OXFORD

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

Michael D. Bauer and Glenn D. Rudebusch

Federal Reserve Bank of San Francisco

Abstract

Most existing macro-finance term structure models (MTSMs) appear incompatible with regression evidence of unspanned macro risk. This "spanning puzzle" appears to invalidate those models in favor of new unspanned MTSMs. However, our empirical analysis supports the previous spanned models. Using simulations to investigate the spanning implications of MTSMs, we show that a canonical spanned model is consistent with the regression evidence; thus, we resolve the spanning puzzle. In addition, direct likelihood-ratio tests find that the knife-edge restrictions of unspanned models are rejected with high statistical significance, though these restrictions have only small effects on cross-sectional fit and estimated term premia.

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 (VAR) for the macroeconomic and yield-curve variables coupled with a reduced-form pricing kernel.[1] In

addition, by incorporating 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 they imply macro spanning, the models used in macro-finance term structure research have recently come under severe criticism. Joslin, Priebsch, and Singleton (2014) (henceforth JPS) argue that previous MTSMs impose "counterfactual restrictions on the joint distribution of bond yields and the macroeconomy" (p. 1197). The criticism is based on regression evidence that suggests the presence of unspanned macro information. First, regressions of macro variables on observed yields can give quite low R^2. For example, JPS find that only 15% of the variation in their measure of economic activity is captured in the first three principal components (PCs) of the yield curve, rather than the 100% predicted by theoretical macro spanning (also see Duffee, 2013b). Second, there is evidence that macroeconomic variables have predictive power for excess bond and stock returns beyond the information contained in yields, as documented by JPS and others (Cooper and Priestley, 2008; Ludvigson and Ng, 2009; Greenwood and Vayanos, 2014). Third, Duffee (2013a, 2013b) documents that forecasts of macroeconomic variables are not spanned by the yield curve as implied by spanned MTSMs.

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." It casts doubt on the validity of essentially all previous macro-finance models used in the literature and is a major road-block to further macro-finance term structure research. 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). In response, JPS and others advocate replacing existing spanned MTSMs with unspanned MTSMs. These new models impose knife-edge restrictions in an otherwise standard MTSM to sever 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

---

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

3   Examples of research using reduced-form unspanned models include Wright (2011); Chernov and Mueller (2012); Priebsch (2014); and Coroneo, Giannone, and Modugno (2015).

macro variables, that would require development of a new class of structural economic models to connect bond yields to the economy.[4]

In this paper, we resolve the spanning puzzle by providing strong empirical support for spanned models and reconciling them with the regression evidence. We first consider a direct statistical test of plausible estimated versions of spanned and unspanned models. We demonstrate how unspanned models are nested by spanned models and perform 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 usual measures of economic activity and inflation. We also directly 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 a simulation-based test which shows that estimated spanned MTSMs are in fact not contradicted statistically by the regression evidence on unspanned macro information. 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 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 generally 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, $\mathcal{N}$ risk factors, macro spanning implies that macro variables will be spanned by $\mathcal{N}$ linear combinations of yields. But the macro variables will not be spanned by less than $\mathcal{N}$ linear combinations, so the validity of the regression evidence for unspanned macro information depends on using a sufficient number of linear combinations of yields.[5] Second, even after incorporating the correct 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. But 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

---

4 It would also require a 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).

5 For 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.

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, very small yield measurement errors in empirical MTSMs is sufficient to generate the appearance of unspanned macro information in the data: The wedge created by measurement error with a standard deviation of six basis points is enough to prevent the spanning regressions from properly identifying the presence or absence of spanning in MTSMs.[6]

Some other studies have also investigated the wedge between information in true and observed risk factors created by measurement error in yield-curve models. Duffee (2011b) 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.

Our results show that the unspanned regression evidence 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. 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.[7] 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.

In addition, 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.[8] 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

---

6 Our finding that plausible, small measurement errors is sufficient to generate the empirically observed patterns of unspanned macro information stands in contrast to claims in the literature, including JPS, Duffee (2013b), and Priebsch (2014). 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.

7 We come to a different conclusion in this regard from JPS because, as we explain in the paper, they in fact compare an unspanned MTSM to a yields-only model rather than a spanned model.

8 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.

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 presents and discusses the conventional, spanned macro-finance models, the spanning puzzle, and the unspanned models proposed by JPS. In Section 3 we test the knife-edge restrictions of unspanned models and show that they are rejected by the data. In Sections 4, 5, and 6 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 7, we investigate the implications of macro-spanning for term premia. Section 8 concludes.

2. Spanning in MTSMs

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, and our specification and estimation of the models.

2.1   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 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, Le, and Singleton (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_t^j and W^j the first j yield portfolios and their weights. We take the first $\mathcal{L}$ linear combination of yields, $P_t^{\mathcal{L}}$, as the yield factors. We use PCs of observed yields, and the corresponding loadings make up the rows of W. Hence, there are $\mathcal{N} = \mathcal{L} + \mathcal{M}$ risk factors, denoted $Z_t = (P_t^{\mathcal{L}'}, M_t')'$, 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:

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

The risk factors are assumed to follow a Gaussian VAR under the risk-neutral probability measure Q:

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

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

$$Y_t = A + BZ_t = A + B_P P_t^{\mathcal{L}} + B_M M_t, \tag{3}$$

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[9]:

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

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 $\mathcal{N}$ yield portfolios. To see this, premultiply Equation (3) with an $(\mathcal{N} \times J)$ matrix, $W^{\mathcal{N}}$, to select $\mathcal{N}$ linear combinations of model-implied yields, $P_t^{\mathcal{N}} = W^{\mathcal{N}} A + W^{\mathcal{N}} B Z_t$. This equation can, outside of knife-edge cases, be inverted for Z_t, and in particular for the macro factors:

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

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

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 PCs, so one regression specification to examine this issue is

$$m_t = \beta_0 + \beta_1' \text{PC}_t^{(3)} + u_t, \tag{6}$$

where m_t is one of the macroeconomic variables and $\text{PC}_t^{(3)}$ 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^2s are on the wrong side of 1/2" (p. 412). We will investigate the regression evidence on unspanned macro variation in Section 4.

---

9   While there is some evidence that additional lags (Cochrane and Piazzesi, 2005; Joslin, Le, and Singleton, 2013a) or moving average terms (Feunou and Fontaine, 2015) can be helpful to capture yield dynamics, we stay in the class of first-order Markov models like the majority of the literature on no-arbitrage MTSMs.

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

$$rx_{t,t+12}^{(n)} = \beta_0 + \beta_1'PC_t^{(3)} + \beta_2'M_t + u_{t+12}, \tag{7}$$

where $rx_{t+12}^{(n)}$ 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 β_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 5.

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

$$m_{t+1} = \beta_0 + \beta_1'PC_t^{(3)} + \beta_2'M_t + u_t \tag{8}$$

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, 2013) provides strong evidence against $\beta_2 = 0$, which we term evidence for unspanned macro forecasts. We revisit this evidence in Section 6.

In sum, there is regression evidence for 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 constitute the spanning puzzle.

2.3 The Unspanned MTSM

We now turn to the alternative model that JPS proposed to address the spanning puzzle, a new macro-finance model specification that incorporates unspanned macro risks.

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

$$r_t = \rho_0 + \rho_P'P_t^{\mathcal{L}} + \underbrace{0_{\mathcal{M}}'}_{\rho_M'} M_t. \tag{9}$$

Furthermore, instead of Equation (2), the yield factors $P_t^{\mathcal{L}}$ follow an autonomous VAR under Q that is independent of the macro factors:

$$P_t^{\mathcal{L}} = \mu_{P0}^Q + \varphi_{PP}^Q P_{t-1}^{\mathcal{L}} + \underbrace{0_{\mathcal{L}\times\mathcal{M}}}_{\varphi_{PM}^Q} M_{t-1} + \Sigma_P \varepsilon_{tP}^Q, \quad \varepsilon_{tP}^Q \overset{iid}{\sim} N(0, I_{\mathcal{L}}). \tag{10}$$

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

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

---

10 However, Bauer and Hamilton (2016) argue that this evidence suffers from small-sample problems and is therefore difficult to interpret. They suggest a bootstrap approach to testing whether spanning holds in the data.

As a consequence of Equations (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

$$Y_t = A + B_P P_t^{\mathcal{L}} + \underbrace{0_{\mathcal{M} \times \mathcal{M}}}_{B_M} M_t. \tag{11}$$

Equation (11) clarifies that in unspanned models, there is no direct link from macro factors to contemporaneous yields. This is a direct consequence of the "knife-edge restrictions"[11]

$$\rho_M = 0_{\mathcal{M}}, \quad \varphi_{PM}^Q = 0_{\mathcal{L} \times \mathcal{M}}. \tag{12}$$

Under the real-world measure, the VAR for Z_t is the same as in the spanned model, given in Equation (4). Since expectations of future yields/returns and of future macro variables are not spanned by model-implied yields, macro variables can have additional predictive power for both. Furthermore, the spanning condition (5) does not hold: one cannot back out the risk factors Z_t from model-implied yields, because the matrix $W^{\mathcal{N}} B$ is singular and cannot be inverted to yield Z_t as a function of $P_t^{\mathcal{N}}$. Instead of Equation (5), we have

$$M_t = \gamma_0 + \gamma_P P_t^{\mathcal{L}} + OM_t, \tag{13}$$

(Equation (11) of JPS) where OM_t captures the orthogonal macroeconomic variation not captured by a projection on model-implied yields.

The zero loadings of model-implied yields on macro variables imply that there are no direct effects of macro variables on interest rates. Instead, macro variables indirectly affect yields through their correlation with the yield factors $P_t^{\mathcal{L}}$. The component of macro variables that is uncorrelated with the yield factors, OM_t, does not affect yields at all. A bond yield is the sum of an expectations (risk-neutral) component and a term premium, hence if shocks to OM_t affect expectations of future short rates, they need to affect term premia with exactly the same magnitude but with opposite sign, so that the two effects offset each other and yields remain unchanged.

When comparing Equations (5) and (13) 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 (13) also holds in spanned models, since a projection on $\mathcal{L} < \mathcal{N}$ linear combination of yields cannot fully explain the macroeconomic variation and naturally leaves an orthogonal projection residual. Because the risk factors are the same in both models, γ_0 and γ_P are also identical across models, as is OM_t.[12] In other words,

---

11 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. 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 that is similar to our analysis of unspanned macro information.

12 JPS claim that conventional, spanned MTSMs impose that OM_t in Equation (13) is zero (p. 1205). But 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 $\mathcal{L} = 3$ and $\mathcal{M} = 2$ to a spanned model with $\mathcal{L} = 1$ and $\mathcal{M} = 2$. These models have different risk factors and vastly

spanned and unspanned models have identical implications for projections of macro variables on $\mathcal{L}$ yield factors (e.g., level, slope, and curvature, when $\mathcal{L} = 3$).

The spanned model nests the unspanned model: If the knife-edge restrictions (12) 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. For testing these knife-edge restrictions, one needs to account for the fact that additional parameters are not identified under the null hypothesis, as we do below in Section 3.1.

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

$$\begin{pmatrix} P_t^{\mathcal{L}} \\ M_t \end{pmatrix} = \begin{pmatrix} \mu_{P0} \\ \mu_{M0} \end{pmatrix} + \begin{pmatrix} \varphi_{PP} & \varphi_{PM} \\ \varphi_{MP} & \varphi_{MM} \end{pmatrix} \begin{pmatrix} P_{t-1}^{\mathcal{L}} \\ M_{t-1} \end{pmatrix} + \Sigma \varepsilon_t.$$

The $\mathcal{L} \times \mathcal{M}$ matrix φ_{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 $P_t^{\mathcal{L}}$ 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 in Section 3.

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 $SM(\mathcal{L}, \mathcal{M})$, and the un-spanned models 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 (see also the discussion in Section 4.5).

Our models are estimated with yield data similar to that used by JPS and like them we use 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, 2016). The yields are unsmoothed zero-coupon Treasury yields, bootstrapped from observed bond prices using the Fama–Bliss methodology.[13] The yield maturities are 3 and 6 months, and 1–10 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 3-month moving average of the Chicago Fed's National Activity Index,[14] and INF, which corresponds to survey expectations of

different economic implications and are not properly comparable. An appropriate comparison requires spanned and unspanned models with the same risk factors.

13 We thank Anh Le for supplying these data.

14 This measure is constructed so that negative values indicate below-average economic growth and positive values indicate above-average growth.

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 the CPI excluding food and energy prices, that is, 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 4.1.

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 necessary to include measurement errors to avoid stochastic singularity. We denote the observed yields $\tilde{Y}_t = Y_t + e_t$, where the J-vector e_t contains serially uncorrelated Gaussian measurement errors. We assume that the errors on each maturity have equal variance, σ_e^2, so that the likelihood tries equally hard to match yields of all maturities. As in Joslin, Singleton, and Zhu (2011), Joslin, Le, and Singleton (2013b), JPS, and other recent studies, we assume in our estimation that yield factors are observable, which substantially simplifies estimation as no filtering is necessary.[15] This assumption is largely inconsequential for parameter estimates because filtered and observed low-order PCs are very similar.[16] The presence of measurement error has important implications for the observability of the theoretical macro-spanning condition, as we will show in Sections 4–6.

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″ models in Joslin, Le, and Singleton (2013b) and in the (unspanned) models in JPS. Of course, measurement errors for the macro variables would create further unspanned macro variation 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.

Estimation is carried out using maximum likelihood. Normalization assumptions are needed for identification of the parameters in a no-arbitrage term structure model. For the spanned model, we use the canonical form and parameterization of Joslin, Le, and Singleton (2013b). This normalization is based on the idea that one can rotate the risk factors into $P_t^{\mathcal{N}}$, and then apply the canonical form of Joslin, Singleton, and Zhu (2011). The

---

15   Our assumption $\tilde{P}_t^{\mathcal{L}} = W^{\mathcal{L}} \tilde{Y}_t = W^{\mathcal{L}} Y_t = P_t^{\mathcal{L}}$ implies that $W^{\mathcal{L}} e_t = 0$, so that e_t effectively contains only $J - \mathcal{L}$ independent errors. If we denote by V an orthonormal basis of the nullspace of $W^{\mathcal{L}}$ (such that $W^{\mathcal{L}} V' = 0$ and $VV' = I_{J-\mathcal{L}}$) the measurement error assumption is $V e_t \sim^{\text{iid}} N(0, \sigma_e^2 I_{J-\mathcal{L}})$. The measurement error variance that maximizes the likelihood function is therefore $\hat{\sigma}_e^2 = \frac{\sum e_t' e_t}{T(J-\mathcal{L})}$.

16   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). See also Joslin, Le, and Singleton (2013b).

fundamental parameters of the model are λ^Q, the eigenvalues of φ^Q, k_∞^Q, which determines the long-run risk-neutral mean of the short rate, the spanning parameters γ_0, γ_1, the VAR parameters μ, φ, and Σ, and the standard deviation of the measurement errors, σ_e.[17] For the unspanned model, we use the canonical form of JPS. In this case, the free parameters are k_∞^Q, λ^Q, μ, φ, Σ, and σ_e. The parameters k_∞^Q, μ, φ, and σ_e can be concentrated out of the likelihood function, and numerical optimization of the maximum likelihood is carried out over the remaining parameters, of which there are thirty-two for SM(3, 2) and eighteen for USM(3, 2).

Our model specifications do not impose any overidentifying restrictions, that is, they are maximally flexible. An alternative is to impose restrictions on risk prices, which typically improves inference about risk premia by making better use of the information in the cross section of interest rates—for an in-depth discussion see Bauer (2016).[18] 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.[19] 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 I reports root-mean-squared errors (RMSEs), calculated for selected individual yields as well as across all yields, as well as estimates for the standard deviation of the yield measurement errors, $\hat{\sigma}_e$. All four models fit yields well, with fitting errors on average around five to six basis points.[20] 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 can capture some additional yield variation, but these improvements are small in economic terms. Our spanned models fit the yield curve notably better than those in Joslin, Le, and Singleton (2013b) or Joslin, Le, and Singleton (2013a), which allow for only one or two yield factors.

3. Testing Knife-Edge Restrictions

We now carry out likelihood ratio tests of the restrictions of unspanned MTSMs. The use of unspanned MTSMs is typically motivated and justified only on the basis of indirect regression evidence for unspanned macro information (see JPS, Wright (2011), and Priebsch

---

17 The parameters ρ_0, ρ_1, μ^Q, and φ^Q are determined by the fundamental parameters according the mapping provided in Appendix A of Joslin, Le, and Singleton (2013b).

18 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 (2016) and JPS, or directly adjust for the small-sample bias, as in Bauer, Rudebusch, and Wu (2012).

19 In addition, they restrict the largest eigenvalue of φ to equal the largest eigenvalue of φ^Q.

20 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.

Table I. Cross-sectional fit of spanned and unspanned MTSMs

This table presents the root-mean-squared fitting errors (RMSEs) 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, $\hat{\sigma}_e$, which by construction equals the RMSE times $\sqrt{J/(J-\mathcal{L})}$.

Model	Macro variables	$\hat{\sigma}_e$	All	3 m	6 m	1 y	2 y	3 y	5 y	7 y	10 y
SM(3, 2)	GRO, INF	5.7	5.0	5.1	4.6	6.8	3.6	3.9	5.4	4.2	6.3
SM(3, 2)	UGAP, CPI	6.2	5.4	5.9	4.6	7.3	3.8	4.8	6.1	4.4	7.4
USM(3, 2)	GRO, INF	6.6	5.7	6.1	4.8	8.4	4.2	4.7	6.1	4.5	7.7
USM(3, 2)	UGAP, CPI	6.6	5.7	6.1	4.8	8.4	4.2	4.7	6.1	4.5	7.7

(2014), among others). In contrast, we conduct direct hypothesis tests of the knife-edge unspanned macro restrictions that underlie these models.[21]

3.1 Model-Based Tests

We first test the knife-edge restrictions that model USM(3, 2) imposes on model SM(3, 2), described in Section 2.3. The log-likelihood values for these two models are shown in Table II for the two different data sets with macro variables GRO/INF and UGAP/CPI. For both data sets, the spanned models fit the data substantially better than the unspanned ones, as reflected in the larger log-likelihood values. A separate comparison of the two components of the log-likelihood for cross-sectional fit (based on measurement errors) and for time series fit (based on VAR forecast errors) reveals that the improved fit is almost exclusively due to smaller yield fitting errors. The knife-edge restrictions barely affect the VAR dynamics—only Σ changes slightly while μ and Φ are unchanged—but they do affect how well the models fit observed yields: Allowing the macro variables to enter as risk factors in the yield Equations (3) reduces the fitting errors of the models (see also Table I). 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.[22]

The last row of each panel of Table II reports the values of the likelihood-ratio test statistics. The null hypothesis imposes eight parameter restrictions, given in Equation (12), but some additional parameters are not identified under the null. Under 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. The canonical unspanned model has fourteen fewer parameters, which means that six parameters are not identified under the null. We cannot simply compare the values of the likelihood-ratio test statistics to the critical values of a χ^2-distribution with eight degrees of freedom, since the usual regularity conditions for the

---

21 Chernov and Mueller (2012) test in an MTSM whether the yield loadings of a partially hidden latent factor are significantly different from zero. In contrast, we directly test the assumption of unspanned macro factors.

22 In models with one or two yield factors, as in Joslin, Le, and Singleton (2013b) and Joslin, Le, and Singleton (2013a), the inclusion of macro variables leads to even more substantial improvements in cross-sectional fit.

Table II. Tests of knife-edge restrictions in MTSMs.

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).

	Cross section	Time series	Total
GRO/INF models			
LLK SM(3, 2)	21,109	393	21,502
LLK USM(3, 2)	20,781	392	21,173
Likelihood-ratio statistic			658
UGAP/CPI models			
LLK SM(3, 2)	20,906	431	21,337
LLK USM(3, 2)	20,780	429	21,210
Likelihood-ratio statistic			254

validity of the asymptotic χ^2-distribution are not satisfied, and the limiting distribution is non-standard (Hansen, 1996). One possible way to address this problem is to approximate the small-sample distribution of the test statistic using bootstrap simulations. But such a procedure would require re-estimating our macro-finance models many times on simulated data sets, which is computationally very costly.

Fortunately, there is a much simpler way to calculate how significant the improvements in the log-likelihood function are.[23] We can come up with conservative critical values for the test statistics based on the insight that every MTSM can be viewed as a restricted version of a reduced-form factor model. MTSMs restrict the loadings of yields on the factors to be consistent with no-arbitrage, whereas factor models leave these loadings unrestricted.[24] Consider the reduced-form model corresponding to SM(3, 2), which consists of a VAR for Z_t and $J - \mathcal{L}$ non-trivial measurement equations.[25] We denote this reduced-form spanned model by RSM(3, 2). Our spanned and unspanned models are restricted versions of RSM(3, 2), in which the loadings of yields on Z_t satisfy no-arbitrage conditions. In particular, USM(3, 2) has fifty less parameters than RSM(3, 2), because RSM(3, 2) has measurement equations with $(J - \mathcal{L})(1 + \mathcal{N}) = 54$ free loadings, while in USM(3, 2) these loadings are determined by just $1 + \mathcal{L} = 4$ structural parameters. We could test the plausibility of these fifty restrictions using a likelihood-ratio test—denote the value of the test statistic by LR_1. Our real goal is to test the restrictions of USM(3, 2) versus SM(3, 2)—denote the value of this test statistic by LR_2. Because RSM(3, 2) is more general than SM(3, 2), we know that $LR_2 < LR_1$. We also know that under the null LR_1 has an approximate

---

23   This was pointed out to us by Jim Hamilton.

24   This connection was emphasized and used for estimation of term structure models by Hamilton and Wu (2012). Simple factor models for the yield curve were used, for example, by Duffee (2011a) to model and forecast bond yields.

25   J model-implied yields are linearly determined by the factors in Z_t, with measurement errors, but $\mathcal{L}$ linear combinations of yields are observed without error.

χ^2-distribution with 50 degrees of freedom. Therefore, we can evaluate our test statistics LR_2 against the critical values of a $\chi^2(50)$-distribution, and be sure that this is a conservative test. The 5% (0.1%) critical value for this distribution is 67.5 (86.7), and consequently we strongly reject the knife-edge restrictions with minuscule p-values.

3.2 Model-Free Tests

Instead of comparing no-arbitrage MTSMs with and without the knife-edge unspanned macro restrictions, we can alternatively test these restrictions in reduced-form factor models for yields. In particular, we can simply estimate models $RSM(\mathcal{L}, \mathcal{M})$ using ordinary least squares (OLS) and test whether the loadings on the macro variables are zero using likelihood-ratio statistics. In addition to its simplicity, the obvious advantage of this procedure is that it does not rely on a particular choice of MTSM specification.

The factor model $RSM(\mathcal{L}, \mathcal{M})$ consists of a VAR for $Z_t = (P_t^{\mathcal{L}}\prime, M_t\prime)\prime$ and measurement equations relating observed yields to Z_t. We estimate models with $\mathcal{L} = 1, \ldots, 5$ yield factors and $\mathcal{M} = 2$ macro factors. The factors in $P_t^{\mathcal{L}}$ are taken to be the first $\mathcal{L}$ PCs of observed yields, as in the MTSMs described in Section 2.4. The measurement equations are

$$P_t^{-\mathcal{L}} = A^* + B_P^* P_t^{\mathcal{L}} + B_M^* M_t + u_t, \tag{14}$$

where $P_t^{-\mathcal{L}}$ contains the remaining $J - \mathcal{L}$ PCs of observed yields $\tilde{Y}_t$, and u_t is a $(J - \mathcal{L})$ iid vector of measurement errors with covariance matrix Ω. We can calculate the likelihood-ratio test statistic for the restriction $B_M^* = 0$ as follows: First, we estimate the unrestricted model using OLS, obtain the residuals $\hat{u}_t$, and calculate $\hat{\Omega}_1 = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t\prime$. Second, we estimate the constrained model by dropping M_t from Equation (14), and estimate the residual covariance matrix $\hat{\Omega}_0$. The likelihood ratio statistic, using Sims' small-sample correction (see Hamilton, 1994, eq. 11.1.34), is $LR = (T - 1 - \mathcal{L} - \mathcal{M})(\log|\hat{\Omega}_1| - \log|\hat{\Omega}_0|)$ and has an approximate χ^2-distribution with degrees of freedom equal to the number of parameter restrictions, $(J - \mathcal{L})\mathcal{M}$. Note that since the VAR for Z_t is unaffected by the restrictions, there is no need to estimate it for calculation of the LR statistic.

Table III shows that for both macro data sets, and for any number of yield factors, the LR statistics are always substantially larger than the 0.1%-critical values. These model-free tests very clearly reject the knife-edge restrictions, and suggest that it is beneficial to include macro variables as factors explaining cross-sectional variation in yields.

To better understand these results we can go one step further and consider the role of each macro variable for these rejections. We estimate restricted versions of Equation (14) where we drop only one macro variable at a time. The resulting LR statistics are reported in the right panel of Table III. For the GRO/INF data set, both macro variables are individually as well as jointly highly significant. For the UGAP/CPI data, the same is true for up to three yield factors, but beyond that UGAP is only marginally significant or insignificant. The likely reason is that UGAP is closely correlated with the slope of the yield curve, as we discuss further below. More generally, however, we conclude that the rejections are not primarily driven by only specific macro variables—exclusion restrictions are typically rejected for any of the considered macro risk factors.

Our evidence casts doubt on the validity of the restrictions imposed by unspanned macro-finance models. The statistical significance of the rejections is very high.[26] At the

---

26 We have also calculated Wald test statistics that allow for the possibility of serial correlation in u_t (using Newey-West covariance estimation) and have obtained very similar results.

Table III. Tests of knife-edge unspanned macro restrictions in reduced-form regressions.

The left panel shows the values of likelihood-ratio test statistics 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 two 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 χ^2-distribution with $2(J - \mathcal{L})$ degrees of freedom, and the table shows the corresponding 0.1%-critical values (c.v.). The right panel shows test results for individually restricting the loadings on macro variables to zero, in which case the degrees of freedom are $J - \mathcal{L}$.

$\mathcal{L}$	GRO/INF	UGAP/CPI	0.1% c.v.	GRO	INF	UGAP	CPI	5% c.v.	0.1% c.v.
1	559.6	375.7	48.3	140.5	358.7	140.4	233.1	19.7	31.3
2	248.3	338.9	45.3	138.2	80.9	139.5	196.9	18.3	29.6
3	119.6	199.7	42.3	48.9	67.0	50.8	143.7	16.9	27.9
4	113.7	81.2	39.3	48.7	61.4	15.3	65.3	15.5	26.1
5	89.8	76.5	36.1	46.3	42.3	15.2	60.7	14.1	24.3

same time, however, the rejections of the knife-edge restrictions do not have high economic significance, as macro variables explain only little cross-sectional variation in the yield curve beyond that explained in the first three PCs of yields. We will revisit the issue of economic significance in Section 7, where we compare estimated term premia obtained from spanned and unspanned models.

3.3 Do JPS Reject Spanned Models?

On the surface, our results 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 "$\mathcal{M}_{us}$," to a restricted model "$\mathcal{M}_{span}$," in which the block of the φ matrix corresponding to the lagged macro variables is set to zero. Their likelihood-ratio test strongly rejects the restricted model, with a reported χ^2-statistic of 1,189. However, the restrictions of $\mathcal{M}_{span}$ do not imply a spanned MTSM. Instead, since $\mathcal{M}_{span}$ restricts φ_{PM} and φ_{MM} to zero, it effectively corresponds to a yields-only model (see Section 2.3).[27] The only difference between $\mathcal{M}_{span}$ 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 the corresponding macro forecast errors. But for yields and risk premia, $\mathcal{M}_{span}$ 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.

---

27 The restrictions of $\mathcal{M}_{span}$ do not imply macro-spanning as in Equation (5) but instead that expectations of macro variables are spanned by yields.

In the data used by JPS, current yields do not completely capture the relevant information for forecasting, as macro variables have additional predictive power.[28] This statistical rejection of model $\mathcal{M}_{\text{span}}$ is simply a reflection of the regression evidence for unspanned macro risks and unspanned macro forecasts in GRO and INF. In Sections 5 and 6 we will show that such regression evidence is consistent with both spanned and unspanned MTSMs.

4. 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. We now investigate whether this regression evidence can discriminate between spanned and unspanned MTSMs. In this section, we focus on the evidence for unspanned macro variation, based on regressions of the form in Equation (6).

4.1 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, that is, core PCE inflation.[29]

Regarding the activity measures, we include measures of the level and the growth of activity in the US 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.[30] 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 3-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 3-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

---

28   The extremely large χ^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 φ the χ^2-statistic is only 74 for the model of JPS.

29   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.

30   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.

as the ones most important for central banks in setting the short-term interest rate.[31] To 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 IV report the R^2 for these policy rules, and the partial R^2 for each macro variable under consideration (i.e., 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 with 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 0.60–0.70, 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 variables.

4.2 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 IV, 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.[32]

To help uncover how the yield curve captures macro variation, the last three columns of Table IV 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

---

31 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).

32 We find that 28% of the variation in GRO is explained by the first three PCs of yields, whereas JPS estimate only 15%. The difference is due to the fact that we use a somewhat different yield data set.

Table IV. Monetary policy rules and unspanned macro variation

The first two columns report monetary policy rule regressions in which each economic activity measure (variables 1, 2, 6–10) are paired with Core 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.

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

small, noisy portion of yield curve variation, this correlation with growth variables could be a sign of overfitting and 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 level/gap measures (policy factors) and growth measures (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.[33] Policy factors—those macro variables that subtantially drive monetary policy—show very little evidence for unspanned macro variation, and are essentially spanned by the yield curve. This is true for measures of resource slack and for core inflation. On the other hand, the non-policy factors—

---

33 This is consistent with the findings in Diebold, Rudebusch, and Aruoba (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.
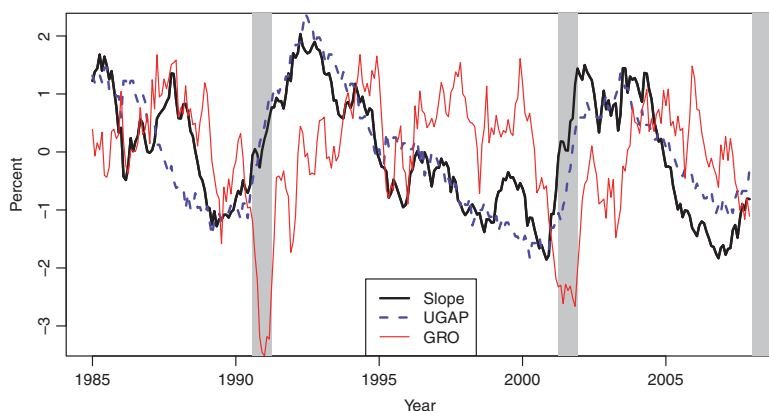
Figure 1. Slope of the yield curve and macroeconomic variables.
This figure plots the slope of the yield curve, which is measured as the second PC of yields; UGAP, which is the unemployment gap; and GRO, which is the economic growth indicator used by Joslin, Priebsch, and Singleton (2014). All variables are standardized to have mean zero and unit variance.

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 appear to have in the Fed's decisions on how to set the short-term interest rate.[34]

4.3 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. In addition to these small-sample simulations, we also calculate and report population moments for each of the regressions in Appendix C.

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. The simulated data sets have the same length as the actual data, which is $T = 276$ months. Our procedure is to simulate yield and macro factors from the VAR, construct fitted yields using the affine factor loadings, and add iid Gaussian measurement error to obtain simulated yields.[35] The standard deviation of the measurement errors, σ, is taken to be equal to six basis points, similar to our maximum-likelihood estimates reported in Table I. We also consider the case with

---

34 Our arguments suggest that growth measures of economic activity would be related more strongly to changes than to levels of the yield curve. We have found that this is indeed the case. For example, a regression with quarterly changes in three PCs explains 42% of the variation in GRO, as opposed to the 28% that Table IV reports for a specification in levels. This underlines the importance of using the right type of macro variables when investigating spanning by the yield curve.

35 Pricing errors of estimated term structure models are often serially correlated, as documented for example by Adrian, Crump, and Moench (2013) and Hamilton and Wu (2014). However, using serially correlated errors in our simulations did not change our results.

$\sigma = 0.$[36] In each simulated yield data set, we construct the PCs using the loadings in W. Then we run the spanning regressions in the simulated macro/yield data, in order to 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 whether the assumptions of the model are consistent with the unspanned regression evidence.

4.4 Spanned MTSMs Are Consistent with Regression Evidence

We first run regression (6) in both actual and simulated data samples. Table V 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 V 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 five PCs using data simulated from the spanned model SM(3, 2)—these results are in the bottom panel of Table V. 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 the model cannot reproduce any unspanned macro variation and appears inconsistent with the data. However, adding small, plausible yield measurement error changes this conclusion: With $\sigma = 6$bps the amount of unspanned macro variation in the simulated data matches that in the actual data, and the p-values are substantially above five percent, so the spanned model cannot be rejected.

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 we have explored a variety of alternative specifications with different variables, numbers of factors, and parameter restrictions and obtained similar results.

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 V. 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

---

36 In the estimation we assumed observable yield factors, whereas in our simulations all yields are measured with error. This is inconsequential for our results, as filtered and observed low-order PCs are virtually indistinguishable (see footnote 16).

Table V. Unspanned macro variation in MTSMs

This table documents unspanned macro variation in actual data and in data that are 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). We simulated 10,000 artificial data samples in each case.

			GRO	INF	CPI	UGAP
Data	3 PCs		0.28	0.81	0.81	0.72
	5 PCs		0.38	0.86	0.81	0.75
USM(3, 2)	3 PCs	$\sigma = 6bps$	0.31	0.74	0.68	0.71
			[0.38]	[0.72]	[0.82]	[0.53]
		$\sigma = 0bps$	0.32	0.74	0.69	0.71
			[0.35]	[0.72]	[0.81]	[0.53]
SM(3, 2)	3 PCs	$\sigma = 6bps$	0.31	0.72	0.68	0.72
			[0.39]	[0.79]	[0.83]	[0.51]
		$\sigma = 0bps$	0.32	0.72	0.68	0.72
			[0.34]	[0.79]	[0.81]	[0.51]
SM(3, 2)	5 PCs	$\sigma = 6bps$	0.51	0.83	0.68	0.74
			[0.10]	[0.68]	[0.82]	[0.54]
		$\sigma = 0bps$	1.00	1.00	1.00	1.00
			[0.00]	[0.00]	[0.00]	[0.00]

macro variables are spanned by $\mathcal{L}$ PCs of yields. Only when we include higher-order PCs, that is, condition on $\mathcal{N}$ 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 V. In this case, we use three PCs of yields, because only the three yield factors enter the affine yield equations—including more PCs does not increase explanatory power of yields but leads to multicollinearity. The results of this exercise show, not surprisingly, that the unspanned models are able to replicate the unspanned macro variation in the data. But note that both spanned and unspanned models have the same implications for the relationship between macro variables and $\mathcal{L}$ linear combinations of yields, because Equation (13) holds in both types of models. Therefore, the results reported in the top panel of Table V are essentially identical to those in the middle panel. Based on spanning regressions on $\mathcal{L}$ yield PCs, there is no evidence to distinguish between spanned and unspanned models.

In Appendix C, we consider large-sample results for the spanning regressions based on model-implied population moments, which provides further detail and intuition. For unspanned macro variation, the small-sample simulation results are confirmed by the population moments.

4.5   The Role of Measurement Error

Measurement error plays a crucial role for reconciling spanned models with the evidence from regressions of macro variables on five yield PCs. Yield-curve models generally require measurement error to fit the data, because their factor structure otherwise would lead to stochastic singularity. For macro-finance models, measurement error has the additional

important effect that it breaks theoretical macro spanning, just as it breaks stochastic singularity. While this is obvious conceptually, our analysis shows that it also works empirically: 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 generally very close to low-order PCs of model-implied yields. The reason that spanned models with small measurement errors can generate substantial amounts of unspanned macro information is that measurement error matters little for level, slope, and curvature, but substantially affects higher-order PCs, since these are relatively smaller and less precisely estimated. In Appendix D we report the share of the variance in each PC that is due to measurement error. This share is large for the higher-order PCs, for example, it is about 99% for the fifth PC. The higher-order yield PCs matter for theoretical macro-spanning, but they are contaminated by noise with the introduction of 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, Le, and Singleton (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, Reinhart, and Sack (2004) and Joslin, Le, and Singleton (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. That is, both reduced-form MTSMs that fit yields very well and have small measurement errors, as well as models that do not match observed yields very accurately, are consistent with the regression evidence on unspanned macro variation.

5. 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 Equation (7), using the average excess returns for bonds of maturities from 2 to 10 years as the dependent variable. Then we investigate whether this evidence rejects spanned MTSMs.

5.1   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 VI 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 eighteen lags.[37] The third column reports the increase in R^2 when including macro

---

37   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).

Table VI. Unspanned macro risk and unspanned macro forecasts

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 eighteen 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 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, that is, 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 twelve lags. The last column shows the relative RMSE of macro-yield versus yields-only forecast.

	Excess returns				Macro forecasts		
	t-stat.	p-value	ΔR^2	RMSE	Autocorr.	t-stat.	RMSE
Policy factors							
1) Unemp. gap	0.61	(0.54)	0.01	1.00	0.98	50.84	0.34
2) Output gap	0.68	(0.50)	0.01	1.00	0.95	26.99	0.46
3) INF (JPS)	4.20	(0.00)	0.16	0.89	0.99	44.29	0.34
4) Core CPI inflation	1.63	(0.10)	0.06	0.96	0.99	55.51	0.29
5) Core PCE inflation	1.73	(0.08)	0.03	0.98	0.98	49.60	0.32
Non-policy factors							
6) GRO (JPS)	2.70	(0.01)	0.05	0.97	0.91	21.31	0.50
7) GDP growth (ma3)	1.83	(0.07)	0.01	0.99	0.47	4.94	0.92
8) GDP growth (yoy)	0.84	(0.40)	0.01	1.00	0.77	11.18	0.71
9) IP growth (ma3)	3.39	(0.00)	0.12	0.92	0.94	36.45	0.42
10) Jobs growth (ma3)	1.60	(0.11)	0.03	0.98	0.87	17.00	0.53

variables, relative the R^2 in predictive regressions with only three yield PCs, which is 20%. The fourth column shows the relative RMSE of the predictions that include the macro variable compared with 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.[38] 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 for the regressions on

---

38 While Cooper and Priestley (2008) find some predictive power of the output gap for excess bond returns, Bauer and Hamilton (2016) show that these results are not robust, consistent with our results here.

unspanned macro variation in Section 4, there is a marked difference in the results 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, that is, 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 is robust to changes in the sample period used. Furthermore, inference about spanning in bond return regressions is problematic from an econometric perspective—see Bauer and Hamilton (2016). Finally, the evidence on the predictive power of macro variables is in sample and may not be reflected in out-of-sample forecasting.[39] 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.

5.2   Spanned MTSMs Are Consistent with Excess Return Regressions

Can spanned models reproduce the regression evidence on bond return predictability? In Table VII 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 VII reports Wald statistics, calculated using Newey–West standard errors with eighteen 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 VII 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 χ^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 VII reports medians of the statistics across simulated samples, as well as the fractions of samples in which these statistics are above the values in the actual data, that is, simulation-based p-values. Without measurement error ($\sigma = 0$) and with five PCs as regressors, the spanned model SM(3, 2) of course implies that the macro data cannot have any additional predictive power—see the bottom two rows of Table VII. 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—there is some unspanned macro risk in the simulated data. The PCs of observed yields do not fully capture the information in the yield curve and as a consequence, macro variables, though theoretically

---

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

Table VII. Unspanned macro risk in MTSMs

This table shows evidence for unspanned macro risk in actual and simulated data, based on predictive regressions for 1-year excess bond returns on yield PCs and macro variables. Statistical significance is measured by the Wald statistic for joint significance of the coefficients on the two macro variables, using Newey–West standard errors with eighteen lags. 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 χ^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 (i.e., p-values). We simulated 10,000 artificial data samples in each case.

			GRO/INF		UGAP/CPI	
			Wald	ΔR^2	Wald	ΔR^2
Data	3 PCs		21.64	0.19	3.00	0.06
			(0.00)		(0.22)	
Data	3 PCs		13.71	0.15	4.23	0.08
			(0.00)		(0.12)	
USM(3, 2)	3 PCs	$\sigma = 6bps$	9.41	0.10	4.48	0.05
			[0.20]	[0.20]	[0.62]	[0.41]
		$\sigma = 0bps$	9.38	0.10	4.51	0.05
			[0.19]	[0.19]	[0.62]	[0.41]
SM(3, 2)	3 PCs	$\sigma = 6bps$	10.03	0.11	4.57	0.05
			[0.21]	[0.21]	[0.62]	[0.41]
		$\sigma = 0bps$	10.00	0.11	4.56	0.05
			[0.21]	[0.22]	[0.63]	[0.41]
SM(3, 2)	5 PCs	$\sigma = 6bps$	8.53	0.06	4.61	0.05
			[0.32]	[0.05]	[0.53]	[0.29]
		$\sigma = 0bps$	0.00	0.00	0.00	0.00
			[0.00]	[0.00]	[0.00]	[0.00]

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. With yield measurement error, the spanned model cannot be rejected by this evidence. The statistical significance for predictive power in the data, measured by the Wald statistic, is easily matched in the simulated data, with p-values above 0.3 for both data sets.

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 three yield PCs, we again see that the spanned and unspanned models have virtually identical implications.

For predictive regressions of asset returns, small-sample econometric issues often arise. Indeed, Appendix C.2 shows that in population, the model-implied predictability of bond

returns is smaller, as is the degree of unspanned macro risk. It is therefore important to simulate small samples, as we do here, in order to correctly assess the models' plausibility in light of the predictive regression evidence. Our small-sample results show that the evidence on unspanned macro risk does not give any reason to prefer unspanned models over spanned models.

6. Empirical MTSMs and Unspanned Macro Forecasts

The third and final dimension of macro spanning that we investigate concerns macroeconomic forecasts. If macro information is spanned by the yield curve, then 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 Equation (8).

6.1   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 VI reports t-statistics (using Newey–West standard errors with twelve lags) for the null hypothesis that the lagged macro variable can be excluded, that is, 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, that is, 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.

6.2   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 versus 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 VIII 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

Table VIII. Unspanned macro forecasts in MTSMs

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. It reports relative RMSE of forecasts based on both the macro variable and yield PCs compared with 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 (i.e., p-values). We simulated 10,000 artificial data samples in each case.

			GRO	INF	CPI	UGAP
Data	3 PCs		0.47	0.30	0.29	0.32
	5 PCs		0.50	0.34	0.29	0.34
USM(3, 2)	3 PCs	$\sigma = 6bps$	0.49	0.36	0.33	0.37
			[0.38]	[0.11]	[0.21]	[0.16]
		$\sigma = 0bps$	0.49	0.36	0.33	0.37
			[0.35]	[0.11]	[0.19]	[0.15]
SM(3, 2)	3 PCs	$\sigma = 6bps$	0.49	0.35	0.33	0.38
			[0.37]	[0.15]	[0.19]	[0.14]
		$\sigma = 0bps$	0.49	0.35	0.34	0.38
			[0.35]	[0.15]	[0.16]	[0.13]
SM(3, 2)	5 PCs	$\sigma = 6bps$	0.55	0.44	0.33	0.39
			[0.11]	[0.01]	[0.18]	[0.16]
		$\sigma = 0bps$	1.00	1.00	1.00	1.00
			[0.00]	[0.00]	[0.00]	[0.00]

above, the evidence for unspanned macro forecasts is very strong in our data. The rest of the table shows 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. The distribution of the median relative RMSEs across the simulations is close to the values in the real-world data. This is true even for the spanned model in a regression with five yield PCs, where measurement error generates a substantial amount of unspanned macro forecasts. and helps to reconcile the models with the regression evidence. With the exception of INF, the small-sample p-values are above 0.1. 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 VI and VII—so the yield curve completely captures all relevant macro forecast information and lags of macro variables are not needed.

We also calculated population moments for these regressions, which are reported in Appendix C.3. For unspanned macro forecasts, just like for unspanned macro variation, the large-sample results closely parallel the small-sample results.

The simulated regression evidence on unspanned macro forecasts mirrors the results for unspanned macro variation and risk: spanned models are generally able to match the regression evidence about as well as unspanned models. In particular, the actual values of the regression test statistics from the real-world data are usually well within the probability distributions for these statistics when generated from plausible estimated spanned MTSMs.
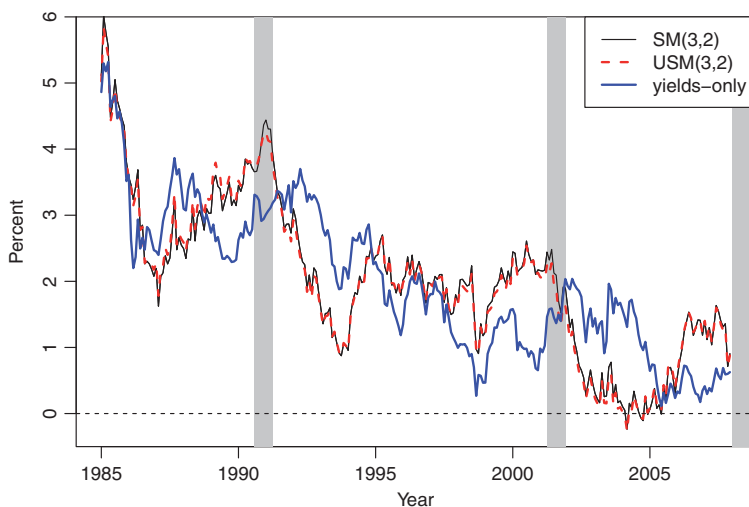
Figure 2. Term premium estimates from MTSMs with GRO and INF.
This figure depicts the two-to-three-year forward term premium estimated from spanned and unspanned macro-finance models—SM(3, 2) and USM(3, 2), respectively—using GRO and INF macro data, as well as from a three-factor yields-only model.

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.

7. 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.[40] 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 $\mathcal{M}_{us}$ model, and our yields-only model corresponds to their $\mathcal{M}_{span}$ model. The comparison shows that our estimated term premia closely resemble those of JPS. This is true 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

---

40 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, Singleton, and Zhu (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 3 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 (Table I). 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. At the same time, spanned and unspanned models contain the same VAR specification for Z_t, and result in almost identical forecasts. 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 spanned models. Second, both types of models allow for essentially the same risk premium dynamics—incorporating unspanned macro risks in an MTSM does not change the term premium implications of the model.

The yields-only model (which corresponds to the $\mathcal{M}_{\text{span}}$ 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 3.3), 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 expansions 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 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.[41]

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

---

41 Bauer, Rudebusch, and Wu (2012, 2014) also discuss the countercyclical behavior of term premia estimated from term structure models.
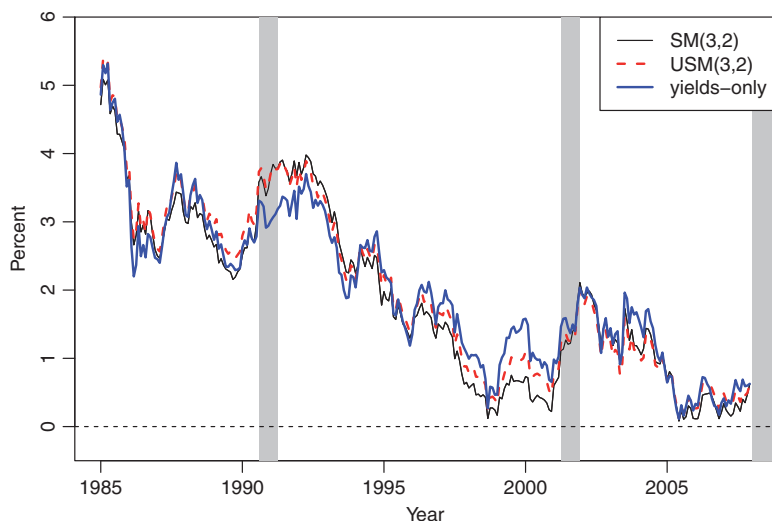
Figure 3. Term premium estimates from MTSMs with UGAP and CPI.
This figure depicts the two-to-three-year forward term premium estimated from spanned and unspanned macro-finance models—SM(3, 2) and USM(3, 2), respectively—using UGAP and CPI macro data, as well as from a three-factor yields-only model.

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 4.2). From a macroeconomic perspective, the term premia in Figure 3 appear 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.[42]

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. One benefit of these models is the lower number of parameters, and this parsimony simplifies estimation and inference. Overall, unspanned

---

42 Duffee (2013b) notes that "the spanning requirement [...] reduces significantly the ability of re-searchers 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.

models may be a useful shortcut in practical applications of MTSMs.[43] 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.

8. Conclusion

In this paper, we have proposed a solution to the macro-finance spanning puzzle that provides support for the conventional MTSMs widely used in the literature. Our findings should reassure the many researchers who have employed conventional spanned models for analyzing macro-finance interactions. 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. The reason is that these regressions cannot distinguish between plausible estimated spanned and unspanned macro-finance models. We also show that the knife-edge restrictions of the alternative unspanned MTSMs are strongly rejected in the data. At the same time, spanned and unspanned models deliver similar estimates of term premia in long-term interest rates and fit observed yields about equally well. 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 (e.g., due to changes in the monetary policy rule), we cannot rule out this possibility.[44] 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.

   Our simulation-based approach can be used to assess the spanning implications of any macro-finance model of the yield curve. We have argued that the canonical reduced-form MTSM considered in this paper is representative of a broad class of spanned macro-finance

---

43 An established example of another such usefully constrained model in the literature is the arbitrage-free Nelson-Siegel (AFNS) model of Christensen, Diebold, and Rudebusch (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.

44 For example, Song (2014) appears to generate some unspanned macro variation using regime-switching in an equilibrium MTSM.

models, and in this sense our conclusions extend to many other models. However, equilibrium models of the yield curve often impose additional restrictions on macro-yield interactions, or are calibrated to match certain asset pricing moments in the data. It may therefore be of interest to investigate whether specific structural macro-finance models are indeed consistent with the unspanned regression evidence. We leave such an investigation to future research.

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. It is unclear whether the indirect link through the interaction of macro and yield factors is sufficient to thoroughly study 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.

References

Adrian, T., Crump, R. K., and Moench, E. (2013) Pricing the term structure with linear regressions, *Journal of Financial Economics* **110**, 110–138.

Ang, A., Bekaert, G., and Wei, M. (2008) The term structure of real rates and expected inflation, *Journal of Finance* **63**, 797–849.

Ang, A., Boivin, J., Dong, S., and Loo-Kung, R. (2011) Monetary policy shifts and the term structure, *Review of Economic Studies* **78**, 429–457.

Ang, A. and Piazzesi, M. (2003) A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**, 745–787.

Bansal, R. and Shaliastovich, I. (2013) A long-run risks explanation of predictability puzzles in bond and currency markets, *Review of Financial Studies* **26**, 1–33.

Bauer, M. D. (2016) Restrictions on risk prices in dynamic term structure models, *Journal of Business & Economic Statistics*, 2016, forthcoming, Journal of Business & Economic Statistics, http://www.tandfonline.com/doi/abs/10.1080/07350015.2016.1164707.

Bauer, M. D. and Hamilton, J. D.(2016) Robust bond risk premia. Working paper 2015-15, Federal Reserve Bank of San Francisco.

Bauer, M. D. and Rudebusch, G. D. (2016) Monetary policy expectations at the zero lower bound, *Journal of Money, Credit and Banking*, forthcoming.

Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2012) Correcting estimation bias in dynamic term structure models, *Journal of Business and Economic Statistics* **30**, 454–467.

Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2014) Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: comment, *American Economic Review* **104**, 1–16.

Bean, C. (2005) Commentary: separating the business cycle from other economic fluctuations, in: *Federal Reserve Bank of Kansas City Proceedings*, pp. 181–186. Proceedings - Economic Policy Symposium - Jackson Hole. Federal Reserve Bank of Kansas City, August 2005. https://www.kansascityfed.org/publications/research/escp/symposiums/escp-2005.

Bekaert, G., Engstrom, E., and Xing, Y. (2009) Risk, uncertainty, and asset prices, *Journal of Financial Economics* **91**, 59–82.

Bernanke, B. S., Reinhart, V. R., and Sack, B. P. (2004) Monetary policy alternatives at the zero bound: an empirical assessment, *Brookings Papers on Economic Activity* **2004**, 1–100.

Bikbov, R. and Chernov, M. (2009) Unspanned stochastic volatility in affine models: evidence from eurodollar futures and options, *Management Science* **55**, 1292–1305.

Bikbov, R. and Chernov, M. (2010) No-arbitrage macroeconomic determinants of the yield curve, *Journal of Econometrics* **159**, 166–182.

Buraschi, A. and Jiltsov, A. (2007) Habit formation and macroeconomic models of the term structure of interest rates, *The Journal of Finance* **62**, 3009–3063.

Chernov, M. and Mueller, P. (2012) The term structure of inflation expectations, *Journal of Financial Economics* **106**, 367–394.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2011) The affine arbitrage-free class of Nelson–Siegel term structure models, *Journal of Econometrics* **164**, 4–20.

Cieslak, A. and Povala, P. (2015) Expected returns in treasury bonds, *Review of Financial Studies* **28**, 2859–2901.

Cochrane, J. H. and Piazzesi, M. (2005) Bond risk premia, *American Economic Review* **95**, 138–160.

Collin-Dufresne, P. and Goldstein, R. S. (2002) Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *The Journal of Finance* **57**, 1685–1730.

Cooper, I. and Priestley, R. (2008) Time-varying risk premiums and the output gap, *Review of Financial Studies* **22**, 2801–2833.

Coroneo, L., Giannone, D., and Modugno, M. (2016) Unspanned macroeconomic factors in the yields curve, *Journal of Business and Economic Statistics*, **34**, No. 3; DOI: 10.1080/07350015.2015.1052456.

Dewachter, H. and Lyrio, M. (2006) Macro factors and the term structure of interest rates, *Journal of Money, Credit and Banking* **38**, 119–140.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006) The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* **131**, 309–338.

Duffee, G. R. (2011a) *Forecasting with the term structure: the role of no-arbitrage*. Working paper January, Johns Hopkins University.

Duffee, G. R. (2011b) Information in (and not in) the term structure, *Review of Financial Studies* **24**, 2895–2934.

Duffee, G. R. (2013a) Bond pricing and the macroeconomy, in: Milton Harris George M. Constantinides and Rene M. Stulz (eds.), *Handbook of the Economics of Finance*, Vol. **2, Part B**: Elsevier, pp. 907–967.

Duffee, G. R. (2013b) Forecasting interest rates, in: Graham Elliott and Allan (Timmermanneds.), *Handbook of Economic Forecasting*, Vol. **2, Part A**: Elsevier, pp. 385–426.

Feunou, B. and Fontaine, J.-S. (2015) *Bond risk premia and Gaussian term structure models*. Working paper, Bank of Canada.

Gallmeyer, M. F., Hollifield, B., Palomino, F., and Zin, S. E. (2007) *Arbitrage-free bond pricing with dynamic macroeconomic models*. Working paper 13245, National Bureau of Economic Research.

Greenwood, R. and Vayanos, D. (2014) Bond supply and excess bond returns, *Review of Financial Studies* **27**, 663–713.

Gürkaynak, R. S. and Wright, J. H. (2012) Macroeconomics and the term structure, *Journal of Economic Literature* **50**, 331–367.

Hamilton, J. D. (1994) *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hamilton, J. D. and Wu, J. C. (2012) Identification and estimation of Gaussian affine term structure models, *Journal of Econometrics* **168**, 315–331.

Hamilton, J. D. and Wu, J. C. (2014) Testable implications of affine term structure models, *Journal of Econometrics* **178**, 231–242.

Hansen, B. E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**, 413–430.

Hördahl, P., Tristani, O., and Vestin, D. (2006) A joint econometric model of macroeconomic and term-structure dynamics, *Journal of Econometrics* **131**, 405–444.

Joslin, S., Le, A., and Singleton, K. J. (2013a) Gaussian macro-finance term structure models with lags, *Journal of Financial Econometrics* **11**, 581–609.

Joslin, S., Le, A., and Singleton, K. J. (2013b) Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs, *Journal of Financial Economics* **109**, 604–622.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014) Risk premiums in dynamic term structure models with unspanned macro risks, *Journal of Finance* **69**, 1197–1233.

Joslin, S., Singleton, K. J., and Zhu, H. (2011) A new perspective on Gaussian dynamic term structure models, *Review of Financial Studies* **24**, 926–970.

Litterman, R. and Scheinkman, J. (1991) Common factors affecting bond returns, *Journal of Fixed Income* **1**, 54–61.

Ludvigson, S. C. and Ng, S. (2009) Macro factors in bond risk premia, *Review of Financial Studies* **22**, 5027–5067.

Orphanides, A. (2003) Historical monetary policy analysis and the Taylor rule, *Journal of Monetary Economics* **50**, 983–1022.

Piazzesi, M. and Schneider, M. (2007) Equilibrium yield curves, in: D. Acemoglu, K. Rogoff, and M. Woodford (eds.), *NBER Macroeconomics Annual 2006*, Vol. **21**. MIT Press, Cambridge, MA, pp. 389–472.

Priebsch, M. (2014) *(Un)conventional monetary policy and the yield curve*. Working paper, Federal Reserve Board, Washington, DC.

Rudebusch, G. D. (2002) Assessing nominal income rules for monetary policy with model and data uncertainty, *The Economic Journal* **112**, 402–432.

Rudebusch, G. D. (2006) Monetary policy inertia: fact or fiction?, *International Journal of Central Banking* **2**, 85–135.

Rudebusch, G. D. and Swanson, E. T. (2012) The bond premium in a DSGE model with long-run real and nominal risks, *American Economic Journal: Macroeconomics* **4**, 105–143.

Rudebusch, G. D. and Wu, T. (2007) Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models, *Journal of Money, Credit and Banking* **39**, 395–422.

Rudebusch, G. D. and Wu, T. (2008) A macro-finance model of the term structure, monetary policy, and the economy, *Economic Journal* **118**, 906–926.

Song, D. (2014) *Bond market exposures to macroeconomic and monetary policy risks*. Unpublished manuscript, Boston College.

Taylor, J. B. (1999) A historical analysis of monetary policy rules, in: J. B. Taylor, *Monetary Policy Rules*. National Bureau of Economic Research, University of Chicago Press, Chicago, IL, pp. 319–348.

Wachter, J. A. (2006) A consumption-based model of the term structure of interest rates, *Journal of Financial Economics* **79**, 365–399.

Wright, J. H. (2011) Term premia and inflation uncertainty: empirical evidence from an international panel dataset, *American Economic Review* **101**, 1514–1534.

Appendix A: Affine Bond Pricing

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

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

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

$$\mathcal{A}_{m+1} = \mathcal{A}_m + (\mu^Q)'\mathcal{B}_m + \frac{1}{2}\mathcal{B}'_m \Sigma\Sigma'\mathcal{B}_m - \delta_0$$

$$\mathcal{B}_{m+1} = (\varphi^Q)'\mathcal{B}_m - \delta_1$$

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

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

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

Appendix 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 4.3. The yield factors are related to yields by the matrix W, which contains the PC loadings. To construct W we start from the orthonormal eigenvectors of actual yields, scale the loadings for the first PC to add up to unity, and then scale all loadings by 1200 so that they correspond to annualized percentages. In this way the yield factors have the same scaling as the macro variables.

Table BI presents the parameters for models SM(3, 2) and USM(3, 2) estimated on the data set with the macro variables GRO and INF, and Table BII presents the estimates obtained using the data set with UGAP and INF. For further discussion of the macro variables, see Section 4.1. We present the parameter estimates without standard errors, as these are not needed for our purpose of simulating and comparing maximally flexible models. In addition, the numerical approximations to the first and second derivatives of the likelihood functions are unreliable for models with this many parameters. While could obtain standard errors using a bootstrap exercise, this would require re-estimating our macro-finance models many times on simulated data sets, involving very high computational costs.

Tables BI and BII report in the last row of each panel the estimated measurement error standard deviation, which is multiplied by 1200 so that the units are annualized percentage points, and indicates the cross-sectional fit to observed yields. Also reported are the maximized value of the log-likelihood function, and the maximum absolute eigenvalue of the VAR mean-reversion matrix Φ, which shows that the estimated VARs are stationary.

Table BI. Parameter estimates for macro-finance models using data set with GRO/INF

k^Q_∞ and σ_e are scaled by 1,200. "Φ ev." is largest absolute eigenvalue of Φ.

Spanned model: SM(3, 2)

k^Q_∞	λ^Q				
0.032	0.991	0.990	0.864	0.864	0.864
γ_0	γ_1				
−0.153	0.031	−0.013	−1.434	5.911	62.960
−1.300	0.398	0.104	0.957	2.822	−40.953
μ	Φ				
0.078	0.937	−0.013	−0.022	0.105	0.100
0.021	0.006	0.978	0.170	−0.167	−0.040
0.100	−0.010	−0.006	0.783	−0.040	0.044
0.069	−0.005	0.018	−0.111	0.882	−0.002
0.050	0.002	0.006	−0.059	0.028	0.986
Σ					
0.284	0	0	0	0	
0.158	0.275	0	0	0	
−0.084	−0.020	0.130	0	0	
0.044	0.001	−0.023	0.187	0	
0.030	0.012	0.014	0.006	0.112	
σ_e	LLK			Φ ev.	
0.057	21501.9			0.984	

Unspanned model: USM(3, 2)

k^Q_∞	λ^Q				
0.033	0.997	0.959	0.872		
μ	Φ				
0.078	0.937	−0.013	−0.022	0.105	0.100
0.021	0.006	0.978	0.170	−0.167	−0.040
0.100	−0.010	−0.006	0.783	−0.040	0.044
0.069	−0.005	0.018	−0.111	0.882	−0.002
0.050	0.002	0.006	−0.059	0.028	0.986
Σ					
0.282	0	0	0	0	
0.185	0.281	0	0	0	
−0.082	−0.016	0.130	0	0	
0.047	0.001	−0.019	0.186	0	
0.036	0.014	0.015	0.006	0.112	
σ_e	LLK			Φ ev.	
0.066	21,172.9			0.984	

Table BII. Parameter estimates for macro-finance models using data set with UGAP/CPI

k_∞^Q and σ_e are scaled by 1,200. "Φ ev." is largest absolute eigenvalue of Φ.

Spanned model: SM(3, 2)

k_∞^Q	λ^Q				
0.028	0.996	0.975	0.843	0.623	0.623
γ_0	γ_1				
−86.891	1.112	−1.416	24.812	203.202	−1293.107
−28.240	0.156	−0.862	5.771	78.644	−345.862
μ	Φ				
0.127	0.965	0.005	−0.074	0.033	0.028
0.042	−0.059	0.951	0.166	0.095	−0.025
0.107	−0.005	−0.008	0.812	0.025	−0.006
−0.058	0.025	0.012	0.046	0.950	0.037
0.092	0.031	−0.024	−0.064	−0.056	0.918
Σ					
0.275	0	0	0	0	
0.138	0.291	0	0	0	
−0.083	−0.024	0.131	0	0	
0.005	0.001	0.010	0.132	0	
0.025	−0.015	−0.001	0.006	0.126	
σ_e	LLK				Φ ev.
0.062	21,336.5				0.970

Unspanned model: USM(3, 2)

k_∞^Q	λ^Q				
0.033	0.997	0.960	0.870		
μ	Φ				
0.127	0.965	0.005	−0.074	0.033	0.028
0.042	−0.059	0.951	0.166	0.095	−0.025
0.107	−0.005	−0.008	0.812	0.025	−0.006
−0.058	0.025	0.012	0.046	0.950	0.037
0.092	0.031	−0.024	−0.064	−0.056	0.918
Σ					
0.286	0	0	0	0	
0.161	0.296	0	0	0	
−0.081	−0.022	0.131	0	0	
0.006	0.002	0.010	0.132	0	
0.025	−0.015	−0.001	0.006	0.126	
σ_e	LLK				Φ ev.
0.066	21,209.5				0.970

Appendix C: Spanning Results in Large Samples

Here we calculate the model-implied population moments for spanning regressions to complement the results for small-sample model simulations in Sections 4–6 and to see the possible role of small-sample issues.

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

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

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

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

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

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

where R_{ur}^2 and R_{r}^2 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

$$\tilde{Y}_t = Y_t + e_t = \mathcal{A} + \mathcal{B} Z_t + e_t,$$

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 PCs 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 $P_t = W Y_t$. With this notation, the first type of spanning regression is

$$m_t = \beta_0 + \beta' \tilde{P}_t + u_t.$$

The population R^2 is

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

The relevant population moments are

Table CI. Unspanned macro variation in MTSMs—population moments results

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

		GRO	INF	CPI	UGAP
Data	3 PCs	0.28	0.81	0.81	0.73
	5 PCs	0.38	0.86	0.81	0.75
USM(3, 2)	3 PCs				
	$\sigma = 6bps$	0.28	0.79	0.76	0.71
	$\sigma = 0$	0.29	0.79	0.77	0.71
SM(3, 2)	3 PCs				
	$\sigma = 6bps$	0.28	0.76	0.76	0.72
	$\sigma = 0$	0.29	0.77	0.77	0.72
SM(3, 2)	5 PCs				
	$\sigma = 6bp$	0.47	0.87	0.77	0.74
	$\sigma = 0$	1.00	1.00	1.00	1.00

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

$$\text{Cov}(P_t^\sim, m_t) = WB\text{Cov}(Z_t)\iota_m$$

$$\text{Var}(m_t) = \iota_m'\text{Cov}(Z_t)\iota_m.$$

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

Table CI shows the model-implied R^2 for each of the cases and models that we considered in Section 4.4. The results are very similar to the small-sample results reported there—model-implied unspanned macro variation is clearly not a small-sample phenomenon. First, in the case of three 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 three 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 five PCs and do not have measurement error. In this case, the regressors are $P_t^{\mathcal{N}}$, which spans m_t, hence the R^2 is 1. Finally, even in the case that theoretical spanning holds, small measurement error with $\sigma = 6bp$ generates substantial unspanned macro variation, usually enough to match the values in the real-world data.

C.2 Unspanned Macro Risk

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

Table CII. Unspanned macro risk in MTSMs—population moments

Unspanned macro risk in large samples, measured by the difference in theoretical R^2 for regressions of 1-year excess bond returns on yield PCs ("Y") and on both yield PCs and macro variables ("Y + M"), and by the relative RMSE of return forecasts with and without macro variables. The first two rows show these metrics for the actual historical data for comparison.

		GRO/INF				UGAP/CPI			
		R^2 Y	R^2 Y+M	ΔR^2	RMSE	R^2 Y	R^2 Y+M	ΔR^2	RMSE
Data	3 PCs	0.28	0.47	0.19	0.86	0.28	0.34	0.06	0.96
	5 PCs	0.33	0.49	0.15	0.88	0.33	0.41	0.08	0.94
USM(3, 2)	3 PCs								
	$\sigma = 6\text{bps}$	0.11	0.20	0.09	0.95	0.16	0.18	0.02	0.99
	$\sigma = 0$	0.11	0.20	0.09	0.95	0.16	0.18	0.02	0.99
SM(3, 2)	3 PCs								
	$\sigma = 6\text{bps}$	0.11	0.20	0.09	0.94	0.16	0.18	0.02	0.99
	$\sigma = 0$	0.11	0.21	0.10	0.94	0.16	0.18	0.02	0.99
SM(3, 2)	5 PCs								
	$\sigma = 6\text{bps}$	0.17	0.20	0.03	0.98	0.16	0.18	0.02	0.99
	$\sigma = 0$	0.21	0.21	0.00	1.00	0.18	0.18	0.00	1.00

Denote the model-implied excess returns for a bond with maturity n and a holding period of h months by $rx^{(n)}_{t,t+h}$.[45] We first consider the expected excess return, for which we can write

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

This follows easily from the definition of the expected excess return—which is $E_t rx^{(n)}_{t,t+h} = E_t p^{(n-h)}_{t+h} - p^{(n)}_t - y^{(1)}_t$, for $p^{(n)}_t$ the log bond price, $\log P^{(n)}_t$ —the affine formulas for log bond prices (see Appendix A), and the VAR specification for Z_t. The surprise component of the excess return is

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

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

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

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

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

Table CIII. Unspanned macro forecasts in MTSMs—population moments

Unspanned macro forecasts in large samples, measured by the relative 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.

		GRO	INF	CPI	UGAP
Data	3 PCs	0.47	0.30	0.29	0.33
	5 PCs	0.50	0.34	0.29	0.34
USM(3, 2)	3 PCs				
	$\sigma = 6$bps	0.47	0.31	0.30	0.33
	$\sigma = 0$	0.47	0.31	0.30	0.33
SM(3, 2)	3 PCs				
	$\sigma = 6$bps	0.47	0.30	0.30	0.34
	$\sigma = 0$	0.47	0.30	0.30	0.34
SM(3, 2)	5 PCs				
	$\sigma = 6$bps	0.52	0.39	0.30	0.35
	$\sigma = 0$	1.00	1.00	1.00	1.00

Table DI. Variance of model-implied yield PCs

		$\mathrm{Var}(\tilde{P}_t)$	$\mathrm{Var}(P_t)$	$\mathrm{Var}(We_t)$	$\frac{\mathrm{Var}(We_t)}{\mathrm{Var}(\tilde{P}_t)}$
SM(3, 2), GRO/INF					
	PC 1	266.30	266.27	0.03	0.01%
	PC 2	225.95	225.59	0.36	0.16%
	PC 3	8.92	8.56	0.36	4.03%
	PC 4	0.93	0.57	0.36	38.92%
	PC 5	0.36	0.00	0.36	98.96%
SM(3, 2), UGAP/CPI					
	PC 1	251.08	251.05	0.03	0.01%
	PC 2	221.80	221.44	0.36	0.16%
	PC 3	8.81	8.45	0.36	4.09%
	PC 4	0.40	0.04	0.36	90.69%
	PC 5	0.36	0.00	0.36	98.82%

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

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

$$\mathrm{Cov}(\bar{rx}_{t,t+h}, \tilde{P}_t) = \mathrm{Cov}(\bar{rx}_{t,t+h}, P_t) = \bar{\beta}'\mathrm{Cov}(Z_t)(WB)', \ \mathrm{Var}(\bar{rx}_{t,t+h})$$

$$= \bar{\beta}'\mathrm{Cov}(Z_t)\bar{\beta} + \bar{\mathcal{B}}'\mathrm{Cov}(\nu_{t,t+h})\bar{\mathcal{B}}, \ \mathrm{Cov}(\nu_{t,t+h}) = \sum_{i=1}^{h} \varphi^{h-i}\Sigma\Sigma'(\varphi^{h-i})'.$$

The first equality is due to the fact that we focus on model-implied returns. For the unrestricted regression the regressors are $X_t = (\tilde{P}_t', M_t')'$. The population R^2 is

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

and the additional required population moments are

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

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

Table CII shows the population R^2 and the relative RMSEs for the return regressions in the data and in population. The predictability of excess returns is smaller in population than in small samples, as is the additional predictability due to the inclusion of macroeconomic variables. Small-sample issues arise due to the lack of strict exogeneity and the high persistence of the regressors—(Bauer and Hamilton (2016)). Hence, when comparing model implications for unspanned macro risk to real-world data, we need to use short simulated samples, as in Section 5 and Table VII. However, while the absolute magnitude of return predictability is affected by small-sample issues, the qualitative conclusions about unspanned macro risk and unspanned versus spanned models remain unchanged: For three PCs, spanned and unspanned models have the same implications, and measurement error has essentially no effect, while in the case of five PCs for the spanned model measurement error is needed to break theoretical spanning.

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 $\tilde{P}_t$ as predictors. For the restricted regression we can calculate the R^2 using results given above and

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

For the unrestricted regression we also need

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

Note that as for the unspanned macro risk regressions, spanning leads to perfect multicollinearity in the unrestricted regression hence in that case $R_{ur}^2 = R_r^2$.

Table CIII shows the data and model-implied population moments for these predictive regressions for macro variables. 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 three yield PCs in the regressions. The results for the spanned model with five yield PCs again show the importance of measurement error. Notably, even very small measurement errors create a substantial amount of unspanned macro forecasts.

Appendix D: Measurement Error and the Variance of Yield PCs

In Section 4.5 we discuss how the introduction of small measurement error can significantly reduce the information content in yield PCs. Here we provide further details.

The effects of measurement error can be seen most clearly by comparing the variances of the PCs of observed yields and true yields. Note that

$$\mathrm{Cov}(\tilde{P}_t) = \mathrm{Cov}(P_t) + \mathrm{Cov}(We_t) = \mathrm{Cov}(P_t) + \sigma^2\, WW'.$$

Since the rows of W are orthogonal, WW' is diagonal and measurement error affects only the variances but not the covariances of $\tilde{P}_t$. Table DI shows these variances for models SM(3, 2), using $\sigma_e = 6\text{bps}$. The absolute magnitudes of these variances are of little importance, since they affected by the scaling of W, so we report the share of the variance of $\tilde{P}_t$ that is due to measurement error.

For the low-order PCs—the level, slope, and curvature—the variances are little affected by measurement error. However, the fourth and fifth PC are small, relative to the error variance, and hence they get overwhelmed by the measurement errors. For example, about 99% of the variation in the fifth PC is due to measurement error.

How does this matter for the spanning regressions in data? In the spanned model, the higher-order PCs "complete" the spanning in the SM(3, 2) models in the following sense: 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—see the results for three PCs in Tables VI, VII, VIII and in Tables CI–CIII. The theoretical spanning condition implies that spanned models capture the remaining macro information in the higher-order PCs, that is, in the fourth and fifth PCs in SM(3, 2). But these have very different properties from low-order PCs—they are smaller on average and estimated less precisely—so adding measurement error essentially wipes out their information content in the higher-order PCs. Since their information content is crucial, but does not survive the introduction of measurement error, even small measurement errors can lead to a substantial degree of unspanned macro information.