Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Machine-Learning-Based Return Predictors and the Spanning Controversy in Macro-Finance

Jing-Zhi Huang, Zhan Shi

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–25
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Machine-Learning-Based Return Predictors and the Spanning Controversy in Macro-Finance

Jing-Zhi Huang,[a,]* Zhan Shi[b]

[a] Department of Finance, Smeal College of Business, Penn State University, University Park, Pennsylvania 16802; [b] PBC School of Finance, Tsinghua University, Beijing 100083, China
*Corresponding author

Contact: jxh56@psu.edu, https://orcid.org/0000-0002-6453-5827 (J-ZH); shizh@pbcsf.tsinghua.edu.cn, https://orcid.org/0000-0002-9151-372X (ZS)

Received: February 16, 2021
Revised: September 15, 2021
Accepted: November 8, 2021
Published Online in Articles in Advance: March 30, 2022

https://doi.org/10.1287/mnsc.2022.4386

Abstract. We propose a two-step machine learning algorithm—the Supervised Adaptive Group LASSO (SAGLasso) method—that is suitable for constructing parsimonious return predictors from a large set of macro variables. We apply this method to government bonds and a set of 917 macro variables and construct a new, transparent, and easy-to-interpret macro variable with significant out-of-sample predictive power for excess bond returns. This new macro factor, termed the SAGLasso factor, is a *linear* combination of merely 30 selected macro variables out of 917. Furthermore, it can be decomposed into three sublevel factors: a novel *housing* factor, an *employment* factor, and an *inflation* factor. Importantly, the predictive power of the SAGLasso factor is robust to bond yields, namely, the SAGLasso factor is not spanned by bond yields. Moreover, we show that the unspanned variation of the SAGLasso factor cannot be attributed to yield measurement error or macro measurement error. The SAGLasso factor therefore provides a potential resolution to the spanning controversy in the macro-finance literature.

History: Accepted by Haoxiang Zhu, finance.
Funding: This work was supported by a grant from the Penn State Institute for Real Estate Studies.
Supplemental Material: Data and the internet appendices are available at https://doi.org/10.1287/mnsc.2022.4386.

Keywords: machine learning • group lasso • macro-based return predictors • spanning controversy • macro-finance term-structure models

1. Introduction

A growing literature has documented that excess returns of U.S. Treasury bonds are predictable. For instance, the predictors found thus far include forward rates (Cochrane and Piazzesi 2005) and yield-based variables constructed by using filtering (Duffee 2011)[1] as well as macroeconomic variables (e.g., Cooper and Priestley 2009, Ludvigson and Ng 2009). One debate in this literature is whether macroeconomic fundamentals have any such predictive power conditionally over bond yields. Among other things, this debate has important implications for macro-finance term structure models (MTSMs; see, e.g., Joslin et al. 2014)

In this paper, we construct a new macro factor with strong and robust predictive power for bond risk premia using a two-step machine learning algorithm, termed the Supervised Adaptive Group LASSO (SAGLasso) method. We obtain the new macro variable (referred to as the SAGLasso factor) by applying the SAGLasso algorithm to a panel of 131 macro variables (along with six of their lags)—a total of 917 (131 × 7) macro variables—that are adjusted for data revisions and publication lags. In addition to its predictive power, the SAGLasso factor has two other noteworthy features. One is that the factor is parsimonious, transparent, and easy to interpret. The SAGLasso factor is a *linear* combination of merely 30 selected variables out of 917. Furthermore, it can be decomposed into three sublevel factors: a novel *housing* factor, an *employment* factor, and an *inflation* factor—which consist of 13, 11, and 6 macro variables, respectively. The other feature is that the SAGLasso factor is unspanned. Intuitively, this means that the SAGLasso factor is not subsumed (spanned) by yield factors in either predictive regressions or MTSMs. As such, the SAGLasso factor can potentially help resolve the spanning controversy in the macro-finance literature—the debate on whether macro-based return predictors are spanned or not.

We begin our analysis by describing the two-step SAGLasso method, followed by its implementation using the panel of 131 macro series. We construct eight sublevel factors—such as the *housing*, *employment*, and *inflation* factors—in the first step and then the SAGLasso factor in the second step of the procedure. Note that we control for contemporaneous yields in both steps to minimize the information overlap between the SAGLasso factor and the yield curve.

Next, we examine the conditional predictive power of the SAGLasso factor for bond risk premia by testing two hypotheses. The first one, Spanning Hypothesis I, states that macro variables have no incremental predictive power over the current yield curve, the first three principal components (PCs) of yields. The second one, Spanning Hypothesis II—a stronger version of the first hypothesis—posits that macro variables have no incremental predictive power over the filtration generated by the yield curve, proxied by the first five yield PCs filtered from a dynamic term structure model. Our results from both in-sample and out-of-sample tests strongly reject the two spanning hypotheses when the SAGLasso factor is the sole macro variable used. These results indicate that the SAGLasso macro factor has significant incremental predictive power, over price-related information in the Treasury market, for future bond returns. Furthermore, we provide evidence that this predictability can generate significant economic gains for investors.

Lastly, as an important application of the SAGLasso factor, we revisit the spanning controversy. Given that the SAGLasso factor has strong predictive power for bond risk premia yet is weakly correlated with the current yield curve, the new macro factor may shed light on the controversy. To this end, we examine three aspects of the controversy using the Joslin et al. (2014) framework for MTSMs. First, we show that the conditional predictive power of the SAGLasso factor is robust to finite sample tests. Second, we focus on part of the spanning controversy formulated under the MTSM framework and test the macro-unspanning hypothesis (MUH), which says that a given MTSM's macro state variables are not spanned by its yield factors.[2] We find that when an $\mathcal{N}$-factor MTSM with $4 \leq \mathcal{N} \leq 6$ includes the SAGLasso factor as its sole macro factor, our likelihood ratio (LR) tests do not reject the MUH, thereby presenting statistical evidence on the relevance of unspanned MTSMs. Third, we provide confirmative evidence that the temporal variation in the SAGLasso factor is not spanned/explained by the current yield curve. Importantly, this result is robust to measurement errors in yields or in the SAGLasso macro variable itself. Taken together, these findings suggest that the SAGLasso factor provides a potential resolution to the spanning controversy.

To summarize, this study contributes to the macro finance literature in three dimensions. First, it is among the first to introduce a machine learning algorithm suitable for constructing parsimonious return predictors from a large set of macro variables and apply the algorithm to the bond market. Second, using this algorithm we construct a new macro factor with strong out-of-sample conditional predictive power for bond risk premia. Moreover, unlike commonly used macro variables in the literature, the SAGLasso factor is unspanned and has tiny measurement error. Third, due to its unique features, the SAGLasso factor can address those concerns raised in Bauer and Rudebusch (2016), Bauer and Hamilton (2018), and Ghysels et al. (2018) in a unified manner and thus can potentially help resolve the spanning controversy.

While this paper focuses on linear models of predictors, two related studies use nonlinear machine learning models to construct bond return predictors (but do not address the spanning controversy). Huang et al. (2016) find that the macro series selected by SAGLasso is robust to various nonlinear models they consider. Bianchi et al. (2021) study bond risk premia using tree-based algorithms as well as neural networks and find that their superb statistical performance translates into large economic gains. Although these highly nonlinear methods can accommodate more complex data, the SAGLasso method can lead to easier-to-interpret return predictors.[3]

The remainder of the paper is organized as follows. Section 2 states Spanning Hypotheses I and II, followed by Section 3 on the data we use. Section 4 presents the SAGLasso algorithm, constructs the SAGLasso factor, and examines its properties. Section 5 revisits the spanning controversy. Section 6 concludes. Appendix A lists some notation and terms frequently used in the paper.

2. Hypotheses on the Predictive Power of Macro Variables

2.1. Basic Setup

We use continuously compounded annual log returns on an n-year zero-coupon Treasury bond in excess of the annualized yield on a one-year zero-coupon Treasury bond. That is, for $t = 1, \cdots, T$, excess returns $rx^{(12n)}_{t,t+12} = r^{(12n)}_{t,t+12} - y^{(12)}_t = ny^{(12n)}_t - (n-1)y^{(12(n-1))}_{t+12} - y^{(12)}_t$, where $r^{(12n)}_{t,t+12}$ is the one-year log holding-period return on an n-year bond purchased at the end of month t and sold at the end of month $t + 12$ and $y^{(12n)}_t$ is the time-t log yield on the n-year bond.

We consider the following predictive regression that is often used to investigate the role of the macroeconomy in shaping bond risk premia (e.g., Ludvigson and Ng 2009 and Joslin et al. 2014):

$$rx^{(12n)}_{t,t+12} = \alpha + \beta'_Z Z_t + \beta'_F F_t + e_{t+12}, \quad (1)$$

where Z represents yield curve factors that are supposed to summarize yield-based information in the Treasury bond market and F denotes macroeconomic factors. For example, Z can be factors constructed from the current yield curve (e.g., yield spreads used in Campbell and Shiller 1991) or return predictors estimated using historical yields (e.g., the Cochrane-Piazzesi forward rate factor). Similarly, F can be either predetermined macroeconomic measures (e.g., the gross domestic product (GDP) growth and National Association of Purchasing Managers' (NAPM) price index) or factors extracted from a set of macroeconomic series, such as the Ludvigson and Ng (2009) factor and the new macro factors constructed in this study. The remainder of this section focuses on null hypotheses about the predictive power of macro variables and whether they are spanned.

2.2. Spanning Hypotheses

The issue of interest is macro factors' conditional predictive power above and beyond that contained in the yield curve. Empirically, this issue can be examined based on the significance of β_F in Equation (1), for a given Z_t.

It is known that the first three principal components of yields explain all but a negligible fraction of the variation in the term structure (Litterman and Scheinkman 1991). If the current yield curve is supposed to contain almost all the information useful for determining term premia, we arrive at Spanning Hypothesis I (a hypothesis formulated and tested by Joslin et al. 2014 and Bauer and Hamilton 2018):

$$H_0^{S1}: \text{in Equation (1), if } Z_t = PC^o_{1\text{–}3,t}, \text{ then } \beta_F = 0,$$

where $PC^o_{1\text{–}3} = (PC^o_1, PC^o_2, PC^o_3)$, the vector of the first three PCs of the *observed* yield curve.

Interestingly, Duffee (2011) finds that the fourth and fifth PCs are also informative about predicting bond returns. These factors need to be estimated using filtering techniques based on both current and historical yields, however, as the effects of such factors on cross-sectional yields are too small to dominate measurement error in observed yields. Nonetheless, a natural question is whether macro variables contain information about future bond returns that is not captured by the filtration generated by the yield curve process. If the "true" yield curve is Markov, as is commonly assumed in term structuring modeling, this question leads to Spanning Hypothesis II:

$$H_0^{S2}: \text{in Equation (1), if } Z_t = PC_{1\text{–}5,t}, \text{ then } \beta_F = 0,$$

where $PC_{1\text{–}5} = (PC_1, \ldots, PC_5)$, the vector of the first five PCs of the *noise-uncontaminated* yield curve. Given the predictive power of filtered $PC_{4\text{–}5}$, H_0^{S2} provides a stronger test of the conditional predictive power of F_t than does H_0^{S1}.[4] We also consider an alternative version of H_0^{S2} where Z_t is the spanned "cycle" factor of Cieslak and Povala (2015) in Internet Appendix IA.F.

Small-sample distortions may also take place in tests of H_0^{S1} and H_0^{S2}. Bauer and Hamilton (2018) demonstrate that estimates of standard errors in the *t*-test of $\beta_F = 0$ can be biased because PCs (Z_t) are typically persistent and autoregressive with innovation terms that are possibly correlated with e_{t+12}. They propose a bootstrap procedure to account for the size distortion and conclude that much of extant "evidence against the spanning hypothesis is much weaker than it originally appeared" (p. 399). Besides the statistical inference about β_F in Equation (1), Bauer and Hamilton (2018) also study the finite-sample distribution of the increase in R^2 when F_t is added to the regression. They show that serially correlated e_{t+12} due to overlapping observations could substantially inflate the incremental R^2 in small samples, even if F_t provides no help in predicting bond returns. We test H_0^{S1} and H_0^{S2} by conducting an asymptotic inference (Section 4.4.2) as well as an MTSM-based finite-sample inference (Section 5.2).

3. Data

We use monthly data on bond returns and macroeconomic variables over the period January 1964 to December 2014 in our analysis. The start of our sample coincides with that of many other studies that also use the Fama-Bliss yield data set (e.g., Cochrane and Piazzesi 2005, Ludvigson and Ng 2009). We also conduct part of the empirical analysis based on the 1985–2014 subsample because, first, several studies, including Joslin et al. (2014) and Bauer and Rudebusch (2016), focus on post-1984 samples; secondly, some studies argue that the predictive power of macro variables weakens in more recent samples, especially post-1984;[5] and thirdly, the vintage data coverage for many time series starts in the early 1980s.

Bond data used in this study consist of monthly prices for 1- through 5-year zero-coupon Treasury bonds from the Center for Research in Security Prices (CRSP) (Fama Risk Free Rates and Fama-Bliss Discount Bond Yields) for the full sample and self-constructed monthly zero yields with maturity beyond 5 but up to 10 years for the post-1984 sample. The latter data set extends the original Fama-Bliss data to longer maturities and is constructed using monthly quotes on individual bonds from the CRSP Master File of Treasury Bonds by following Le and Singleton (2013).[6] Zero yields can then be used to construct annual excess returns as defined in Section 2.1.

Our macro data set consists of a balanced panel of 131 monthly macroeconomic times series and is an updated and "real-time" version of the macro data set used in Stock and Watson (2002, 2005) and Ludvigson and Ng (2009) that includes one more economic series no longer available. The main source of our real-time macro data is the Archival Federal Reserve Economic Data (ALFRED) database at the Federal Reserve Bank of St. Louis, which is a collection of vintage versions of U.S. economic data and contains more monthly sampled series than does the Philadelphia Fed's Real-Time Data Set. Appendix B includes the list of the 131 series in Table B.1 and describes how our macro data are compiled. The 131 series are organized in a hierarchical manner. Such a cluster structure of macro variables turns out to be useful to model selection. To that end, following Ludvigson and Ng (2011), we group the 131 series into eight categories: (i) output (17 series); (ii) labor market (32 series); (iii) housing sector (10 series); (iv) orders and inventories (14 series); (v) money and credit (11 series); (vi) bond and foreign exchange (FX)—interest rates or financial (22 series); (vii) prices or price indices (21 series); and (viii) stock market (4 series). Column (2) of Table B.1 reports the group ID of each series. Section 4.2 shows that some of the eight groups have stronger predictive power than the others.

4. Adaptive-Lasso-Based Model Selection

In this section, we first describe the Supervised Adaptive Group LASSO algorithm. We next use the algorithm to construct a macro factor with low correlations to the yield curve. We then examine the predictive power of this new macro factor for future bond returns as well as economic gains of such bond return predictability.

4.1. Supervised Adaptive Group LASSO

For a $T \times 1$ response vector $\mathbf{y}$, consider the following penalized least squares (PLS) function:

$$f^{\text{PLS}}(\beta) = \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_{i=1}^{N} |\beta_i|, \tag{2}$$

where $\lambda \geq 0$ is a tuning parameter used to penalize the complexity of the model and $\|\cdot\|$ is the ℓ_2-norm, namely, $\|\eta\| := (\eta'\eta)^{1/2}$, $\forall \eta \in \mathbb{R}^T$. The ℓ_1-norm penalty $|\beta_i|$ used here induces sparsity in the solution and defines the "least absolute shrinkage and selection operator" (Tibshirani 1996, p. 267)—this method is usually referred to as *lasso* rather than *LASSO* in the statistics literature. The lasso estimate is given by $\hat{\beta}^{lasso} = \arg\min_{\beta} f^{\text{PLS}}(\beta)$.

If λ is zero, then $\hat{\beta}^{lasso}$ equals the ordinary least squares (OLS) estimate, $\hat{\beta}^{ols}$, provided that the OLS estimation is feasible. Although none of $\hat{\beta}^{ols}$'s components are zero, some components of $\hat{\beta}^{lasso}$ will shrink to zero as λ increases; as a result, the corresponding "useless" explanatory variables will be dropped and the resulting regression model will become more parsimonious.

Lasso has several advantages over OLS. First, by construction, lasso reduces the variance of the predicted value and thus improves the overall (out-of-sample) forecasting performance. Second, OLS is known to have poor finite sample properties when the dimension of parameters to be estimated is comparable with the number of observations. For instance, in our case there are 131 macro series along with six of their lags—917 (131 × 7) macro variables in total—with only 600 observations for each series. Lasso is developed to handle such problems. Third, lasso leads to a much more parsimonious and easier-to-interpret model than OLS. In fact, the parsimonious or sparse feature of lasso distinguishes it from ridge regression, another shrinkage method.

Despite lasso's popularity, one limitation of the method is that lasso estimates can be biased. Zou (2006) shows that this problem can be fixed by using Adaptive Lasso, which minimizes the following objective function:

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_{i=1}^{N} \lambda_i \, |\beta_i|, \tag{3}$$

where different tuning parameters $\{\lambda_i\}$ are introduced to penalize different β_is separately.

We construct a macro-based return predictor in two steps. In the first step, we utilize the cluster structure of our macroeconomic panel and consider variable selection separately within each of the eight groups/clusters formed in Section 3; that is, we screen out less important or irrelevant individual economic series and identify informative ones within each cluster using adaptive lasso. This is done for three reasons. First, even variables within the same group may represent certain quantitative measurements of different economic sectors. For instance, the Industrial Production (IP) Index of Consumer Goods and the IP Index of Materials (in group i) might be connected to bond risk premia in a different manner. Second, we want to select macroeconomic measures that are jointly significantly associated with bond risk premia. Third, adaptive lasso selects only a small number of macro variables within each cluster and thus allows us to construct parsimonious models, including easy-to-interpret group macro factors if necessary.

In the second step, we consider all the groups together, each of which now consists of only those macro variables selected in step one, and then conduct variable selection at the group level. We implement this idea using the group lasso of Yuan and Lin (2006) to deal with situations in which covariates are assumed to be clustered in groups (see Appendix C). That is, we select important clusters using group lasso, thereby identifying influential economic sectors in addition to individual variables selected in the first step.[7]

We refer to this two-step procedure as the Supervised Adaptive Group LASSO algorithm.[8] Its key feature is to consider penalized time-series selection at both the within-cluster level and the cluster level. We construct bond return predictors by applying SAGLasso to a large set of macro series in this study. SAGLasso should also be useful in similar big data applications in finance and economics.

4.2. A Macro-Based Return-Forecasting Factor

This subsection implements the two-step SAGLasso procedure using the average excess return (the bond market return), $arx_{t,t+12} = \frac{1}{(n_b-1)}\sum_{n=2}^{n_b} rx_{t,t+12}^{(n)}$, as the dependent variable, where n_b equals 5 (10) when the full (post-1984) sample is used.

First, we perform model selection in each of the eight groups of macro series separately, using only macro variables within the same group along with their six lagged values. To minimize the information overlap with respect to yield curve factors, we include the first three yield PCs in our variable selection but do not penalize the associated coefficients. Put differently, in the regression framework of Equation (1), Z_t is $PC^o_{1-3,t}$ but β_Z are not penalized; F_t includes contemporaneous and lagged macro variables in a given group and β_F are subject to shrinkage. Therefore, at the intragroup level of group j, we minimize the following objective function:

$$\left\|\mathbf{arx} - \mathbf{Z}\boldsymbol{\beta}^{\langle 1\rangle}_{Z,j} - \mathbf{F}\boldsymbol{\beta}^{\langle 1\rangle}_{F,j}\right\|^2 + \sum_{i=1}^{7N_j} \lambda_i^j \left|\beta^{\langle 1\rangle}_{F,j,i}\right|,$$

where λ_i^j is the tuning parameter; N_j denotes the number of economic series contained in group j; $\beta^{\langle 1\rangle}_{F,j,i}$ is the i-th component of $\boldsymbol{\beta}^{\langle 1\rangle}_{F,j}$; and the superscript "$\langle 1\rangle$" emphasizes that these beta coefficients are obtained in the first step of the SAGLasso procedure.

This first step allows us to screen out a large portion of candidate predictors within each group.[9] In total, only 39 out of 131 series remain and have nonzero coefficients on their contemporaneous and/or lagged values after the adaptive lasso is applied; the number of the selected macro variables is only 58 out of 917 (131×7). Let $\hat{X}_j^{\langle 1\rangle}, j = i, \ldots, viii$, denote the set of macro variables, in group j, that survive from the first stage.

In the second step, we select those relevant $\hat{X}_j^{\langle 1\rangle}$ using group lasso. Yield PCs are included as control variables as before. The results from the group lasso show that the coefficients of groups i, iv, v, vi, and viii are shrunk to exactly zero; particularly, group vi (bond and FX) is not selected as a result of controlling for yield factors. For each of the three selected groups—labor market (group ii), housing (group iii), and price indices (group vii)—the group lasso solution obtained from Equation (C.5) in Appendix C yields its corresponding group macro factor:

$$\hat{g}_{\mathrm{j}} = \hat{X}_{\mathrm{j}}^{\langle 1\rangle}\hat{\beta}_{\mathrm{j}}^{\langle 2\rangle}, \ \mathrm{j} = ii, iii, vii, \tag{4}$$

where j denotes the index of group j whose beta coefficient in step two, $\hat{\beta}_{\mathrm{j}}^{\langle 2\rangle}$, is not zero. For ease of reference, we relabel $\{\hat{g}_{\mathrm{j}}\}$ as $\{\hat{g}_h; h = 1,2,3\}$. They each have a clear economic interpretation by construction and represent the employment, housing, and inflation factors, respectively.

Unlike inflation and employment, which are commonly incorporated in MTSMs and are well motivated by certain equilibrium term structure models (e.g., Wachter 2006), the housing sector has received little attention in the term structure literature. Given that $\hat{g}_2$ is a reflection of the share of aggregate consumption devoted to housing, the link between our housing factor and the term premium may be motivated using the idea of Piazzesi et al. (2007) that the expenditure share on housing can drive the equity risk premium.

Note that each of $\{\hat{g}_{\mathfrak{h}}\}$ is parsimonious: $\hat{g}_1$ includes five series (11 variables); $\hat{g}_2$ eight series (13 variables); and $\hat{g}_3$ six series (6 variables). In total, out of the original 131 series (917 variables), we identify 19 series (30 variables) associated with labor market, housing, and prices that have strongest connection with bond risk premia but the least overlap with yield PCs. Moreover, 21 selected variables (out of 30) are lagged, indicating that many series have a lagged effect on bond risk premia. In particular, certain types of shocks to consumer prices or the labor market seem to require a long lag to manifest their impact on the bond market. The SAGLasso method allows us to select those important lagged variables and capture their lag effect on bond risk premia (e.g., $\hat{g}_3$ includes no current Consumer Price Index (CPI) and Producer Price Index (PPI) variables).

Figure 1 provides a visualization of the selected macro variables. To illustrate the words most relevant to bond return prediction, the word cloud font is drawn proportional to the number of selected macro series (including lagged variables) in which the word appears. The most notable finding is that new housing units started and authorized are highly informative about bond risk premia. In addition to the group level information, the word cloud also reveals the most important subsectors within each selected group. For example, housing market conditions in the western and northeastern states seem to play a more important role than that in the Midwest. Also, commodity price indices appear to be more useful than more general price indices for bond return prediction.

For purposes of forecasting, term structure modeling, and model comparison, we construct a single aggregate macro predictor using the aforementioned three group factors:

$$\hat{G} \equiv \sum_{h=1}^{3} \hat{g}_h. \tag{5}$$

We refer to this predictor as the SAGLasso (single) macro factor hereafter. Note that this factor is a linear combination of only 30 macro variables belonging to merely 19 different series, yet it has strong predictive power for bond risk premia as shown below.

4.3. A Recursively Constructed SAGLasso Factor

The SAGLasso factor constructed in Section 4.2 is based on the full sample and is an unconditional/static factor.

Figure 1. (Color online) Word Cloud from Selected Macroeconomic Series

Notes. This figure reports the list of words constituting the names of macroeconomic series that are selected from the SAGLasso algorithm. Font size of a word is proportional to the frequency with which the word appears in selected macroeconomic variables and their lags.

Below we construct a dynamic SAGLasso factor recursively. To avoid forward-looking bias, we estimate everything using only the information available at the time of the forecast; namely, we recursively re-estimate *both* factors and parameters when the new information becomes available. We denote a recursively constructed factor by a tilde (e.g., $\tilde{G}$) to differentiate it from its unconditional counterpart, denoted by a hat (e.g., $\hat{G}$).

Suppose we want to construct $\tilde{G}$ at month t based on covariate observations from $t-R-12$ to $t-13$ and use the predictor to help forecast annual excess bond returns at time $t+12$, where $R>1$ denotes the number of monthly observations included in the training period. Namely, in month $t=R+12$, we have the following information set of monthly observations available: $\mathcal{F}_R=\{X_{s-12},\{rx^{(12n)}_{s-12,s},\ 2\le n\le 5\},s=13,\ldots,R+12\}$.

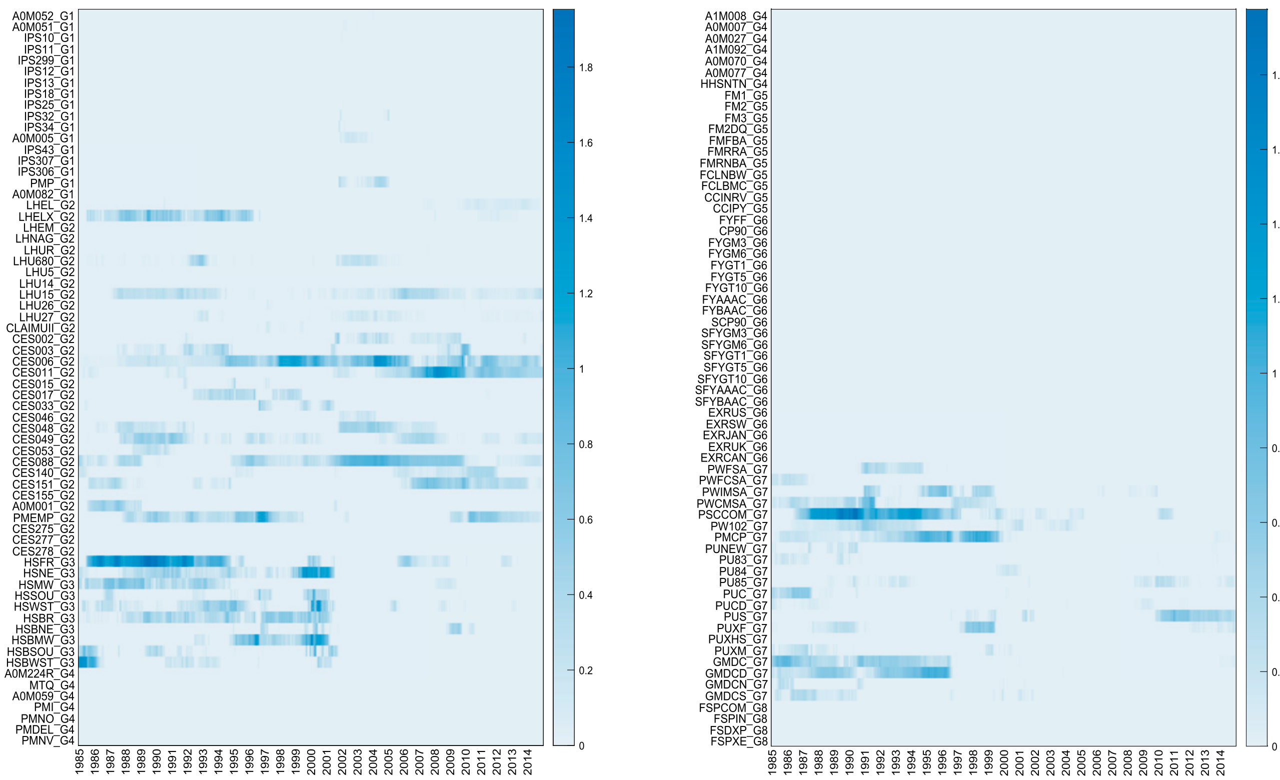
To examine the importance of macro variables over time, we focus on rolling-window estimations.[10] That is, we construct $\tilde{G}$ at, say, $t+1$ using observations from $t-R+1$ to t. We use $R=240$ (a 20-year training period) in this exercise. Figure 2 depicts the importance of individual macro variables over time. From the rolling-window prediction at time t, we extract coefficients of standardized macro variable k and their lagged values $\beta_{k,l,t}(1\le k\le 131, 0\le l\le 6)$, and map their norm $\sqrt{\sum_l\beta^2_{k,l,t}}$ to the color gradients displayed on the right side of the figure. At the group level, the selection results are fairly stable over time: The labor, housing, and inflation groups are selected in most months. The only exception is the 2002–2005 period, during which macro variables in housing and inflation groups diminished in importance and a couple of variables on industrial production are selected instead.[11] At the individual level, the selected macro series are consistent with the results in Figure 1. Within the labor market group, nonfarm payrolls in the manufacturing and financial sectors play crucial roles in bond return predictions. In the inflation group, the commodity price index appears the most prominent determinant of bond risk premiums.

4.4. Predictive Power of the SAGLasso Factor

4.4.1. In-Sample Evidence. Figure 3 plots the SAGLasso factor (in darker line) and excess returns on the five-year bond (in lighter line) in the full sample period, where shaded areas indicate the periods designated by the National Bureau of Economic Research as recession periods. As expected, $\hat{G}$ captures the countercyclic component in risk premia and leads movements in the realized bond returns. Indeed $\hat{G}$ generally starts rising at the early stage of economic downturns and peaks during recessions; accordingly, excess bond returns follow and tend to peak toward the end of (or even after) recessions.

Panel A of Table 1 presents results on the in-sample predictive power of $\hat{G}$, for 2-, 3-, 4-, and 5-year bonds, over the full sample. Test statistics are reported for two different standard errors: Hansen and Hodrick (1980) method of moments (GMM) (in parentheses) and Newey and West (1987) (in brackets).[12] Columns (1)–(4) show that $\hat{G}$ alone has significant predictive power for excess returns, with the R^2 ranging from 0.35 for the two-year bond to 0.39 for the five-year one. Columns (5)–(20) indicate that the significance of $\hat{G}$ is robust to each of the following four factors: (a) a modified Ludvigson and Ng (2009) factor ($\widehat{LN}^m$), (b)

Figure 2. (Color online) Time Variation of Macro Variable Importance



Notes. This figure presents the norm of coefficients associated with the 131 macroeconomic series and their lagged values in the rolling-window bond return prediction. The 131 series are divided into eight groups: (G1) output (17 series); (G2) labor market (32 series); (G3) housing sector (10 series); (G4) orders and inventories (14 series); (G5) money and credit (11 series); (G6) bond and FX—interest rates or financial (22 series); (G7) prices or price indices (21 series); and (G8) stock market (4 series). In each month since January 1984, the macroeconomic panel data over the past 20 years is input into the SAGLasso algorithm to forecast one-year-ahead excess bond returns, and each macroeconomic series could have at most seven nonzero coefficients (on their contemporaneous and lagged values). The x-axis corresponds to the observation date of excess bond returns, and color gradients within each column indicate the most impactful (dark blue) to least important (white) variables.

the Cochrane and Piazzesi (2005) forward-rate factor (CP), the Duffee (2011) hidden factor ($\hat{H}$), and the convergence gap ($\widehat{CG}$) defined by Berardi et al. (2021).[13] The $\hat{G}$ factor, however, does not completely subsume any of these four factors. The main reason is that whereas $\hat{G}$ is a pure macro factor by construction, $\widehat{LN}^m$ includes Treasury and FX variables (group vi), $\widehat{CG}$ exploits information in the Federal Funds rate market, and both $\widehat{CP}$ and $\hat{H}$ are purely yield-curve-based factors. For example, $\hat{G}$ does not subsume $\widehat{CG}$ for the 2-year bond in the bivariate regression. This result is intuitive given that by construction, $\widehat{CG}$ is expected to be most informative about short-term bond premiums, whereas $\hat{G}$ is trained on the aggregate bond market returns rather than a specific-maturity bond. As another example, if yield PCs are not controlled for in the second step of the construction of $\hat{G}$, then the resultant $\hat{G}$ subsumes $\widehat{LN}^m$ (Huang and Shi 2010).

Panel B reports the results for 2-, 5-, 7-, and 10-year bonds for the post-1984 subsample. Although the results on $\hat{G}$ are generally similar to their counterparts in Panel A, the predictive power of the other return predictors all becomes weaker except for $\widehat{CG}$. For instance, $\hat{G}$ now subsumes $\widehat{LN}^m$ under the Hansen and Hodrick (1980) correction, but $\widehat{CG}$ has increased values of both the t-statistics and incremental R^2s.

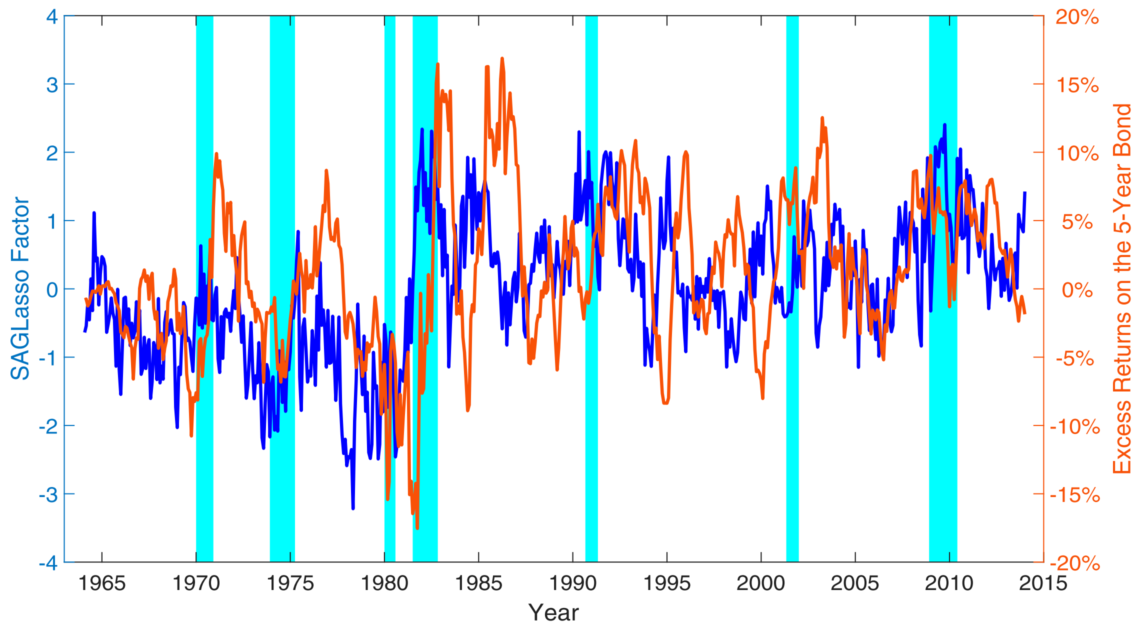
In summary, Table 1 shows that $\hat{G}$ has both significant unconditional and conditional predictive power for bond risk premia. Additionally, $\hat{G}$ subsumes other macro-based predictors post-1984. In Internet Appendix IA.B, we also conduct in-sample spanning tests and find that both H_0^{S1} and H_0^{S2} are overwhelmingly rejected.

4.4.2. Out-of-Sample Accuracy.

We next examine the out-of-sample performance of the SAGLasso factor, focusing on its incremental power above and beyond yield-curve factors.

We divide the sample into training/estimating and out-of-sample (testing) portions. The former consists of $R > 1$ observations. We use fixed rolling-windows with $R = 240$ ($R = 180$) for the full sample (subsample) analysis. If P denotes the number of one-step-ahead predictions, then $T = R + P + 12$, where T is the total number of observations of macro series. We construct

Figure 3. (Color online) The SAGLasso Factor and Excess Returns on the Five-Year Bond



Notes. This figure presents time variation in the normalized SAGLasso factor as well as excess returns on the five-year bond over the sample period from January 1964 to December 2013. Shaded bars denote months designated as recessions by the National Bureau of Economic Research.

$\tilde{G}$ recursively month by month using only information available at the time of estimation as described in Section 4.3. Similarly, we recursively re-estimate the yield-curve factors $PC^o_{1-3,t}$ and $PC_{1-5,t}$, whose dynamic versions are denoted $\widetilde{PC}^o_{1-3,t}$ and $\widetilde{PC}_{1-5,t}$.[14]

Given the dynamic macro and yield-curve factors, we form our out-of-sample tests of H^{S1}_0 as follows: Consider a "restricted" benchmark model and an "unrestricted" model, where the former is the return forecasting model solely based on $\widetilde{PC}^o_{1-3,t}$ and the latter includes $\widetilde{PC}^o_{1-3,t}$ and $\tilde{G}_t$. Given this pair of nested specifications, we can obtain their time series of realized forecast errors over the entire (out-of-sample) testing period and then conduct a model comparison. In other words, the statistical significance of $\tilde{G}$'s incremental predictive power can be assessed by testing the null hypothesis that the restricted model encompasses the unrestricted one. We form tests of H^{S2}_0 similarly by replacing $\widetilde{PC}^o_{1-3,t}$ with $\widetilde{PC}_{1-5,t}$.

Panel A of Table 2 assesses the out-of-sample performance of $\tilde{G}$ with three metrics: the out-of-sample R^2 (Campbell and Thompson 2008) along with its incremental changes due to $\tilde{G}_t$ (R^2_{oos} and ΔR^2_{oos}) and two encompassing tests for nested models—the Ericsson (1992) ENC-REG and Clark and McCracken (2001) ENC-NEW tests.[15] The R^2_{oos} levels of $\tilde{G}_t$ show that $\tilde{G}_t$ alone captures nontrivial real-time information on bond risk premiums. Also, the R^2_{oos} increases with the bond maturity. In fact, the R^2_{oos} for the 2-year bond is substantially lower than that for the 5-year (10-year) bond in the full sample (subsample).[16]

Panel A1 (A2) shows that incorporating $\tilde{G}_t$ into the restricted model based on $\widetilde{PC}^o_{1-3,t}$ ($\widetilde{PC}_{1-5,t}$) and a constant improves the model performance substantially in either the full or sub sample. First, both the ENC-REG and ENC-NEW test statistics greatly exceed their asymptotic critical values, regardless of how the asymptotic ratio of P/R is specified, thereby rejecting both H^{S1}_0 and H^{S2}_0. Second, including $\tilde{G}_t$ also raises R^2_{oos} substantially. For instance, when $\widetilde{PC}^o_{1-3,t}$ is augmented with $\tilde{G}_t$, ΔR^2_{oos} ranges from 0.271 for the five-year bond to 0.349 for the two-year bond in the full-sample analysis. Note that the high values of ΔR^2_{oos} here are partially attributable to the negative R^2_{oos} values under the restricted models. To summarize, Panel A shows that the improvement in forecasting accuracy due to $\tilde{G}$ is statistically significant.

4.4.3. Economic Values.

We now examine economic gains of $\tilde{G}$'s out-of-sample predictive power. We follow Campbell and Thompson (2008) and assess a mean-variance investor's utility gains from trading on $\tilde{G}$ against a benchmark. The investor is assumed to dynamically allocate his or her portfolio between an N-year bond ($N \geq 2$) and a one-year bond (the risk-free asset) at a monthly basis, based on the standard optimal (timing) strategy (e.g., Thornton and Valente 2012). Given his or her risk aversion coefficient (γ) and the N-year bond return volatility at time t, the investor implements the strategy based on his or her out-of-sample forecasts of the N-year bond risk premium.

We consider three return predictors: $\tilde{G}_t$, $\widetilde{PC}^o_{1-3}$, and $\widetilde{PC}^o_{1-3,t} + \tilde{G}_t$. The timing strategies based on these

Table 1. Unconditional and Conditional Predictive Power of the SAGLasso Macro Factor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel A: Full sample, 1964–2014																				
Maturity n (year)	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
$\hat{G}_t$	1.047	1.982	2.810	3.498	0.924	1.671	2.301	2.807	0.889	1.677	2.314	2.929	1.038	1.962	2.783	3.471	0.924	1.836	2.667	3.368
	(5.911)	(6.431)	(6.753)	(7.003)	(4.704)	(4.960)	(4.980)	(5.013)	(5.039)	(5.737)	(6.276)	(6.466)	(6.679)	(7.715)	(7.840)	(7.825)	(4.652)	(5.044)	(5.277)	(5.557)
	[6.427]	[6.954]	[7.323]	[7.607]	[5.206]	[5.483]	[5.542]	[5.595]	[5.477]	[6.172]	[6.736]	[6.989]	[7.277]	[8.357]	[8.550]	[8.546]	[5.152]	[5.589]	[5.866]	[6.184]
$\widehat{LN}_t^m$					0.071	0.179	0.293	0.398												
					(1.451)	(2.076)	(2.609)	(3.026)												
					[1.601]	[2.300]	[2.904]	[3.355]												
$\widehat{CP}_t$									0.235	0.453	0.737	0.845								
									(2.850)	(2.818)	(3.246)	(2.972)								
									[3.187]	[3.186]	[3.681]	[3.361]								
$\hat{H}_t$													−0.234	−0.531	−0.713	−0.713				
													(−2.363)	(−3.255)	(−3.159)	(−2.515)				
													[−2.464]	[−3.336]	[−3.250]	[−2.599]				
$\hat{CG}_t$																	−0.172	−0.203	−0.199	−0.181
																	(−2.681)	(−1.635)	(−1.133)	(−0.861)
																	[−2.835]	[−1.743]	[−1.218]	[−0.921]
R^2	0.352	0.375	0.388	0.392	0.365	0.398	0.420	0.431	0.398	0.425	0.457	0.451	0.396	0.441	0.449	0.432	0.401	0.395	0.398	0.397
Panel B: Subsample, 1985–2014																				
Maturity n (year)	2	5	7	10	2	5	7	10	2	5	7	10	2	5	7	10	2	5	7	10
$\hat{G}_t$	0.729	2.900	4.096	5.225	0.706	2.634	3.740	4.743	0.651	2.611	3.657	4.536	0.735	2.894	4.011	5.150	0.676	2.794	4.006	5.147
	(5.278)	(7.584)	(6.297)	(5.680)	(5.550)	(6.027)	(4.991)	(4.574)	(4.698)	(6.234)	(5.731)	(5.252)	(5.346)	(7.569)	(6.677)	(5.979)	(5.278)	(7.062)	(5.781)	(5.327)
	[5.731]	[7.945]	[6.585]	[5.989]	[5.876]	[6.392]	[5.288]	[4.857]	[5.085]	[6.717]	[6.118]	[5.638]	[5.813]	[7.942]	[6.958]	[6.298]	[5.605]	[7.379]	[6.085]	[5.642]
$\widehat{LN}_t^m$					0.030	0.333	0.447	0.604												
					(0.426)	(1.916)	(1.801)	(1.945)												
					[0.462]	[2.107]	[1.988]	[2.148]												
$\widehat{CP}_t$									0.149	0.551	0.837	1.315								
									(1.405)	(1.738)	(1.993)	(2.708)								
									[1.583]	[1.953]	[2.194]	[2.953]								
$\hat{H}_t$													−0.001	0.002	0.021	0.019				
													(−0.347)	(0.121)	(1.241)	(1.065)				
													[−0.378]	[0.131]	[1.345]	[1.139]				
$\hat{CG}_t$																	−0.340	−0.675	−0.575	−0.500
																	(−3.612)	(−2.548)	(−1.344)	(−0.894)
																	[−3.979]	[−2.749]	[−1.442]	[−0.963]
R^2	0.301	0.397	0.393	0.398	0.303	0.416	0.410	0.416	0.331	0.431	0.432	0.457	0.303	0.397	0.410	0.406	0.413	0.434	0.406	0.404

Notes. The return to an n-year zero-coupon bond from month t to month $t+12$ less the month-t yield on a one-year bond is regressed on the SAGLasso macro factor $\hat{G}_t$ alone (columns (1)–(4)) or $\hat{G}_t$ and another return predictor together. The latter predictor considered includes the modified Ludvigson and Ng (2011) factor $\widehat{LN}_t^m$ (columns (5)–(8)); the Cochrane and Piazzesi (2005) forward-rate factor $\widehat{CP}_t$ (columns (9)–(12)); the hidden component $\hat{H}_t$ in the one-dimensional term premium factor as constructed in Duffee (2011) (columns (13)–(16)); and the Convergence Gap $\hat{CG}_t$ as defined in Berardi et al. (2021) (columns (17)–(20)). Test statistics are computed using the Hansen and Hodrick (1980) GMM covariance estimator (in parentheses), or the Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) covariance estimator (in brackets). Bond maturity n equals 2, 3, 4, and 5 for the full sample spanning the period January 1964–December 2014 (Panel A), and equals 2, 5, 7, and 10 for the post-1984 sample period January 1985–December 2014 (Panel B).

Table 2. Out-of-Sample Performance Assessment

	Full sample, 1964–2014				Subsample, 1985–2014			
Maturity (year)	2	3	4	5	2	5	7	10
Panel A: Statistical significance								
R^2_{oos}	0.123	0.187	0.226	0.246	0.033	0.248	0.236	0.205
Panel A1: $\tilde{G}_t + \widetilde{PC}^o_{1-3,t}$ vs. $\widetilde{PC}^o_{1-3,t}$								
ENC-REG	4.764	4.987	4.831	4.871	3.539	4.570	4.804	5.258
ENC-NEW	191.91	180.91	162.44	147.10	95.33	138.46	128.64	109.49
ΔR^2_{oos}	0.349	0.335	0.296	0.271	0.704	1.029	0.922	0.661
Panel A2: $\tilde{G}_t + \widetilde{PC}_{1-5,t}$ vs. $\widetilde{PC}_{1-5,t}$								
ENC-REG	4.781	5.118	4.823	4.526	3.654	4.831	5.218	4.829
ENC-NEW	180.94	173.49	151.82	130.10	73.93	134.07	130.97	99.17
ΔR^2_{oos}	0.353	0.340	0.292	0.256	0.809	1.026	0.886	0.543
Panel B: Economic significance								
Panel B1: Trading on $\tilde{G}_t$ vs. buy-and-hold								
$\gamma = 3$	0.343	1.267	2.702	4.478	0.308	2.293	4.083	8.745
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
$\gamma = 5$	0.565	2.481	5.289	8.622	0.340	4.053	7.858	16.630
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
Panel B2: Trading on $\widetilde{PC}^o_{1-3,t} + \tilde{G}_t$ vs. Trading on $\widetilde{PC}^o_{1-3,t}$								
$\gamma = 3$	0.432	0.510	0.504	0.449	0.579	1.131	1.133	0.750
	⟨0.008⟩	⟨0.019⟩	⟨0.016⟩	⟨0.012⟩	⟨0.018⟩	⟨0.018⟩	⟨0.009⟩	⟨0.004⟩
$\gamma = 5$	0.292	0.289	0.277	0.239	0.407	0.682	0.685	0.450
	⟨0.022⟩	⟨0.028⟩	⟨0.022⟩	⟨0.019⟩	⟨0.057⟩	⟨0.031⟩	⟨0.009⟩	⟨0.004⟩

Notes. Panel A reports accuracy of out-of-sample forecasts from models with and without the real-time macro factor $\tilde{G}$ as a return predictor. Benchmark predictors considered include the first three principal components of observed yields ($\widetilde{PC}^o_{1-3,t}$) and the first five PCs of the noise-uncontaminated yield curve ($\widetilde{PC}_{1-5,t}$). The rows labeled "ENC-REG" report the out-of-sample t-statistics proposed by Ericsson (1992), and those labeled "ENC-NEW" report a variant of the ENC-REG statistic proposed by Clark and McCracken (2001); both tests share the same null hypothesis that the benchmark model encompasses the unrestricted model with excess predictors. "R^2_{oos}" denotes the out-of-sample R^2 of Campbell and Thompson (2008), and the rows labeled "ΔR^2_{oos}" represent the incremental R^2_{oos} due to $\tilde{G}$. Panel B reports the certainty equivalent gains (in percentage) for a mean-variance investor who selects an N-year bond ($N \geq 2$) along with a one-year bond and who uses portfolios weights potentially depending on $\tilde{G}$-based forecasts. The investor's risk aversion coefficient γ is assumed to be either three or five. The p-values of certainty equivalent returns (in angle brackets) are based on an extended version of Diebold and Mariano (1995) test. All out-of-sample forecasts are formed recursively, with a "training" period of 20 years for the entire sample or that of 15 years in the subsample analysis.

predictors are denoted $\mathcal{S}^G$, $\mathcal{S}^Y$, and $\mathcal{S}^{G+Y}$, respectively. In addition, we consider a buy-and-hold strategy, denoted $\mathcal{S}^{BH}$. We then compare $\mathcal{S}^G$ against $\mathcal{S}^{BH}$, as well as $\mathcal{S}^{G+Y}$ against $\mathcal{S}^Y$, to examine incremental welfare gains due to $\tilde{G}$. Specifically, we calculate the certainty equivalent return (CER) values for each month in the testing sample and then estimate the following regression: $u_{g,t} - u_{0,t} = \nu + \varepsilon_t$, where $u_{g,t}$ and $u_{0,t}$ represent realized utilities generated by strategies $\mathcal{S}^G$ and $\mathcal{S}^{BH}$ or $\mathcal{S}^{G+Y}$ and $\mathcal{S}^Y$, respectively. To examine whether the incremental utility gains are significant or not, we test the null hypothesis that $\nu = 0$ (denoted H_0^ν) using a variant of the Diebold and Mariano (1995) test, proposed by Harvey et al. (1997), that accounts for autocorrelation in the forecasting errors.

Panel B of Table 2 reports the annualized CER values along with the corresponding p-values for H_0^ν (in angel brackets) with $N = 2, 3, 4, 5$ for the full sample or $N = 2, 5, 7, 10$ for the post-1984 subsample. In each panel, we consider two risk version levels: $\gamma = 3$ as adopted by Campbell and Thompson (2008) and Gu et al. (2020) and $\gamma = 5$ as adopted by Thornton and Valente (2012) and Bianchi et al. (2021). We also follow these studies to limit the portfolio weight on the N-year bond to lie between 0% and 150%.

Results for $\mathcal{S}^G$ versus $\mathcal{S}^{BH}$, reported in Panel B1, indicate that the out-of-sample predictive power of $\tilde{G}$ can generate sizable welfare benefits relevant for investors. For example, in the case of $\gamma = 5$ with $n = 5$, $\mathcal{S}^G$ leads to certainty equivalent gains of 8.62% (4.05%) relative to $\mathcal{S}^{BH}$ for the full (post-1984) sample. Campbell and Thompson (2008) show that the investor's welfare gain depends on the relative magnitude of predictive R^2 and the buy-and-hold Sharpe ratio. Because the R^2_{oos} values of $\tilde{G}$ increase with the bond maturity and the Sharpe ratio decreases with the maturity, it is not surprising to find that CER values become greater as the bond maturity increases.

Results for $\mathcal{S}^{G+Y}$ versus $\mathcal{S}^Y$, reported in Panel B2, show that the hypothesis H_0^ν is rejected at the 5% significance level in all but one case (with $n = 2$ and

$\gamma = 5$). In other words, incorporating $\tilde{G}$ into the out-of-sample forecasting of the bond risk premium can lead to significant utility gains relative to trading on $\widetilde{PC}^{o}_{1-3}$ alone. Because these utility differences have the units of expected annualized return, they can be roughly interpreted as the differences in portfolio management fees. We find that a mean-variance investor with $\gamma = 3$ is prepared to pay extra 43–113 basis points (bps) per year to exploit the additional information as contained in factor $\tilde{G}$.

4.4.4. Additional Evidence. We further examine the predictive power of the SAGLasso factor in Internet Appendix IA.B and summarize the main findings here.

Given that $\widehat{LN}^m$ is constructed using the same set of 131 macro series and includes all 131 series as well as squares and cubes of these macro variables, $\widehat{LN}^m$ serves as a natural benchmark for $\hat{G}$ (a linear combination of 19 series and some of their lagged variables). We find that $\hat{G}$ shows stronger predictive ability than $\widehat{LN}^m$ in both in-sample and out-of-sample analyses.

As mentioned before, the set of 131 macro series we use is adjusted for both data revisions and publication lags. One relevant question is the impact of these two adjustments on bond return predictability. We find that the return predictability evidence based on $\hat{G}$ is not sensitive to the vintage of macro data used. In contrast, publication lags pose much greater "danger" than data revisions in forecasting future bond returns based on macro variables, at least in our sample. This problem can be mitigated straightforwardly, however, because it is easier to make an adjustment for publication lags than to figure out preliminary macro data releases and adjust for data revisions.

To better understand the source of the predictive power of the SAGLasso factor ($\hat{G}_t$), we also examine properties of its three components: the employment ($\hat{g}_{1t}$), housing ($\hat{g}_{2t}$), and inflation ($\hat{g}_{3t}$) factors. As expected, $\hat{g}_{1t}, \hat{g}_{2t}$, and $\hat{g}_{3t}$ all have low correlations with the yield curve factors; as a result, $\hat{G}$ is weakly correlated with $PC^{o}_{1-3,t}$ and hardly correlated with $PC_{4,t}$ and $PC_{5,t}$. The three group factors also show significant predictive power, both individually and jointly. Following Joslin et al. (2014), we also examine the relative importance of the three group factors across bond maturity. Our results indicate that relatively speaking, among the three group factors, $\hat{g}_{1t}$ is the most important, followed by $\hat{g}_{3t}$, and then by $\hat{g}_{2t}$, regardless of the bond maturity.[17]

The SAGLasso algorithm is implemented using 131 macro variables along with six of their lags. One question that arises is the following: are lags of macro variables essential to the predictive power of the SAGLasso factor? If yes, what is the optimal number of lags to be included in our supervised learning? We repeat the baseline analysis using the 131 macro variables along with N_L of their lags, where $N_L = 0, 3, 9, 12$. We find that the evidence of the return predictability is robust to the use of no lags ($N_L = 0$). Nonetheless, our results suggest that the SAGLasso factor constructed using the 131 macro variables along with three or six of their lags has the best performance in both the in-sample and out-of-sample predictions. This finding reflects a trade-off between including more information in the supervised learning and imposing a denser data structure to enhance the estimation stability. Although the baseline SAGLasso factor (with $N_L = 6$) seems to capture more information on long-term bond premiums, the alternative SAGLasso factor with $N_L = 3$ outperforms for short-term bonds.

To summarize, Section 4 provides strong evidence against H_0^{S1} and H_0^{S2}. It also shows that rejection of the these two hypothesis carries significant economic values.

5. The SAGLasso Factor and the Spanning Controversy

As an important application of the SAGLasso factor, we revisit the spanning controversy in this section. We focus on the three main aspects of the controversy: first, whether a macro factor's predictive power is robust to finite samples as discussed in Section 2; second, whether a macro factor is an unspanned pricing factor in an MTSM; third, whether a macro factor's temporal variation can be captured by the yield curve. We show that the SAGLasso factor can address all three aspects of the controversy by using the dynamic term-structure modeling framework.

5.1. The Modeling Framework

Following Joslin et al. (2014), we assume that all risks in the economy are encompassed by an $\mathcal{N}$-dimensional state vector $X_t = (\mathcal{P}_t, F_t)$, where $\mathcal{P}_t$ denotes $\mathcal{L}$ linear combinations of (noise-free) zero yields and the ($\mathcal{N}$-$\mathcal{L}$)-vector F_t represents macro factors as before. The short rate is an affine function of X_t:

$$r_t = \delta_0 + \delta_1' X_t = \delta_0 + \delta_{1p}' \mathcal{P}_t + \delta_{1f}' F_t. \tag{6}$$

The dynamics of X_t under the risk-neutral measure $\mathbb{Q}$ are assumed to follow a Gaussian process:

$$X_t = \begin{bmatrix} \mathcal{P}_t \\ F_t \end{bmatrix} = \begin{bmatrix} \mu_p^{\mathbb{Q}} \\ \mu_f^{\mathbb{Q}} \end{bmatrix} + \begin{bmatrix} \Phi_{pp}^{\mathbb{Q}} & \Phi_{pf}^{\mathbb{Q}} \\ \Phi_{fp}^{\mathbb{Q}} & \Phi_{ff}^{\mathbb{Q}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_{t-1} \\ F_{t-1} \end{bmatrix} + \Sigma_x \epsilon_{x,t}^{\mathbb{Q}},$$
$$\epsilon_t^{\mathbb{Q}} \sim MVN(0, I). \tag{7}$$

It follows from Duffie and Kan (1996) that the yield of an m-period zero-coupon bond is

$$y_t^{(m)} = A_m + B_m' X_t, \tag{8}$$

where the expressions for A_m and B_m are given in Internet Appendix IA.C.1. The market price of risk follows the "essentially affine" structure of Duffee (2002, p. 405):

$$\Sigma\Lambda_t = \mu_x^{\mathbb{P}} - \mu_x^{\mathbb{Q}} + (\Phi^{\mathbb{P}} - \Phi^{\mathbb{Q}})X_t = \lambda_0 + \lambda_1 X_t, \quad (9)$$

where $\{\mu^{\mathbb{P}}, \Phi^{\mathbb{P}}\}$ are the $\mathbb{P}$-measure counterparts of $\{\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}\}$.

5.2. Finite Sample Analysis

The statistical inference done in Section 4.4 is based on asymptotic distributions. We now examine H_0^{S1} and H_0^{S2} using a finite-sample analysis. This is necessary because, first, our dependent variables involve overlapping observations by construction and, secondly, the first and second PCs of yield curves are highly persistent in our sample, with first-order autoregressive coefficients (ACF) of 0.99 and 0.94, respectively (whereas the ACF of the SAGLasso factor is only 0.82). Below we first specify the underlying data-generating processes (DGPs) for H_0^{S1} and H_0^{S2} within the framework described in Section 5.1. We then construct finite-sample distributions of test statistics from return-forecasting regressions and conduct finite-sample inference based on such distributions.

5.2.1. Data-Generating Processes for Null Hypotheses. DGPs under H_0^{S1} or H_0^{S2} impose no restrictions on model parameters and allow them to be estimated freely. That is, as long as the $\mathcal{N} \times \mathcal{N}$ yield loading matrix $\mathcal{B} \equiv (B_{m_1}, \ldots, B_{m_{\mathcal{N}}})'$ is invertible, the fraction of variations in term premia that are associated with macro factors is also attributable to certain linear combinations of these yields. This type of MTSMs are referred to as spanned models and denoted by $SM(\mathcal{L}, \mathcal{N})$. If $\mathcal{B}$ is not invertible, then the model is no longer spanned.

Given that $F_t = G_t$, the DGP for H_0^{S1} is model $SM(2,3)$. To see why, suppose that yield PCs are defined in terms of k zero-coupon bonds with maturities $\mathcal{M} = \{m_1, \ldots, m_k\}$ as follows:

$$PC_{1-\mathcal{N},t} = WY_t^{\mathcal{M}} \equiv W(\mathcal{A}_{\mathcal{M}} + \mathcal{B}'_{\mathcal{M}} X_t), \quad W \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}.$$

Because $SM(2,3)$ is a spanned model, $\text{rank}(\mathcal{B}_{\mathcal{M}}) = \mathcal{N} = 3$. The resultant invertibility of $W\mathcal{B}'_{\mathcal{M}}$ implies

$$E_t\left(rx_{t,t+12}^{(12n)}\right) = \text{constant} + \psi'_{12n,12}(W\mathcal{B}'_{\mathcal{M}})^{-1} PC_{1-3,t}, \quad (10)$$

where $\psi_{m,12} = mB'_m - (m-12)B'_{m-12}(\Phi^{\mathbb{P}})^{12} - 12B_{12}$ for $m > 12$. This result means that G_t has no incremental predictive power for annual excess returns in the presence of $PC_{1-3,t}$, consistent with H_0^{S1}. Similarly, the DGP for H_0^{S2} is model $SM(4,5)$.

At the heart of Equation (10) is the theoretical spanning of G_t by any three zero yields. In other words, as long as $k \geq \mathcal{N}$, the covariance matrix of $Y_t^{\mathcal{M}}$ (stacked bond yields) has a rank of three. However, empirically the sample covariance matrices are nonsingular regardless of the choice of maturities $\mathcal{M}$. The standard interpretation in the literature is that observed yields (denoted $Y_t^{o\mathcal{M}}$) are contaminated by small transitory noise, modeled as idiosyncratic "measurement error" (representing a catch all term for model misspecification and other imperfections) as follows:

$$Y_t^{o\mathcal{M}} = \mathcal{A}_{\mathcal{M}} + \mathcal{B}'_{\mathcal{M}} X_t + \eta_{yt}, \quad \eta_{yt} \sim MVN\left(0, \sigma_{\eta_y}^2 I\right). \quad (11)$$

The presence of η_{yt} is also important in terms of accommodating hidden yield factors in spanned models with $\mathcal{N} > 3$. For instance, consider model $SM(4,5)$, where $PC_{1-5,t}$ fully determine the term premia and absorb the role of G_t. If at least five zero yields (or their linear combinations) are assumed to be measured without error, the full-rank $\mathcal{B}'_{\mathcal{M}}$ indicates that the entire state vector can be perfectly extracted from the five yields. Consequently, H_0^{S2} degenerates into a version of H_0^{S1} that involves more than three yield PCs. Alternatively, if measurement error is ubiquitous, it becomes difficult to extract higher-order PCs, say, $PC_{4,t}$, from the cross section of yields. As such, Equation (11) opens up the possibility that bond risk premia contain a component attributable to higher-order PCs, yet hidden from the observed yield curve, namely, a hidden factor.

5.2.2. Finite-Sample Inference. This subsection reports finite-sample properties of test statistics under H_0^{S1} or H_0^{S2}, whose underlying DGPs are $SM(2,3)$ and $SM(4,5)$, respectively. We estimate these spanned models using the full-sample zero-coupon yields with maturities $\mathcal{M} = \{0.25, 1, 2, 3, 4, 5\}$ to generate samples over the period 1964–2014 or using extended Fama-Bliss zero yield data with $\mathcal{M} = \{0.5, 1, 2, 3, 4, 5, 7, 10\}$ to generate samples for the post-1984 period.

As the inference about H_0^{S2} requires all yields to be measured with errors, we implement the model estimation with maximum likelihood using the Kalman filter. To facilitate the interpretation of the sources of risk compensation, we normalize yield-based state variables $\mathcal{P}_t$ to the first $\mathcal{L}$ PCs of zero yields, namely, $X_t = (PC_{1-\mathcal{L},t}, G_t)$. This rotation also offers OLS-based starting values in the estimation of $\mathbb{P}$-dynamics of X_t. When estimating $\mathbb{Q}$-measure parameters, we rotate X_t to X_t^*, a state vector that satisfies the canonical form of Joslin et al. (2013).[18]

Under each spanning hypothesis, we generate 5,000 artificial data sets from its underlying DGP estimated with the full or post-1984 sample. In the in-sample analysis, we obtain the distributions for two t-statistics (based on HH and NW standard errors, respectively) and R^2.[19] In the out-of-sample analysis, we consider the ENC-REG and ENC-NEW tests and R_{oos}^2.[20] We calculate the 5% critical value and p-value for each set of statistics, the latter being defined as the frequency of bootstrap replications in which the test statistics are at least as large as in the real data.

Panel A of Table 3 reports finite-sample properties of test statistics for the full sample. Note from Panels A1 (in-sample) and A2 (out-of-sample) that small-sample distortions appear more severe under H_0^{S1}. For in-sample t-statistics, the "true" 5% critical value ranges from 3.46–4.47, depending on the bond maturity and standard errors used; for ΔR^2 (the incremental in-sample R^2 due to G_t), the 95th percentile of its small-sample distribution is higher than 9%. However, all of these critical values are substantially lower than actual statistics obtained from our data sample. Similarly, note from Panel A2 that there is strong evidence against H_0^{S1}. In particular, all statistics have bootstrapped p-values less than 1%. Also, the critical value of ΔR^2_{oos} ranges from 11.7% for the five-year bond to 13.6% for the two-year bond. Results reported in Panels A3 (in-sample) and A4 (out-of-sample) of Table 3 illustrate that under H_0^{S2}, small-sample distributions of test statistics show even greater deviations from their asymptotic distributions. For instance, the critical value for the HH t-statistics under H_0^{S2} (Panel A3) is at least 0.8 higher than its counterpart under H_0^{S1} (Panel A1), with the biggest difference of 1.29 (= 4.75 − 3.46) for the five-year bond. For out-of-sample tests, the ENC-REG critical value ranges from 4.02 to 4.36, and the ENC-NEW critical value can be as high as 52.18 in small samples (Panel A4); but the critical values are still not large enough to overturn the asymptotic analysis-based rejection of H_0^{S2} concluded in Section 4.4.2.

We find similar results for the post-1984 sample (Panel B of Table 3), although statistics estimated from the subsample are subject to less severe distortions than those from the full sample. Particularly, the asymptotic analysis-based evidence against H_0^{S1} and H_0^{S2} post 1984 (Panel B of Table 2) is robust to small samples.

Overall, we draw three conclusions from Table 3. First, small-sample bias tends to decrease with the

Table 3. Finite-Sample Properties of Test Statistics under Spanning Hypotheses I and II

	Panel A: Full sample, 1964–2014				Panel B: Subsample, 1985–2014			
Maturity (year)	2	3	4	5	2	5	7	10
	Panel A1: In-sample under H_0^{S1}				Panel B1: In-sample under H_0^{S1}			
HH	4.937	4.896	4.712	4.509	4.080	3.910	3.784	3.594
	⟨0.010⟩	⟨0.003⟩	⟨0.001⟩	⟨0.001⟩	⟨0.003⟩	⟨0.003⟩	⟨0.005⟩	⟨0.005⟩
NW	5.064	5.010	4.839	4.654	3.984	3.867	3.714	3.509
	⟨0.006⟩	⟨0.003⟩	⟨0.001⟩	⟨0.000⟩	⟨0.001⟩	⟨0.000⟩	⟨0.001⟩	⟨0.001⟩
ΔR^2	0.108	0.105	0.099	0.091	0.076	0.080	0.066	0.053
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
	Panel A2: Out-of-sample under H_0^{S1}				Panel B2: Out-of-sample under H_0^{S1}			
ENC-REG	4.285	4.195	4.095	3.940	3.421	3.282	3.158	2.996
	⟨0.026⟩	⟨0.018⟩	⟨0.019⟩	⟨0.015⟩	⟨0.045⟩	⟨0.012⟩	⟨0.008⟩	⟨0.004⟩
ENC-NEW	51.03	50.39	47.716	43.622	18.710	17.392	15.596	13.439
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
ΔR^2_{oos}	0.167	0.163	0.153	0.139	0.147	0.147	0.122	0.095
	⟨0.000⟩	⟨0.001⟩	⟨0.001⟩	⟨0.001⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
	Panel A3: In-sample under H_0^{S2}				Panel B3: In-sample under H_0^{S2}			
HH	5.054	4.987	4.909	4.783	4.190	3.947	3.782	3.727
	⟨0.005⟩	⟨0.002⟩	⟨0.002⟩	⟨0.002⟩	⟨0.004⟩	⟨0.003⟩	⟨0.005⟩	⟨0.007⟩
NW	5.202	5.149	5.045	4.962	4.110	3.851	3.745	3.654
	⟨0.003⟩	⟨0.001⟩	⟨0.001⟩	⟨0.000⟩	⟨0.001⟩	⟨0.000⟩	⟨0.002⟩	⟨0.003⟩
ΔR^2	0.117	0.113	0.109	0.103	0.083	0.078	0.070	0.063
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
	Panel A4: Out-of-sample under H_0^{S2}				Panel B4: Out-of-sample under H_0^{S2}			
ENC-REG	4.326	4.184	4.120	4.046	3.416	3.325	3.229	3.162
	⟨0.027⟩	⟨0.017⟩	⟨0.019⟩	⟨0.024⟩	⟨0.038⟩	⟨0.009⟩	⟨0.006⟩	⟨0.004⟩
ENC-NEW	56.68	54.53	52.38	49.387	18.516	16.281	15.098	14.050
	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
ΔR^2_{oos}	0.185	0.177	0.170	0.160	0.161	0.149	0.131	0.119
	⟨0.002⟩	⟨0.003⟩	⟨0.004⟩	⟨0.007⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩

Notes. This table presents results based on finite-sample distributions of the statistics that are involved in tests of Spanning Hypotheses I and II (H_0^{S1} and H_0^{S2}). Five thousand bootstrapped samples are generated from spanned term structure models, $SM(\mathcal{L},\mathcal{N})$, specified in Section 5.2.1; the length of each bootstrapped sample is set to be consistent with either the entire data sample (Panel A) or the post-1984 data sample (Panel B). Results in Panels A1 through B2 (Panels A3 through B4) are obtained from model $SM(2,3)$ (model $SM(4,5)$) that satisfies H_0^{S1} (H_0^{S2}). Test statistics considered include those computed using the Hansen and Hodrick (1980) GMM covariance estimator (HH) and the Newey and West (1987) HAC covariance estimator (NW) with 18 lags and the out-of-sample ENC-REG test of Ericsson (1992) and ENC-NEW test of Clark and McCracken (2001). For each set of test statistics, the 95th percentile of the bootstrap distribution is reported as the 5% critical value and the p-values (in angle brackets) are the frequency of bootstrap replications in which the test statistics are at least as large as the statistic in the data. The "ΔR^2" and "ΔR^2_{oos}" measures denote the incremental R^2 and out-of-sample R^2 of Campbell and Thompson (2008), respectively.

bond maturity. Second, the asymptotic analysis-based evidence against H_0^{S1} and H_0^{S2} (Table 2 and also Internet Appendix IA.B) is too strong to be overturned. Third, results on descriptive statistics show that none of the 5,000 artificial samples are able to generate a ΔR^2 or ΔR^2_{oos} that exceeds the actual incremental R^2.

We present more robustness analyses in the internet appendix. Section IA.D shows that model $SM(2,3)$ provides a more robust test of H_0^{S1} than does the DGP proposed in Bauer and Hamilton (2018). Section IA.E conducts the Ibragimov and Müller (2010) test of H_0^{S1} and H_0^{S2} that is robust to heteroscedasticity, autocorrelation, and structural breaks and finds that among the five yield factors and the SAGLasso factor, the latter is the only robust bond return predictor. Finally, Section IA.F examines an alternative version of H_0^{S2} where the conditioning variable Z_t is the "cycle" factor of Cieslak and Povala (2015) given that this factor is spanned. We find that this hypothesis is rejected as well.

To summarize, the results from our finite-sample analysis strongly reject the two spanning hypotheses, suggesting that it is very unlikely for a spanned MTSM to account for the additional predictive power of the SAGLasso factor as observed in our sample.

5.3. Testing the Macro-Unspanning Hypothesis

The rejection of the spanning hypotheses with $F_t = \hat{G}_t$ implies that MTSMs incorporating $\hat{G}_t$ may be preferable to "yields-only" term structure models (YTSMs), say, for term premium inference. Then a follow-up question is: Should $\hat{G}_t$ be used as a bond-pricing factor in an MTSM and if yes, is $\hat{G}_t$ a spanned pricing factor? We address this question by formulating and testing the "macro-unspanning hypothesis" (MUH), which intuitively says that in spite of its predictive power for bond risk premia, $\hat{G}_t$ is not a spanned pricing factor.

5.3.1. The Macro-Unspanning Hypothesis. In the MTSM framework described in Section 5.1, the MUH (arising from the conditions specified in Joslin et al. (2014) and Bauer and Rudebusch (2016) for unspanned macro risks) can be stated as follows:

$$H_0^{US}: \quad \delta_{1f} = 0 \quad \text{and} \quad \Phi_{pf}^{\mathbb{Q}} = 0. \tag{12}$$

Under these so-called "knife-edge" restrictions, the short rate depends only on $\mathcal{P}_t$ ($\mathcal{L}$ linear combinations of zero yields) and the $\mathbb{Q}$-dynamics of F_t as represented by $\{\mu_f^{\mathbb{Q}}, \Phi_{fp}^{\mathbb{Q}}, \Phi_{ff}^{\mathbb{Q}}\}$ are not identifiable without information from other asset markets. It follows that only risks of yield PCs are priced in the Treasury market. Namely, the one-period risk premium, $\Sigma\Lambda_t$, given below, is $\mathcal{L}$-dimensional:

$$\Sigma\Lambda_t = \mu_p^{\mathbb{P}} - \mu_p^{\mathbb{Q}} + \left[\Phi_{pp}^{\mathbb{P}} - \Phi_{pp}^{\mathbb{Q}}, \Phi_{pf}^{\mathbb{P}}\right] X_t = \lambda_0 + \lambda_1 X_t. \tag{13}$$

For convenience, such an $\mathcal{N}$-factor MTSM that satisfies H_0^{US} is termed an unspanned model and denoted $USM(\mathcal{L},\mathcal{N})$.

Note that when $\mathcal{L} = 3$, H_0^{US} represents the standard version of the MUH: macro-based forecasts are not spanned by the contemporaneous yield curve (equivalent to the case focused on in the Bauer and Rudebusch (2016) likelihood-ratio tests). When $\mathcal{L} > 3$, H_0^{US} denotes a more general version of the MUH and states that the predictive ability of macro factors is not spanned by the filtration generated by the yield dynamics. We examine both versions of the MUH and thereby estimate both models $SM(\mathcal{L},\mathcal{N})$ and $USM(\mathcal{L},\mathcal{N})$ with $\mathcal{L} = 3,4,5$ in this analysis. To match the data sample used in Joslin et al. (2014) and Bauer and Rudebusch (2016), we estimate each of these six models using zero yields with $\mathcal{M} = \{0.5,1,2,3,4,5,7,10\}$ over the period 1985–2007.

Note also that H_0^{US} is *not* simply the opposite of H_0^{S1} or H_0^{S2}. First, although H_0^{US} concerns whether a given macro factor with some explanatory power for term premia is a pricing factor, H_0^{S1} and H_0^{S2} focus on whether variables outside of the bond market provide additional explanatory power for bond risk premia. Second, term structure modeling implications from the outcome of testing H_0^{S1} or H_0^{S2} are different from those from the outcome of testing H_0^{US}. For instance, suppose $\mathcal{N} = 5$. Rejecting H_0^{US} implies a rejection of model $USM(4,5)$, where the alternative model is $SM(4,5)$; namely, it is $SM(4,5)$ versus $USM(4,5)$. In contrast, rejecting H_0^{S2} implies that $USM(5,6)$ ought to be used to infer the risk premium component in long-term yields; accepting H_0^{S2} means that $SM(4,5)$ (or $YTSM(5)$) should be used; that is, it is $SM(4,5)$ versus $USM(5,6)$.[21]

5.3.2. Statistical Tests of the Macro-Unspanning Hypothesis. We conduct two tests of H_0^{US}. One is a model-based likelihood ratio test. As there is no analytic expression available for the limiting distribution under H_0^{US}, we compute the critical values of the test statistic based on the approximation method used by Bauer and Rudebusch (2016). However, the approximation is done conservatively; as a result, this LR test tends to under-reject H_0^{US}.[22] To circumvent this problem and make a more robust inference, we perform another test of H_0^{US} (a model-free test in the spirit of Bauer and Rudebusch 2016) by directly testing the yield loadings on the SAGLasso factor without imposing no-arbitrage restrictions. Given the assumption that all yields are observed with measurement error, we can focus on the loading matrix $\mathcal{B}'_{\mathcal{M}} = (\mathcal{B}'_{\mathcal{L},p}, \mathcal{B}'_{\mathcal{L},f})$ in Equation (11) in this model-free test. To implement the test, we first estimate Equation (11) with the OLS and then conduct LR tests of $\mathcal{B}_{\mathcal{L},f} = 0$.

Panel A of Table 4 reports the results from both the model-based (column (2)) and model-free (column (3)) tests of H_0^{US}, for $\mathcal{L} = \mathcal{N} - 1 = 3,4,5$. Note from column (2) that the LR statistics are always smaller than the 10% critical values, $\forall \mathcal{L}$. An unreported decomposition of the log-likelihood function reveals that the

Table 4. Statistical Inference About Unspanned Macro Risks

(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Tests of unspanning restrictions			Panel B: Regressions of $\hat{G}_t$ on $PC^o_{1-\mathcal{N},t}$		
$\mathcal{N}$	Model-based	Model-free	Macro M.E.	R^2	AR(1) of residuals
4	28.69	10.05		0.145	
	⟨0.122⟩	⟨0.074⟩	No ($\eta_f = 0$)	[0.593 0.847]	
			Yes ($\eta_f \neq 0$)	[0.587 0.769]	0.667
5	24.29	8.23		0.145	
	⟨0.185⟩	⟨0.083⟩	No ($\eta_f = 0$)	[0.506 0.833]	
			Yes ($\eta_f \neq 0$)	[0.459 0.784]	0.667
6	17.55	6.17		0.146	
	⟨0.287⟩	⟨0.104⟩	No ($\eta_f = 0$)	[0.263 0.651]	
			Yes ($\eta_f \neq 0$)	[0.239 0.630]	0.666

Notes. Panel A reports results from likelihood-ratio tests of the macro-unspanning restrictions (H_0^{US}), given in Equation (12), that are imposed on an $\mathcal{N}$-factor unconstrained macro-finance term structure model. Its underlying state vector is $X_t = (PC_{1-\mathcal{L},t}, \hat{G}_t)$, where $PC_{1-\mathcal{L},t}$ denotes the vector of the first $\mathcal{L}$ principal components of the noise-uncontaminated yield curve and $\hat{G}_t$ represents the SAGLasso macro factor. Model-based test statistics (column (2)) are evaluated against the critical values of a χ^2-distribution with degrees of freedom equal to $(k - \mathcal{N})(\mathcal{N} + 1) - 1$, where k is the number of bonds involved. Model-free test statistics (column (3)) are evaluated based on the $\chi^2(k)$-distribution. The p-values appear in angle brackets immediately beneath. Panel B considers the projection of the SAGLasso macro factor ($\hat{G}_t$) onto the first $\mathcal{N}$ PCs of the yield curve ($PC^o_{1-\mathcal{N},t}$). Column (5) shows regression R^2s along with two sets of 95% confidence intervals based on 5,000 artificial samples simulated from model $CSM(\mathcal{L},\mathcal{N})$ as specified in Section IA.G.1 in the internet appendix (which denotes the $\mathcal{N}$-factor constrained MTSM with a spanned $\hat{G}_t$ and whose state vector $X_t = (PC_{1-\mathcal{L},t}, \hat{G}_t)$). The confidence intervals in brackets beneath are obtained under either the assumption that there is no macro measurement error ($\eta_f = 0$) or that there is macro measurement error ($\eta_f \neq 0$), as indicated in column (4) where η_f denotes macro measurement error ("Macro M.E."). Column (6) reports the first-order serial correlation of residuals.

difference between $SM(\mathcal{L},\mathcal{N})$ and $USM(\mathcal{L},\mathcal{N})$ mainly derives from the $\mathbb{Q}$-likelihood. This result, as documented by Bauer and Rudebusch (2016) for $\mathcal{L} = 3$ with two macro factors, is not surprising as the restrictions in H_0^{US} are not placed on the $\mathbb{P}$-dynamics of $USM(\mathcal{L},\mathcal{N})$. However, our test results show that the improved yield curve fitting of $SM(\mathcal{L},\mathcal{N})$ over $USM(\mathcal{L},\mathcal{N})$ is statistically insignificant, in contrast to Bauer and Rudebusch's (2016) finding. The p-values reported in column (3) indicate that H_0^{US} is not rejected by the model-free test either at the conventional significance level of 5%, $\forall\mathcal{L}$.

Results in Panel A also suggest that the negative effect of excluding $\hat{G}$ from fitting the yield curve becomes weaker when $\mathcal{N}$ increases. This finding is not surprising: Although the higher-order PCs are considered to be unimportant in explaining cross-sectional variations in yields, they help fit the term structure more or less. Thus, when an additional yield factor is included in the model, the already limited role of G_t in the cross section becomes more redundant.

To summarize, when the SAGLasso factor is used as the sole macro factor of an unspanned model, both the model-based and model-free tests fail to reject the MUH. As mentioned before, the main reason for this finding is that in spite of its strong predictive power for excess bond returns, the SAGLasso variable is weakly correlated with yield PCs and is unspanned (see Section 5.4). See Internet Appendix IA.G for more applications of unspanned models.

5.4. Is the SAGLasso Factor Unspanned?

To examine whether the yield curve can explain the temporal variation in the SAGLasso factor, we follow Joslin et al. (2014) and regress G_t on $\mathcal{N}$ observed yield PCs:

$$G_t = \gamma_0 + \gamma_1 \cdot PC^o_{1-\mathcal{N},t} + \varepsilon_t. \tag{14}$$

To see whether the regression R^2 is low enough to invalidate spanned models, we follow Bauer and Rudebusch (2016) and evaluate it against its distribution implied from an $\mathcal{N}$-factor spanned model rather than against unity. To this end, we consider distributions implied by "unconstrained" models as well as "constrained" ones and also allow for macro measurement error, denoted by η_f with a standard deviation of σ_{η_f}. In contrast, Bauer and Rudebusch (2016) focus on unconstrained models with zero η_f. Unconstrained models here refer to MTSMs imposing no constraints on the Sharpe ratio (SR) of bond returns; such models may imply unrealistic SRs, as noted in Duffee (2010) and Joslin et al. (2011). MTSMs with the selected zero restrictions on $\{\lambda_0, \lambda_1\}$ are referred to as constrained models and denoted $CSM(\mathcal{L},\mathcal{N})$ for spanned models and $CUSM(\mathcal{L},\mathcal{N})$ for unspanned models, with $\mathcal{L}$ being the number of yield factors included in the model (see Internet Appendices IA.C and IA.G).

Panel B of Table 4 reports the empirical R^2 value and its 95% confidence interval (in brackets underneath) in column (5), where the interval is based on 5,000 data sets simulated from constrained model $CSM(\mathcal{N}\text{-}1,\mathcal{N})$, estimated with and without macro measurement errors, for $\mathcal{N} = 4, 5, 6$. First, consider the case without macro measurement errors ($\eta_f = 0$), a commonly made assumption in the macro finance literature (see, e.g., Joslin et al. 2014 and Bauer and Rudebusch 2016). The results show that $\forall\mathcal{N}$, the empirical R^2 is around 14.5% and outside of its 95% confidence interval with a *p*-value (defined as the fraction of the simulated samples in which the R^2 is below the value in the actual data) lower than 2.5%. That is, the SAGLasso factor indeed has R^2 values too low to be reconcilable with spanned models. We also evaluate empirical R^2s against their distributions implied from unconstrained models $SM(\cdot)$ and find that the results are similar to those reported in Panel B. Because we assume in our model estimation that bond yields are all measured with error, the aforementioned results provide evidence that yield measurement error does *not* account for the large proportion of unspanned macro variation as observed in the real data in our sample.[23]

Next, we assume that $\eta_f \neq 0$. Intuitively, allowing for macro measurement errors would create a further unspanned variation of G_t and thus make it more likely for spanned models to reproduce documented regression evidence. We re-estimate model $CSM(\mathcal{N}\text{-}1, \mathcal{N})$ assuming $\eta_f \neq 0$ and find that the resulting implied R^2 distributions are barely distinguishable from their counterparts with zero η_f. For example, the 95% confidence intervals implied from model $CSM(3,4)$ with and without macro measurement error are $[0.587, 0.769]$ and $[0.593, 0.847]$, respectively (column (5) of Table 4). As a result, even if including η_f shifts the model-implied R^2 distribution to the left, the net impact is minimal; that is, unspanned macro variation observed in our sample cannot be attributed to macro measurement errors either. Behind this finding is the tiny standard deviation of the measurement error in $\hat{G}_t$: $\hat{\sigma}_{\eta_f} < 3$ bps for $3 \leq \mathcal{N} \leq 6$. Note that as $\hat{G}_t$ is standardized under the SAGLasso procedures (Section 4.2), $\hat{\sigma}_{\eta_f}$ is negligible compared with the total standard deviation of $\hat{G}_t$.

Panel B of Table 4 also includes the results from a spanning test applicable to macro factors allowed to contain "noise" (Duffee 2013): if yields span the true state vector, the regression in Equation (14) should produce serially uncorrelated residuals even though the estimated R^2 could substantially deviate from one. The estimated first-order correlation of residuals of the regression is around 0.67, $4 \leq \mathcal{N} \leq 6$ (column (6)). Given that the serial correlation of G_t is 0.71, the above result suggests that whatever the regression is missing cannot be explained by white-noise shocks.

Overall, the results of Section 5.4 provide strong evidence that much of the variation in G_t is not captured by the yield curve. This unspanned nature of the SAGLasso factor reinforces our earlier conclusion that the factor carries term premium information independent of the yield curve. Moreover, this macro variable has very small measurement error even when it is included as a spanned factor in a low-dimensional MTSM.

6. Conclusion

There is no consensus in the literature on whether macro variables have incremental predictive power for future excess bond returns over contemporaneous bond yields. However, macro variables considered in the empirical literature are typically standard ones, such as measures of real growth and inflation. These variables either show little unconditional predictive power for bond risk premia or are highly correlated with contemporaneous yields and thus have insignificant conditional predictive power. In this study, we construct a new macro variable using the Supervised Adaptive Group LASSO, a machine learning algorithm, from a panel of 917 macro variables (131 macro series along with six of their lags) that are adjusted for both data revisions and publication lags. We show that this new macro variable, termed the SAGLasso (macro) factor, has strong out-of-sample predictive power for bond risk premia conditional on the yield curve. Additionally, this predictability can provide investors with significant economic gains.

Importantly, the SAGLasso factor is parsimonious, intuitive, and easy to interpret. Specifically, it is a *linear* combination of merely 30 selected variables out of 917 and consists of a novel housing factor, an employment factor, and an inflation factor. In addition, in spite of its strong predictive power, the SAGLasso factor has low correlations with contemporaneous yields by construction; thus, it is a "pure" macro-based bond return predictor.

The SAGLasso macro factor also provides a potential resolution to the spanning controversy in the macro-finance literature. First, the SAGLasso factor is not spanned by contemporaneous yields. Second, in an MTSM with the SAGLasso factor as its sole macro factor, the hypothesis that it is unspanned by the yield factors is not rejected. Third, incorporating the unspanned SAGLasso factor into an MTSM with realistic Sharpe ratios has nontrivial economic benefits. Fourth, the importance of the SAGLasso factor cannot be attributed to measurement errors in yields or itself. Furthermore, its measurement error is small.

To summarize, using a machine learning algorithm we are able to construct a new, parsimonious, and easy-to-interpret macro variable with strong and robust predictive power for bond risk premia. In addition, this new macro factor can potentially help resolve the spanning controversy in the macro-finance literature. We use the algorithm to construct macro-based bond return predictors in this study, but SAGLasso should also be useful in similar big data applications in finance and economics. For instance, we may construct a real-time expectation factor using the SAGLasso algorithm and examine if the implied bond risk premia are consistent with those demanded by investors in history (Piazzesi et al. 2015). This would allow us to explore an alternative explanation for the spanning controversy: It is due to the discrepancy between the short-rate expectation of real-time investors and the ex post estimates of an econometrician (Cieslak 2018).[24] We may also expand the macro panel data to incorporate survey forecasts of macro variables, which are shown to provide additional information in term structure modeling (see, e.g., Chernov and Mueller 2012 and Kim and Orphanides 2012). We leave these questions to future research.

Acknowledgments

The authors are grateful to Haoxiang Zhu (finance department editor), an associate editor, and an anonymous referee for their extensive and detailed comments that helped to improve the article substantially. The authors also thank Yakov Amihud, Gurdip Bakshi, Charles Cao, Long Chen, Mike Chernov, Ed Coulson, Heber Farnsworth, Peter Feldhütter (American Finance Association Conference discussant), Laura Field, Itay Goldstein, Olesya Grishchenko, Raymond Kan, Anh Le, Hong Liu, Stefan Nagel, Matt Pritsker (Darla Moore Fixed Income Conference discussant), Marco Rossi, Ilona Shiller (Financial Management Association Conference discussant), Andrea Tamoni, Dan Thornton, Joel Vanden, Hong Yan (Summer Institute of Finance Conference discussant), Weina Zhang (China International Conference in Finance discussant), Wei Zhong, and Hao Zhou; seminar participants at Fordham, National University of Singapore, Penn State Smeal College, Penn State Mathematics Department, Federal Reserve Bank of Philadelphia, Singapore Management University, Federal Reserve Bank of St. Louis, Temple University, University of Rhode Island, University of Waterloo, and University of Wisconsin-Milwaukee; and participants at the 20th Federal Deposit Insurance Corporation Derivatives Securities and Risk Management Conference, the 2010 China International Conference in Finance, the 2010 Financial Management Association, the 2010 Summer Institute of Finance Conference, the 2011 American Finance Association, and the 2012 Darla Moore Fixed Income Conference for valuable comments and suggestions. The authors also thank Aaron Henrichsen and Terrence O'Brien for editorial assistance. This paper is a new incarnation of the previous work circulated under the titles "Determinants of Bond Risk Premia" and "Determinants of Bond Risk Premia: A Machine-Learning-Based Resolution of the Spanning Controversy."

Appendix A. Notation and Frequently Used Terms

Spanning Hypothesis I (H_0^{S1})	Macro variables have no additional predictive power for excess bond returns over the first three principal components of the observed yield curve
Spanning Hypothesis II (H_0^{S2})	Macro variables have no additional predictive power for excess bond returns over the first five PCs of the noise-uncontaminated yield curve
Macro-unspanning hypothesis (H_0^{US})	So-called knife-edge restrictions given in Equation (12) for a macro-finance term-structure model to be unspanned
$\widehat{CG}$	The convergence gap defined by Berardi et al. (2021)
$\widehat{CP}$	The Cochrane and Piazzesi (2005) forward rate factor
$\hat{G}$	The (unconditional) Supervised Adaptive Group LASSO macro factor constructed in this study
$\tilde{G}$	The recursive SAGLasso macro factor constructed in this study
$\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$	(Unconditional) SAGLasso group factors constructed in this study, representing *employment*, *housing*, and *inflation*, respectively
$\tilde{g}_1$, $\tilde{g}_2$, and $\tilde{g}_3$	Recursively constructed $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$
$\hat{H}$	The hidden factor proposed by Duffee (2011)
$\widehat{LN}^m$	A modified Ludvigson and Ng (2009) macro-based return predictor
$PC_{1-3}^o = (PC_1^o, PC_2^o, PC_3^o)$	Vector of the first three PCs of the observed yield curve
$\widetilde{PC}_{1-3}^o$	Recursively constructed PC_{1-3}^o
$PC_{1-5} = (PC_1, \ldots, PC_5)$	Vector of the first five PCs of the noise-uncontaminated yield curve
$\widetilde{PC}_{1-5}$	Recursively constructed PC_{1-5}
$CSM(\mathcal{L},\mathcal{N})$	An $\mathcal{N}$-factor constrained, spanned MTSM—Model $SM(\mathcal{L},\mathcal{N})$ with restrictions on the model-implied Sharpe ratios of bond returns
$CUSM(\mathcal{L},\mathcal{N})$	An $\mathcal{N}$-factor constrained, unspanned MTSM—Model $USM(\mathcal{L},\mathcal{N})$ with restrictions on the model-implied Sharpe ratios of bond returns
$SM(\mathcal{L},\mathcal{N})$	An $\mathcal{N}$-factor spanned model—an $\mathcal{N}$-factor MTSM with $\mathcal{L}$ ($\mathcal{N}-1$) yield factors and one macro factor (the SAGLasso factor G) that does not satisfy the macro-unspanning hypothesis H_0^{US}
$USM(\mathcal{L},\mathcal{N})$	An $\mathcal{N}$-factor unspanned model—an $\mathcal{N}$-factor MTSM with $\mathcal{L}$ ($\mathcal{N}-1$) yield factors and one macro factor (the SAGLasso factor G) that satisfies H_0^{US}
$YTSM(\mathcal{N})$	An $\mathcal{N}$-factor *yields-only* term-structure model

Appendix B. Macroeconomic Series Used in the Analysis

Two sets of 131 macroeconomic series are used in our empirical analysis. The first, the standard one used in the literature, includes revised macroeconomic data. The second set consists of real-time macroeconomic data only—the macro series adjusted for data revisions and publication lags.

Table B.1 lists the 131 macroeconomic series and contains the full name (column (4)) of each series, along with its series number (column (1)), group number (column (2)), mnemonic—the series label used in the source database (column (3)), short name (column (5)), and data transformation flag (column (6)). The transformation *flag* = 1: no transformation applied to the series; *flag* = 2: the first difference applied; *flag* = 3: the second difference;

Table B.1. Macro Data Description

Series No.	Group	Mnemonic	Description	Short name	Tran	$\hat{G}_t$	Lag	Vintage
1	1	a0m052	Personal income (AR, bil. chain 2000 $)	PI	5		1	*
2	1	A0M051	Personal income less transfer payments (AR, bil. chain 2000 $)	PI less transfers	5		1	*
3	4	A0M224R	Real consumption (AC) A0m224/gmdc	Consumption	5		1	*
4	4	A0M057	Manufacturing and trade sales (mil. Chain 1996 $)	M & T sales	5		1	
5	4	A0M059	Sales of retail stores (mil. Chain 2000 $)	Retail sales	5		1	
6	1	IPS10	Industrial production index - total index	IP: total	5		1	*
7	1	IPS11	Industrial production index - products, total	IP: products	5		1	*
8	1	IPS299	Industrial production index - final products	IP: final prod	5		1	*
9	1	IPS12	Industrial production index - consumer goods	IP: cons gds	5		1	*
10	1	IPS13	Industrial production index - durable consumer goods	IP: cons dble	5		1	*
11	1	IPS18	Industrial production index - nondurable consumer goods	iIP:cons nondble	5		1	*
12	1	IPS25	Industrial production index - business equipment	IP:bus eqpt	5		1	*
13	1	IPS32	Industrial production index - materials	IP: matls	5		1	*
14	1	IPS34	Industrial production index - durable goods materials	IP: dble mats	5		1	*
15	1	IPS38	Industrial production index - nondurable goods materials	IP:nondble mats	5		1	*
16	1	IPS43	Industrial production index - manufacturing (SIC)	IP: mfg	5		1	*
17	1	IPS307	Industrial production index - residential utilities	IP: res util	5		1	*
18	1	IPS306	Industrial production index - fuels	IP: fuels	5		1	*
19	1	PMP	Napm production index (PERCENT)	NAPM prodn	1		1	
20	1	A0m082	Capacity utilization (Mfg)	Cap util	2		1	
21	2	LHEL	Index of help-wanted advertising in newspapers (1967 = 100; SA)	Help wanted indx	2		1	
22	2	LHELX	Employment: ratio; help-wanted ads:no. unemployed clf	Help wanted/ emp	2		1	
23	2	LHEM	Civilian labor force: employed, total (thous.,sa)	Emp CPS total	5		1	*
24	2	LHNAG	Civilian labor force: employed, nonagric.industries (thous., sa)	Emp CPS nonag	5		1	*
25	2	LHUR	Unemployment rate: all workers, 16 years & over (%,sa)	U: all	2		1	*
26	2	LHU680	Unemploy.by duration: average(mean)duration in weeks (SA)	U: mean duration	2		1	*
27	2	LHU5	Unemploy.by duration: persons unempl.less than 5 wks (thous., SA)	U < 5 wks	5		1	
28	2	LHU14	Unemploy.by duration: persons unempl.5 to 14 wks (thous., SA)	U 5–14 wks	5		1	
29	2	LHU15	Unemploy.by duration: persons unempl.15 wks + (thous., SA)	U 15+ wks	5		1	
30	2	LHU26	Unemploy.by duration: persons unempl.15 to 26 wks (thous., SA)	U 15-26 wks	5		1	
31	2	LHU27	Unemploy.by duration: persons unempl.27 wks + (thous, SA)	U 27+ wks	5		1	
32	2	A0M005	Average weekly initial claims, unemploy. insurance (thous.)	UI claims	5		1	*
33	2	CES002	Employees on nonfarm payrolls - total private	Emp: total	5		1	*
34	2	CES003	Employees on nonfarm payrolls - goods-producing	Emp: gds prod	5		1	*
35	2	CES006	Employees on nonfarm payrolls - mining	Emp: mining	5		1	*
36	2	CES011	Employees on nonfarm payrolls - construction	Emp: const	5		1	*
37	2	CES015	Employees on nonfarm payrolls - manufacturing	Emp: mfg	5		1	*

Table B.1. (Continued)

Series No.	Group	Mnemonic	Description	Short name	Tran	$\hat{G}_t$	Lag	Vintage
38	2	CES017	Employees on nonfarm payrolls - durable goods	Emp: dble gds	5		1	*
39	2	CES033	Employees on nonfarm payrolls - nondurable goods	Emp: nondbles	5		1	*
40	2	CES046	Employees on nonfarm payrolls - service-providing	Emp: services	5		1	*
41	2	CES048	Employees on nonfarm payrolls - trade, transportation, and utilities	Emp: TTU	5	5, 6	1	*
42	2	CES049	Employees on nonfarm payrolls - wholesale trade	Emp: wholesale	5		1	*
43	2	CES053	Employees on nonfarm payrolls - retail trade	Emp: retail	5		1	*
44	2	CES088	Employees on nonfarm payrolls - financial activities	Emp: FIRE	5	0, 1, 2, 3	1	*
45	2	CES140	Employees on nonfarm payrolls - government	Emp: Govt	5		1	*
46	2	CES151	Average weekly hours of production or nonsupervisory workers on private nonfar	Avg hrs	1	0, 2	1	*
47	2	CES155	Average weekly hours of production or nonsupervisory workers on private nonfar	Overtime: mfg	2		1	*
48	2	aom001	Average weekly hours, mfg. (hours)	Avg hrs: mfg	1	0, 2	1	*
49	2	PMEMP	Napm employment index (PERCENT)	NAPM empl	1	0	1	*
50	3	HSFR	Housing starts:nonfarm(1947-58);total farm & nonfarm(1959-)(thous.,SA	HStarts: Total	5	5	1	*
51	3	HSNE	Housing starts:northeast (thous.u.)S.A.	HStarts: NE	4	2, 6	1	*
52	3	HSMW	Housing starts:midwest(thous.u.)S.A.	HStarts: MW	4	2	1	*
53	3	HSSOU	Housing starts:south (thous.u.)S.A.	HStarts: South	4		1	*
54	3	HSWST	Housing starts:west (thous.u.)S.A.	HStarts: West	4	1, 3	1	*
55	3	HSBR	Housing authorized: total new priv housing units (thous.,SAAR)	BP: total	4	3, 6	1	*
56	3	HSBNE	Houses authorized by build. permits:northeast(thou.u.)S.A	BP: NE	4	0	1	*
57	3	HSBMW	Houses authorized by build. permits:midwest(thou.u.)S.A.	BP: MW	4		1	*
58	3	HSBSOU	Houses authorized by build. permits:south(thou.u.)S.A.	BP: South	4	0,6	1	*
59	3	HSBWST	Houses authorized by build. permits:west(thou.u.)S.A.	BP: West	4	3,6	1	*
60	4	PMI	Purchasing managers' index (SA)	PMI	1		1	
61	4	PMNO	Napm new orders index (PERCENT)	NAPM new ordrs	1		2	
62	4	PMDEL	Napm vendor deliveries index (PERCENT)	NAPM vendor del	1		2	
63	4	PMNV	Napm inventories index (PERCENT)	NAPM Invent	1		2	
64	4	A0M008	Mfrs' new orders, consumer goods and materials (bil. chain 1982 $)	Orders: cons gds	5		2	
65	4	A0M007	Mfrs' new orders, durable goods industries (bil. chain 2000 $)	Orders: dble gds	5		2	
66	4	A0M027	Mfrs' new orders, nondefense capital goods (mil. chain 1982 $)	Orders: cap gds	5		2	
67	4	A1M092	Mfrs' unfilled orders, durable goods indus. (bil. chain 2000 $)	Unf orders: dble	5		1	
68	4	A0M070	Manufacturing and trade inventories (bil. chain 2000 $)	M & T invent	5		2	
69	4	A0M077	Ratio, mfg. and trade inventories to sales (based on chain 2000 $)	M & T invent/ sales	2		2	
70	5	FM1	Money stock: m1(curr,trav.cks,dem dep,other ck'able dep)(BIL$,SA)	M1	6		1	*
71	5	FM2	Money stock:m2(M1+o'nite rps,euro$,g/p&b/d mmmfs&sav&sm time dep(BIL$,	M2	6		1	*
72	5	FM3	Money stock: m3(M2+lg time dep,term rp's&inst only mmmfs)(bil$,SA)	M3	6		1	*
73	5	FM2DQ	Money supply - M2 in 1996 dollars (BCI)	M2 (real)	5		1	*
74	5	FMFBA	Monetary base, adj for reserve requirement changes(MIL$,SA)	MB	6		1	
75	5	FMRRA	Depository inst reserves:total,adj for reserve req chgs(MIL$,SA)	Reserves tot	6		1	
76	5	FMRNBA	Depository inst reserves:nonborrowed,adj res req chgs(MIL$,SA)	Reserves nonbor	6		1	

Table B.1. (Continued)

Series No.	Group	Mnemonic	Description	Short name	Tran	$\hat{G}_t$	Lag	Vintage
77	5	FCLNQ	Commercial & industrial loans oustanding in 1996 dollars (BCI)	C&I loans	6		1	
78	5	FCLBMC	Wkly rp lg com'l banks:net change com'l & indus loans(BIL$,SAAR)	C&I loans	1		1	
79	5	CCINRV	Consumer credit outstanding - nonrevolving(G19)	Cons credit-Nonrevolving	6		1	
80	5	A0M095	Ratio, consumer installment credit to personal income (pct.)	Inst cred/PI	2		1	
81	8	FSPCOM	S&P's common stock price index: composite (1941-43 = 10)	S&P 500	5		0	
82	8	FSPIN	S&P's common stock price index: industrials (1941-43 = 10)	S&P: indust	5		0	
83	8	FSDXP	S&P's composite common stock: dividend yield (% PER ANNUM)	S&P div yield	2		0	
84	8	FSPXE	S&P's composite common stock: price-earnings ratio (%,NSA)	S&P PE ratio	5		0	
85	6	FYFF	Interest rate: federal funds (effective) (% PER ANNUM,NSA)	FedFunds	2		1	
86	6	CP90	Commercial Paper Rate (AC)	Commpaper	2		1	
87	6	FYGM3	Interest rate: u.s.treasury bills,sec mkt,3-mo. (% PER ANN,NSA)	3 mo T-bill	2		1	
88	6	FYGM6	Interest rate: u.s.treasury bills,sec mkt,6-mo. (% PER ANN,NSA)	6 mo T-bill	2		1	
89	6	FYGT1	Interest rate: u.s.treasury const maturities,1-yr. (% PER ANN,NSA)	1 yr T-bond	2		1	
90	6	FYGT5	Interest rate: u.s.treasury const maturities,5-yr. (% PER ANN,NSA)	5 yr T-bond	2		1	
91	6	FYGT10	Interest rate: u.s.treasury const maturities,10-yr. (% PER ANN,NSA)	10 yr T-bond	2		1	
92	6	FYAAAC	Bond yield: moody's aaa corporate (% PER ANNUM)	Aaabond	2		1	
93	6	FYBAAC	Bond yield: moody's baa corporate (% PER ANNUM)	Baa bond	2		1	
94	6	scp90	cp90-fyff	CP-FF spread	1		1	
95	6	sfygm3	fygm3-fyff	3 mo-FF spread	1		1	
96	6	sFYGM6	fygm6-fyff	6 mo-FF spread	1		1	
97	6	sFYGT1	fygt1-fyff	1 yr-FF spread	1		1	
98	6	sFYGT5	fygt5-fyff	5 yr-FFspread	1		1	
99	6	sFYGT10	fygt10-fyff	10 yr-FF spread	1		1	
100	6	sFYAAAC	fyaaac-fyff	Aaa-FF spread	1		1	
101	6	sFYBAAC	fybaac-fyff	Baa-FF spread	1		1	
102	6	EXRUS	United states;effective exchange rate(MERM) (INDEX NO.)	Ex rate: avg	5		2	
103	6	EXRSW	Foreign exchange rate: switzerland (swiss franc per U.S.$)	Ex rate: Switz	5		1	
104	6	EXRJAN	Foreign exchange rate: japan (yen per U.S.$)	Ex rate: Japan	5		1	
105	6	EXRUK	Foreign exchange rate: united kingdom (cents per pound)	Ex rate: UK	5		1	
106	6	EXRCAN	Foreign exchange rate: canada (canadian *perU.S.*)	EX rate: Canada	5		1	
107	7	PWFSA	Producer price index: finished goods (82 = 100, SA)	PPI: fin gds	6		1	*
108	7	PWFCSA	Producer price index:finished consumer goods (82 = 100, SA)	PPI: cons gds	6		1	*
109	7	PWIMSA	Producer price index:intermed mat.supplies & components(82 = 100, SA)	PPI: int matls	6		1	*
110	7	PWCMSA	Producer price index:crude materials (82 = 100, SA)	PPI: crude matls	6		1	*
111	7	PSCCOM	Spot market price index:bls & crb: all commodities(1967 = 100)	Commod: spot price	6	0	1	
112	7	PSM99Q	Index of sensitive materials prices (1990 = 100) (BCI-99A)	Sens matls price	6	0	1	
113	7	PMCP	Napm commodity prices index (PERCENT)	NAPM com price	1	0	1	

Table B.1. (Continued)

Series No.	Group	Mnemonic	Description	Short name	Tran	$\hat{G}_t$	Lag	Vintage
114	7	PUNEW	CPI-U: all items (82-84 = 100, SA)	CPI-U: all	6		1	*
115	7	PU83	CPI-U: apparel & upkeep (82-84 = 100, SA)	CPI-U: apparel	6		1	*
116	7	PU84	CPI-U: transportation (82-84 = 100, SA)	CPI-U: transp	6		1	*
117	7	PU85	CPI-U: medical care (82-84 = 100, SA)	CPI-U: medical	6		1	*
118	7	PUC	CPI-U: commodities (82-84 = 100, SA)	CPI-U: comm.	6		1	*
119	7	PUCD	CPI-U: durables (82-84 = 100, SA)	CPI-U: dbles	6		1	*
120	7	PUS	CPI-U: services (82-84 = 100, SA)	CPI-U: services	6		1	*
121	7	PUXF	CPI-U: all items less food (82-84 = 100, SA)	CPI-U: ex food	6		1	*
122	7	PUXHS	CPI-U: all items less shelter (82-84 = 100, SA)	CPI-U: ex shelter	6	6	1	*
123	7	PUXM	CPI-U: all items less medical care (82-84 = 100, SA)	CPI-U: ex med	6		1	*
124	7	GMDC	PCE,IMPL PR DEFL:PCE (1987 = 100)	PCE defl	6		2	*
125	7	GMDCD	PCE,IMPL PR DEFL:PCE; durables (1987 = 100)	PCE defl: dlbes	6	4	2	*
126	7	GMDCN	PCE,IMPL PR DEFL:PCE; nondurables (1996 = 100)	PCE defl: nondble	6		2	*
127	7	GMDCS	PCE,IMPL PR DEFL:PCE; services (1987 = 100)	PCE defl: services	6	6	2	*
128	2	CES275	Average hourly earnings of production or nonsupervisory workers on private no	AHE: goods	6		1	
129	2	CES277	Average hourly earnings of production or nonsupervisory workers on private no	AHE: const	6		1	
130	2	CES278	Average hourly earnings of production or nonsupervisory workers on private no	AHE: mfg	6		1	
131	4	HHSNTN	U. of mich. index of consumer expectations (BCD-83)	Consumer expect	2		1	

flag = 4: the logarithm; *flag* = 5: the first difference of logarithm; and *flag* = 6: the second difference of logarithm.[25]

We compile our macro data in three steps. First, we match the panel of 131 series with ALFRED and find that 70 of them are included in the latter. For each of the 70 matched series, we collect its latest *nine* real-time observations at the end of each month (we do this because some macro variables need to be transformed to their second-order log-differences). However, vintage versions of these 70 series are not balanced and go back to 1964 for only 25 series. Nonetheless, only 3 out of the 19 macro variables eventually selected by SAGLasso do *not* have their vintage data available going back to January 1985. Therefore, the look-forward biases should have a minimum impact, at least on our results obtained from the post-1984 sample.

Second, for the 45 incomplete series in ALFRED, we fill in their missing observations using data over 1964–2007 provided by Ludvigson and Ng (2011) and our manually updated observations from the Federal Reserve Economic Data and The Conference Board over the post-2007 period. As for the 61 series not included in ALFRED, these variables are presumably not subject to revision.[26] We obtain observations for these 61 series from the aforementioned two sources. We then adjust all these macro variables for their publication lags; that is, for each of these time series, we calculate the integer number of months in the time interval between the end of the period over which it is measured and its release date. As shown later, such adjustments matter in our predictability analysis.

Finally, we investigate the time-series properties of these 131 series and determine transformations needed to stationarize each of these series. Table B.1 provides a complete list of the 131 series and, for each series, its data transforms applied, its publication lag, and the availability of its vintage data.

Column (7) labeled "$\hat{G}_t$" of Table B.1 shows the values of a flag indicating which of the 131 macroeconomic series has a nonzero coefficient for its contemporaneous and/or lagged values (up to six) in the SAGLasso regression. The flag value of "0" corresponds to the contemporaneous variable; the value of "ℓ" denotes lag ℓ (in months), $\ell = 1, \ldots, 6$. For instance, macro series #41 (CES048) in group 2—which measures the employment situation in the industry sector "Trade, Transportation, and Utilities"—is selected by the SAGLasso approach and has two variables (out of seven), the lag-5 and lag-6 values of the series, included in the SAGLasso macro factor $\hat{G}$. In total, 19 out of the 131 series (30 out of the 917 macro variables) enter the $\hat{G}$ factor. Column (8) labeled "Lag" reports each series' publication lag (in months), which is defined as the time between the end of the period over which the series is measured and its first release date. Note that out of the 131 series, the four in group 8 "stock market" (#81 through #84) are the only ones without a publication delay. The last column, labeled "Vintage," indicates which macro series has vintage data available, where an asterisk denotes those series whose real-time series are available and used in our empirical analysis. Note that out of the 19 series included in the $\hat{G}$ factor and two additional series (#42 and #53) included in $\tilde{G}$ (the out-of-sample version of $\hat{G}$), the three commodity price indices (#111 through #113) are the only series that have no vintage data available in the ALFRED database. However, given the nature of these three series, they should not be subject to revision.

Appendix C. Supervised Adaptive Group LASSO Method

We first briefly review the group lasso (Yuan and Lin 2006). We begin with the following model:

$$\mathbf{Y} = \mathbf{X}\beta^{\mathbf{0}} + e, \tag{C.1}$$

where e is assumed to be a T-dimensional vector of independent and identically distributed errors (we will relax this assumption later). The main assumption of the group lasso is that some subvectors of the true coefficients $\beta^{\mathbf{0}}$ are zero. Let $\mathcal{H}$ be an index set representing a class of linear subspaces of $\mathbb{R}^T$, each subspace being spanned by the columns of X_h, where $h \in \mathcal{H}$. We denote by $h \in \mathcal{H}_1$ the unknown index set of non-zero subvectors of $\beta^{\mathbf{0}}$ (i.e., $\mathcal{H}_1 = \{h : \beta^{\mathbf{0}}_{\mathbf{h}} \neq 0\}$). Hence, the group lasso involves identifying $\mathcal{H}_1$ and estimating $\beta^{\mathbf{0}}$.

The method is usually implemented by estimating the following restrictive form:

$$\min_{\beta \in \mathbb{R}^N} \left\{ \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_h \|\beta_h\| \right\}. \tag{C.2}$$

Note that expression (C.2) reduces to the lasso when $|\mathcal{H}| = N$ and each h corresponds to the one-dimensional subspace of $\mathbb{R}^T$ spanned by the corresponding column of the design matrix **X**. In our implementation, we consider the general group lasso and, more specifically, the adaptive group lasso, as follows:

$$\min_{\beta \in \mathbb{R}^N} \left\{ \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_h w_h \|\beta_h\| \right\}. \tag{C.3}$$

Next, we describe the Supervised Adaptive Group LASSO algorithm proposed in Section 4.1. The method consists of two steps.

Step I: For cluster $h \in \mathcal{H}$, compute $\hat{\beta}^h$—the cluster-wise adaptive lasso estimate of β^h, namely,

$$\hat{\beta}^h = \arg\min_{\beta^h} \left\{ \|\mathbf{arx} - \mathbf{X}_h \beta^h\|^2 + \sum_j \lambda_h * \hat{w}_{hj} |\beta^h_j| \right\}, \tag{C.4}$$

where **arx** is a vector of average excess bond returns across maturity and $\hat{w}_{hj}$ the j-th component of $\hat{\mathbf{w}}_h$, the vector of the (adaptive) weights. Zou (2006) recommends using $\hat{\beta}^{OLS}$ to construct $\hat{\mathbf{w}}_h$. As collinearity is a concern in our case, we set $\hat{\mathbf{w}}_h = 1/|\hat{\beta}^{RID}_h|^{\gamma_h}$, where $\hat{\beta}^{RID}_h$ is the best ridge regression fit of **arx** on $\mathbf{X}_h$. That is, for cluster h we only use macroeconomic variables within that cluster to construct predictive models. The optimal pairs of (γ_h, λ_h) are determined using five-fold cross-validations. It is worth noting that tuning parameters λ_h are selected for each cluster separately in order to have different degrees of regularization for different clusters. This flexibility allows us to uncover subtle structures that otherwise will be missed when applying the (adaptive) lasso method to all the series/clusters at the same time.

Note that for each cluster $h \in \mathcal{H}$, the adaptive lasso $\hat{\beta}^h$ has only a small number of nonzero components. Let $\tilde{\beta}^h = \hat{\beta}^h \setminus \mathbf{0}$, the vector of nonzero estimated components of $\hat{\beta}^h$ given by the cluster-wise model (C.4), and denote the corresponding part of $\mathbf{X}_h$ by $\tilde{X}_h$. In our case, a typical cluster size ($\dim(\mathbf{X}_h)$) of 80 variables may reduce to a $\dim(\tilde{X}_h)$ of $8 \sim 10$. Namely, the number of macro variables selected in Step I is significantly smaller than the original number to begin with.

Step II: Construct the joint predictive model under the group lasso constraint as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \|\mathbf{arx} - \tilde{\mathbf{X}}\beta\|^2 + \lambda \sum_{h \in \mathcal{H}} w_h \|\beta_h\| \right\}, \tag{C.5}$$

where $\tilde{\mathbf{X}}$ is formed by concatenating the design matrices $\tilde{X}_h$. The parameter λ is also chosen by five-fold cross-validation. With $\lambda \to \infty$, estimates of some components of $\tilde{\beta}_h$ s can be exactly zero. Following Yuan and Lin (2006), we obtain the solution in Equation (C.5) efficiently by using the modified least angle regression selection algorithm of Efron et al. (2004).

In out-of-sample tests conducted in our analysis, tuning parameters $\{\lambda_h, \lambda\}$ are selected recursively starting from the beginning of the test period using cross-validation as well as information only available at the time of estimation. However, to reduce the bias due to the limited training sample size, we use 10-fold cross-validation for the first five years of the out-of-sample testing period (e.g., the period 1985–1989 for the full sample). After that, we go back to standard five-fold cross-validation to restore the balance between bias and variance. Also, to reduce the computational burden in the finite-sample analysis (Section 5.2.2), we select $\{\lambda_h, \lambda\}$ once for each quarter rather than for each month; that is, $\{\lambda_h, \lambda\}$ selected in January are also used to perform SAGLasso model selection in February and March, until they are reselected in April.

Note that the SAGLasso algorithm differs from the supervised principal component analysis (SPCA)—another two-step supervised learning approach—proposed by Bair et al. (2006) in a biological setting, which has been applied to inflation forecasts in Bai and Ng (2008).[27] For instance, the former takes into account the underlying cluster structure of candidate variables, whereas the SPCA does not consider all the candidates simultaneously. Also, variables selected in the SPCA are the PCs whose economic interpretations may not be obvious even though they may have satisfactory prediction performance. Factors constructed using SAGLasso, however, are easier to interpret.

Group lasso is also applied by Freyberger et al. (2020) to identify firm characteristics in shaping expected equity returns. In their analysis, each group consists of 20 portfolios associated with (a polynomial function of) one characteristic, and model selection is done at the group level only. In our analysis, each group consists of macro variables supposed to capture the same economic concept, and adaptive lasso is used within each group (before model selection at the group level) to further mitigate the curse of dimensionality and boost the out-of-sample performance.

Endnotes

[1] See also Fama and Bliss (1987), Stambaugh (1988), and Campbell and Shiller (1991).

[2] Such models are referred to as unspanned MTSMs. Models with spanned macro risks are called spanned models.

[3] Several other studies focus on the application of machine learning in the other finance markets. Freyberger et al. (2020) use group lasso to study the impact of characteristics on expected stock returns. Gu et al. (2020) compare group lasso with other machine learning

methods in the context of stock return prediction. Bali et al. (2021) and He et al. (2021) apply nonlinear machine learning models to inferring corporate bond risk premiums.

[4] The use of PC_{4-5} rather than PC^o_{4-5} in H_0^{S2} is because the latter's predictive power is weaker (see Internet Appendix IA.A). The version of H_0^{S2} based on PC^o_{4-5} is examined in Joslin et al. (2014), Bauer and Rudebusch (2016), and Bauer and Hamilton (2018).

[5] For instance, Bauer and Hamilton (2018) find that the predictive power of macro variables is substantially weaker in extended samples that include observations in 2010s; Bauer and Hamilton (2018) also question the stability of Ludvigson and Ng's (2009) results for their macro return predictors across different subsample periods, especially over the post-1984 sample. Additionally, Duffee (2013, p. 952) notes that "the predictability associated with Ludvigson and Ng's real activity factor may be sample-specific." Our main results are also robust to a backward sample extension to 1952, the starting year of the original Fama-Bliss data (Internet Appendix IA.B).

[6] Similarly extended Fama-Bliss data are used in Joslin et al. (2014) and Bauer and Rudebusch (2016). An alternative data set used in the literature is constructed by Gürkaynak et al. (2007).

[7] Using high-dimension model selection (e.g., Huang et al. (2015)), Huang et al. (2016) find that the variables selected under the SAGLasso procedure are robust to a variety of nonlinear models. Bianchi et al. (2021) also emphasize that it is important to exploit the cluster structure of the macroeconomic panel and do selection within groups and across groups. As such, different machine learning methods seemingly can capture the "common" cluster structure of the same macro data, at least for the purpose of bond return predictions.

[8] In statistical learning, a problem is considered to be supervised if the goal is to predict the value of an outcome measure based on a variety of input measures. See Appendix C for more details of the SAGLasso procedure.

[9] For instance, consider the largest group, the "labor market," that originally contains 32 series and thus 32×7 (= 224) variables. Column (7) of Table B.1 indicates that only 5 series (out of 32), #41, #44, #46, #48, and #49, are selected and that only 11 out of the original 224 variables are selected, including lag-5 and lag-6 of #41; #44 along with its lag-1, lag-2, and lag-3; #46 along with its lag-2; #49 along with its lag-2; and #49 itself.

[10] See, for example, Lewellen (2015) who uses a 10-year rolling window to form OLS-based forecasts of individual stock returns and finds that the importance of many characteristics diminishes over time. The procedure using an expanding window to construct $\tilde{G}$ has higher stability than that using the rolling window: $\hat{g}_1, \hat{g}_2$ and $\hat{g}_3$ are the only groups selected. At the individual level, variables #42 (belonging to "labor market") and #53 (belonging to "housing sector") are the only new variables selected in certain months (and not included in the unconditional $\hat{G}$ factor). The predictive power of $\tilde{G}$ with the expanding window is closely comparable to that with the rolling-window.

[11] Given that the housing market boom after the early 2000s recession makes the share of housing consumption less of a concern, it is unsurprising that variables in the housing sector become less important in this period. By the same logic, the decline in the importance of inflation indices can be attributable to the stable inflation uncertainty in 2000s (e.g., Wright (2011)).

[12] In an earlier version, we also report the t-statistics with Hodrick (1992) 1B covariance estimator, which is constructed using the approximate method of Wei and Wright (2013). The results for $\hat{G}$ are qualitatively similar, but other return predictors tend to lose their significance with the Hodrick standard errors.

[13] In an untabulated analysis, we also consider the output gap factor (*gap*) of Cooper and Priestley (2009); the new-order factor (*NOS*) of Jones and Tuzel (2013); the Cieslak and Povala (2015) "cycle" factor based on yield curves and inflation; and a realized jump-mean factor constructed by Wright and Zhou (2009) (the latter two for the post-1984 sample only). We find that $\hat{G}$ subsumes *gap* and *NOS* and is not driven out by the other two factors. Chernov and Mueller (2012) uncover a hidden factor that captures inflation expectations as well as bond risk premia; however, this "survey" factor is present only in models estimated with survey-based information.

[14] To reduce the computational burden, we estimate the parameters in model $YTSM(5)$ only once using the full sample and then extract $\widetilde{PC}_{1-5,t}$ using filtering (not smoothing) from the estimated model. That is, $\widetilde{PC}_{1-5,t} = \hat{PC}_{1-5,t}$ in Section 4.4.2. Using $\hat{PC}_{1-5,t}$, however, is biased against the predictive power of $\tilde{G}_t$. Indeed, we find that using $(\widetilde{PC}^o_{1-3,t}, \hat{PC}_{4-5,t})$ instead of $\hat{PC}_{1-5,t}$ slightly strengthens $\tilde{G}_t$'s predictive power (untabulated).

[15] The precise asymptotic distribution of the test statistics in these two tests depends on the asymptotic ratio of P/R, denoted by $\pi \equiv \lim_{P,R\to\inf} P/R$. The Ericsson test critical values from a standard normal distribution are conservative if $\pi > 0$. Given that $\pi \geq 1$, the simulation results of Clark and McCracken (2001) show that the 95% critical value ranges from 1.584–2.685 for testing one additional predictor.

[16] Bianchi et al. (2021) find that the performance of their macro factors is also relatively weak for short-term bonds.

[17] Bianchi et al. (2021) consider more categories and find that variables related to the stock and labor market (the output and income, and orders and inventories) are more important for the short-end (long-end) of the yield curve. Note that the aggregate bond market is used to train the group factors $\{\hat{g}_h\}$ here.

[18] In other words, instead of directly estimating parameters in Equations (6) and (7), we estimate another (and shorter) parameter vector $\Theta_M^{\mathbb{Q}}$ (defined in Internet Appendix IA.C.1) that encompasses all bond pricing information.

[19] We do not consider the t-statistic based on the Hodrick (1992) standard errors here because it tends to under-reject the null. Also, Ang and Bekaert (2007) show that it has desirable small-sample properties.

[20] In our baseline finite-sample inference, there is no distinction between the in-sample factor $\hat{G}_t$ and the real-time factor $\tilde{G}_t$. To make the out-of-sample inference truly out of sample, we perform full-scale simulations in which the time series of 131 individual macro variables are generated together with the $\mathcal{N}$-$\mathcal{L}$ yield factors. In each trial, the SAGLasso estimator is implemented on the generated macro variables to construct macro factors $\hat{G}_t$ and $\tilde{G}_t$. These re-simulated $\hat{G}_t$ s and $\tilde{G}_t$ s are then used to infer the finite-sample distribution of test statistics. This exercise guards against the data mining concerns being translated into the finite-sample analysis. Unreported results indicate that the properties of test statistics under the full-scale simulations are similar to those under our baseline simulations.

[21] As a result, a test of H_0^{US} corresponds to a test of equal forecast accuracy for non-nested models in the regression setting in Equation (1). Suppose that $Z_t = PC_{1-5,t}$ and $F_t = G_t$. The question of interest is whether the additional predictive power of G_t is captured by the six yield factors (i.e., $PC_{1-6,t}$) or any other six linear combinations of "true" yields, similar to an encompassing test for comparing non-nested models: $(PC_{1-5,t}, G_t)$ versus $PC_{1-6,t}$.

[22] As discussed in Bauer and Rudebusch (2016), although H_0^{US} imposes four zero restrictions for the case of $\mathcal{L} = 3$, a comparison of test statistics with the critical values of a $\chi^2(4)$-distribution would be misleading. Under the approximation adopted by Bauer and Rudebusch (2016) (detailed in their Section 3.1), test statistics are evaluated against a χ^2-distribution with $(k - \mathcal{N})(\mathcal{N} + 1) - 1$ degrees of freedom when only one macro variable is used, where k is the number of bonds involved.

[23] Bauer and Rudebusch (2016) consider regressions similar to Equation (14) albeit with *GRO* or *INF* as the dependent variable; their simulation results, based on unconstrained models, indicate that adding small yield measurement error makes spanned models capable of generating the appearance of unspanned macro information in the real data. In an untabulated analysis, we show that the main reason for such simulation results is, however, that when a macro variable with a low correlation to the yield curve is used as a spanned factor, most variation in this macro factor is captured by high-order yield factors *by construction*; as a result, a spanned model with small yield measurement error can reproduce a large amount of unspanned macro variation even if the macro variable under consideration is unspanned.

[24] In an earlier version of this paper (Huang and Shi 2010), we provide evidence consistent with the potential mechanism suggested by Duffee (2011). As noted in Cieslak (2018, p. 3269), these different explanations of the spanning controversy are not, however, mutually exclusive because its resolution "depends on the particular variables that the econometrician assumes a part of his/her information set."

[25] Second-order log-differences are the reason for keeping the latest nine observations at each point of historical time for each of the 70 matched series in ALFRED (see Section 3). To see that, let $x_{s|t}$ denote the value of a particular macro variable collected for calendar month s at the end of month $t \geq s$. Suppose that this variable is released with a one-month lag and needs to be log-differenced twice to attain stationarity. The final data to be included in the SAGLasso procedures would be $\{\Delta^2 \ln x_{t-1|t}, \Delta^2 \ln x_{t-2|t}, \ldots, \Delta^2 \ln x_{t-8|t}\}$, where $\Delta^2 \ln x_{t-1|t} = \ln x_{t-1|t} - 2\ln x_{t-2|t} + \ln x_{t-3|t}$.

[26] This conjecture is partially confirmed by checking observations of these macro series around the end of 2007. The logic is as follows. The Ludvigson and Ng (2009) data set ceases its coverage of macro time series in December 2007. If a specific macroeconomic measure (not included in ALFRED) is subject to data revision, its observations for the last couple of months in their data set are likely from the first (preliminary) and second releases. These observations are then compared with corresponding ones collected in 2015, which are definitely from the third (final) release. We find that they are identical. Regardless, the main findings of this study are not affected by this conjecture. As mentioned earlier, it turns out that among those macro series included in the SAGLasso factor, only three commodity price indices have no vintage data available; these indices should not be subject to revision.

[27] Gibson and Pritsker (2000) use partial least squares to choose risk factors of fixed-income portfolios. Goto and Xu (2015) apply the graphical lasso to portfolio selection.

References

Ang A, Bekaert G (2007) Stock return predictability: Is it there? *Rev. Financial Stud.* 20(3):651–707.

Bai J, Ng S (2008) Forecasting economic time series using targeted predictors. *J. Econometrics* 146(2):304–317.

Bair E, Hastie T, Paul D, Tibshirani R (2006) Prediction by supervised principal components. *J. Amer. Statist. Assoc.* 101(473):119–137.

Bali TG, Goyal A, Huang D, Jiang F, Wen Q (2021) Different strokes: Return predictability across stocks and bonds with machine learning and big data. Working paper, Georgetown University, Washington, DC.

Bauer MD, Hamilton JD (2018) Robust bond risk premia. *Rev. Financial Stud.* 31(2):399–448.

Bauer MD, Rudebusch GD (2016) Resolving the spanning puzzle in macro-finance term structure models. *Rev. Finance* 21(2): 511–553.

Berardi A, Markovich M, Plazzi A, Tamoni A (2021) Mind the (convergence) gap: Bond predictability strikes back! *Management Sci.* 67(12):7888–7911.

Bianchi D, Büchner M, Tamoni A (2021) Bond risk premia with machine learning. *Rev. Financial Stud.* 34(2):1046–1089.

Campbell J, Shiller R (1991) Yield spreads and interest rate movements: A bird's eye view. *Rev. Econom. Stud.* 58(3):495–514.

Campbell J, Thompson S (2008) Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financial Stud.* 21(4):1509–1531.

Chernov M, Mueller P (2012) The term structure of inflation expectations. *J. Financial Econom.* 106(2):367–394.

Cieslak A (2018) Short-rate expectations and unexpected returns in treasury bonds. *Rev. Financial Stud.* 31(9):3265–3306.

Cieslak A, Povala P (2015) Expected returns in treasury bonds. *Rev. Financial Stud.* 28(10):2859–2901.

Clark T, McCracken M (2001) Tests of equal forecast accuracy and encompassing for nested models. *J. Econometrics* 105(1): 85–110.

Cochrane J, Piazzesi M (2005) Bond risk premia. *Amer. Econom. Rev.* 95(1):138–160.

Cooper I, Priestley R (2009) Time-varying risk premiums and the output gap. *Rev. Financial Stud.* 22(7):2801–2833.

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. *J. Bus. Econom. Statist.* 13(3):253–263.

Duffee GR (2002) Term premia and interest rate forecasts in affine models. *J. Finance* 57(1):405–443.

Duffee GR (2010) Sharpe ratios in term structure models. Working paper, Johns Hopkins University, Baltimore.

Duffee GR (2011) Information in (and not in) the term structure. *Rev. Financial Stud.* 24:2895–2934.

Duffee GR (2013) Bond pricing and the macroeconomy. Constantinides GM, Harris M, Stulz RM, eds. *Handbook of the Economics of Finance*, vol. 2B (North Holland, Amsterdam), 907–968.

Duffie D, Kan R (1996) A yield-factor model of interest rates. *Math. Finance* 6:379–406.

Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. *Ann. Statist.* 32(2):407–451.

Ericsson N (1992) Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *J. Policy Model.* 14(4):465–495.

Fama E, Bliss R (1987) The information in long-maturity forward rates. *Amer. Econom. Rev.* 77:680–692.

Freyberger J, Neuhierl A, Weber M (2020) Dissecting characteristics non-parametrically. *Rev. Financial Stud.* 33(5):2326–2377.

Ghysels E, Horan C, Moench E (2018) Forecasting through the rearview mirror: Data revisions and bond return predictability. *Rev. Financial Stud.* 31(2):678–714.

Gibson MS, Pritsker M (2000) Improving grid-based methods for estimating value at risk for fixed income portfolios. *J. Risk* 3(Winter):65–89.

Goto S, Xu Y (2015) Improving mean variance optimization through sparse hedging restrictions. *J. Financial Quant. Anal.* 50(06): 1415–1441.

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *Rev. Financial Stud.* 33(5):2223–2273.

Gürkaynak R, Sack B, Wright JH (2007) The U.S. Treasury yield curve: 1961 to present. *J. Monetary Econom.* 54:2291–2304.

Hansen L, Hodrick R (1980) Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *J. Political Econom.* 88(5):829–853.

Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. *Internat. J. Forecast.* 13(2):281–291.

He X, Feng G, Wang J, Wu C (2021) Predicting individual corporate bond returns. Working paper, City University of Hong Kong, Kowloon.

Hodrick R (1992) Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Rev. Financial Stud.* 5(3):357–386.

Huang J-Z, Shi Z (2010) Determinants of bond risk premia. AFA 2011 Denver Meetings Paper. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.615.8517&rep=rep1&type=pdf.

Huang J-Z, Shi Z, Zhong W (2015) Model selection for high-dimensional problems. Lee C-F, Lee JC, eds. *Handbook of Financial Econometrics and Statistics*, vol. 4 (Springer-Verlag, New York), 2093–2118.

Huang J-Z, Li R, Ni J, Shi Z (2016) Forecasting bond returns using high-dimensional model selection. MFA 2017 Chicago Meetings Paper, Pennsylvania State University, State College.

Ibragimov R, Müller UK (2010) t-Statistic based correlation and heterogeneity robust inference. *J. Bus. Econom. Statist.* 28(4):453–468.

Jones CS, Tuzel S (2013) New orders and asset prices. *Rev. Financial Stud.* 26(1):115–157.

Joslin S, Le A, Singleton KJ (2013) Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. *J. Financial Econom.* 109(3):604–622.

Joslin S, Priebsch M, Singleton KJ (2014) Risk premiums in dynamic term structure models with unspanned macro risks. *J. Finance* 69(3):1197–1233.

Joslin S, Singleton KJ, Zhu H (2011) A new perspective on Gaussian dynamic term structure models. *Rev. Financial Stud.* 24(3):926–970.

Kim D, Orphanides A (2012) Term structure estimation with survey data on interest rate forecasts. *J. Financial Quant. Anal.* 47(1):241–272.

Le A, Singleton K (2013) The structure of risks in equilibrium affine models of bond yields. Working paper, Kenan-Flagler Business School, University of North Carolina, Chapel Hill.

Lewellen J (2015) The cross-section of expected stock returns. *Critical Finance Rev.* 4(1):1–44.

Litterman RB, Scheinkman J (1991) Common factors affecting bond returns. *J. Fixed Income* 1(1):54–61.

Ludvigson S, Ng S (2009) Macro factors in bond risk premia. *Rev. Financial Stud.* 22(12):5027–5067.

Ludvigson S, Ng S (2011) A factor analysis of bond risk premia. Ullah A, Giles DEA, eds. *Handbook of Empirical Economics and Finance* (CRC Press, Boca Raton, FL), 313–372.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

Piazzesi M, Salomao J, Schneider M (2015) Trend and cycle in bond premia. Working paper, Stanford University, Stanford, CA.

Piazzesi M, Schneider M, Tuzel S (2007) Housing, consumption and asset pricing. *J. Financial Econom.* 83(3):531–569.

Stambaugh R (1988) The information in forward rates: Implications for models of the term structure. *J. Financial Econom.* 21(1):41–70.

Stock J, Watson M (2002) Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* 97(460):1167–1179.

Stock J, Watson M (2005) Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467, National Bureau of Economic Research, Cambridge, MA.

Thornton DL, Valente G (2012) Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Rev. Financial Stud.* 25(10):3141–3168.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* 58(1):267–288.

Wachter J (2006) A consumption-based model of the term structure of interest rates. *J. Financial Econom.* 79(2):365–399.

Wei M, Wright JH (2013) Reverse regressions and long-horizon forecasting. *J. Appl. Econometrics* 28(3):353–371.

Wright JH (2011) Term premia and inflation uncertainty: Empirical evidence from an international panel data set. *Amer. Econom. Rev.* 101(4):1514–1534.

Wright JH, Zhou H (2009) Bond risk premia and realized jump risk. *J. Banking Finance* 33(12):2333–2345.

Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. B* 68(1):49–67.

Zou H (2006) The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101(476):1418–1429.