_1 P^N_t. \tag{5}
\]

That is, \( M_t \) is a deterministic function of \( P^N_t \), or equivalently of any other \( N \) linear combinations of yields.

However, while conventional macro-finance models theoretically imply that all relevant information about the economy is captured by the current yield curve, there are three strands of regression evidence suggesting otherwise. The first strand is a straightforward direct examination of spanning that simply regresses macro variables on yields. If macro variables are indeed spanned by yields, then this regression should have an \( R^2 \) near one. The information in yields is often considered well summarized by three principal components (PCs), so one regression specification to examine this issue is

\[
m_t = \beta_0 + \beta_1 P^{(3)}_t + u_t, \tag{6}
\]

where \( m_t \) is one of the macroeconomic variables and \( P^{(3)}_t \) are the first three PCs of observed yields. There is unspanned macro variation if the \( R^2 \) in such regressions is low. Evidence for unspanned macro variation is documented by JPS, Duffee (2013a), and others. For example, in referring to these regressions, Duffee (2013b) finds that for “typical variables included in macro-finance models, the \( R^2 \)s are on the wrong side of 1/2” (p. 412). We will investigate the regression evidence on unspanned macro variation in Section 3.

A second implication of macro-spanning is that only current yield curve predicts excess bond returns, because it completely captures the predictive power that macro variables may have. Specifically, under macro spanning, \( \beta_2 = 0 \) in the predictive regression

\[
r_{x_{t,t+12}}^{(n)} = \beta_0 + \beta_1 P^{(3)}_t + \beta_2 M_t + u_{t+12}, \tag{7}
\]

where \( r_{x_{t,t+12}}^{(n)} \) is a one-year holding-period excess return on a bond with \( n \) years maturity and \( M_t \) contains one or more macro variables. Finding that \( \beta_2 \) is significantly different from zero is evidence for unspanned macro risk. Such evidence has been described by JPS, Cooper
and Priestley (2008), Ludvigson and Ng (2009), and others.¹⁰ We will consider this type of evidence in Section 4.

A third implication of macro-spanning is that current yields completely capture the persistence of macro variables. More precisely, in the regression

\[ m_{t+1} = \beta_0 + \beta_1' P_{t}^{(3)} + \beta_2' M_t + u_t \]  

(8)

macro spanning implies that \( \beta_2 = 0 \), meaning that macro variables have no predictive power for future macro variables after conditioning on the current yield curve. However, Duffee (2013a,b) provides strong evidence against \( \beta_2 = 0 \), which we term evidence for unspanned macro forecasts. We revisit this evidence in Section 5.

In sum, there is regression evidence suggesting unspanned macro variation, and this unspanned variation does not appear to just be noise as it seems to help predict future bond returns and macro variables. The apparent inconsistency between this tripartite regression evidence of unspanned macro information and the predictions of the standard macro-finance model constitutes the spanning puzzle. As noted in the introduction, this puzzle is one of the major road-blocks for further macro-finance term structure research.

### 2.3 The unspanned MTSM

We now turn to the alternative model that JPS proposed to address the spanning puzzle. JPS argue that the regression evidence and the spanning puzzle is a definitive empirical rejection of all spanned MTSMs used by previous authors. Therefore, they introduce a new macro-finance specification that incorporates unspanned macro risks. JPS also report a likelihood-ratio test that they view as evidence against the spanned model, which we will revisit below.

For the short rate equation, instead of (1), the unspanned model assumes that the short rate depends only on the yield factors and not the macro factors:

\[ r_t = \rho_0 + \rho_P P_t^L + \underbrace{0_{M}}_{\rho_M} M_t. \]  

(9)

Furthermore, instead of equation (2), the yield factors \( P_t^L \) follow an autonomous VAR under

¹⁰However, Bauer and Hamilton (2015) argue that this evidence suffers from severe small-sample problems and is overturned when appropriate econometric methods are used.
Q that is independent of the macro factors:

\[ P_t^c = \mu^Q_P + \phi^Q_P P_{t-1}^c + 0_{\mathcal{L} \times \mathcal{M}} M_t + \Sigma_P \varepsilon_{tP}, \quad \varepsilon_{tP} \overset{iid}{\sim} N(0, I_L). \]  

(10)

That is, macro factors do not affect the risk-neutral expectations of future yield factors:

\[ E^Q(P_{t+h}^c | Z_t) = E^Q(P_{t+h}^c | P_t^c), \quad \forall h. \]

As a consequence of (9) and (10), yields depend only on the yield factors but not on the macro factors. That is, instead of equation (3) with a full-rank loading matrix, we have

\[ Y_t = A + B_P P_t^c + 0_{\mathcal{M} \times \mathcal{M}} M_t. \]  

(11)

Equation (11) 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 spanning condition (5) does not hold in these models because of the knife-edge restrictions in equations (9) and (10)—we cannot back out the true risk factors from model-implied yields. Formally, the matrix \( W^N B \) is singular and cannot be inverted to yield \( Z_t \) as a function of \( P_t^N \). Instead of (5), we have

\[ M_t = \gamma_0 + \gamma_P P_t^c + OM_t, \]  

(12)

(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 variation 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 can have predictive power for future yields and returns.

When comparing equations (5) and (12) 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 (12) 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, \( \gamma_0 \) and \( \gamma_P \) are also identical across models, as is \( OM_t \).\textsuperscript{11} In other words, spanned and unspanned models...

\textsuperscript{11}JPS claim that conventional, spanned MTSMs impose that \( OM_t \) in equation (12) is zero (p. 1205). But
models have identical implications for projections of macro variables on $L$ yield factors (e.g., level, slope, and curvature, when $L = 3$).

Importantly, the spanned model is more flexible than, and in fact nests the unspanned model. If the knife-edge restrictions

$$
\rho_M = 0_M, \quad \phi^{Q}_PM = 0_{L\times M}
$$

are imposed on the spanned model, we obtain the corresponding unspanned model in which yields do not load on macro factors. For the spanned model with three yield factors and two macro factors, eight zero restrictions are required to obtain the unspanned model. Note that in the unspanned model, the risk-neutral distribution of the macro factors is not identified, since there are no payoffs that depend on macro factors. Due to lack of identification, the difference in number of parameters in the canonical spanned and unspanned models is larger than the number of parameter restrictions on the spanned model—as described in Section 2.5.

The only way macro variables enter the unspanned model is as additional predictors in the VAR in (4), so that they affect real-world expectations of future interest rates and term premia. Expanding the VAR parameters, we can write

$$
\begin{pmatrix}
P_L^t \\
M_t
\end{pmatrix}
= \begin{pmatrix}
\mu_{P0} \\
\mu_{M0}
\end{pmatrix}
+ \begin{pmatrix}
\phi_{PP} & \phi_{PM} \\
\phi_{MM} & \phi_{MP}
\end{pmatrix}
\begin{pmatrix}
P_L^{t-1} \\
M_{t-1}
\end{pmatrix}
+ \Sigma \varepsilon_t.
$$

The $L \times M$ matrix $\phi_{PM}$ plays a crucial role, as it determines the effects of macro variables on expectations of yields. If it is restricted to zero, macro variables drop out completely from the model, as they then affect neither real-world nor risk-neutral expectations of future yields. In that case we obtain a yields-only model, in which only $P_L^t$ are the risk factors. We see that the canonical spanned model nests the unspanned model, which in turn nests the corresponding yields-only model. These nesting relations will be important to understand and interpret the different likelihood-ratio tests of these models.

### 2.4 Empirical spanned and unspanned MTSMs

We will assess the empirical relevance of spanning puzzle using estimated spanned and unspanned models. We denote the spanned models by $SM(L, M)$, and the unspanned models this statement is based on a comparison of spanned and unspanned models with a different number of risk factors—for example, comparing an unspanned model with $L = 3$ and $M = 2$ to a spanned model with $L = 1$ and $M = 2$. 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.
by $USM(\mathcal{L}, \mathcal{M})$. 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 period, 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.\footnote{We thank Anh Le for supplying these data.} 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 $GRO$, the three-month moving average of the Chicago Fed’s National Activity Index\footnote{This measure is constructed so that negative values indicate below-average economic growth and positive values indicate above-average growth.}, and $INF$, which corresponds to survey expectations of inflation in the Consumer Price Index (CPI) over the coming year (from the Blue Chip Financial Forecasts). The second set includes measures of economic activity and inflation that are more standard in the context of monetary policy analysis. For economic activity, this is the unemployment gap, $UGAP$, calculated as the difference between the actual unemployment rate and the estimate of the natural rate of unemployment from the Congressional Budget Office (CBO), and for inflation this is year-over-year growth in CPI excluding food and energy prices, i.e., core CPI inflation, which we denote by $CPI$. While $INF$ and $CPI$ are highly correlated (with a correlation coefficient of 0.89), the two activity indicators $GRO$ and $UGAP$ are essentially uncorrelated (with a correlation coefficient of -0.07). We will discuss differences between the activity indicators in Section 3.1.

An important element for estimation of any term structure model is the choice of the measurement error specification. Because a low-dimensional factor model cannot perfectly match the entire yield curve, it is always necessary—even in yields-only models—to include measurement errors to avoid stochastic singularity. We denote the observed yields $Y^o_t = Y_t + e_t$, where the $J$-vector $e_t$ contains serially uncorrelated Gaussian measurement errors. We assume that the errors on each maturity have equal variance, $\sigma^2_e$, so that the likelihood tries equally hard to match yields of all maturities. As in Joslin et al. (2011), Joslin et al. (2013b), JPS, and other recent studies, we assume that yield factors are observable, which substantially
simplifies estimation as no filtering is necessary.\textsuperscript{14} This assumption is largely inconsequential for parameter estimates because filtered and observed low-order PCs are virtually identical (Joslin et al., 2013b). Importantly, 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.\textsuperscript{15}

As is usual in the macro-finance term structure literature, macroeconomic variables are assumed to be observed without error. Notably, 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 errors for the macro variables would create further unspanned macro variation and would break macro spanning and 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, which have no macro measurement errors and 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 \(P_t^N\), and then apply the canonical form of Joslin et al. (2011). The fundamental parameters of the model are \(r^Q_\infty\), the long-run risk-neutral mean of the short rate, \(\lambda^Q\), the eigenvalues of \(\phi^Q\), the spanning parameters \(\gamma_0\), \(\gamma_1\), the VAR parameters \(\mu\), \(\phi\), and \(\Sigma\), and the standard deviation of the measurement errors, \(\sigma_e\).\textsuperscript{16}

For the unspanned model, we use the canonical form of JPS. In this case, the free parameters are \(r^Q_\infty\), \(\lambda^Q\), \(\mu\), \(\phi\), \(\Sigma\), and \(\sigma_e\). Conveniently, for both spanned and unspanned models \(\mu\) and \(\phi\), as well as \(\sigma_e\) 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

\textsuperscript{14}This implies that \(W^L e_t = 0\), so that \(e_t\) effectively contains only \(J - L\) independent errors.

\textsuperscript{15}Autocorrelated yield measurement errors, as in Hamilton and Wu (2014) and Adrian et al. (2013), would still introduce unspanned macro variation.

\textsuperscript{16}The parameters \(\rho_0\), \(\rho_1\), \(\mu^Q\), and \(\phi^Q\) are determined by the fundamental parameters according the mapping provided in Appendix A of Joslin et al. (2013b).
interest rates—for an in-depth discussion see Bauer (2015).  In their estimation 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. We also conducted our analysis after carrying out similar model selection exercises with very similar parameter restrictions. However, including such restrictions did not affect our results because the restrictions mainly alter the VAR dynamics, which are not important for assessing macro spanning. Therefore, 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.

We report individual parameter estimates in Appendix B, and focus here 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. 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, but as noted above, this difference has no bearing on our resolution of the spanning puzzle. Table 1 also reports estimates for the standard deviation of the yield measurement errors, $\hat{\sigma}_e$.

2.5 Testing knife-edge restrictions

We now carry out direct likelihood ratio tests of the restrictions of unspanned MTSMs. These restrictions sever the direct link from macro variables to yields. This feature 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 for unspanned factors to exist. However, unlike unspanned volatility

---

17 Estimating models with short samples of highly persistent interest rate data can result in a small-sample parameter bias. To address this problem, one can take advantage of the information in the cross section of interest rates with plausible restrictions on risk pricing, as in Bauer (2015) and JPS, or directly adjust for the small-sample bias, as in Bauer et al. (2012).

18 In addition, they restrict the largest eigenvalue of $\phi$ to equal the largest eigenvalue of $\phi^Q$.

19 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.

20 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, end up being exactly zero. Bikbov and Chernov (2009) conduct an analysis of unspanned stochastic 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 estimated unspanned MTSMS—including JPS, Wright (2011), and Priebsch (2014), among others—we are the first to conduct direct hypothesis tests of the knife-edge restrictions that underlie these models.

To assess the plausibility of the knife-edge restrictions, we use likelihood-ratio tests based on the fact that $SM(3, 2)$ nests $USM(3, 2)$, as described in Section 2.3. The log-likelihood values for these two models are shown in Table 2 for the two different pairs of macro variables, $GRO/INF$ and $UGAP/CPI$. For either macro data set, the spanned models fit the data substantially better than the unspanned ones.\textsuperscript{21} The results reveal that the improved fit is exclusively due to improvements in the cross-sectional fit, and not to smaller VAR forecast errors.\textsuperscript{22} Allowing the macro variables to enter as risk factors in the yield equations (3) reduces the fitting errors of the models—see also Table 1. While the improvements in cross-sectional fit are modest relative to the magnitude and variability of yields, they are substantial in the sense that they translate into large increases in the log-likelihood value. The last row of each panel of Table 2 reports the likelihood-ratio test statistics, which are very sizable due to the improved cross-sectional fit.

To interpret these test statistics we have to account for the issue that some parameters are not identified under the null hypothesis. The reason is that under the null of the knife-edge restrictions, bond prices are not directly linked to macro variables and hence the risk-neutral dynamics of $M_t$ are not identified. Specifically, the null imposes $\mathcal{M}(1+\mathcal{L})$ (here eight) parameter restrictions—see (13)—but the canonical unspanned model has $\mathcal{M}(2+\mathcal{N})$ (14) fewer parameters than the canonical spanned model, which means that $\mathcal{M}^2 + \mathcal{M}$ (six) parameters are not identified under the null.\textsuperscript{23} As a consequence, the usual regularity conditions for the validity of the asymptotic $\chi^2$-distribution of the likelihood-ratio test statistics are not satisfied, and the limiting distribution is non-standard (Hansen, 1996). So we cannot calculate $p$-values using a $\chi^2$-distribution with eight degrees of freedom.

\textsuperscript{21}Our results are even stronger when considering models with one or two yield factors, as in Joslin et al. (2013b) and Joslin et al. (2013a), since in those models the inclusion of macro variables leads to even more substantial improvements in cross-sectional fit.

\textsuperscript{22}The time series fit remains essentially unchanged by the knife-edge restrictions, since the VAR parameters $\mu$ and $\phi$ are identical in the spanned and unspanned models (see also JPS, p. 1207, on this point).

\textsuperscript{23}There are $\mathcal{M}$ fewer parameters in the short-rate equation (9) than in equation (1), and the $\mathcal{M}(1+\mathcal{N})$ spanning parameters in $\gamma_0$ and $\gamma_1$ are absent in the unspanned model.
However, we can provide upper bounds to the $p$-values for these test statistics. Every MTSM can be viewed as a restricted version of a reduced-form factor model, as emphasized by Hamilton and Wu (2012). In particular, consider the reduced-form model corresponding to $SM(3,2)$, which consists of a VAR for $Z_t$ and $J - L$ non-trivial measurement equations, and denote this reduced-form spanned model by $RSM(3,2)$. Both $SM(3,2)$ and $USM(3,2)$ can be obtained from $RSM(3,2)$ under some nonlinear parameter restrictions. Specifically, $USM(3,2)$ has 50 less parameters than $RSM(3,2)$, since the latter model has measurement equations with $(J - L)(1 + N) = 54$ free loadings, which in the former model are determined by just $1 + L = 4$ structural parameters. We could test the plausibility of these restrictions using a likelihood-ratio test. Now the likelihood-ratio statistic comparing $USM(3,2)$ to $SM(3,2)$ is of course always strictly smaller than the test statistic comparing $USM(3,2)$ to $RSM(3,2)$, since $SM(3,2)$ is more restrictive than $RSM(3,2)$. So we can evaluate our test statistics against a $\chi^2$-distribution with 50 degrees of freedom, and be sure that this is a conservative test in the sense that the resulting $p$-values are necessarily higher than the true $p$-values. The five-percent critical value for this distribution is 67.5, and we strongly reject the null hypothesis with minuscule $p$-values.

An alternative to testing the knife-edge unspanned macro restrictions in no-arbitrage models is to test these restrictions in the reduced-form factor models for yields, in the spirit of Hamilton and Wu (2014). This is straightforward as such models are estimated instantly using least squares and the restrictions can be tested using likelihood-ratio statistics. We estimate models $RSM(L,M)$ and the corresponding unspanned models, using $L = 1, \ldots, 5$ yield factors and $M = 2$ macro factors, and again considering the two different macro data sets. The yield factors are the first $L$ PCs of yields, and the dependent variables in the measurement equations are the next $J - L$ PCs of yields. There are $2(J - L)$ knife-edge restrictions, and the likelihood-ratio statistic has an asymptotic $\chi^2_{2(J - L)}$-distribution. Table 3 shows the results. For both macro data sets, and for any number of yield factors we considered, these tests strongly reject the knife-edge restrictions, with $p$-values that are always much smaller than 0.1%.

This evidence casts much doubt on the unspanned model proposed by JPS. The use of unspanned MTSMs has typically been motivated and justified only indirectly on the basis of the regression evidence for unspanned macro information. In contrast, our direct tests of the knife-edge unspanned macro restrictions show that unspanned models are rejected by the data with with high statistical significance.

24Taking $J$ individual yields as dependent variables would lead to a near-singular covariance matrix for the residuals, as $L$ linear combinations of yields are perfectly explained by the PCs.
2.6 Do JPS reject spanned models?

On the surface, our results completely contradict the test result in JPS (p. 1214) that appears to show a spanned MTSM rejected in favor of an unspanned MTSM. Just as puzzling is the fact that JPS test the spanned model as a restricted version of the unspanned model, which is precisely the opposite of the specification nesting demonstrated above. Here we reconcile these differences and reinterpret the test result in JPS.

JPS compare an unspanned MTSM, labeled “$M_{us}$,” to a restricted model “$M_{span}$,” in which the block of the $\phi$ matrix corresponding to the lagged macro variables is set to zero. Their likelihood-ratio test strongly rejects the restricted model, with a reported $\chi^2$-statistic of 1,189. However, the restrictions of $M_{span}$ do not imply a spanned MTSM. Instead, since $M_{span}$ restricts $\phi_{PM}$ and $\phi_{MM}$ to zero, it effectively corresponds to a yields-only model (see Section 2.3). The only difference between $M_{span}$ and a pure yields-only model is that the former includes two VAR equations for forecasting macro variables using yield factors, and its likelihood function includes macro forecast errors. But for yields and risk premia, $M_{span}$ and a yields-only model have the exact same observational implications. The correct interpretation of the likelihood-ratio test in JPS is that in their specific macro-finance data set including GRO and INF, yields-only models are rejected in favor of macro-finance models, whether they are spanned or unspanned MTSMs. Importantly, this is not a rejection of spanned MTSMs.

In the data used by JPS, current yields do not completely capture the relevant information for forecasting, as macro variables have additional predictive power. The statistical rejection is simply a reflection of the regression evidence for unspanned macro risks and unspanned macro forecasts in GRO and INF. In Sections 4 and 5 we will show that such regression evidence is consistent with both spanned and unspanned MTSMs.

3 Empirical MTSMs and unspanned macro variation

In Section 2.2, we described three strands of regression evidence for unspanned macro information that appear to reject spanned macro-finance models. Such results are in direct contradiction to the likelihood ratio test results we provided in the previous section. We now reconcile this regression evidence with our test results by delineating the ability of such regression evidence to discriminate between spanned and unspanned MTSMs. In this section, we

---

25 The restrictions of $M_{span}$ do not imply macro-spanning as in equation (5) but instead that expectations of macro variables are spanned by yields.

26 The extremely large $\chi^2$-statistic that is reported by JPS is mostly driven by unspanned macro forecasts: If one does not zero out own lags of macro variables in $\phi$ the $\chi^2$-statistic is only 74 for the model of JPS.
focus on the evidence for unspanned macro variation. We first describe our macro data set and the real-world regression evidence, which is based on regressions of the form in (6). We then outline our model-based simulation experiment and provide results. In subsequent sections, we consider unspanned macro risk and unspanned macro forecasts in a similar fashion.

3.1 Macro data and the distinction between levels and growth

For robustness, we consider ten different macroeconomic inflation and economic activity variables. Our sample period, which coincides with that used by JPS, extends from January 1985 to December 2007. Our measures of inflation include $INF$, the survey-based measure used by JPS, as well as $CPI$ (defined in Section 2.4), and year-over-year growth in the Personal Consumption Expenditure (PCE) Price Index excluding food and energy prices, i.e., core PCE inflation.27

Regarding the activity measures, we include measures of the level and the growth of activity in the U.S. economy, which differ greatly. 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 ($UGAP$). 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.28 We consider five measures of growth in economic activity: $GRO$, the measure used by JPS; 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 nonfarm payroll employment (the last two are measured as three-month moving averages).

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 rather than growth variables as the ones most important for central banks in setting the short-term interest rate.29

27 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 statements of monetary policymakers.

28 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.

29 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 (1999), Orphanides (2003), Bean (2005), and Rudebusch (2006). The low weight on growth in monetary policy rules can also be optimal (e.g. Rudebusch, 2002).
see this difference, we estimate standard policy rules that regress 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 4 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 variable 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 as demonstrated by their high explanatory power in simple policy rules. When paired with core CPI, they achieve an $R^2$ of 0.80. In contrast, growth measures do not show a close association with the policy rate. In particular, with $GRO$, the policy rule $R^2$ barely edges up to 0.53 (compared with the $R^2$ of 0.51 for the univariate fed funds rate regression on inflation). 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 $R^2$ values ranging from 0.74 to 0.80, and are therefore included in our set of policy factors. The partial $R^2$ estimates using the policy factors are all 0.60 to 0.70, whereas the partial $R^2$ estimates using the non-policy factors are at most 0.20. This sharp difference clearly shows the dichotomy between the policy and non-policy groups of variables.

### 3.2 Regression evidence for unspanned macro variation

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—see equation (6). The $R^2$ estimates from these regressions, displayed in the third column of Table 4, show that most of the variation in each of the policy factors is explained by the yield curve, with $R^2$ values from 0.60 to 0.70. This is true 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 $R^2$ estimates are all below 0.30.

To help uncover how the yield curve captures macro variation, the last three columns of Table 4 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, 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 of yields (the slope of the yield curve), $GRO$, and $UGAP$, with all three series 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 yield slope is very clear—the correlation coefficient is 0.84. In other words, this policy factor is closely related to (i.e., essentially spanned by) the yield curve. On the other hand, $GRO$ is basically uncorrelated with the slope.

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. Policy factors—those macro variables that substantially 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 Fed.

### 3.3 Simulating data from empirical MTSMs

To examine the statistical relevance of the spanning puzzle, we use the empirical spanned and unspanned models that we estimated as described in Section 2.4, and generate simulated data samples. We then estimate the three spanning regressions using these simulated samples, and assess whether spanned models can reproduce the regression evidence in the actual data or are rejected by that evidence.

Our simulation design is the following: With parameter estimates for four different MTSMs in hand—$SM(3,2)$ and $USM(3,2)$, each for two different macro data sets—we simulate 10,000 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

---

30This 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.
measurement error to obtain simulated yields.\footnote{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.} The measurement errors have a standard deviation of $\sigma$, which is the same across maturities.\footnote{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 for our simulation results, as filtered and observed low-order PCs are virtually indistinguishable (Joslin et al., 2013b).} We consider the effects of changing $\sigma$ from the maximum likelihood estimates $\hat{\sigma}_e$ (reported in Table 1) to a small value of 1 basis point and to 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 the spanning regressions in the simulated macro/yield data. In this way, we obtain the model-based distribution of the regression statistics. By comparing the regression statistics obtained in the actual data to these distributions, we can then test the assumptions of the model and, in particular, can evaluate the claim by JPS that spanned models are at odds with the regression evidence.

### 3.4 Spanned MTSMs are consistent with regression evidence

To assess whether the regression evidence on unspanned macro variation can discriminate between plausible estimated spanned and unspanned models, we run regression (6) in both actual and simulated data samples. Table 6 reports the $R^2$ for regressions using the actual data in the top two rows, with either three or five PCs of yields, which provide the comparison benchmark for the model-based simulated $R^2$ distribution. As noted above, GRO exhibits much unspanned variation, while UGAP and both inflation measures are largely spanned by yields. The remaining rows of Table 6 show results for simulated data, namely, medians of the $R^2$ across simulations, as well as the fractions of simulated samples in which the $R^2$ is below the value in the actual data. These fractions can be interpreted as model-based, small-sample $p$-values for the joint restrictions imposed by the MTSM, and using the conventional significance level, values below 0.05 would indicate evidence against a particular model.

First, we consider regressions of macro variables on 5 PCs using data simulated from the spanned model $SM(3, 2)$—these results are in the bottom panel of Table 6. The 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 and regress macro variables on five PCs, we find an $R^2$ of 1 in every simulated data set (the final two lines in the table with $\sigma = 0$). 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, adding small, plausible yield measurement error completely changes this
conclusion. Using a standard deviation for the measurement error equal to the maximum likelihood estimate from the real-world data ($\sigma = \hat{\sigma}_e$), the amount of unspanned macro variation in the simulated data closely matches that in the actual data. This is true even for $GRO$, the variable with the largest degree of unspanned macro variation in the data: the median $R^2$ in the simulated data is 0.37, which is almost identical to the $R^2$ of 0.38 in the actual data. The simulation-based $p$-values are all above 0.5, indicating that there is no evidence against the spanned model. We also obtain sufficiently low $R^2$ with just a minuscule measurement error of $\sigma = 1bp$ (with the exception of $GRO$). In sum, small yield measurement error can generate substantial unspanned macro variation and reconcile spanned models with the regression results in the data.

Of course, it is not surprising that with some amount of measurement error, a spanned model has data-generating properties that are consistent with the data-generating properties of an unspanned model. Indeed, with enough stochastic variation, any two economic models are indistinguishable. But our results show there is an empirically plausible spanned model, with the usual, tiny yield measurement error, that cannot be distinguished from the corresponding estimated unspanned version using the spanning regression. These results certainly depend on the specific models and samples employed, but as noted above, we have explored a variety of alternative specifications with different variables, numbers of factors, and parameter restrictions and obtained similar results.\textsuperscript{33}

Theoretical spanning only holds if the number of PCs in the spanning regression is at least as large as the number of risk factors. For $SM(3, 2)$, regressions using only three PCs of simulated yields give $R^2$'s that 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 6. Evidently, even in a spanned MTSM with three yield factors, which fits the yield curve well, three PCs of yields leave a substantial amount of macro variation unspanned. This further lessens the significance of the theoretical spanning condition: spanned models do not impose that macro variables are spanned by only $L$ PCs of yields. Only when we include higher-order PCs, i.e., condition on $N$ linear combinations of yields, does the spanning condition in conventional MTSMs have testable implications. But higher-order PCs have less explanatory power for the cross section of yields and are measured much more imprecisely.

Finally, we consider the unspanned models, $USM(3, 2)$, in the top panel of Table 4. 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

\textsuperscript{33}In Appendix C, we consider large-sample results for the spanning regressions based on model-implied population moments, which provides further detail and intuition.
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 same implications for the relationship between macro variables and linear combinations of yields, because equation (12) holds in both types of models! Therefore, the results reported in the top panel of Table 6 are essentially identical to those in the middle panel. Based on spanning regressions on yield PCs, there is no evidence to distinguish between spanned and unspanned models.

3.5 The role of measurement error

We have documented that regressions of macro variables on yields provide no reason to prefer unspanned over spanned models. For regressions of macro variables on three yield PCs, both spanned and unspanned models equally match the evidence. For regressions using five yield PCs, a spanned MTSM with three yield factors and two macro factors can also match the regression evidence—with measurement error being a crucial factor in this congruence. Of course, yield-curve models cannot fit the data without allowing for measurement error, because their low-dimensional factor structure never holds exactly in the data. The addition of measurement error avoids stochastic singularity and makes these models empirically useful. For macro-finance yield-curve models, measurement error has another important effect: It breaks theoretical macro spanning, just as it breaks stochastic singularity. While this is obvious conceptually, the key question is whether it works empirically. We have shown that the small amount of measurement error needed to fit observed yields at the same time also generates sufficient unspanned macroeconomic information to reconcile macro-finance models with the spanning regression evidence.

It may appear puzzling that a large amount of unspanned macro information can be generated by very small measurement errors. After all, “the measurement error of yields is tiny relative to the variability of yields” (Duffee, 2013b, p. 412). Furthermore, low-order PCs of observed yields are very close to PCs of model-implied yields, even if all yields are observed with error, as in our simulations, because not only is measurement error small, but it also gets washed out when using low-order PCs of yields. How, then, can spanned models with small measurement errors generate substantial amounts of unspanned macro information?

---

34 Duffee goes on to assert 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.

35 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).
The reason is that measurement error matters little for low-order PCs but 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 Table 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-order PCs: they are smaller on average and estimated much less precisely. Therefore, adding measurement error essentially wipes out the information in the higher-order PCs, and in this way reduces the information content in the yield PCs. Appendix C shows this intuition mathematically. In other words, the level and slope (and to some extent the curvature) of the yield curve are large relative to measurement errors and little affected by it, but the higher-order yield PCs, which matter for theoretical macro-spanning, become noisy with the introduction even of small yield measurement error.

While the results and analysis in this paper pertain to MTSMs with three yield factors, we have found that spanned models with fewer yield factors also reproduce the regression evidence on unspanned macro information. For models with two yield factors, as in Joslin et al. (2013a), the intuition is similar to the case of $SM(3, 2)$, because the yield fit is reasonable and the measurement error modest. For MTSMs with only one yield factor, like those in Bernanke et al. (2004) and Joslin et al. (2013b), the fit to the cross section of yields is much poorer and large measurement errors are required, which in turn generate substantial unspanned macroeconomic information.

Some other studies have also investigated the effects of measurement error in yield-curve models on the information content of measured yield-curve factors. Duffee (2011) shows that this can hide important information from the yield curve factors, which can be recovered using Kalman-filtering. In a very specific macro-finance model with trend inflation, Cieslak and Povala (2015) show that the presence of yield measurement error makes it difficult to recover a risk-premium factor. These studies do not investigate the effects of measurement error in commonly used macro-finance models or address whether those models can be reconciled with the spanning regression evidence, as we do in this paper.

4 Empirical MTSMs and unspanned macro risk

We now consider 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? We use predictive regressions of the form in (7), using the average excess returns for bonds of maturities from two through ten years as the dependent variable. Then we investigate whether this evidence rejects spanned MTSMs.

4.1 Regression evidence for unspanned macro risk

We first provide the regression evidence on unspanned macro risk using as predictors the first three PCs of yields and each of our ten macroeconomic variables. The first and second columns of Table 5 show t-statistics and p-values for testing the hypothesis that the macro variable can be excluded from this regression, using Newey-West standard errors with 18 lags. The third column reports the increase in $R^2$ when including macro variables, relative the $R^2$ in predictive regressions with only three yield PCs, which is 20%. The fourth column shows 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-tests indicate statistical significance, the changes in $R^2$ and 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. 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 significantly improve forecast accuracy. 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 was true above with our spanning regressions, there is a marked difference in predictive power between policy and non-policy factors (although $INF$ is an exception). However, even for those macro variables that significantly improve bond return forecasts, the size of forecast gains is quite modest—the median RMSE is improved by only 3% by inclusion of a macro predictor.

Overall, there is some evidence for predictive power of macro variables for bond returns, i.e., for unspanned macro risk, consistent with the findings of Cooper and Priestley (2008), Ludvigson and Ng (2009), and JPS. But the evidence appears somewhat weak and variable across different macroeconomic data series. Duffee (2013b) is also skeptical that such evidence

---

36 This is the usual lag choice for predictive regressions for annual returns with overlapping monthly observations—see, for example, Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009).

37 While Cooper and Priestley (2008) find some predictive power of the output gap for excess bond returns, Bauer and Hamilton (2015) show that these results are not robust, consistent with our results here.
is robust to changes in the sample period used. Furthermore, inference about predictive accuracy in bond return regressions is problematic from an econometric perspective—for an in-depth analysis, see Bauer and Hamilton (2015). Finally, the evidence on the predictive power of macro variables is in sample and may not be reflected in out-of-sample forecasting.\footnote{Note that GRO, the Chicago National Activity Index, was not available to investors in real time, but only became available in 1999.} In sum, there are many question marks about the evidence on unspanned macro risk. However, even if this evidence is taken at face value, we show below that it does not invalidate the spanned models.

### 4.2 Spanned MTSMs are consistent with excess return regressions

Can spanned models reproduce the regression evidence on bond return predictability? In Table 7 we compare the results for predictive regressions in actual and simulated data. In this case the predictors are PCs of yields and either \( GRO \) and \( INF \) or \( UGAP \) and \( CPI \), depending on the model. Table 7 reports Wald statistics, calculated using Newey-West standard errors with 18 lags, for the joint significance of the two macro variables in the predictive regressions, and the changes in \( R^2 \) when the macro variables are added to the predictive regressions. The former measures statistical significance, while the latter measures economic significance, and high values indicate the presence of unspanned macro risk.

The first two rows in Table 7 show the results in the actual data when we condition on either three or five PCs of yields. For the Wald statistics, we report \( p \)-values using the asymptotic \( \chi^2 \)-distribution. 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 \( CPI \) improve forecast accuracy only marginally, and this improvement is not statistically significant—with \( p \)-values of 0.22 and 0.12, respectively.

The rest of Table 7 reports medians of the statistics across simulated samples, as well as the fractions of samples in which these statistics are above the value in the actual data, i.e., simulation-based \( p \)-values. Without measurement error (so \( \sigma = 0 \)) and with five PCs used as regressors, the spanned model \( SM(3,2) \) of course implies that the macro data cannot have any additional predictive power (shown at the bottom of Table 7). However, this theoretical spanning result is 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 positive increases in \( R^2 \)—that is, there is evidence for unspanned macro risk in the simulated data. Intuitively, 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. The key question is, can the spanned model reproduce the amount of unspanned macro risk we find in the data?

It turns out that the answer is yes, even for the case with $GRO/INF$ where JPS have found a large degree of unspanned macro risk. The spanned models can reproduce the gains in return forecast accuracy of macro variables, even if we condition on five yield PCs. For yield measurement errors of the usual size ($\sigma = \hat{\sigma}_e$), the medians of the statistics measuring predictive power of macro variables are only slightly below the values in the data, and the $p$-values are always above 0.2. That is, there is no evidence against the spanned models from excess-return regressions. Even for the case with minuscule measurement error, the Wald-test $p$-values for the $GRO/INF$ model are above 0.2.

Furthermore, unspanned models do not have any advantage over spanned models in capturing the predictability of excess bond returns. Comparing the results for predictive regressions using 3 yield PCs, we again see that the spanned and unspanned models have virtually identical implications.

These results show that the evidence on unspanned macro risk is not at all inconsistent with conventional spanned macro-finance models. 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. (2015), 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.

5 Empirical MTSMs and unspanned macro forecasts

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 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 regression-based analysis “more direct evidence that the problem is misspecification” (p. 412). To study unspanned macro forecasts, we take a similar approach as Duffee (2013b), using the predictive regressions of the form in (8).
5.1 Regression evidence for unspanned macro forecasts

We first consider the evidence for unspanned macro forecasts for our ten macro variables. We obtain yields-only macro forecasts by regressing each macro variable in month \( t + 1 \) on three PCs of yields dated at \( t \), and compare these to "macro-yields" forecasts in which the predictors are augmented by a lag of the macro variable. Table 5 reports \( t \)-statistics (using Newey-West standard errors with 12 lags) for the null hypothesis that the lagged macro variable can be excluded, i.e., that macro forecasts are spanned, as well as the relative RMSE comparing macro-yields forecasts to yields-only forecasts. We also calculate first-order autocorrelations of the macro variables to help elucidate the role of persistence in this context.

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 measures of slack and (year-over-year) core inflation are more persistent than Duffee’s growth and quarterly headline inflation measures.

5.2 Spanned MTSMs are consistent with macro forecast regressions

We now investigate whether spanned MTSMs are inconsistent with the empirical findings of unspanned macro forecasts. 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. Again, we compare the values obtained for regressions using the real-world data to the distribution of values in regressions using simulated data.

Table 8 provides 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 we noted above, the evidence for unspanned macro forecasts is very strong in our data. The rest of the table show medians and \( p \)-values based on the simulations from our MTSMs. Importantly, the spanned MTSMs can match the strong regression evidence on unspanned macro forecasts in the real-world data. In almost all specifications, the distribution of the relative RMSEs across the simulations is centered near the value in the real-world data. For the usual measurement error specifications (\( \sigma = \hat{\sigma}_e \)), a substantial fraction of the simulations exhibit at least as much unspanned macro forecast information as in the real-world data, that is, the \( p \)-values are far...
above 0.05 and there is no evidence to reject the spanned model. The only case in which data simulated from a spanned MTSM does not exhibit any unspanned macro forecasts is when we condition on all five PCs and do not allow for any yield measurement errors. In this case, macro spanning holds exactly in the simulated data—as in Tables 6 and 7—so the yield curve completely captures all relevant macro forecast information and lags of macro variables are not needed.

The simulated regression evidence on unspanned macro forecasts mirrors the results for unspanned macro variation and risk: spanned models are able to match the regression evidence as well as unspanned models. In particular, the actual values of the regression test statistics from the real-world data are well within the probability distributions for these statistics when generated from plausible estimated spanned MTSMs. That is, statistically, even if the null hypothesis of theoretical macro spanning holds, the regressions appear to reject macro-spanning based on the presence of unspanned macroeconomic information—the regression tests are oversized as they reject the true null too often. As a result, the regression evidence cannot discriminate between these spanned and unspanned MTSMs. Importantly, we have shown that spanned models are consistent with the regression evidence, which resolves the spanning puzzle.

6 Term premia in spanned and unspanned MTSMs

We now examine the economic implications of macro spanning 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.\(^{39}\) Forward term premia are defined as differences between model-implied forward rates and risk-neutral forward rates. With one addition, this figure essentially reproduces Figure 1 of JPS: Our USM(3, 2) model corresponds to their M\(_{\text{us}}\) model, and our yields-only model corresponds to their M\(_{\text{span}}\) model. The comparison shows that our estimated term premia closely resemble those of JPS. This is true even though our models are maximally flexible while those in JPS have various overidentifying restrictions (see our discussion in Section 2.4) and despite our use of a slightly different yields data set.

Figure 2 also 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

\(^{39}\)Our yields-only model is a maximally flexible affine model estimated using maximum likelihood with the first three PCs of observed yields as risk factors and the normalization of Joslin et al. (2011).
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 in Section 2.5 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.

The intuition for our finding of essentially identical term premia from spanned and unspanned models is straightforward. Fitted yields and forward rates are very similar from both types of models, as indicated by the similar cross-sectional fit of these models shown in Table 1. At the same time, spanned and unspanned models contain the same VAR specification for \( Z_t \), and result in almost identical VAR parameter estimates and VAR forecasts (see Table 2). Therefore not only the fitted rates but also the risk-neutral rates, which aside from convexity terms correspond to VAR-based forecasts of future short rates, are very similar for spanned and unspanned model. Term premia, the differences between these two series, are therefore essentially identical as well.

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 \( M_{\text{span}} \) model in JPS) implies a very different term premium than the macro-finance models with \( GRO \) and \( INF \). Just like the rejections of the VAR restrictions emphasized by JPS (see Section 2.6), 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; Rudebusch and Swanson, 2012). 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 clear 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.\textsuperscript{40}

An MTSM with more conventional macroeconomic variables delivers term premia that 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 3.2). 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. This evidence can be viewed as a caution against including macro variables in MTSMs that are selected on the basis of high in-sample predictive power for excess returns. Such variables can substantially alter estimated risk premia and can reduce their plausibility.\textsuperscript{41}

Our key point here is that spanned and unspanned models imply essentially identical 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 at least some 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 the fewer number of parameters, which has practical advantages. In addition to simplifying estimation and inference, the parameter parsimony may also prove useful in forecasting, where more tightly parameterized

\textsuperscript{40}Bauer et al. (2012) and Bauer et al. (2014) also discuss the countercyclical behavior of term premia estimated from term structure models.

\textsuperscript{41}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.

30
models typically do better. Overall, unspanned models may be a useful shortcut in practical applications of MTSMs.\footnote{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.} 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.

7 Conclusion

In this paper, we have resolved the spanning puzzle in macro-finance and salvaged the conventional MTSMs widely used in the literature. We show that the theoretical spanning of macro variables by model-implied yields in MTSMs does not have any practical significance in terms of the regression evidence. Specifically, a spanned MTSM with conventional small yield measurement errors does not conflict with the regression evidence on unspanned macro variation, macro risk, or macro forecasts. That is, the regression-based tests in JPS and other studies, when properly evaluated, do not lead to a spanning puzzle. The reason is that they cannot distinguish between plausible estimated spanned and unspanned macro-finance models. These findings should reassure the many researchers who have employed 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, spanned and unspanned models appear to deliver similar estimates of term premia in long-term interest rates. 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.

One could imagine alternative solutions to the spanning puzzle. For example, some structural and reduced-form MTSMs imply nonlinear mappings from risk factors to bond yields. In theory, such nonlinearity breaks the (linear) spanning condition, but it remains an empirical question as to 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 the monetary policy rule), we cannot rule out this possibility. 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. 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


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. regime switches are the cause of unspanned macro information.


A Affine bond pricing

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

\[ B^m_t = e^{A^m + B^m X_t}, \]

and the loadings \( A^m = A^m(\mu, \phi, \delta_0, \delta_1, \Sigma) \) and \( B^m = B^m(\phi, \delta_1) \) follow the recursions

\[
A^m_{t+1} = A^m_t + (\mu^Q)^t B^m + \frac{1}{2} B^m \Sigma^t \Sigma^t B^m - \delta_0 \\
B^m_{t+1} = (\phi^Q)^t B^m - \delta_1
\]

with starting values \( A^0 = 0 \) and \( B^0 = 0 \). Model-implied yields are determined by \( y^m_t = -m^{-1} \log B^m_t = A^m_t + B^m X_t \), with \( A^m = -m^{-1} A^m \) and \( B^m = -m^{-1} B^m \). Risk-neutral yields, the yields that would prevail if investors were risk-neutral, can be calculated using

\[
\tilde{y}^m_t = \tilde{A}^m_t + \tilde{B}^m X_t, \quad \tilde{A}^m = -m^{-1} A^m(\mu, \phi, \delta_0, \Sigma), \quad \tilde{B}^m = -m^{-1} B^m(\phi, \delta_1).
\]

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, \( y_t p^m_t = y^m_t - \tilde{y}^m_t \).

B Parameter estimates

Here we report parameter estimates for our spanned and unspanned models, which were obtained using maximum likelihood as described in Section 2.4 and then used for generating simulated macro-finance data sets as described in Section 3.3. TBC

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

$$\frac{\text{RMSE}_{\text{ur}}}{\text{RMSE}_{\text{r}}} = \sqrt{1 - R^2_{\text{ur}}} \over 1 - R^2_{\text{r}},$$

where $R^2_{\text{ur}}$ and $R^2_{\text{r}}$ are the population $R^2$ of the unrestricted and restricted models, respectively.

### C.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 $\text{Cov}(e_t) = \sigma^2 I_J$. 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 $\tilde{P}_t = W\tilde{Y}_t$, 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 $L = 3$ or $N = 5$ yield PCs. With this notation, the spanning regressions are

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

The population $R^2$ is

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

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

$$\text{Cov}(\tilde{P}_t) = W\text{Cov}(\tilde{Y}_t)W' = WBCov(Z_t)(BW)' + \sigma^2 WW',$$

$$\text{Cov}(P_t, m_t) = WBCov(Z_t)\iota_m,$$

$$\text{Var}(m_t) = \iota_m'Cov(Z_t)\iota_m.$$
Table C.1: Unspanned macro variation in MTSMs – large-sample results

<table>
<thead>
<tr>
<th></th>
<th>GRO</th>
<th>INF</th>
<th>CPI</th>
<th>UGAP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs</td>
<td>0.28</td>
<td>0.81</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td>5 PCs</td>
<td>0.38</td>
<td>0.86</td>
<td>0.81</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>USM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{σ}_e</td>
<td>0.28</td>
<td>0.79</td>
<td>0.76</td>
<td>0.71</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.29</td>
<td>0.79</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>σ = 0</td>
<td>0.29</td>
<td>0.79</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>SM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{σ}_e</td>
<td>0.28</td>
<td>0.77</td>
<td>0.75</td>
<td>0.72</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.29</td>
<td>0.78</td>
<td>0.76</td>
<td>0.73</td>
</tr>
<tr>
<td>σ = 0</td>
<td>0.29</td>
<td>0.78</td>
<td>0.76</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>SM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{σ}_e</td>
<td>0.48</td>
<td>0.88</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.77</td>
<td>0.96</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>σ = 0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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.

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

Table C.1 shows the model-implied $R^2$ for each of the cases and models that we considered in Section 3.4. 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 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 $P_t^N$, which spans $m_t$, hence the $R^2$ is 1. Finally, even in the case that theoretical spanning holds, small measurement error with $\sigma = \hat{σ}_e$ 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.
C.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 \( r\bar{x}_{t,t+h}^{(n)} \). We first consider the expected excess return, for which we can write

\[
E_t r\bar{x}_{t,t+h}^{(n)} = \beta'_n Z_t, \quad \beta'_n = B'_{n-h} \phi^h - B'_n + B'_h.
\]

This follows easily from the definition of the expected excess return, which is

\[
E_t r\bar{x}_{t,t+h}^{(n)} = E_t \bar{p}_{t+h}^{(n-h)} - p_t^{(n)} - y_t^{(1)} (p_t^{(n)} \text{ is the log bond price, } \log P_t^{(n)}) \text{, the affine formulas for log bond prices (see Appendix A), and the VAR specification for } Z_t. \text{ The surprise component of the excess return is}
\]

\[
r\bar{x}_{t,t+h}^{(n)} - E_t r\bar{x}_{t,t+h}^{(n)} = B'_{n-h} \nu_{t,t+h}, \quad \nu_{t,t+h} = \sum_{i=1}^{h} \phi^{h-i} \Sigma \epsilon_{t+i},
\]

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

\[
r\bar{x}_{t,t+h} = K^{-1} \sum_{n} r\bar{x}_{t,t+h}^{(n)},
\]

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 \( r\bar{x}_{t,t+h} = \tilde{\beta} Z_t + \tilde{B} \nu_{t,t+h} \), where \( \tilde{\beta} \) denotes the average of \( \beta_h \) and \( \tilde{B} \) denotes the average of \( B_{n-h} \) across these \( K \) maturities.

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

\[
R^2_r = \frac{\text{Cov}(r\bar{x}_{t,t+h}, \tilde{P}_t)\text{Cov}^{-1}(\tilde{P}_t)\text{Cov}(\tilde{P}_t, r\bar{x}_{t,t+h})}{\text{Var}(r\bar{x}_{t,t+h})},
\]

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

\[
\text{Cov}(r\bar{x}_{t,t+h}, \tilde{P}_t) = \text{Cov}(r\bar{x}_{t,t+h}, P_t) = \tilde{\beta} \text{Cov}(Z_t)(WB)' ,
\]

\[
\text{Var}(r\bar{x}_{t,t+h}) = \tilde{\beta} \text{Cov}(Z_t)\tilde{\beta} + \tilde{B} \text{Cov}(\nu_{t,t+h})\tilde{B},
\]

\[
\text{Cov}(\nu_{t,t+h}) = \sum_{i=1}^{h} \phi^{h-i} \Sigma \phi^{h-i}'.
\]

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

---

44To 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.
Table C.2: Unspanned macro risk in MTSMs – large-sample results

<table>
<thead>
<tr>
<th></th>
<th>( GRO/INF )</th>
<th>( UGAP/CPI )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 ) Y</td>
<td>( R^2 ) Y+M</td>
</tr>
<tr>
<td>Data 3 PCs</td>
<td>0.28</td>
<td>0.47</td>
</tr>
<tr>
<td>5 PCs</td>
<td>0.33</td>
<td>0.49</td>
</tr>
<tr>
<td>USM(3, 2) 3 PCs</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>( \sigma = \hat{\sigma}_e )</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>( \sigma = 1bp )</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>( \sigma = 0 )</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>SM(3, 2) 3 PCs</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = \hat{\sigma}_e )</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = 1bp )</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = 0 )</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>SM(3, 2) 5 PCs</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = \hat{\sigma}_e )</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = 1bp )</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>( \sigma = 0 )</td>
<td>0.21</td>
<td>0.21</td>
</tr>
</tbody>
</table>

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.

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

\[
R^2_{ur} = \frac{Cov(\tilde{r}_{x,t+t+h}, V_t)Cov^{-1}(V_t)Cov(V_t, \tilde{r}_{x,t+t+h})}{Var(\tilde{r}_{x,t+t+h})},
\]

and the additional required population moments are

\[
Cov(\tilde{r}_{x,t+t+h}, M_t) = \tilde{\beta}' Cov(Z_t)_{t_M},
\]
\[
Cov(\tilde{P}_t, M_t) = WB Cov(Z_t)_{t_M}, \text{ and}
\]
\[
Cov(M) = \iota_{M}^{'} Cov(Z_t)_{t_{M}}.
\]

where \( \iota_{M} \) 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 \( R^2_{ur} = R^2 \).

Table C.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 C.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 7. 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.

C.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

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

For the unrestricted model we also need

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

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 C.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%.

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

In Section 3.5 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 covariance matrix of the PCs of observed yields and true yields. We focus on the case with $N = 5$ PCs, since this is the case where measurement error matters. Note that $\text{Cov}(\tilde{P}_t) = \text{Cov}(P_t) + \sigma^2WW'$. Since the rows of $W$ are orthogonal, $WW'$ is diagonal and we can focus on the diagonal elements of $\text{Cov}(\tilde{P}_t)$ and $\text{Cov}(P_t)$. For model $SM(3,2)$ with UGAP/CPI, the numerical values of the diagonal elements of these matrices and of $\sigma^2WW'$ are shown in Table C.4. The absolute magnitudes of these variances are of little importance, since they are
Table C.3: Unspanned macro forecasts in MTSMs – large-sample results

<table>
<thead>
<tr>
<th></th>
<th>GRO</th>
<th>INF</th>
<th>CPI</th>
<th>UGAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data 3 PCs</td>
<td>0.47</td>
<td>0.30</td>
<td>0.29</td>
<td>0.33</td>
</tr>
<tr>
<td>Data 5 PCs</td>
<td>0.50</td>
<td>0.34</td>
<td>0.29</td>
<td>0.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>GRO/INF Model</th>
<th>UGAP/CPI Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>USM(3,2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{\sigma}_e</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>σ = 0</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>SM(3,2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{\sigma}_e</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>σ = 0</td>
<td>0.47</td>
<td>0.31</td>
</tr>
<tr>
<td>SM(3,2) 5 PCs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = \hat{\sigma}_e</td>
<td>0.53</td>
<td>0.40</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>0.68</td>
<td>0.63</td>
</tr>
<tr>
<td>σ = 0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>$Var(P_t^N)$</th>
<th>$Var(P^N_t)$</th>
<th>$Var(W e_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC 1</td>
<td>247.58</td>
<td>247.54</td>
<td>0.03</td>
</tr>
<tr>
<td>PC 2</td>
<td>222.93</td>
<td>222.53</td>
<td>0.40</td>
</tr>
<tr>
<td>PC 3</td>
<td>8.87</td>
<td>8.47</td>
<td>0.40</td>
</tr>
<tr>
<td>PC 4</td>
<td>0.44</td>
<td>0.04</td>
<td>0.40</td>
</tr>
<tr>
<td>PC 5</td>
<td>0.41</td>
<td>0.01</td>
<td>0.40</td>
</tr>
</tbody>
</table>

determined by our scaling of $W$. 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.

---

45To 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Macro Variables</th>
<th>$\hat{\sigma}_e$</th>
<th>All</th>
<th>3m</th>
<th>6m</th>
<th>1y</th>
<th>2y</th>
<th>3y</th>
<th>5y</th>
<th>7y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SM(3,2)$</td>
<td>$GRO, INF$</td>
<td>5.8</td>
<td>5.1</td>
<td>5.2</td>
<td>4.8</td>
<td>6.9</td>
<td>3.7</td>
<td>3.9</td>
<td>5.2</td>
<td>6.0</td>
<td>6.3</td>
</tr>
<tr>
<td>$SM(3,2)$</td>
<td>$UGAP, CPI$</td>
<td>6.3</td>
<td>5.5</td>
<td>6.0</td>
<td>5.0</td>
<td>7.6</td>
<td>4.0</td>
<td>4.0</td>
<td>5.4</td>
<td>6.1</td>
<td>7.3</td>
</tr>
<tr>
<td>$USM(3,2)$</td>
<td>$GRO, INF$</td>
<td>6.6</td>
<td>5.7</td>
<td>6.1</td>
<td>4.8</td>
<td>8.4</td>
<td>4.2</td>
<td>4.4</td>
<td>5.6</td>
<td>6.1</td>
<td>7.7</td>
</tr>
<tr>
<td>$USM(3,2)$</td>
<td>$UGAP, CPI$</td>
<td>6.6</td>
<td>5.7</td>
<td>6.1</td>
<td>4.8</td>
<td>8.4</td>
<td>4.2</td>
<td>4.4</td>
<td>5.6</td>
<td>6.1</td>
<td>7.7</td>
</tr>
</tbody>
</table>

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, as well as the estimated standard deviation of the yield measurement errors ($\hat{\sigma}_e$).

Table 2: Tests of knife-edge restrictions in MTSMs

<table>
<thead>
<tr>
<th>Models</th>
<th>cross section</th>
<th>time series</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$GRO/INF$ models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLK $SM(3,2)$</td>
<td>21,083</td>
<td>1,660</td>
<td>22,744</td>
</tr>
<tr>
<td>LLK $USM(3,2)$</td>
<td>20,781</td>
<td>1,659</td>
<td>22,439</td>
</tr>
<tr>
<td>Likelihood-ratio statistic</td>
<td></td>
<td></td>
<td>609</td>
</tr>
<tr>
<td>$UGAP/CPI$ models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLK $SM(3,2)$</td>
<td>20,875</td>
<td>430</td>
<td>21,306</td>
</tr>
<tr>
<td>LLK $USM(3,2)$</td>
<td>20,780</td>
<td>429</td>
<td>21,210</td>
</tr>
<tr>
<td>Likelihood-ratio statistic</td>
<td></td>
<td></td>
<td>192</td>
</tr>
</tbody>
</table>

For each pair of MTSMs, the table presents the log-likelihood (LLK) values for spanned and unspanned MTSMs—$SM(3,2)$ and $USM(3,2)$, respectively—and the likelihood-ratio statistic that $USM(3,2)$ is an acceptable restricted version of $SM(3,2)$. The first two columns report the contribution to the log-likelihood of the cross-sectional fitting errors and the VAR forecast errors. The 5% (0.1%) critical value for a $\chi^2(50)$-distributed random variable is 67.5 (86.7).
Table 3: Tests of knife-edge unspanned macro restrictions in reduced-form regressions

<table>
<thead>
<tr>
<th>( \mathcal{L} )</th>
<th>df</th>
<th>5% c.v.</th>
<th>0.1% c.v.</th>
<th>GRO/INF</th>
<th>UGAP/CPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>33.9</td>
<td>48.3</td>
<td>567.8</td>
<td>381.2</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>31.4</td>
<td>45.3</td>
<td>252.9</td>
<td>345.2</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>28.9</td>
<td>42.3</td>
<td>122.3</td>
<td>204.2</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>26.3</td>
<td>39.3</td>
<td>116.6</td>
<td>83.3</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>23.7</td>
<td>36.1</td>
<td>92.5</td>
<td>78.7</td>
</tr>
</tbody>
</table>

Values of likelihood-ratio test statistics (in the last two columns) for the null hypothesis that macro variables have zero loadings in the system of cross-sectional regressions of yields on risk factors. The explanatory variables in the unrestricted regression are \( \mathcal{L} \) PCs of yields and 2 macro variables. The dependent variables are the remaining \( J - \mathcal{L} \) PCs of yields, with \( J = 12 \) the number of yield maturities. The test statistics have an asymptotic \( \chi^2 \)-distribution with \( df = 2(J - \mathcal{L}) \) degrees of freedom, and the table shows the corresponding critical values (c.v.).

Table 4: Monetary policy rules and unspanned macro variation

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

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 5: Unspanned macro risk and unspanned macro forecasts

<table>
<thead>
<tr>
<th>Policy factors</th>
<th>Excess returns</th>
<th>Macro forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-stat.</td>
<td>p-value</td>
</tr>
<tr>
<td>1.) Unemp. gap</td>
<td>0.61</td>
<td>(0.54)</td>
</tr>
<tr>
<td>2.) Output gap</td>
<td>0.68</td>
<td>(0.50)</td>
</tr>
<tr>
<td>3.) INF (JPS)</td>
<td>4.20</td>
<td>(0.00)</td>
</tr>
<tr>
<td>4.) Core CPI inflation</td>
<td>1.63</td>
<td>(0.10)</td>
</tr>
<tr>
<td>5.) Core PCE inflation</td>
<td>1.73</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Non-policy factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.) GRO (JPS)</td>
<td>2.70</td>
<td>(0.01)</td>
</tr>
<tr>
<td>7.) GDP growth (ma3)</td>
<td>1.83</td>
<td>(0.07)</td>
</tr>
<tr>
<td>8.) GDP growth (yoy)</td>
<td>0.84</td>
<td>(0.40)</td>
</tr>
<tr>
<td>9.) IP growth (ma3)</td>
<td>3.39</td>
<td>(0.00)</td>
</tr>
<tr>
<td>10.) Jobs growth (ma3)</td>
<td>1.60</td>
<td>(0.11)</td>
</tr>
</tbody>
</table>

The first four columns assess unspanned macro risk via the predictive power of macro variables for one-year excess bond returns. The first and second columns show the t-statistic and the p-value for the coefficient on the macro variable, using Newey-West standard errors with 18 lags. The third column shows the increase in $R^2$ 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 0.20. The fourth 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 fifth column reports the first-order autocorrelation of the macro variables. The sixth column shows the t-statistics for testing the null hypothesis that macro variables can be excluded from the forecasting regressions, using Newey-West standard errors with 12 lags. The last column shows the relative RMSE of macro-yield vs. yields-only forecast.
Table 6: Unspanned macro variation in MTSMs

<table>
<thead>
<tr>
<th></th>
<th>GRO</th>
<th>INF</th>
<th>CPI</th>
<th>UGAP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs</td>
<td>0.28</td>
<td>0.81</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td>5 PCs</td>
<td>0.38</td>
<td>0.86</td>
<td>0.81</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>GRO/INF Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UGAP/CPI Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>USM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs</td>
<td>0.22</td>
<td>0.71</td>
<td>0.61</td>
<td>0.66</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>[0.67]</td>
<td>[0.80]</td>
<td>[0.91]</td>
<td>[0.66]</td>
</tr>
<tr>
<td>σ = 0</td>
<td>[0.42]</td>
<td>[0.75]</td>
<td>[0.83]</td>
<td>[0.55]</td>
</tr>
<tr>
<td><strong>SM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs</td>
<td>0.19</td>
<td>0.71</td>
<td>0.59</td>
<td>0.68</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>[0.72]</td>
<td>[0.81]</td>
<td>[0.92]</td>
<td>[0.64]</td>
</tr>
<tr>
<td>σ = 0</td>
<td>[0.25]</td>
<td>[0.72]</td>
<td>[0.86]</td>
<td>[0.56]</td>
</tr>
<tr>
<td><strong>SM(3, 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 PCs</td>
<td>0.37</td>
<td>0.76</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>σ = 1bp</td>
<td>[0.54]</td>
<td>[0.88]</td>
<td>[0.87]</td>
<td>[0.60]</td>
</tr>
<tr>
<td>σ = 0</td>
<td>[0.00]</td>
<td>[0.39]</td>
<td>[0.71]</td>
<td>[0.26]</td>
</tr>
</tbody>
</table>

This table documents unspanned macro variation in actual data and in data that is simulated from MTSMs, measured by the $R^2$’s from regressions of macro variables on contemporaneous yield PCs. Low $R^2$ indicate a large degree of unspanned macro variation. The first two rows show results for the actual data while the rest of the table reports median $R^2$ across simulations and in square brackets the fractions of the simulated samples in which the $R^2$ is below the value in the actual data (that is, $p$-values).
Table 7: Unspanned macro risk in MTSMs

<table>
<thead>
<tr>
<th></th>
<th>GRO/INF Model</th>
<th>UGAP/CPI Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wald</td>
<td>$\Delta R^2$</td>
</tr>
<tr>
<td>Data 3 PCs</td>
<td>21.64</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Data 5 PCs</td>
<td>13.71</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>$USM(3,2)$ 3 PCs σ = $\hat{\sigma}_e$</td>
<td>9.45</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>[0.20]</td>
<td>[0.21]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>8.97</td>
</tr>
<tr>
<td></td>
<td>[0.21]</td>
<td>[0.20]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>10.12</td>
</tr>
<tr>
<td></td>
<td>[0.20]</td>
<td>[0.21]</td>
</tr>
<tr>
<td>$SM(3,2)$ 3 PCs σ = $\hat{\sigma}_e$</td>
<td>10.37</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>[0.22]</td>
<td>[0.24]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>9.66</td>
</tr>
<tr>
<td></td>
<td>[0.20]</td>
<td>[0.22]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>7.87</td>
</tr>
<tr>
<td></td>
<td>[0.18]</td>
<td>[0.17]</td>
</tr>
<tr>
<td>$SM(3,2)$ 5 PCs σ = $\hat{\sigma}_e$</td>
<td>10.08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>[0.38]</td>
<td>[0.22]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>[0.20]</td>
<td>[0.01]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>[0.00]</td>
<td>[0.00]</td>
</tr>
</tbody>
</table>

This table shows evidence for unspanned macro risk in actual and simulated data, based on predictive regressions for one-quarter excess bond returns on yield PCs and macro variables. Statistical significance is measured by the Wald statistic—calculated using Newey-West standard errors using 18 lags—for joint significance of the coefficients on the two macro variables. Economic significance is measured by the change in $R^2$ when macro variables are added as predictors. High values of the statistics indicate a large degree of unspanned macro risk. The top panel reports results for the actual data, and numbers in parentheses are $p$-values using the asymptotic $\chi^2$-distribution. The rest of the table reports medians of the statistics across simulated samples, and numbers in square brackets are the fractions of the simulated samples in which each statistic is above the value in the actual data (that is, $p$-values).
Table 8: Unspanned macro forecasts in MTSMs

<table>
<thead>
<tr>
<th>Data</th>
<th>3 PCs</th>
<th>5 PCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GRO</td>
<td>INF</td>
</tr>
<tr>
<td>Data 3 PCs</td>
<td>0.47</td>
<td>0.30</td>
</tr>
<tr>
<td>Data 5 PCs</td>
<td>0.50</td>
<td>0.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>GRO/INF Model</th>
<th>UGAP/CPI Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>USM(3,2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs σ = ˆσ_e</td>
<td>0.46</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>[0.61]</td>
<td>[0.17]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>[0.41]</td>
<td>[0.11]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>[0.36]</td>
<td>[0.10]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>[0.24]</td>
<td>[0.09]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>[0.18]</td>
<td>[0.15]</td>
</tr>
<tr>
<td>SM(3,2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 PCs σ = ˆσ_e</td>
<td>0.46</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>[0.59]</td>
<td>[0.16]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>[0.24]</td>
<td>[0.09]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>[0.18]</td>
<td>[0.15]</td>
</tr>
<tr>
<td>SM(3,2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 PCs σ = ˆσ_e</td>
<td>0.50</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>[0.48]</td>
<td>[0.24]</td>
</tr>
<tr>
<td></td>
<td>σ = 1bp</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>[0.00]</td>
<td>[0.01]</td>
</tr>
<tr>
<td></td>
<td>σ = 0</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>[0.00]</td>
<td>[0.00]</td>
</tr>
</tbody>
</table>

This table shows the degree of unspanned macro forecasts in actual and simulated data, based on one-step-ahead macro forecasts using linear projections on current yield PCs and the relevant macro variable. The table reports relative root-mean-squared error (RMSE) of forecasts based on both the macro variable and yield PCs compared to forecasts based only on yield PCs. The first two rows show the relative RMSEs for the actual data, and the rest of the table shows medians across simulated samples and in square brackets the fractions of the simulated samples in which the relative RMSE is below the value in the actual data (that is, p-values).
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.
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 \textit{UGAP} and \textit{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 \textit{UGAP} and \textit{CPI} macro data, as well as from a three-factor yields-only model.