An Affine Term Structure Model with a Large Number of Factors and Machine Learning

Ga-Young Jang
Department of Finance, Hanyang University Business School, 222 Wangsimni-ro,
Seongdong-gu, Seoul, Korea 04763; 1 858 209 3515; kathy.g.jang@gmail.com

Hyoung-Goo Kang (C.A.)
Department of Finance, Hanyang University Business School, 222 Wangsimni-ro,
Seongdong-gu, Seoul, Korea 04763; +82 2 2220 1177; hyoungkang@hanyang.ac.kr

Daejin Kim
Department of Fintech, Sungkyunkwan University, 25-2, Sungkyunkwan-ro,
Jongno-gu, Seoul, Korea 03063; +82 2 760 0847; daejin@skku.edu

Dongjoon Lee
Department of Fintech, Sungkyunkwan University, 25-2, Sungkyunkwan-ro,
Jongno-gu, Seoul, Korea 03063; +82 10 2113 8119; chasae94@gmail.com

Abstract

This study proposes an affine term structure model (ATSM), which incorporates 129 factors with their interactions (five standard yield factors plus 124 macro-financial factors), and implements machine learning with no-arbitrage conditions. First, our empirical model fits yields and predicts future excess returns with fast computation - the larger the number of macro-financial factors, the better the performance. Through Lasso regression combined with principal component analysis, we illustrate how machine learning helps identify 23 macro-financial variables to predict bond return. The results yield specific economic implications such that yield curve dynamics explicitly covary with housing permits, short-term rates, stock prices, labor market, and inflation. Our data augmentation facilitates machine learning and enhances model performance. In sum, our ATSM mitigates the long-standing challenges of affine models: the small sample size, computational complexity to process numerous macro-financial factors, and the use of latent variables that makes economic interpretations difficult.

Keywords

Affine term structure model (ATSM); bond pricing; machine learning; macro-financial factors

JEL classifications

G12 E43 E44

I. Introduction

This paper proposes an affine term structure model (ATSM). Our model presents nearly instant computation, a straightforward empirical setting, enhanced predictability, and the capability to incorporate an *infinite number of factors* hypothetically and machine learning without breaching *economic restrictions*. Through the identification of relevant macro-financial factors and data-augmentation idea, our simple design extends HJM (Heath, Jarrow & Morton, 1992), Fama-MacBeth regression (Fama & MacBeth, 1973) and ACM (Adrian et al., 2013; Jang et al., 2021) while linking the empirical outcomes to specific economic implications.

Bonds are issued in different tranches and maturities. This requires modeling a yield curve under the assumption of no-arbitrage. Litterman and Scheinkman (1991) use a principal component analysis (PCA) and find three principal components (level, slope, and curvature factors) to explain how yield curves change. An ATSM explores yield-curve dynamics and predicts future bond returns with simple linear factors while explicitly imposing no-arbitrage conditions. Thus, ATSMs have become the main technique of bond pricing studies, some of which characterize affine structure with latent factors (Duffie & Kan, 1996), develop a multi-factor model of the term structure of interest-rate swap yields (Duffie & Singleton, 1997), investigate an additional return-forecasting factor as the fourth factor (Cochrane & Piazzesi, 2008), and use a computationally simple regression method (Adrian et al., 2013; Jang et al., 2021).

Nevertheless, due to the economic restrictions and computational challenges, existing ATSM studies seldom (i) extend the number of factors to explore the best specifications or (ii) exploit machine-learning methods to extend the model's use. Besides, (iii) a small sample issue exists to consider more factors.

First, in terms of the number of factors, asset pricing models for equities investigate up to 316 factors (Harvey et al., 2016). For bonds, the most up to date ATSM may be the discovery of the five principal factors through the regression method ("ACM": Adrian et al., 2013). Furthermore, ATSMs before the ACM model primarily use maximum likelihood methods to estimate principal factors (Chen & Scott, 1993), which make computation hard as the number of latent variables increases. The ACM's regression incorporates higher-order factors such as the fourth and fifth principal components without compromising the model's predictive power or increasing computational difficulty. However, the ACM model does not resolve the other two problems, that is, the inability to incorporate machine learning or a number of factors beyond five, as well as to provide economic interpretations on observable macro-financial data. This is a practical challenge too because it is unclear how to incorporate numerous observable macro-financial time series into interest-rate derivatives pricing where no-arbitrage conditions are crucial.

Second, in terms of machine learning, an increasing number of academics and practitioners use it in asset pricing even for bonds (Bianchi et al., 2020) and extensively review its applications (Israel et al., 2020). However, the related literature seldom bases an investigation on important asset pricing assumptions, such as no arbitrage or affine structure. How to constrain machine-learning models with economic theories to construct an ATSM remains an important challenge in the literature, especially for bonds.

Third, the small sample size issue prevents the existing asset pricing literature from expanding its implications. The issue is often highlighted in asset pricing studies in predicting returns and therefore should be carefully addressed (Dai & Singleton, 2002; Nelson and Kim, 1993). In addition to addressing the small sample problems ex-post, such as by running additional out-of-sample tests to check robustness, one can perhaps mitigate this issue by expanding the size of the data using data augmentation.

To fill the gaps in the literature and address these challenges, we investigate a new ATSM, which extends and simplifies the ACM model. Hypothetically, our proposed affine model can incorporate infinite factors without compromising the model's predictability or exploding computational capacity. The ability to exploit numerous factors then enables implementing data-craving machine learning techniques. Our empirical analysis illustrates a useful approach using data-augmentation to expand the underlying data and Lasso regression, which explains why the ACM model can be regarded as a special case of our ATSM.

To do so, we estimate the ATSM with four, five, 129 factors, plus their interactions, including 5 yield factors and 127 macro-financial variables, during the sample period of 1972:8 to 2021:2. The new model generates forward rates (Heath, Jarrow & Morton, 1992), which are compared to observed rates for yield fitting and predicting future bond returns. The empirical results strongly support the model's fitting and predictive power even when compared to the ACM model with four and five factors. For bonds with certain maturities, the prediction improves as the number of factors increases.

In addition, we run Lasso regression in combination with PCA to select a set of the most relevant and important macro-financial variables. This reduces the number of macro-financial variables from 127 to 23, making economic interpretations possible. For example, we find that the most important systematic factors are related to housing markets (e.g., New Housing Permits and Housing Starts), short-term rates (e.g., effective Federal Funds Rate and one-year Treasury rates) (i.e., a level factor), stock markets (e.g., S&P's common stock Index and its price-to-earnings ratio), and inflation (e.g., the Producer Price Index (PPI) and the Consumer Price Index (CPI).

Our methods also feature how to expand the underlying data using linear interpolation. This allows our ATSM to incorporate machine learning techniques. As an example. we perform Lasso regression for the out-of-sample empirical tests, of which the results confirm that the predictive power of the model remains intact in the out-of-sample analysis.

For robustness, we use the ACM model with four or five factors for performance comparison in terms of yield fitting and predicting future bond returns. The outcomes from the comparison strongly support the predictive power of our model with minimal root mean square errors (RMSEs) among other measures. An additional robustness check for performance, we use the COVID-19 pandemic as an out-of-sample period. The results are reported in comparison to the performance of the ACM model with four and five factors and again strongly support the proposed model's performance.

Our contributions are summarized as follows. First, this study extends the bond literature by introducing an innovative model, yet strictly grounded on the traditional assumption of asset pricing. We do so by empirically showing how to combine big data, machine learning and an ATSM.

Second, we develop an ATSM that incorporates a large number of factors including yield and macro-financial factors. Macro-financial variables are essential for predicting bond returns in practice. Prior studies shed light on the significant effect of the changes in macro-financial variables on bond returns (Estrella & Hardouvelis, 1991; Ludvigson & Ng, 2009; Wachter, 2006). However, the prior models' limited capacity to accommodate a large number of factors limits investigating the influence of macro-financial variables to yield curves. Such capability of incorporating extensive macro-financial variables would enrich the empirical analysis if aligned with specific economic implications.

Third, we identify economically significant macro-financial factors to explain the term structure of interest rates out of 127 variables, or 8,778 variables including the interaction variables. In practice, fixed income managers closely follow the changes in the set of macro-financial variables they consider essential and would want to investigate their influence on bond returns. By using latent factors only, one cannot specify what macro-financial variables contribute to the outcomes of their investment; however, using our model, a manager can choose the most relevant and important sets of macro-financial variables with much flexibility and align the outcomes of their investment to specific variables. In addition, prior studies show that a three-factor model explains over 90% of the return variations (Fama & French, 1993). Our study extends the asset pricing literature by creating an explainable affine model using macro-financial variables with enhanced yield fitting.

Fourth, we show how to augment underlying data for empirical tests so that the model can implement machine learning techniques and big data without facing a small sample problem in finance. We specifically propose using linear interpolation on the original data points for data augmentation, which ultimately affects the empirical outcomes. One could use other generative models instead of linear interpolation. Our examples can provide a useful reference for future studies in asset pricing or other areas that suffer from a small sample issue.

Lastly, we achieve all of these while significantly reducing the computational burden throughout the process. We illustrate how we implement the regression method and machine learning techniques in the model to achieve faster yet computationally easier estimation process step-by-step.

The rest of the paper is organized as follows: Section 2 discusses affine models in general and introduces our new model. Section 3 explains the data and empirical methods of this study. Section 4 reports the empirical results. Section 5 concludes.

2. Literature review

Early factor model studies are data-driven to examine the relationship between bond returns and forward or yield spreads (Campbell & Shiller, 1991; Fama & Bliss, 1987). Three principal components, *level*, *slope*, and *curvature*, are found to drive most of the variations in bond yields (Litterman & Scheinkman, 1991). Duffie and Kan (1996) advance the literature and provide necessary and sufficient conditions for an arbitrage-free multifactor model of the term structure of the interest rates for the affine representation of the zero-coupon bond prices. Then, Duffee (2002) distinguishes complete affine models from essentially affine models, which have more freedom to predict expected excess bond returns and alleviate numerical challenges,

While the search for a complete characterization of affine models continues, Cochrane and Pizzesi (2009) find a linear combination of forward rates ("return forecasting factor") that predicts one-year excess bond returns of each maturity with precision (an R-squared as high as 0.35). Adrian et al. (2013) extend the CP model by investigating an affine model with higher-order factors, i.e., the fourth, and fifth principal components. Despite such developments, the asset pricing literature on fixed income assets has been limited in scope, compared to that on equities, especially under the no-arbitrage condition. For example, equity pricing models investigate up to 316 factors (Harvey and Liu, 2016) while the most well-known and up to date factor model for bonds accounts for five principal factors (Adrian et al., 2013) (to which our new approach will add 124 macro-financial factors for illustration).

One of the main reasons for such a shortfall would be related to the complex nature of bond pricing due to the wide array of features (e.g., maturities, embedded-options) that make each bond unique and tightly connected. More importantly, the treasury yield curve at the time of issuance is a major determining factor in bond pricing while macro-financial factors significantly determine the yield curve changes. For example, prior studies find that bond risk premia are significantly driven by shocks to inflation and aggregate consumption (Brandt & Wang, 2003; Wachter, 2006) or even counter-cyclical movements caused by macro-financial uncertainty (Bansal & Yaron, 2004; Bansal et al., 2005). Ludvigson and Ng (2009), who use more than a hundred macro-financial indicators to investigate their effect of cyclical fluctuations in bond pricing albeit without imposing restrictions of no-arbitrage, criticize the existing affine models for being constructed with non-cyclical financial factors only and, thus, not truly reflecting the reality.

To overcome this issue, an affine model would need to be able to accommodate a large number of factors without breaching the no-arbitrage condition. There are studies that investigate the use of machine learning techniques in asset pricing. For example, Bianchi, Buchner, and Tamoni (2020) investigate bond risk premia using various machine learning methods, or so-called *non-linear* methods, in the regression-based forecasting. They find that non-linear methods are effective in predicting bond returns especially in the out-of-sample test. Using deep neural networks as a complex non-linear feature, they also show macroeconomic variables have incremental value in enhancing the prediction of bond returns. Chen, Pelger, and Zhu (2019) also use deep neural networks to investigate an asset pricing model for individual stock returns. They show that their model outperforms all referenced approaches in the out-of-sample analysis and attribute the superior predictability to the use of no-arbitrage condition and macroeconomic information. They write, “Including the no-arbitrage constraint in the learning algorithm significantly improves the risk premium signal and makes it possible to explain individual stock returns.” As such, asset pricing studies that use machine learning methods, partially or entirely, tend to use a careful approach in addressing traditional economic restrictions not only to avoid the criticisms of foregoing fundamental economic assumptions to support their findings but also to enhance their empirical results.

Furthermore, there is another aspect in ATSM studies, to which machine learning techniques can provide partial answers. ATSM studies up to date indirectly interpret their outcomes economically because they rely on latent factors. The use of latent factors for bond pricing simplifies the process of model construction but also reduces economic interpretability. We argue that, exploiting the advantages of machine learning, one can test different sets of macroeconomic variables to check their relative significance with significantly reduced time and computational complexity. Moreover, using observable macro-financial factors instead of latent factors has big advantages. One can use capital market assumptions to predict the former, but not the latter. It is generally more straightforward to formulate a prediction model for the former than for the latter. It is also practical to use observable macro-financial factors because one can incorporate the factors into interest-rate derivatives pricing.

Lastly, data augmentation is mostly mentioned in machine learning studies and is relatively new in asset pricing studies. However, it has numerous advantages that can help overcome some of the long-standing issues in the asset pricing literature such as a small sample problem that not only affects the inference of the outcomes (Nelson and Kim, 1993), but also limits the use of machine learning techniques. For example, one can use data augmentation to enhance the size and quality of training datasets for enhanced model performance (Shorten & Khoshgoftaar, 2019). By using data augmentation methods, empirical studies that often suffer from the small sizes of underlying data can exploit the advantages of big data and machine learning techniques, thereby leading to enhanced model performance.

In sum, our new model can help overcome the traditional challenges in asset pricing studies by accommodating machine learning while adhering to the no-arbitrage rule. In fact, this is what differentiates our study from ACM, which primarily uses regression to estimate state variables or five linear factors.

3. Methodology

3.1. Empirical methods

As follows, we describe our empirical strategy step by step. A complete mathematical description[1] of the model is included in Appendix I.

Step 1: Generate five yield factors using PCA with the correlation matrix on Liu and Wu's (2021) yield curve data.

Step 2: Download 127 monthly macro-financial data from FRED. Then, use the *tcode* from McCracken and Ng (2015) for data transformation. Following the method described in Appendix II, transform the unbalanced panel into balanced panel data.

[1] Because we avoid adding another ATSM in the zoo, we aim to update existing affine models just enough to incorporate numerous macro-financial factors while consuming a little computational resources.

Step 3: Use the obtained 127 macro-financial variables for another PCA.[2] Exclude principal components that have eigenvalues smaller than 2e-16. (If eigenvalues are too small, one cannot invert the covariance matrix to produce $dz_{v,t}$ in equation (2) later.).

Step 4: Using the final data, produce innovation terms. Start with $\boldsymbol{v}_{t+1}$, a K-dimensional vector of state variables, i.e., five yield factors and macro-financial PCA variables at time t+1. Our simplified equation is as follows although one can use machine learning here:

$$\nu_{t+1} = \gamma \cdot \nu_t + \alpha.$$

Then, generate dv_{t+1} and $dz_{v,t}$, which denotes the source of risk, using the following equations (dt = 1/252):

$$d\nu_{t+1} = \nu_{t+1} - (\hat{\gamma}\nu_t + \hat{\alpha}).$$
$$dz_{\nu,t} \equiv \Sigma_t^{-.5} d\nu_t \sim N(0, I)\sqrt{dt}.$$

Σ_t is the covariance matrix of dv_t, and becomes similar to I when dv_t is obtained as a result of conducting PCA on the time-series data.

Step 5: Collect the sigma and beta (the regression coefficients) using the following equations.

$$\sigma^2_{f,n,t} \equiv std(df_{n,t})^2/dt.$$

[2] One can combine step 1 with step 3 which simplifies the process, but we distinguish them to highlight the five factors of the prior ATSMs.

$$\vec{\beta}_{n,t}dt \equiv cov(dz_{\nu,t}, df_{n,t})/\sigma^2_{f,n,t}.$$

Step 6: Formulate a regression equation (1), which is rearranged as equation (2) as follows:

$$(1) \qquad df_{n,t}\Delta = -\lambda_t' dz_{\nu,t} + \sigma^2_{f,n,t}(\Delta/2 + \lambda_t'\vec{\beta}_{n.t})\Delta dt + \sigma_n \Delta dz_{n,t}.$$
$$\longrightarrow (2) \quad df_{n,t} = constant + \lambda_t'(\sigma^2_{f,n,t}\vec{\beta}_{n,t}dt - dz_{\nu,t}/\Delta) + \epsilon_{n,t}.$$
$$constant \equiv \sigma^2_{f,n,t}\Delta dt/2.$$

λ_t denotes the price of risks and is a linear function of risks, ν_t.

Step 7: Use equation (2) to obtain λ_t for each *t* in Fama-MacBeth regression (Fama & MacBeth, 1973) (possibly Lasso regression when selecting macro-financial factors). One can model λ_t as an affine function of risks and estimate it.

The above empirical method has several advantages over that used for the ACM model. First, the estimation process is more intuitive and simpler because it applies HJM and Fama-MacBeth regression and linearizes the pricing kernel with Ito's lemma. For example, the Python code length for the estimation process for the proposed model is only half of what it is for the ACM model. Second, it takes less time to conduct out-of-sample tests using the 178-month sample period.[3] While it takes 70 seconds on average to conduct the tests with the ACM model, it takes 51 seconds with the proposed model. Third, our method allows the intuitive and direct use of machine learning techniques. One can use the techniques at both first and second stages of our

[3] Our computer has the following specifications: Apple MI (8-core CPU, 7-core GPU), memory (16 GB), SSD (256 GB).

methods similar to that of Fama & MacBeth (1973). Last but not least, it is much easier to add 124 observable macro-financial factors to our ATSM. It is unclear how to add a large number of observable factors to the ACM approach without computational challenges.

3.2. Data

We use Liu and Wu's (2021) yield curve data which has annualized continuously compounded zero-coupon yields.[4] Bianchi et al. (2020) also use the data to implement neural networks in asset pricing. For macro-financial variables, we collect monthly macro-financial data from the Federal Reserve Economic Data (FRED).[5] We use *tcode* to transform the data before generating macro-financial factors on McCracken and Ng (2015); more specifically, due to frequent missing values, we use the five-step procedure to balance the unbalanced panel of macro-financial variables (see Appendix II for details).

We start with the pool of 127 macro-financial variables to collect macro-financial data, following Ludvigson and Ng (2009). PCA is used to generate 127 principal components, but three of the 127 factors have too small eigenvalues to find the inverse of the matrix; therefore, a final set of 124 factors is used for in-sample tests and a set of 121 factors for out-of-sample tests in addition to five yield factors. This makes the total number of factors 129 and 126 for in-sample and out-of-sample tests, respectively. The list of 127 macro-financial variables is attached in Appendix III.

---

[4] https://sites.google.com/view/jingcynthiawu/yield-data
[5] https://research.stlouisfed.org/econ/mccracken/fred-databases

Table 1 compares the summary statistics on the model-generated and observed yields. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12.

###Insert Table 1 about here###

The results support the key result: *increasing the number of factors from five to 129 enhances the fitting of the yields*. The statistics of the model-implied yields generated by 129 factors are nearly identical to those of the observed ones. For example, the average observed yields range from 4.5% to 6.1%, which is the same as the average model-implied yields. The yield variations measured by standard deviations across different maturities are also identical for the observed and model-implied yields as they range from 3.1% to 3.6%. Such results are also plotted in Figure 1 and 2, where observed and model-implied yields overlap almost perfectly.

What if incorporating a machine-learning model which our ATSM can utilize easily? Machine learning can make the 129-factor model simpler, more intuitive and more powerful. For instance, we employ simple machine learning to select relevant macro-financial variables; Lasso regression selects only relevant factors while zeroing out irrelevant ones. It identifies a final set of 23 macro-financial variables out of 127 macro-financial variables in our test while not compromising the fit. A discussion of the results is included in subsection 4.3.

4. Empirical Tests

4.1. Yield fitting

To check the fitting of the forward rates implied by our model, we conduct time series and cross-sectional analysis for the period of 1972:8 to 2021:12. Figure 1 plots the time series yield fitting and estimates of term premia as well as one-month holding returns of the observed and model-implied yields.

###Insert Figure 1 about here###

All four graphs in Figure 1 plot the results generated by using all 129 factors, including five yield factors and 127 macro-financial variables.[6] The upper two graphs show that the observed and model-implied yields are almost perfectly matched when measured over time.[7] The bottom two graphs also show that the proposed model can describe one-month holding excess returns of the bonds almost perfectly while the term premia remain stable.

###Insert Figure 2 about here###

Figure 2 plots cross-sectional regression results using the means and standard deviations of the observed and model-implied yields across different maturities ranging from 3 to 120 months. The graph on the left-hand side plots the unconditional means while the graph on the right-hand side plots the unconditional standard deviations of the observed and model-implied yields. The observed and model-implied yields are nearly perfectly matched.[8]

4.2. Forecasting bond returns using yield and macro-financial factors

Existing studies about affine term structure models primarily focus on how a model generates yields that are close to observed yields. However, examining whether the proposed model can predict future returns during out-of-sample periods would also be equally important, especially for practitioners trying to generate excess returns using the model. Jang et al. (2021) examine the

---

[6] Our model is used to generate four and five yield factors for comparison with prior affine models that use the same number of factors (e.g., CP and ACM models). The results are reported in Appendix IV.

[7] The results are available for sharing upon request.

[8] Results generated from using four and five yield factors are reported in Appendix V.

predictive power of the ACM model using Korean bond data. They use forward rates generated by the ACM model and empirically show that the difference between ACM-implied and observed forward rates predicts the future forward rate changes even with a simple univariate regression. We replicate the tests for the in-sample and out-of-sample periods.

To explore whether our proposed model can predict future returns with forward rates, similar to Jang et al. (2021), we use the following regression:

$$\Delta f_{observed,t+1} = \alpha + \beta_t (f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

To isolate the predictability for each maturity, we convert the observed spot rates in our sample into forward rates. The independent variable is the difference between the observed and model-implied forward rates at t that mature in n months with its coefficient $\boldsymbol{\beta}$. The dependent variable is the difference between the forward rates at t and $t + 1$ that mature in n months. The difference between the forward rates at t-1 and t is used as a control variable and its coefficient is denoted $\boldsymbol{\gamma}$.

###Insert Table 2 about here###

Table 2 reports the regression results generated by our ATSM with four, five, and 129 factors (five yield factors + 124 macro-financial factors). The coefficients and t-values for the independent and control variables are reported. The predictability of the model varies across the bonds' maturities but tends to improve as the number of factors increases from four to 129 for the bonds with up to 18-month maturities. For example, for the bonds with 12-month maturities, the model predicts the returns with much stronger significance when 129 factors are used (t-value of

-3.541) than when four factors are used (t-value of 0.041). However, the results become inconclusive for the bonds with longer maturities.

###Insert Figure 3 about here###

Figure 3 plots t-values in absolute terms from the same regression. The difference between Figure 3 and Table 2 is that, while Table 2 reports the regression results performed with control variables, Figure 3 displays the results *both* with and without control variables. The upper two graphs of Figure 3 show that our model with four or five factors has a significant forecasting power for future bond returns using forward rates at t. However, when the number of factors increases to 129 as in the graphs in the second row, our model without control variables displays no forecasting ability as shown in the flat lines. This may be related to the small sample size, which our new data augmentation approach will address. In fact, we show that the predictive power becomes significant with expanded data, of which the results are presented and discussed in subsection 4.5. The statistical significance of the predictive power of the ACM model with four and five factors is nearly identical whether including or excluding the control variable and similar to our models with the same number of factors.

4.3. Selection of macro-financial variables

Among the 129 factors used in the tests, five are yield factors and 124 are macro-financial factors. It would be challenging for fixed income researchers to interpret and run tests on the data with numerous variables. Conducting unsupervised learning such as PCA helps examine the relative importance of the variables. In addition, Lasso regression helps select only the relevant factors, simplify the model construction process, and enable economic interpretation. Table 3 shows how

Lasso regression helps narrow the number of factors for the empirical tests in combination with our ATSM.

###Insert Table 3 about here###

For this, we first conduct Lasso regression instead of OLS in order to generate lambda coefficients using the following equation derived in the appendix 1:

$$df_{n,t} \cdot \Delta \ = -\lambda_t' dz_{\nu,t} + \sigma_{f,n,t}^2 \left(\Delta/2 + \lambda_t' \cdot \vec{\beta_{n,t}} \right) \Delta dt + \sigma_n \Delta dz_{n,t}$$

We exploit the flexibility of our data augmentation to fill the yield curve with 1,920 data points per month; otherwise, the Lasso coefficients become insignificant. This indicates that data augmentation is crucial in identifying relevant macro-financial factors and generating rich economic implications. Using the same L1 hyperparameter ("Lasso alpha" = 0.001) within the preset range, we generate RMSEs for the out-of-sample tests. Then, we estimate the Lasso coefficients (λ) for the five yield factors and 124 macro-financial factors. We multiply a Lasso coefficient with the eigenvector for each macro-financial variable, of which the time-series average and standard deviation are used to calculate the (Fama-MacBeth) *t*-value = mean / std /$T^{.5}$. In this case, we choose only the variables that have (Fama-MacBeth) *t*-values greater than 1.960. This leaves only 23 macro-financial variables out of 127 variables collected from FRED.

Narrowing down the number of macro-financial variables from 127 to 23 is crucial for making economic inferences. For example, we can make the following inferences from the results in Table 3. First, bond returns most significantly covary with the movements in the housing market, represented by the positive and statistically significant coefficients of the new private housing permits (*t*-value of 2.721) and housing starts (*t*-value of 2.879 or 3.161 for Total and West, respectively). These key economic indicators reflect the number of privately owned new houses on which construction has been initiated in a given period. Any increase from the previous period for both indicators would mean that the housing market is relatively strong.

Second, yield changes are more influenced by the changes in the financial indicators that are directly related to bond pricing, such as the Effective Federal Funds Rate (FFR) (t-value of 2.394) and 1-year Treasury rate (t-value of 2.558). This is the most important factor out of level, slope and curvature, hence an intuitive result. The changes in the stock-related indicators such as the S&P's common stock price index (t-value of -2.017) and its price-to-earnings ratio (PER) (t-value of -2.029) also influence bond returns significantly. Furthermore, while we set Lasso alpha as 0.001 to generate the results in Table 3, when we change the hyperparameter to 0.01 as in Figure 8, the statistical significance of S&P and PER indices as relevant economic indicators disappears. This implies that the stock market factors are sensitive to an empirical setting.

Third, some inflation indicators significantly and negatively affect bond prices. The Producer Price Index (PPI) and the Consumer Price Index (CPI): Commodities show the t-values of -2.648 and -2.796, respectively. Thus, the change in the term structure explicitly moves with inflation. Labor market conditions also signal the level of inflation.

In sum, such findings using our model can make the job easier for data-driven practitioners who consider macro-financial variables important and relevant for the prediction of bond returns. For example, one can brainstorm a wide range of her initial data consisting of macro-financial variables, adjust or reduce the number of the constituents of the dataset easily after repeating tests without consuming excessive time or computational resources, and make specific economic interpretations linked to the prediction of bond returns while imposing economic restrictions such as the no-arbitrage condition.

4.4. Model comparison

Yield fitting

To check the robustness of our model, we compare the yield-fitting results of our proposed model to those of the ACM model. First, Figure 4 presents six graphs plotting the estimates of the percentage valuation errors (ϵ) of the predicted value of forward rate changes (*dfwd*) and RMSEs for the in-sample performance comparison. The percentage valuation errors measure the accuracy in predicting *dfwd* rate and are defined as $\epsilon \equiv \widehat{dfwd} \,/\, dfwd - 1$ where *dfwd* is the difference of forward rate and $\widehat{dfwd}$ is the corresponding model estimate. RMSEs measure the distance between the actual value and the predicted value.

###Insert Figure 4 about here###

The results in Figure 4 indicate that our model has superior predictability to the ACM model using 4, 5, or 129 factors, measured by various methods. The flat lines represent the nearly perfect fit of the new model with 129 factors although it could also imply an overfitting problem for the in-sample tests. For comparison, we also report the summary statistics of the yields generated by the ACM model with four and five factors for the same sample period of 1972:8 to 2021:2. Table 4 reports the results.

###Insert Table 4 about here###

The results in Table 4 are comparable to those in Table 1. The average yields generated by the ACM model with five factors range from 4.5% to 6.1%, similar to the yields generated by the proposed models and the observed yields. The standard deviations also range from 3.1% to 3.6% as was the case for the observed yields and the implied yields from the proposed models. Nevertheless, 129 factor model almost perfectly reproduces the observed yields in in-sample analysis. These results confirm the robustness of the proposed model in generating bond yields using an equal or a large number of factors.

Predictability

We examine the predictive power of the prior affine models in order to check the robustness of our model performance in terms of predicting future bond returns using model-implied forward rates. The results are reported in Figure 5 and also in Table 5. As mentioned earlier, we use 121 macro-financial PCA variables, having eliminated three variables that have too small eigenvalues to make an inverse of the matrix. Therefore, for out-of-sample tests, 126 factors are used including 5 yield factors and 121 macro-financial factors.

Figure 5 presents six graphs plotting the out-of-sample results for model performance comparison measured by the percentage valuation errors of the predicted value of *dfwd* rate (ϵ) and root mean squared errors (RMSEs). The estimates are calculated in the same way as they were for Figure 4. To conduct out-of-sample tests, we use the in-sample data during the first 414 months, which is from 1972:8 to 2007:2 (i.e., 70% of the entire sample period), for the estimation of prediction values in a rolling method. The following steps describe the process:

Step 1. Use the data from 1972:8 to 2007:2 for yield fitting

Step 2. Calculate $d\widehat{fwd}$ by using model-implied and observed rates in the equation,

$$\Delta f_{observed,t+1} = \alpha + \beta_t (f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

Step 3. Expand the period for the estimation of prediction values to 1972:8 to 2007:3, which adds an additional month at the end of the previous estimation period. Repeat Steps 1 and 2.

###Insert Figure 5 about here###

In the bottom right graph, the RMSEs of all models range from zero to 0.005, which means that the forecasting error is nearly zero across all maturities. Therefore, the forecasting power of the proposed model does not decay but even improves at some maturities. Some may argue that using only a few factors in the conventional affine model to predict future returns may be more convenient if the predictive power does not significantly improve; however, our proposed model can accommodate a large number of pricing factors, allowing for the use of any number of *observable* macro-financial variables that affect bond pricing, without compromising the predictive ability, while enabling economic interpretations. They are big advantages. Lastly, in the out-of-sample tests, no longer exists the perfect fit in Figure 4 possibly due to overfitting problem for the proposed model with a large number of factors.

##Insert Table 5 about here###

Table 5 reports the predictive power of the ACM model with four and five factors for comparison and confirms the predictive power of the ACM model for future bond returns with statistical significance across most maturities. When compared to the results generated from our proposed model with 129 factors shown in Table 2, the statistical significance is relatively weaker for bonds with up to 18-maturities. More specifically, the beta coefficients for the ACM model with five factors have similar, mixed statistical significance across almost all maturities to those for the proposed model with five factors; however, when incorporating 129 factors in the proposed model, the statistical significance of the beta coefficients, measured by *t*-values, significantly improves and outperforms that for the ACM model for especially short-term maturities. While inconclusive, such results confirm that the proposed model has robust predictive power that is comparable to the ACM model when the number of factors are the same, but that the proposed model's ability to incorporate a large number of factors is linked to the model's improved predictability.

4.5. Data Augmentation

We perform additional tests to examine the effect of data augmentation on yield fitting. For earlier yield fitting tests, of which the results are reported in Figure 1 and Figure 2, the dependent variable is the difference between the observed and model-implied forward rates. The forward rates have n-month maturities in the multiples of three so that we have 40 data points in total (e.g. 3, 6, 9, …, 117, and 120-month) per cross-sectional analysis. To resolve the data limit, we introduce a data-augmentation method and its intuition. **Appendix VI** presents its details.

To expand the data, we further divide the maturities into multiples of one, 0.25, and 0.125. For example, with the current data, we do not have the data on forward rates at n=1, 2, 4, 5, …, 118, and 119. For example, to expand the data to have monthly forward rates, we use linear interpolation on forward rates at n=3 and 6 to obtain forward rates at n=4 and 5 and use linear interpolation on forward rates at n=117 and 120 to obtain forward rates at n=118 and 119. The resulting data have 120 monthly forward rates in total instead of 40. When we further augment the data to have forward rates with n/4- and n/8-month maturities, we gain a total of 480 and 1,920 data points, respectively. Our model's predictive power with such expanded data is reported in Figure 6.

###Insert Figure 6 about here###

The difference between Figure 3 and Figure 6 shows that the data augmentation significantly enhances our model's prediction power especially when all 129 factors are used. The t-values stay between zero and one in Figure 3 whereas they become more significant and vary across different maturities with augmented data. In sum, data augmentation, a commonly used technique in computer vision, can help mitigate the problem associated with a small sample size in asset pricing studies.

Table 6 also reports the results from the Lasso regression, which is conducted after data augmentation, to test the predictive power of the proposed model with five yield factors and 124 macro-financial factors. The estimates of coefficients and *t*-values for the independent variable, and the control variable are reported for when the underlying data are expanded to have 120, 480, and 1,920 data points.

###Insert Table 6 about here###

For all data points, the results are most significant for bonds with 3-, 24-, and 72-maturities and least significant for bonds with 18-month maturities. The results vary across different maturities and numbers of data points, and do not necessarily strengthen or weaken the results previously obtained for the smaller sample (before data augmentation) as is the case in Panel D of Figure 6; however, they would confirm whether the small sample leads to biased outcomes or not, helping to check robustness of the original results.

4.6. Model performance during the COVID-19 pandemic

We examine the proposed ATSM's testing performance during the period of a pandemic shock. The purpose is to examine whether the model generates superior performance amid extreme market turbulence. Specifically, Figure 7 plots the estimates of the RSMEs of the performance of our model and the ACM's during the COVID-19 pandemic for out-of-sample tests. The original sample data are constructed using Liu and Wu (2021)'s yield curve data. Using this data during the first 414 months from 1972:8 to 2007:2 (i.e., 70% of the entire sample period), we estimate the prediction values in a rolling method and follow the steps used for Figure 5. For performance comparison, we use the reference date of 11 March 2020, when the COVID-19 pandemic was declared by the World Health Organization (WHO), to set the out-of-sample period. Therefore, the upper five lines in Figure 7 represent the RMSEs during the period of 2007:3 to 2021:12 while the bottom five lines represent the RMSEs during the out-of-sample period of 2020:4 to 2021:12.

Insert Figure 7 about here

Figure 7 shows that the RSMEs of our model during the period of COVID-19 are smaller than the other period and range from zero to 0.003 for all specifications reported. The bigger training set due to expanding rolling windows may have led to improved learning, which could be the cause of the smaller RMSEs during the out-of-sample period (represented by the bottom five lines). The result confirms that the predictability stays intact during the period of market distress caused by the breakout of COVID-19, supporting the robust performance of our proposed model.

4.7. Lasso regression and out-of-sample tests

To check the robustness of the results further, we test the model with Lasso regression. The sample period for the initial set is from 1972:8 to 2007:2 with 6:4 splits for training and validation sets to fix the hyperparameter (alpha = 0.01) throughout the paper. The out-of-sample results are not sensitive to the choice of alpha. We generate prediction values, denoted as $\widehat{dfwd}$, for following training set periods in the same steps introduced for Figure 5 (rolling estimation).

Figure 8 presents graphs plotting the RMSE estimates to compare the performance of our ATSM to that of the ACM model with four and five factors. The model construction also follows the same steps as those used to generate the results in Figure 3. The difference is that we use Lasso regression instead of OLS, which is also described in Appendix VI.

Insert Figure 8 about here

Overall, the results indicate that our proposed model performance is not compromised during the out-of-sample periods. The results in Panel A shows that the performance of our model is closer to the performance of the ACM model with four and five factors when Lasso alpha is set to 0.01. The results in Panel B confirm that the predictability of our proposed model is not compromised after data augmentation during the out-of-sample periods.

4.8. Nonlinear relationship between macro-financial variables and bond returns

In reality, macro-financial indicators are often reported on similar dates, and thus their impact on asset prices cannot be perfectly isolated from one another. To reflect this, we run an additional empirical test and show which macro-financial variables in interaction with other variables have the most significant impact on forward rates.

More specifically, first, we use five yield factors and 127 macro-financial variables to construct interaction variables, which results in 8,646 interaction variables or 8,778 independent variables, including 5 yield factors, 127 macro-financial variables, and 8,646 interaction variables, for the empirical test. Then, we apply Steps 4 to 6 introduced in **3.1 Empirical methods** with some variations. For example, in Step 4, we calculate the inverse of the diagonal matrix of Σ_t to generate $dz_{v,t}$, which denotes the source of risk. We conduct Lasso regression with a Lasso alpha equal to 0.001 in Step 6. The results are summarized and reported in Table 7.

Insert Table 7 about here

Consequently, all 284 variables are reported with *t*-values greater than 1.960, which is used to select the previous 23 macro-financial variables reported most significant in Table 3 (see Appendix VII for the full list). Table 7 reports the top 50 variables with the most statistical significance. The results can yield numerous economic interpretations and the following are some examples.

First, the indicators associated with the housing market have significant and positive coefficients in Table 3, and the same indicators when interacted with one another continue to bear positive coefficients with even stronger significance in Table 7. This is an intuitive result that implies that the growth expectation for the housing market, or the economy in general, would lead to lower bond returns as riskier assets draw more attention and capital from the investors.

Second, an increase in average weekly hours, weather in manufacturing and goods producing, positively and significantly influences the changes in forward rates when interacted with the variable for S&P's dividend yields (t-values of 2.735 and 2.701, respectively). The average weekly hours variable is considered a leading indicator, of which an increase would signal the beginning of economic growth. On the other hand, the growth in dividend yields can send either a lagging signal that more profits are being paid out to the shareholders or a leading signal that the company's growth perspective is positive. Therefore, the positive and significant coefficients of the interaction variables using the variables of average weekly hours and dividend yields indicate that the rise in labor productivity and dividend yields lead to higher forward rates, hence lower bond returns. The results are intuitive since when investors expect economic expansion or growth in corporate fundamentals, they become more risk-seeking, shifting their focus from relatively safe assets like bonds towards equities and alternative assets that are considered riskier but yielding higher returns.

Third, the Baa-rated corporate bond yields variable bears a significant and negative coefficient when interacted with the effective federal funds rate variable (the t-value of -2.835). Baa-rated bonds are considered investment grade bonds, which are relatively safer than junk bonds but riskier than other investment grade bonds. The effective federal funds rate variable has a significant and positive coefficient when tested alone (*see* Table 3), implying that an increase of the short term interest rates leads to higher forward rates, hence lower returns. When both variables are interacted with one another, however, the coefficient becomes negative, implying that the negative effect of the increase of short term interest rates on bond returns would be negated by a significant increase in the Baa-rated corporate bond yields.

Lastly, an increase of the difference between 10-year treasury minus federal funds rate indicates yield curve steepening, reflecting the investors' positive economic outlook or an overheated market. When the US dollar appreciates against other major currencies such as the UK sterling pound, it usually signals a positive economic outlook as well. Therefore, the significant and negative coefficient of the interaction variable using 10-Year Treasury C Minus FedFunds and U.S./UK Foreign Exchange Rate (t-value of -2.525) could indicate that bond returns deteriorate when the investors have a positive or even overly optimistic economic outlook, although it is not known how the yields on shorter maturity Treasury bonds or the value of US dollar against other major currencies move simultaneously.

Overall, the results in Table 7 and Appendix VII show how the economic interpretation could be affected by using more than one variable to test their influence on bond returns, highlighting our novel approach in the ATSM literature.

5. Conclusion

This paper shows that the larger the number of macro-financial factors, the better the performance of an affine model in both in-sample and out-of-sample. It is hard to incorporate a large number of observable macro-financial factors into existing no-arbitrage affine models. Therefore, we develop a new affine model and an empirical strategy that makes it easy to integrate numerous factors or even utilize machine learning methods under no-arbitrage. In addition, using the proposed methods in combination with PCA and Lasso helps identify a specific set of relevant macro-financial variables easily. We show that the bond returns in our sample are significantly related to the changes in the indicators about the housing market (e.g., housing permits and housing starts), short-term rates, stock market, and inflation. Reducing the number of macro-financial variables not only helps make specific economic inferences but also saves the users from the wearing job of collecting exhaustive amounts of data or choosing ad-hoc variables without ground. Furthermore, we show that a small sample size problem can be mitigated by data augmentation, which ultimately enables machine learning methods to be employed. The forecasting ability of the model stays robust even after data augmentation.

In academia and practice, there is growing attention to the use of machine learning in asset pricing. Despite the advantages of using machine learning, such as generating outcomes with increased velocity, precision and less manpower, skeptics are often concerned about the possibility of data mining and breach of fundamental asset pricing assumptions such as the no-arbitrage condition (Bianchi et al., 2020; Chen et al., 2019). Our proposed model enables the use of machine learning with much flexibility while adhering to the assumption of no arbitrage. In that sense, our study adds some insight to the ongoing conversations about machine learning and asset pricing, especially for those who value the fundamental assumptions of asset pricing but also recognize the importance of expanding the field by adopting innovations.

Our study has several limitations that can be addressed by future studies. First, one can expand the investigation of the proposed affine model with machine learning by testing it in the international markets. The current empirical setting of our study is the U.S. market. Evidence suggests that the existing affine models can be used to predict bond returns in the financial markets other than the U.S. although implications may differ (Jang et al., 2021; Sekkel, 2011).

Second, one can investigate the proposed model on other assets, such as equities, interest-rate derivatives or alternative assets. Although affine models are primarily used for predicting bond returns, some studies investigate affine models for equities or equity index options (Christoffersen et al., 2006; Lemke & Werner, 2009). It would extend the literature to use machine learning techniques with our proposed approaches to analyze such assets. In particular, since we explicitly link macro-financial variables with yield curves under no-arbitrage conditions, one can also connect the observable economic variables with the prices of interest-rate derivatives.

Third, one can investigate different machine learning techniques in our model that are not introduced in this study. This study shows how applying Lasso in the proposed model can help make estimation easier and faster. The study also shows how using PCA and Lasso can identify a few relevant macro-financial variables to be used in the model. Using autoencoders or deep learning instead of PCA or Lasso could be intuitive and straightforward in our context and contribute to both academic and practical researchers.

References

Adrian, T., Crump, R. K., & Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics*, *110*(1), 110-138.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299.

Bai, J., & Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8, 53-80.

Bansal, R., Khatchatrian, V., and Yaron, A., 2005, Interpretable asset markets?, *European Economic Review*, 49:531–60.

Bansal, R., & Yaron, A., 2004, Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance*, 59:1481–509.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2), 1046-1089.

Brandt, M. W., and Wang, K. Q., 2003, Time-varying risk aversion and unexpected inflation. *Journal of Monetary Economics*, 50:1457–98.

Campbell, J. Y., & Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, *58*(3), 495-514.

Chen, L., Pelger, M., & Zhu, J. (2019). Deep learning in asset pricing. arXiv preprint arXiv:1904.00745.

Chen, R., & Scott, L. (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income*, 3 (3) 14-31.

Christoffersen, P., Jacobs, K., & Mimouni, K. (2006). An empirical comparison of affine and non-affine models for equity index options. Available at SSRN 891127.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1), 138-160.

Cochrane, J. H., & Piazzesi, M. (2009). Decomposing the yield curve. In AFA 2010 Atlanta Meetings Paper.

Dai, Q., & Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of financial Economics*, *63*(3), 415-441.

Duffee, G. R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance*, 57, pp. 405–443.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4), 379-406.

Duffie, D., & Singleton, K. J. (1997). An econometric model of the term structure of interest rate swap yields. *The Journal of Finance*, *52*(4), 1287-1321.

Estrella, A., & Hardouvelis, G. A., 1991, The term structure as a predictor of real economic activity. *The Journal of Finance*, 46(2), 555-576.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *The Journal of Finance*, *45*(4), 1089-1108.

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 680-692.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), 607-636.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of financial economics, 33(1), 3-56.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates:

A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77-105.

Israel, R., Kelly, B. T., & Moskowitz, T. J. (2020). Can Machines Learn Finance? *Journal of Investment Management*.

Harvey, C. R., Liu, Y., and Zhu, H., 2016, … and the cross-section of expected returns, *The Review of Financial Studies* 29(1), pp. 5-68.

Jang, G. Y., Kang, H. G., & Lee, D. J. (2021). An Extension of the Five-factor Affine Term Structure Model: Predicting Future Bond Returns. *Asia-Pacific Journal of Financial Studies*, *50*(6), 659-689.

Lemke, W., & Werner, T. (2009). The term structure of equity premia in an affine arbitrage-free model of bond and stock market dynamics. *Working paper*

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

Liu, Y., & Wu, J. C. (2021). Reconstructing the yield curve. Journal of Financial Economics, 142(3), 1395-1425.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12), 5027-5067.

Nelson, C. R., & Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *The Journal of Finance*, *48*(2), 641-661.

Piazzesi, M. (2010). Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques* (pp. 691-766). North-Holland.

Sekkel, R., 2011, International evidence on bond risk premia, *Journal of Banking and Finance*, 35, pp. 174–181.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1-48.

Wachter, J., (2006). A consumption based model of the term structure of interest rates. *Journal of Financial Economics*, 79:365–99.

Wüthrich, K., & Zhu, Y. (2020). Omitted variable bias of Lasso-based inference methods: A finite sample analysis. *The Review of Economics and Statistics*, 1-47.

Figure 1. Time-series fit and term premia of model-implied yields using 129 factors

This figure plots the yield fitting and term premium estimates, as well as predictability for one-month holding period excess returns of zero-coupon yield curve data for Treasuries with two- and ten-year maturities, as observed and implied by our model with 129 pricing factors. Of the 129 factors, five are yield factors and 124 are macro-financial factors. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. For all graphs, solid lines represent observed yields and returns, dashed green lines represent model-implied yields or returns, and dashed red lines represent the model-implied term premia.

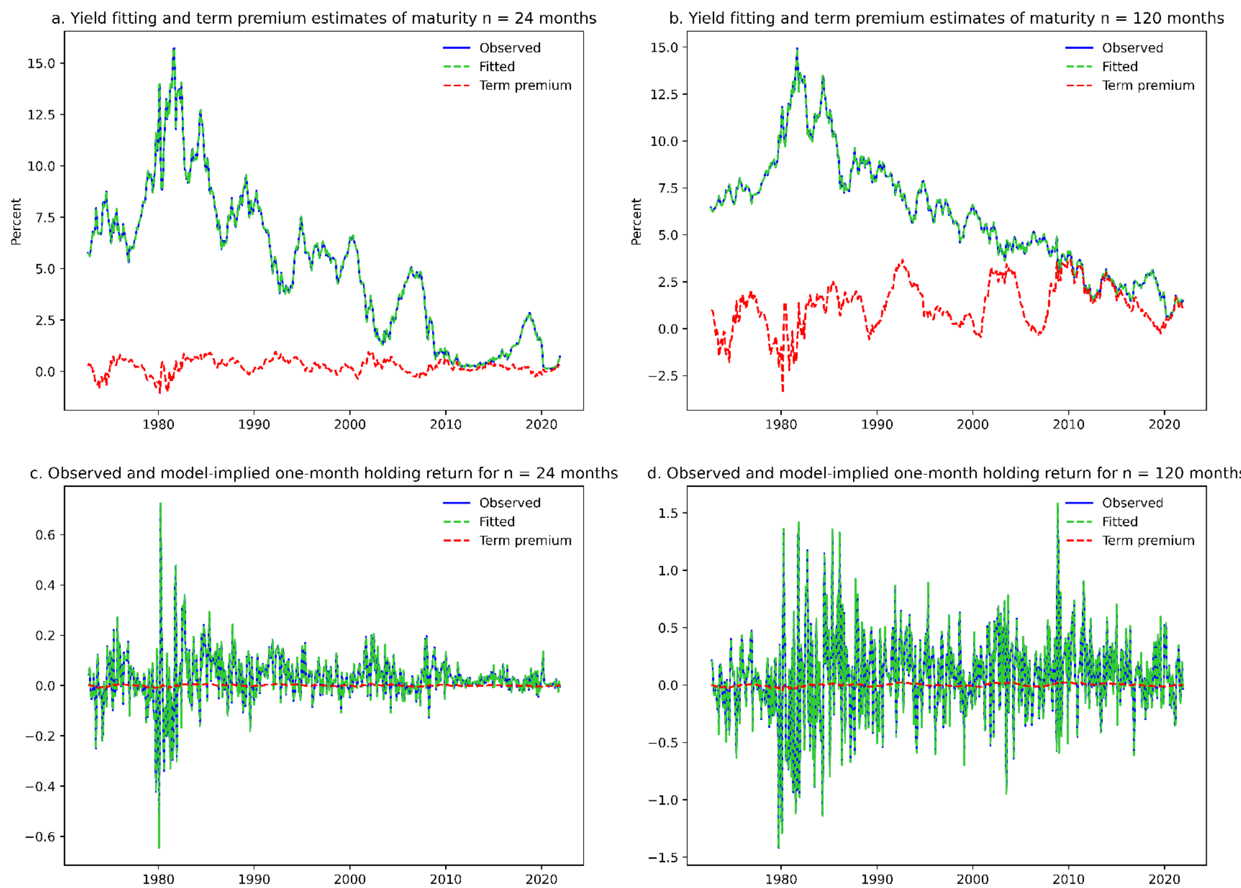


Figure 2. Cross-sectional diagnostics of the model-implied yields using 129 factors

This figure plots graphs exhibiting the cross-sectional fit of the yields generated by using our model with 129 pricing factors, including five yield factors and 124 macro-financial factors. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. The graph on the left-hand side plots the unconditional means while the graph on the right-hand side plots the unconditional standard deviations of the observed and model-implied yields.

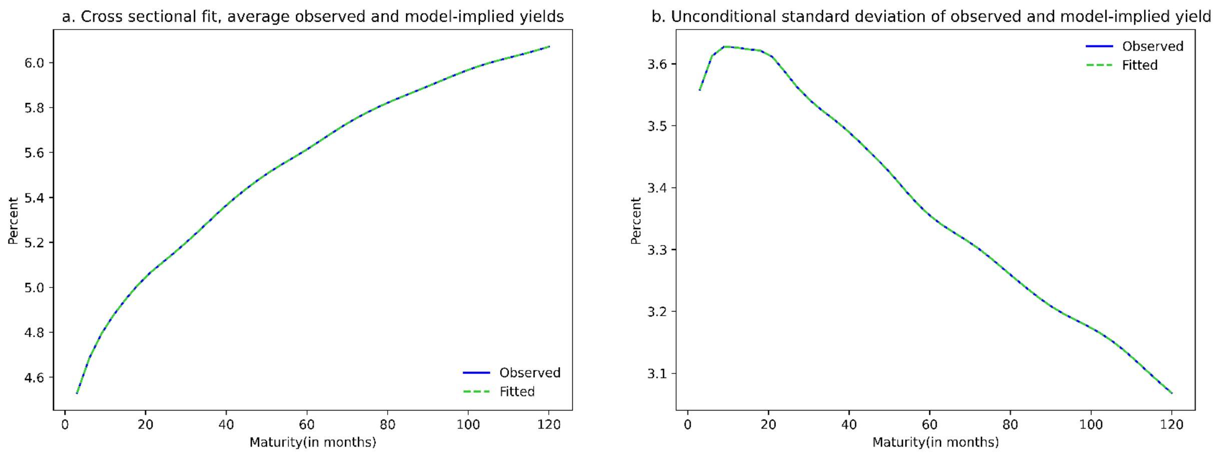


Figure 3. Predictive power across different maturities

This figure plots the absolute *t*-values. |*t*|, for $\boldsymbol{\beta}$ generated from our regression to test the predictive power of our proposed model. The following equations are used:

1) without control variables

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \epsilon_t.$$

2) with control variables

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

The independent variable is the difference between the observed and our model-implied forward rates at *t* that mature in *n* months. The dependent variable is the difference between the forward rates at *t* and *t* + 1 that mature in *n* months. The difference between the forward rates at *t*-1 and *t* is used as a control variable and the coefficient is denoted, $\boldsymbol{\gamma}$. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12.

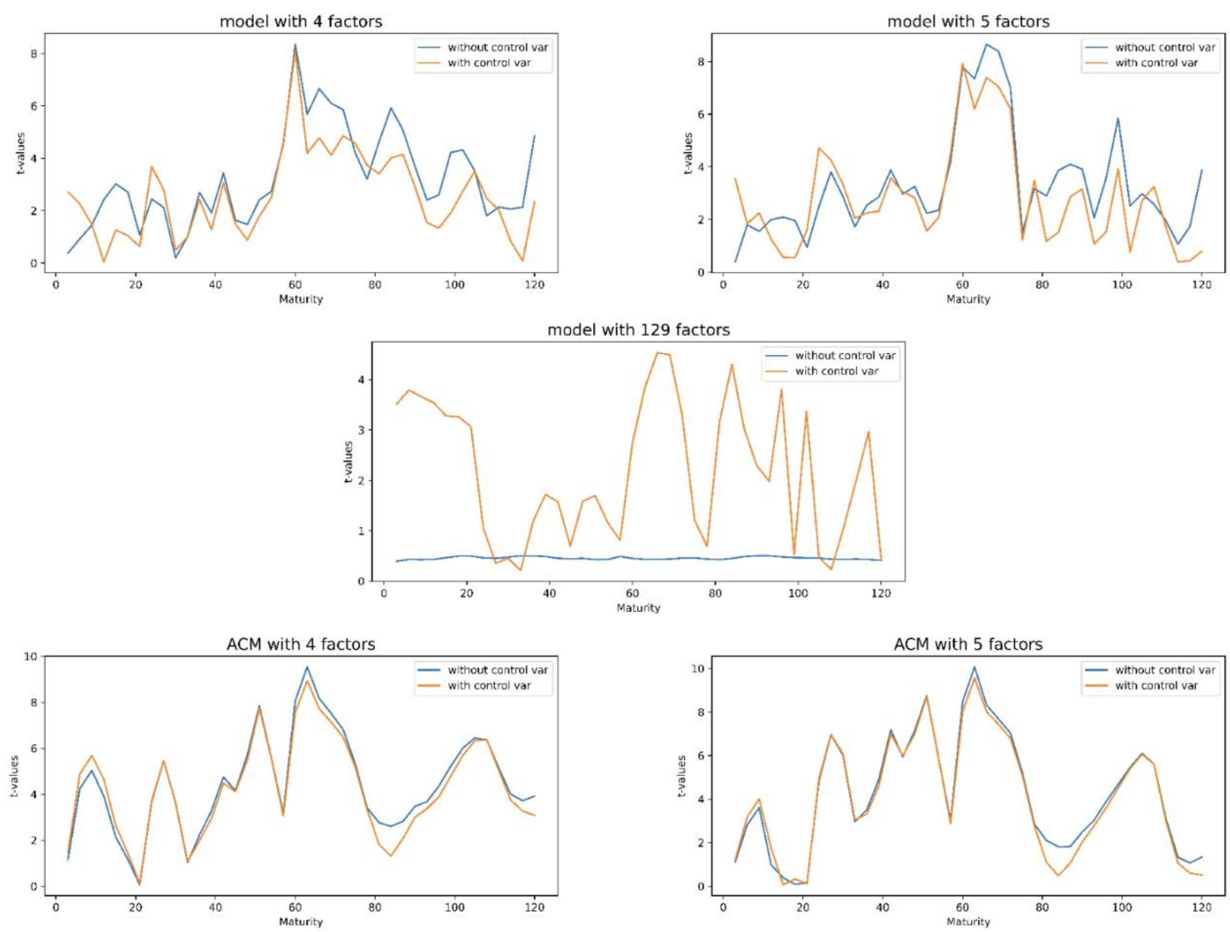


Figure 4. In-sample model performance comparison

The figure plots in-sample model performance for comparison measured by the percentage valuation errors (ϵ) of the predicted value of *dfwd* rate and root mean squared errors (RMSEs). The percentage valuation errors measure the accuracy in predicting *dfwd* rate and are defined as $\epsilon \equiv \widehat{dfwd} \,/\, dfwd - 1$ where $dfwd$ is the difference of forward rate and $\widehat{dfwd}$ is the corresponding model estimate. RMSE measures the difference between the actual value and predicted value, and it is defined as $RMSE(\widehat{dfwd}) \equiv (\sum \quad (dfwd - \widehat{dfwd})^2/N)^{1/2}$

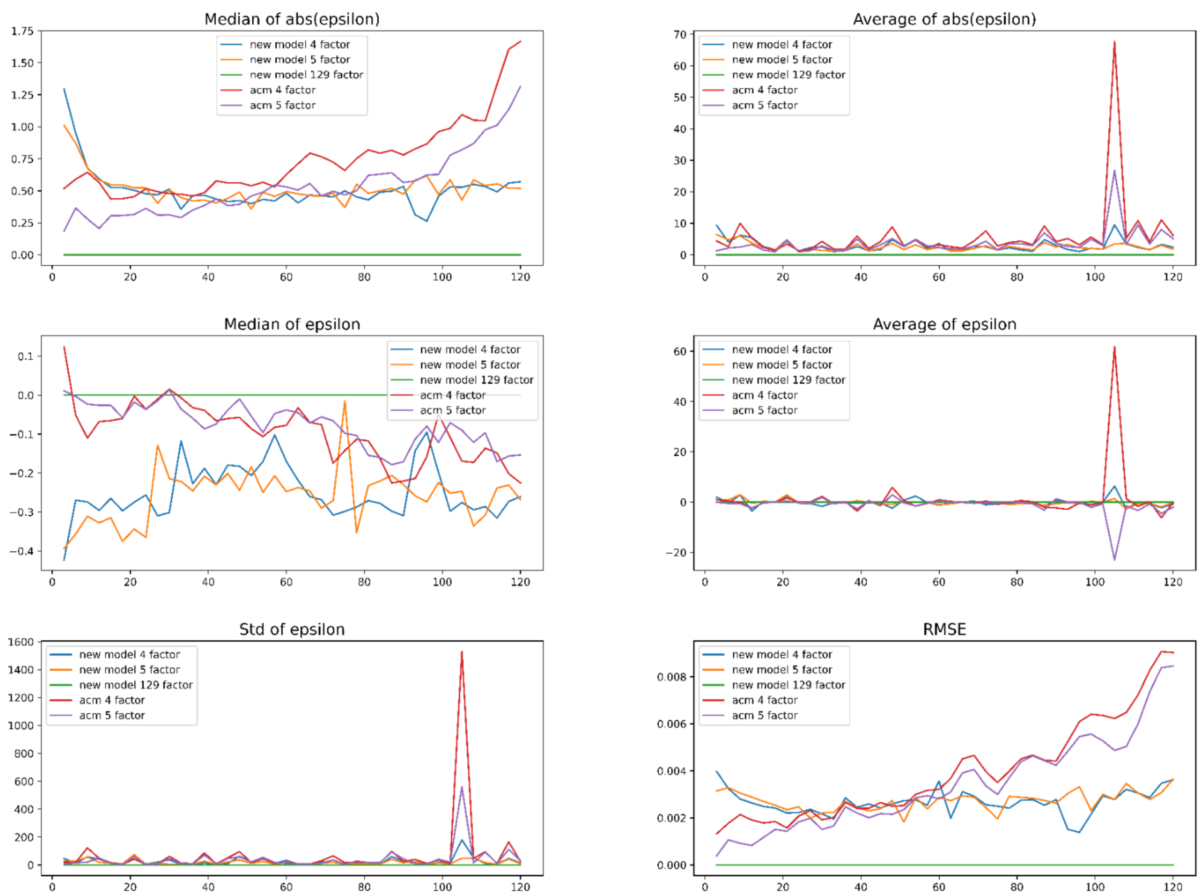


Figure 5. Out-of-sample model performance comparison

This figure presents the out-of-sample results for model performance comparison measured by the percentage valuation errors (ϵ) of the predicted value of *dfwd* rate and root mean squared errors (RMSEs). The percentage valuation errors measure the accuracy in predicting *dfwd* rate and are defined as $\epsilon \equiv \widehat{dfwd} \ / \ dfwd - 1$ where *dfwd* is the difference of forward rate and $\widehat{dfwd}$ is the corresponding model estimate. RMSE measures the difference between the actual value and predicted value, and it is defined as $RMSE(\widehat{dfwd}) \equiv \sqrt{\sum \quad (dfwd - \widehat{dfwd})^2/N}$. For out-of-sample tests, we use the original data constructed by using Liu and Wu (2021)'s yield curve data during the first 414 months, from 1972:8 to 2007:2 (i.e., 70% of the entire sample period), for the estimation of prediction values in a rolling method. The following describes the process:

Step 1. Use the data from 1972:8 to 2007:2 for yield fitting

Step 2. Calculate $\widehat{dfwd}$ by using model-implied and observed rates in the equation,

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

Step 3. Expand the period for the estimation of prediction values to 1972:8 to 2007:3, which adds an additional month at the end of the previous estimation period. Repeat Steps 1 and 2.

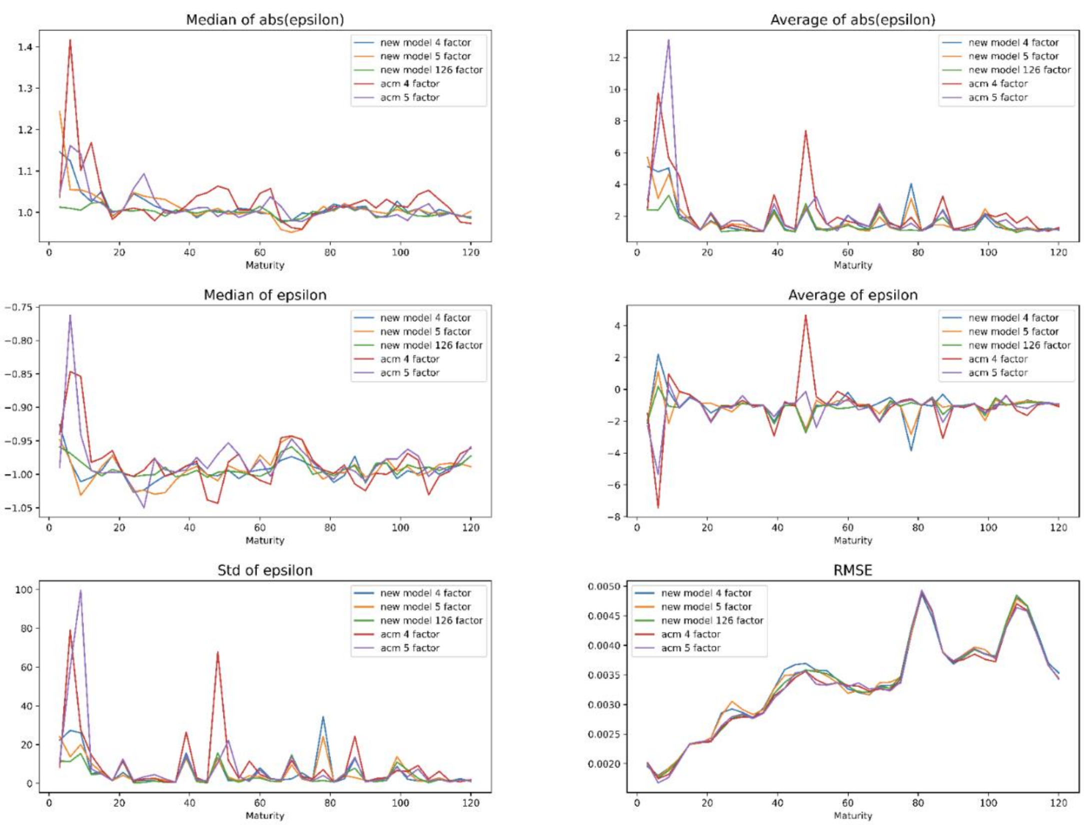


Figure 6. Predictive power of the model using data augmentation

This figure is comparable to Figure 3 and plots the |*t*-values| of beta coefficients generated from regressions to test the predictive power of the model using expanded data. The following equations are used:

1) without control variables

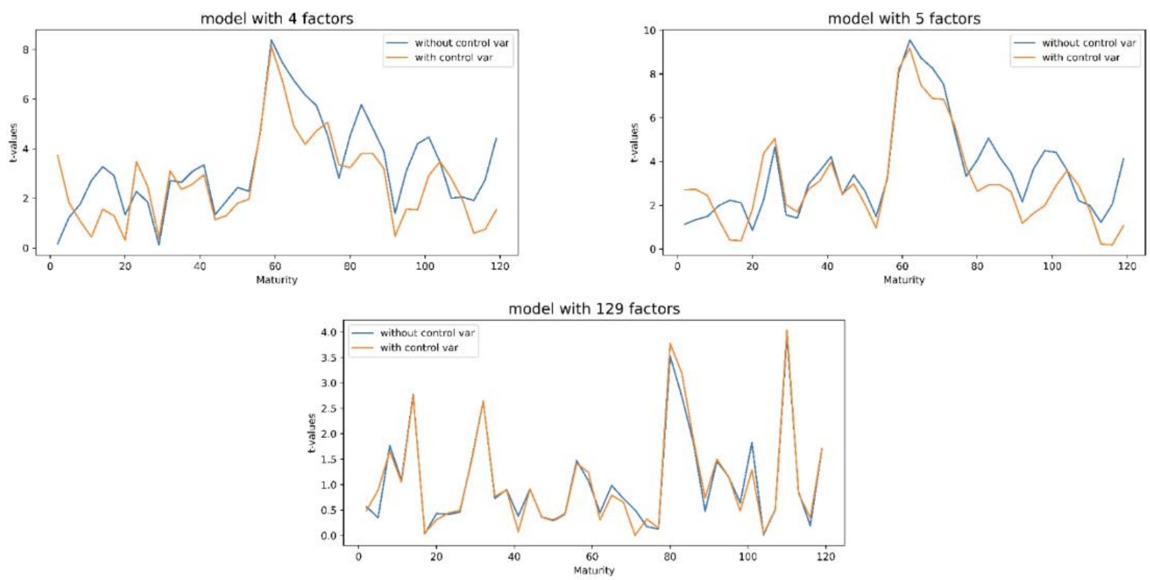
$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \epsilon_t.$$

2) with control variables

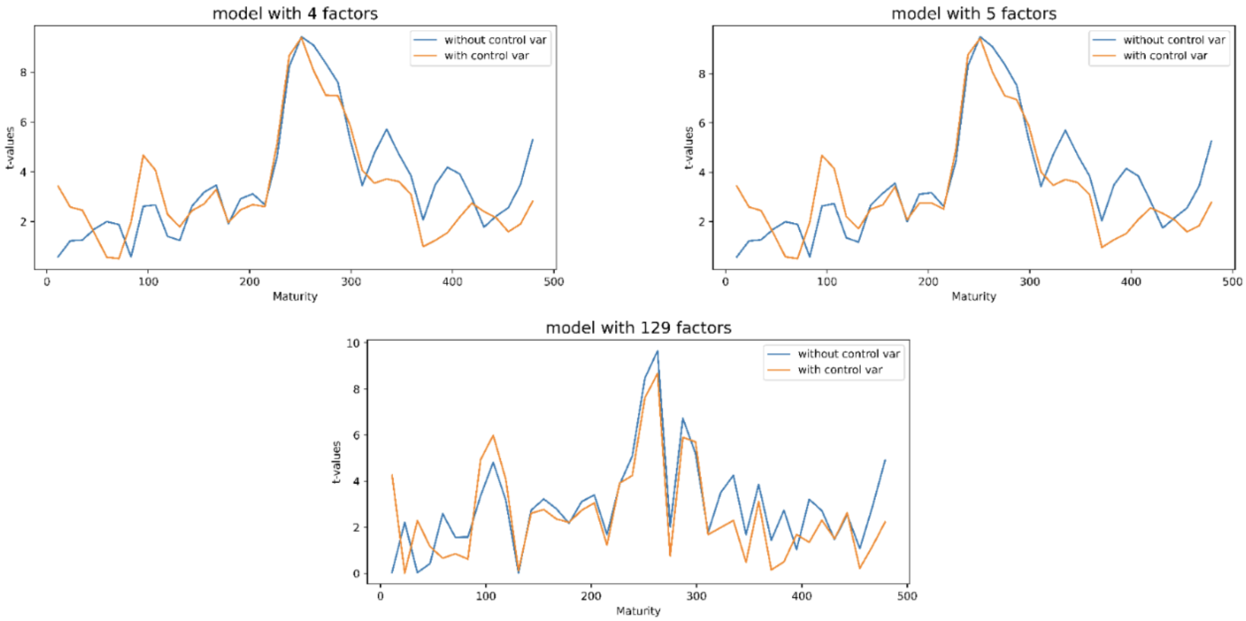
$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

The sample data has a total of 40 data points, consisting of forward rates with *n*-month maturities that are the multiples of 3-months. Through linear interpolation, we augment the data to have 120, 480, and 1,920 data points that are forward rates with 1/*n*-, 0.25/*n*-, and 0.125/*n*-month maturities, respectively. Panel A plots OLS regression results using the data of 120 forward rates implied by our model with four, five, and 129 factors (five yield factors + 124 macro-financial factors) and data augmentation. Panel B and Panel C plot OLS regression results using the data of 480 and 1,920 forward rates implied by our model with four, five, and 129 factors and data augmentation. Panel D plots Lasso regression results using the data of 120, 480 and 1,920 forward rates implied by our model with four, five, and 129 factors and data augmentation. (For Panel D, Lasso alpha (i.e., L1 hyperparameter) is set to 0.01 in line with Figure 8.

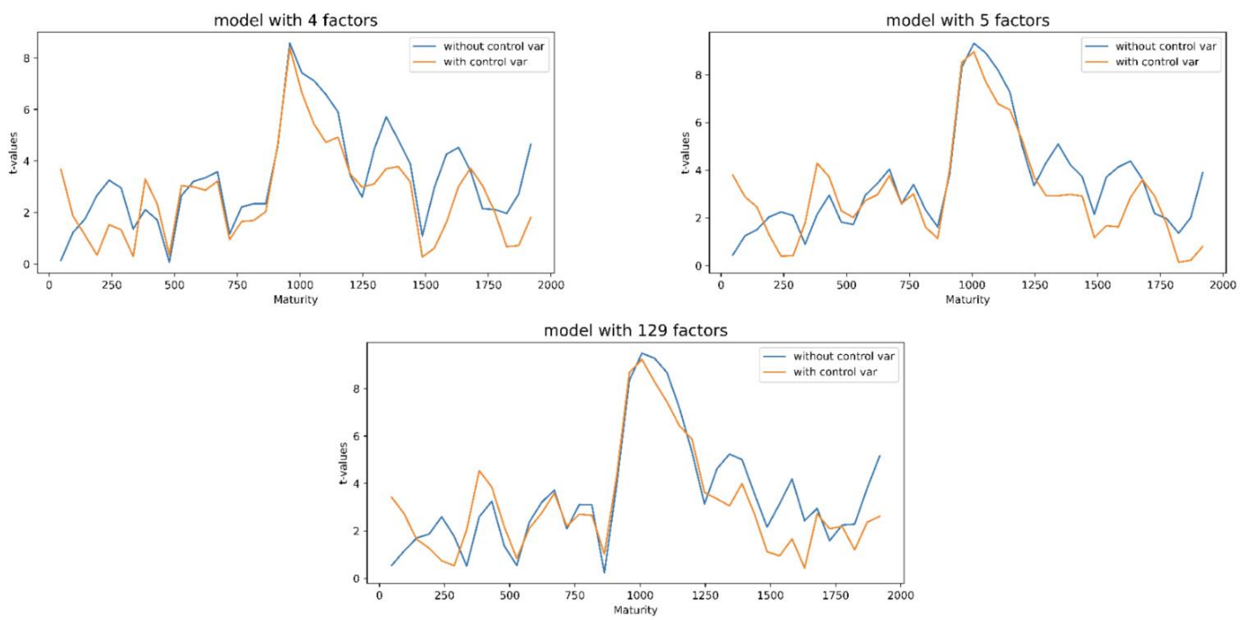
Panel A. |*t*-values| of beta coefficients when using 120 data points



Panel B. |*t*-values| of beta coefficients when using 480 data points



Panel C. |*t*-values| of beta coefficients when using 1,920 data points



Panel D. |*t*-values| of beta coefficients when adding lasso regression with data augmentation

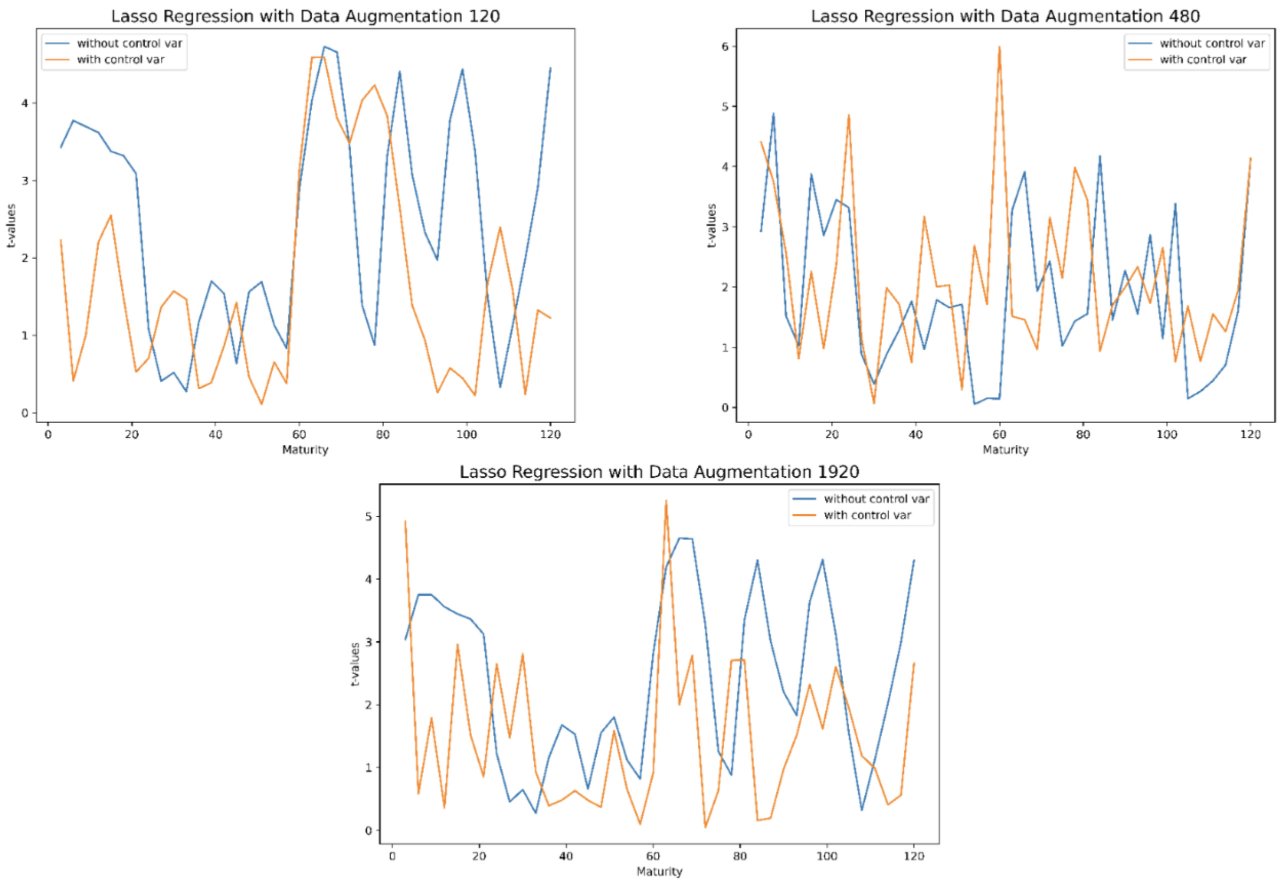


Figure 7. Out-of-sample model performance comparison during COVID-19

This figure plots the estimates of root mean squared error (RMSE) during the COVID-19 pandemic for out-of-sample tests. The original sample data are constructed using Liu and Wu (2021)'s yield curve data. Using this data during the first 414 months, from 1972:8 to 2007:2 (i.e., 70% of the entire sample period), we estimate the prediction values in a rolling method. The following steps describe the process:

Step 1. Use the data from 1972:8 to 2007:2 for yield fitting

Step 2. Calculate $\widehat{dfwd}$ by using model-implied and observed rates in the equation,

$$\Delta f_{observed,t+1} = \alpha + \beta_t (f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

Step 3. Expand the period for the estimation of prediction values to 1972:8 to 2007:3, which adds an additional month at the end of the previous estimation period. Repeat Steps 1 and 2.

For performance comparison, we use the reference date of 11 March 2020, when the COVID-19 pandemic was declared by the World Health Organization (WHO), to set the out-of-sample period. Therefore, the upper five lines in the graph represent the RMSEs during the period of 2007:3 to 2021:12 while the bottom five lines represent the RMSEs during the out-of-sample period of 2020:4 to 2021:12.

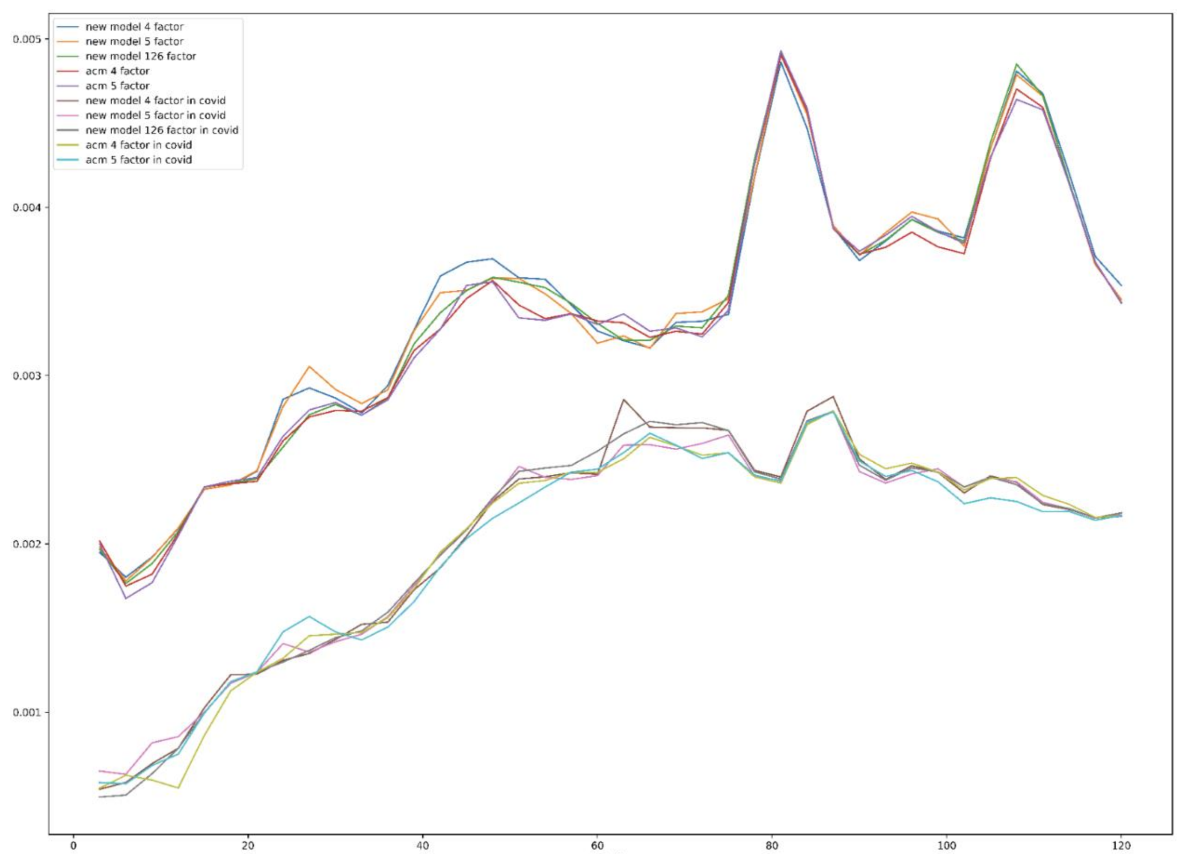


Figure 8. Out-of-sample model performance comparison using Lasso regression

This figure compares the predictability of the ACM and our ATSM during the out-of-sample period using Lasso regression. Panel A plots the RMSEs using the original dataset of 40 maturities in the ACM model with four and five factors and Lasso. Panel B plots the RMSEs using the augmented data, consisting of 120, 480, and 1,920 data points, in the ACM with four and five factors and our proposed model with four, five, and 129 factors. 129 factors include five yield factors as well as 124 macro-financial factors. In Lasso regression, the L1 term (alpha) is a regularizing hyperparameter. The sample period for the initial set is from 1972:8 to 2007:2 with 6:4 splits for training and validation sets to fix the hyperparameter (alpha = 0.01) throughout the tests. We generate prediction values, denoted as $\widehat{dfwd}$, for training set periods in the following steps for rolling estimation:

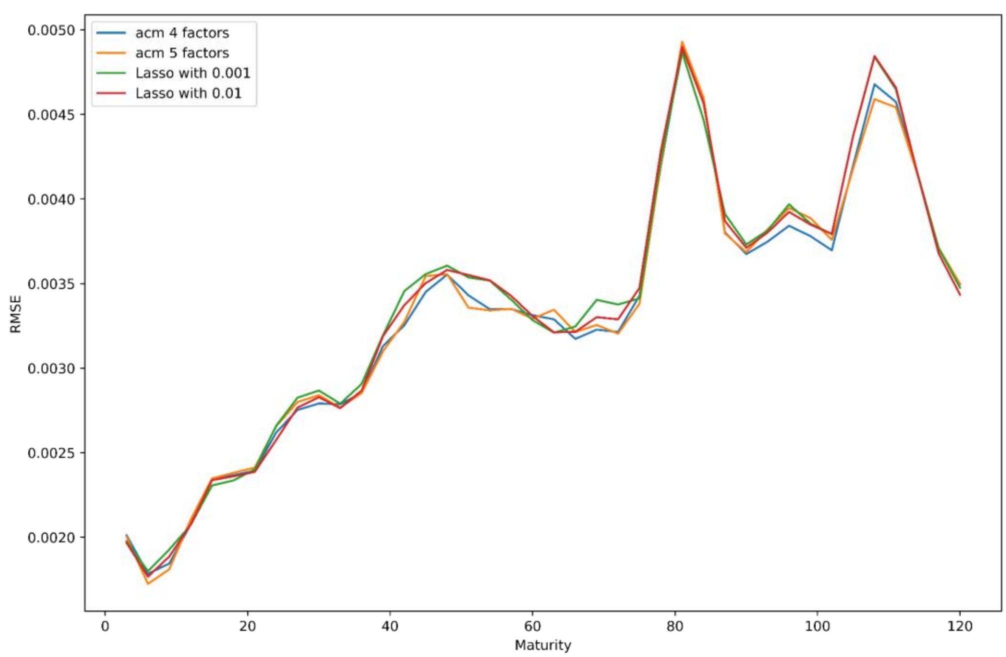
Step 1. Use the data from 1972:8 to 2007:2 for yield fitting.

Step 2. Calculate $\widehat{dfwd}$ by using model-implied and observed rates in the equation,

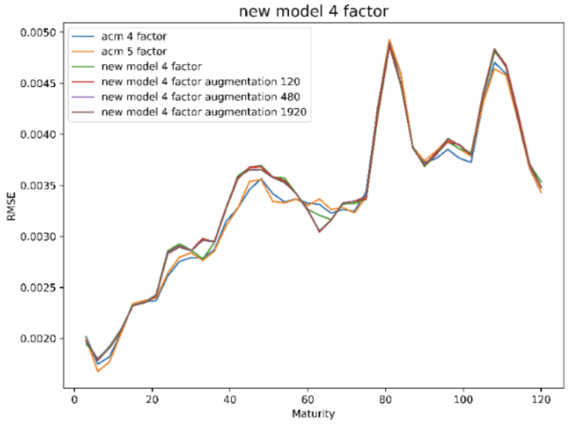
$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

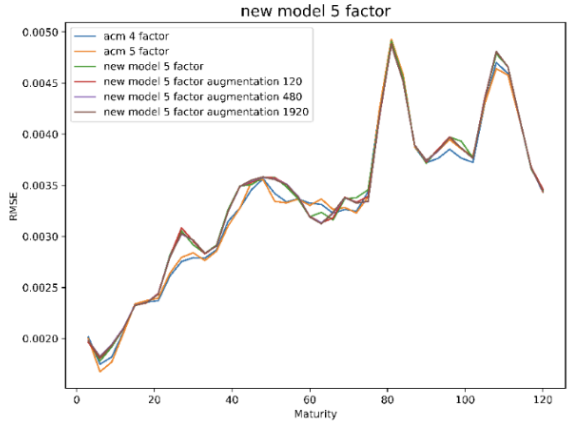
Step 3. Expand the period of the training set to 1972:8 to 2007:3, which adds one month after the end of the previous training set period, and repeat Steps 1 and 2.

Panel A. With the original data of 40 maturities



Panel B. With data augmentation





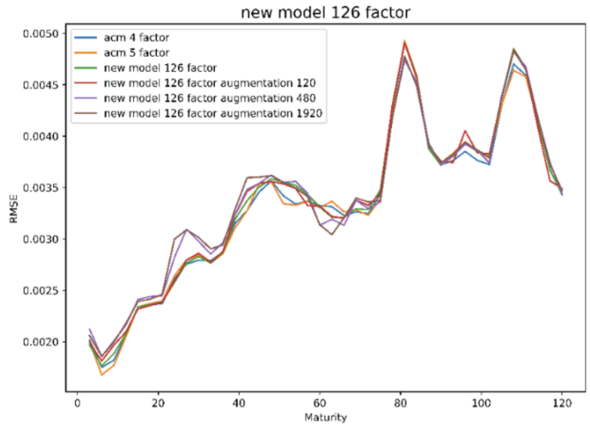


Table 1. Summary statistics of observed and model-implied yields

This table reports the summary statistics of the observed and model-implied yields. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. Panel A shows the summary statistics of the spot rates observed for the same period. Panel B reports the summary statistics of the spot rates generated by five yield factors in our proposed model. Five factors include yield, slope, curvature, and two additional higher-order factors as generated by Adrian, Crump, and Moench (2013). Panel C reports the summary statistics of the spot rates generated by five yield factors in addition to 124 macro-financial factors in our proposed model. The fitted forward rates generated by the model are transformed to spot rates, which are used for the summary statistics to compare with the observed spot rates. For all panels, the number of observations (*count*), average values (*mean*), standard deviations (*std*), minimum values (*min*), 25% (*25%)*, 50% (*50%)*, and 75% (*75%)* percentile values, and maximum values (*max*) are reported.

Panel A. Observed yields

maturities	3	6	9	12	18	24	30	36	60	72	84	120
count	592	592	592	592	592	592	592	592	592	592	592	592
mean	0.045	0.047	0.048	0.049	0.050	0.051	0.052	0.053	0.056	0.058	0.059	0.061
std	0.036	0.036	0.036	0.036	0.036	0.036	0.035	0.035	0.034	0.033	0.032	0.031
min	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.004	0.005
25%	0.012	0.012	0.014	0.015	0.015	0.016	0.018	0.020	0.025	0.028	0.030	0.037
50%	0.049	0.050	0.051	0.052	0.053	0.054	0.055	0.055	0.058	0.058	0.060	0.061
75%	0.066	0.069	0.071	0.071	0.073	0.074	0.075	0.076	0.077	0.078	0.079	0.079
max	0.159	0.161	0.161	0.160	0.159	0.157	0.155	0.156	0.152	0.150	0.150	0.149

Panel B. Model-implied yields using 5 factors

maturities	3	6	9	12	18	24	30	36	60	72	84	120
count	592	592	592	592	592	592	592	592	592	592	592	592
mean	0.045	0.047	0.048	0.049	0.050	0.051	0.052	0.053	0.056	0.057	0.059	0.061
std	0.036	0.036	0.036	0.036	0.036	0.036	0.035	0.035	0.034	0.033	0.032	0.031
min	- 0.008	- 0.006	- 0.004	- 0.003	- 0.003	- 0.002	- 0.002	- 0.000	0.002	0.003	0.004	0.005
25%	0.012	0.013	0.014	0.015	0.016	0.017	0.018	0.020	0.025	0.028	0.031	0.037
50%	0.048	0.050	0.051	0.052	0.054	0.054	0.055	0.056	0.058	0.059	0.059	0.061
75%	0.066	0.069	0.071	0.072	0.073	0.074	0.075	0.076	0.077	0.078	0.079	0.079
max	0.159	0.163	0.164	0.163	0.164	0.161	0.159	0.160	0.156	0.154	0.153	0.149

Panel C. Model-implied yields using 129 factors

maturities	3	6	9	12	18	24	30	36	60	72	84	120
count	592	592	592	592	592	592	592	592	592	592	592	592
mean	0.045	0.047	0.048	0.049	0.050	0.051	0.052	0.053	0.056	0.058	0.059	0.061
std	0.036	0.036	0.036	0.036	0.036	0.036	0.035	0.035	0.034	0.033	0.032	0.031
min	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.004	0.005
25%	0.012	0.012	0.014	0.015	0.015	0.016	0.018	0.020	0.025	0.028	0.030	0.037
50%	0.049	0.050	0.051	0.052	0.053	0.054	0.055	0.055	0.058	0.058	0.060	0.061
75%	0.066	0.069	0.071	0.071	0.073	0.074	0.075	0.076	0.077	0.078	0.079	0.079
max	0.159	0.161	0.161	0.160	0.159	0.157	0.155	0.156	0.152	0.150	0.150	0.149

Table 2. Predictive power of the proposed model

This table presents the ordinary least squares regression results for the predictive power of our proposed model for future bond returns using the following equation:

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

The independent variable is the difference between the observed and model-implied forward rates at t that mature in n months with its coefficient denoted as β. The dependent variable is the difference between the forward rates at t and $t + 1$ that mature in n months. The difference between the forward rates at t-1 and t is used as a control variable and the coefficient is denoted $\boldsymbol{\gamma}$. In the table, we report the t-values for each variable for our model with four, five, and 129 factors including five yield factors and 124 macro-financial factors. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12.

	Four factors		Five factors		129 factors	
maturity	β	γ	β	γ	β	γ
3	-2.718	4.435	-3.552	5.000	3.507	3.507
6	-2.256	4.319	-1.833	3.807	3.789	3.789
9	-1.473	3.669	-2.251	4.012	-3.662	3.662
12	0.041	2.568	-1.247	3.176	-3.541	3.541
18	1.051	2.08	-0.549	2.655	3.261	3.261
24	-3.681	2.934	-4.708	4.107	-1.05	1.05
30	-0.496	0.642	-3.335	1.754	-0.453	0.453
36	-2.421	0.064	-2.251	0.124	1.186	-1.186
60	-8.119	2.035	-7.927	3.102	-2.749	-2.749
72	-4.853	0.683	-6.202	0.866	3.285	-3.285
84	-4.009	0.224	-1.505	-2.406	-4.299	-4.299
120	-2.348	-1.502	-0.792	-2.416	-0.463	-4.502

Table 3. Relative importance of macro-financial variables

This table reports the relative importance of macro-financial variables using Lasso regression. With eigenvalues obtained from conducting PCA on the data of 127 macro-financial variables, we calculate a Lasso coefficient multiplied by the eigenvector for each macro-financial variable, of which the time-series average and standard deviation are used to calculate a *t*-value of each macro-financial variable using the general equation, (Fama-MacBeth) *t*-value = mean / std / $T^{.5}$. Lasso alpha (L1 hyperparameter) is set to 0.001. We fill the yield curve with 1,920 data points per month by our data augmentation; otherwise, the coefficients become zero.

No.	FRED	Category	Description	t-value
1	IPDMAT	Output and Income	IP: Durable Materials	-1.988
2	CES1021000001	Labor Market	All Employees: Mining and Logging: Mining	-2.716
3	USWTRADE	Labor Market	All Employees: Wholesale Trade	2.129
4	AWHMAN	Labor Market	Avg Weekly Hours : Manufacturing	2.216
5	HOUST	Consumption and Orders	Housing Starts: Total New Privately Owned	2.879
6	HOUSTW	Consumption and Orders	Housing Starts, West	3.161
7	PERMIT	Consumption and Orders	New Private Housing Permits (SAAR)	2.721
8	PERMITMW	Consumption and Orders	New Private Housing Permits, Midwest (SAAR)	2.719
9	PERMITW	Consumption and Orders	New Private Housing Permits, West (SAAR)	2.465
10	ACOGNO	Orders and Inventories	New Orders for Consumer Goods	-2.974
11	ISRATIOx	Orders and Inventories	Total Business: Inventories to Sales Ratio	2.056
12	REALLN	Money and Credit	Real Estate Loans at All Commercial Banks	2.102
13	S&P 500	Stock Market	S&P's Common Stock Price Index: Composite	-2.017
14	S&P PE ratio	Stock Market	S&P's Composite Common Stock: Price-Earnings Ratio	-2.029
15	FEDFUNDS	Interest rate and Exchange Rates	Effective Federal Funds Rate	2.394
16	GS1	Interest rate and Exchange Rates	1-Year Treasury Rate	2.558
17	TB6SMFFM	Interest rate and Exchange Rates	6-Month Treasury C Minus FEDFUNDS	-2.036
18	AAAFFM	Interest rate and Exchange Rates	Moody's Aaa Corporate Bond Minus FEDFUNDS	-2.615
19	EXSZUSx	Interest rate and Exchange Rates	Switzerland / U.S. Foreign Exchange Rate	2.320
20	WPSFD49502	Prices	Producer Price Index by Commodity: Final Demand: Personal Consumption Goods	-2.648
21	CUSR0000SAC	Prices	CPI : Commodities	-2.796
22	CUSR0000SAS	Prices	CPI : Services	-2.002
23	PCEPI	Prices	Personal Cons. Expend: Chain Index	-2.703

Table 4. Summary statistics of model-implied rates using alternative models

This table is comparable to Table 2 and reports the summary statistics of the model-implied yields using alternative models for the period of 1972:8 to 2021:2. Panel A reports the summary statistics of the fitted spot rates generated by four yield factors in the model created by Adrian, Crump, and Moench (2013). Four factors include yield, slope, curvature, and CP factor generated by the ACM model. Panel B reports the summary statistics of the fitted spot rates generated by using five yield factors in the ACM model. Five factors include yield, slope, curvature, and two additional higher-order factors as generated by the ACM model. For all panels, the number of observations (*count*), average values (*mean*), standard deviations (*std*), minimum values (*min*), 25% (*25%)*, 50% (*50%)*, and 75% (*75%)* percentile values, and maximum values (*max*) are reported.

Panel A. ACM model with 4 factors

	3	6	9	12	18	24	30	36	60	72	84	120
count	592	592	592	592	592	592	592	592	592	592	592	592
mean	0.045	0.046	0.047	0.048	0.050	0.051	0.052	0.053	0.056	0.057	0.058	0.060
std	0.036	0.036	0.036	0.036	0.036	0.036	0.035	0.035	0.034	0.033	0.032	0.030
min	-0.001	-0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.002	0.003	0.004	0.006
25%	0.012	0.012	0.014	0.014	0.016	0.017	0.018	0.020	0.025	0.028	0.030	0.037
50%	0.049	0.049	0.050	0.051	0.053	0.054	0.055	0.056	0.058	0.058	0.059	0.060
75%	0.067	0.068	0.070	0.070	0.072	0.074	0.075	0.076	0.077	0.078	0.078	0.079
max	0.160	0.160	0.159	0.158	0.157	0.156	0.155	0.155	0.153	0.151	0.149	0.145

Panel B. ACM model with 5 factors

	3	6	9	12	18	24	30	36	60	72	84	120
count	592	592	592	592	592	592	592	592	592	592	592	592
mean	0.045	0.047	0.048	0.049	0.050	0.051	0.052	0.053	0.056	0.057	0.058	0.061
std	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.035	0.034	0.033	0.032	0.031
min	-0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.004	0.006
25%	0.012	0.013	0.013	0.014	0.015	0.016	0.018	0.020	0.025	0.028	0.030	0.037
50%	0.049	0.049	0.050	0.051	0.053	0.054	0.055	0.056	0.058	0.058	0.059	0.061
75%	0.066	0.069	0.071	0.071	0.073	0.074	0.075	0.076	0.077	0.078	0.078	0.079
max	0.159	0.160	0.160	0.160	0.158	0.157	0.157	0.156	0.152	0.150	0.149	0.149

Table 5. Predictive power of alternative models

This table presents regression results for the predictive power of the ACM model with four and four and five factors for future bond returns using the following equation:

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

The independent variable is the difference between the observed and model-implied forward rates at *t* that mature in *n* months. The dependent variable is the difference between the forward rates at *t* and *t* + 1 that mature in *n* months. The difference between the forward rates at *t*-1 and *t* is used as a control variable and the coefficient is denoted, $\boldsymbol{\gamma}$. In the table, we report the *t*-values for each variable for the ACM model with four and five factors. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12.

	ACM with four factors		ACM with five factors	
maturity	β	γ	β	γ
3	-1.462	3.620	-1.207	3.536
6	-4.883	4.483	-3.174	4.064
9	-5.688	4.504	-4.012	4.062
12	-4.650	4.336	-1.753	3.828
18	-1.437	3.363	-0.335	3.274
24	-3.725	1.227	-4.946	1.470
30	-3.643	0.655	-6.079	0.956
36	-1.992	-0.649	-3.317	-0.324
60	-7.564	-0.731	-7.997	-0.851
72	-6.488	-2.608	-6.786	-2.773
84	-1.312	-3.648	-0.480	-3.915
120	-3.087	-3.794	-0.512	-4.318

Table 6. Predictive power of the proposed model after data augmentation

This table presents Lasso regression results for the predictive power of the ACM model with five yield factors and 124 macro-financial factors for future bond returns after data augmentation. The following equation is used:

$$\Delta f_{observed,t+1} = \alpha + \beta_t(f_{observed,t} - f_{model,t}) + \gamma_t \Delta f_{observed,t} + \epsilon_t.$$

The independent variable is the difference between the observed and model-implied forward rates at *t* that mature in *n* months. The dependent variable is the difference between the forward rates at *t* and *t* + 1 that mature in *n* months. The difference between the forward rates at *t*-1 and *t* is used as a control variable. The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. The estimates of the *t*-values of the independent and control variables are reported for the samples with 120, 480, and 1,920 data points after data augmentation. Lasso alpha (L1 hyperparameter) is set to 0.01.

	120 points		480 points		1,920 points	
maturity	β	γ	β	γ	β	γ
3	-2.228	2.349	-4.412	4.825	-4.922	5.230
6	-0.412	0.539	3.747	-2.164	-0.583	0.779
9	1.003	-0.879	-2.573	4.218	1.791	-1.595
12	2.203	-2.078	-0.811	3.481	0.360	-0.131
18	1.487	-1.356	-0.978	1.838	1.513	-1.270
24	0.704	-0.662	-4.863	3.681	2.651	-2.575
30	1.571	-1.550	-0.068	0.234	2.813	-2.773
36	0.312	-0.366	-1.716	1.641	0.392	-0.468
60	-3.129	3.007	5.995	-6.628	-0.910	0.707
72	-3.479	3.324	3.158	-3.857	-0.044	-0.228
84	-2.648	2.467	-0.937	-1.379	-0.158	-0.114
120	1.223	-1.395	4.142	-4.505	2.657	-2.968

Table 7. Nonlinear relationship between macro-financial variables and bond returns (Top 50)

This table reports the *t*-values for the 50 most significant macro-financial variables obtained from Lasso regression using 8,778 variables. The variables include five yield factors, 127 macro-financial variables, and 8,646 interaction variables. The interaction variables are constructed using five yield factors and 127 macro-financial variables. We follow Steps 4-6 introduced under **3.1 Empirical methods** with some variations such as, in Step 4, we calculate the inverse of the diagonal matrix of Σ_t to generate $dz_{v,t}$. With eigenvalues obtained from conducting PCA on the data of 127 macro-financial variables, we calculate a Lasso coefficient multiplied by the eigenvector for each macro-financial variable, of which the time-series average and standard deviation are used to calculate a *t*-value of each macro-financial variable using the general equation, (Fama-MacBeth) *t*-value = mean / std / $T^{.5}$. Lasso alpha (L1 hyperparameter) is set to 0.001.

No.	Variable Name	Description	t-value
1	pc3_IPB51222S	(3rd Yield Factor) * (IP: Residential Utilities)	3.182
2	CMRMTSPLx_CES1021000001	(Real Manu. and Trade Industries Sales) * (All Employees: Mining and Logging: Mining)	3.114
3	HOUSTNE_PERMITNE	(Housing Starts, Northeast) * (New Private Housing Permits, Northeast (SAAR))	3.108
4	HOUSTNE_HOUSTW	(Housing Starts, Northeast) * (Housing Starts, West)	3.029
5	pc3_IPNCONGD	(3rd Yield Factor) * (IP: Nondurable Consumer Goods)	2.970
6	RETAILx_CES1021000001	(Retail and Food Services Sales) * (All Employees: Mining and Logging: Mining)	2.935
7	HWIURATIO_EXCAUSx	(Ratio of Help Wanted/No. Unemployed) * (Canada / U.S. Foreign Exchange Rate)	-2.899
8	HOUSTNE_PERMITW	(Housing Starts, Northeast) * (New Private Housing Permits, West (SAAR))	2.894
9	CES1021000001_PPICMM	(All Employees: Mining and Logging: Mining) * (PPI: Metals and metal products)	2.854
10	FEDFUNDS_BAA	(Effective Federal Funds Rate) * (Moody's Seasoned Baa Corporate Bond Yield)	-2.835
11	HOUSTW_PERMITNE	(Housing Starts, West) * (New Private Housing Permits, Northeast (SAAR))	2.812
12	CE16OV_GS1	(Civilian Employment) * (1-Year Treasury Rate)	2.771
13	CMRMTSPLx_T10YFFM	(Real Manu. and Trade Industries Sales) * (10-Year Treasury C Minus FEDFUNDS)	-2.761
14	BUSINVx_CES0600000008	(Total Business Inventories) * (Avg Hourly Earnings : Goods-Producing)	-2.757
15	PERMITS_S&P div yield	(New Private Housing Permits, South (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.739
16	UEMP15T26_SRVPRD	(Civilians Unemployed for 15-26 Weeks) * (All Employees: Service-Providing Industries)	-2.736

17	AWHMAN_S&P div yield	(Avg Weekly Hours : Manufacturing) * (S&P's Composite Common Stock: Dividend Yield)	2.735
18	CMRMTSPLx_T5YFFM	(Real Manu. and Trade Industries Sales) * (5-Year Treasury C Minus FEDFUNDS)	-2.734
19	pc5_TB3MS	(5th Yield Factor) * (3-Month Treasury Bill)	-2.716
20	BUSLOANS_EXJPUSx	(Commercial and Industrial Loans) * (Japan / U.S. Foreign Exchange Rate)	2.711
21	CMRMTSPLx_AAAFFM	(Real Manu. and Trade Industries Sales) * (Moody's Aaa Corporate Bond Minus FEDFUNDS)	-2.709
22	UEMP27OV_WPSID62	(Civilians Unemployed for 27 Weeks and Over) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	2.708
23	pc2_EXJPUSx	(2nd Yield Factor) * (Japan / U.S. Foreign Exchange Rate)	-2.705
24	NONREVSL_EXCAUSx	(Total Nonrevolving Credit) * (Canada / U.S. Foreign Exchange Rate)	-2.702
25	CES0600000007_S&P div yield	(Avg Weekly Hours : Goods-Producing) * (S&P's Composite Common Stock: Dividend Yield)	2.701
26	USFIRE_BUSLOANS	(All Employees: Financial Activities) * (Commercial and Industrial Loans)	-2.700
27	PERMIT_S&P div yield	(New Private Housing Permits (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.682
28	USGOVT_WPSID62	(All Employees: Government) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	2.682
29	HOUSTS_S&P div yield	(Housing Starts, South) * (S&P's Composite Common Stock: Dividend Yield)	2.680
30	ISRATIOx_EXUSUKx	(Total Business: Inventories to Sales Ratio) * (U.S. / U.K. Foreign Exchange Rate)	-2.672
31	TB3SMFFM_CUSR0000SA0L5	(3-Month Treasury C Minus FEDFUNDS) * (CPI : All items less medical care)	-2.667
32	PERMITW_S&P div yield	(New Private Housing Permits, West (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.667
33	HOUSTNE_PERMITMW	(Housing Starts, Northeast) * (New Private Housing Permits, Midwest (SAAR))	2.663
34	CE16OV_TB6MS	(Civilian Employment) * (6-Month Treasury Bill)	2.663
35	UEMP15T26_PAYEMS	(Civilians Unemployed for 15-26 Weeks) * (All Employees: Total nonfarm)	-2.661
36	PERMITMW_S&P div yield	(New Private Housing Permits, Midwest (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.643
37	HOUSTNE_PERMIT	(Housing Starts, Northeast) * (New Private Housing Permits (SAAR))	2.623

38	CES1021000001_ISRATIOx	(All Employees: Mining and Logging: Mining) * (Total Business: Inventories to Sales Ratio)	-2.622
39	CE16OV_UEMP15OV	(Civilian Employment) * (Civilians Unemployed - 15 Weeks & Over)	-2.616
40	PPICMM_VIXCLSx	(PPI: Metals and metal products) * (CBOE S&P 100 Volatility Index: VXO)	-2.603
41	HOUST_S&P div yield	(Housing Starts: Total New Privately Owned) * (S&P's Composite Common Stock: Dividend Yield)	2.603
42	HOUSTW_S&P div yield	(Housing Starts, West) * (S&P's Composite Common Stock: Dividend Yield)	2.598
43	T10YFFM_EXUSUKx	(10-Year Treasury C Minus FEDFUNDS) * (U.S. / U.K. Foreign Exchange Rate)	-2.595
44	TB3SMFFM_OILPRICEx	(3-Month Treasury C Minus FEDFUNDS) * (Crude Oil, spliced WTI and Cushing)	-2.591
45	BUSINVx_TOTRESNS	(Total Business Inventories) * (Total Reserves of Depository Institutions)	-2.583
46	ACOGNO_DSERRG3M086SBEA	(New Orders for Consumer Goods) * (Personal Cons. Exp: Services)	-2.580
47	PERMITNE_PERMITW	(New Private Housing Permits, Northeast (SAAR)) * (New Private Housing Permits, West (SAAR))	2.573
48	HOUSTNE_S&P div yield	(Housing Starts, Northeast) * (S&P's Composite Common Stock: Dividend Yield)	2.569
49	IPDMAT_CES3000000008	(IP: Durable Materials) * (Avg Hourly Earnings : Manufacturing)	2.558
50	IPDCONGD_USGOVT	(IP: Durable Consumer Goods) * (All Employees: Government)	2.551

Appendix I. The proposed affine model

In line with HJM (Heath, Jarrow & Morton, 1992), let us specify forward rates with maturity *n* at time *t* as:

$$F_{n,t} \equiv F(t, n, n+\Delta) = \exp(-f_{n,t}\Delta).$$
$$F_{n,t+dt} \equiv F(t+dt, n, n+\Delta) = \exp(-f_{n,t+dt}\Delta).$$

$F\,(t, n, n+\Delta)$ is the value of the forward contract maturing at n, determined at *t* (*n* > *t*) that its underlying bond pays one dollar at $n+\Delta$. Therefore, the following relationship holds:

$$F(t, n, n+\Delta) = Z(t, n+\Delta)/Z(t, n).$$

Z(*t*,*n*) is the value of a zero-coupon bond at *t* that pays one dollar at *n* (*n* > *t*).

In addition, our pricing kernel is assumed to be affine as:

$$M_{t+dt} = \exp(-r_t dt - \lambda_t' \lambda_t/2dt - \lambda_t' dz_{\nu,t}).$$
$$dz_{\nu,t} \equiv \Sigma_t^{-.5} d\nu_t \sim N(0, I)\sqrt{dt}.$$
$$\lambda_t \equiv (\lambda_0 + \lambda_1 \nu_t)\Delta.$$

if $d\nu_t$ is from PCA of time-series data, Σ_t is *I*. $dz_{nu,t}$ is the source of risks. Therefore, λ_t denotes the price of the risks. λ_0, λ_1 are estimated with the following regression.

$$\hat{\lambda} = (V^T \cdot V)^{-1} \cdot V^T \cdot L.$$
$$L:\ T^*K\ matrix\ by\ row\text{-}stacked\ \lambda_t.$$
$$V:\ T^*(1{+}K)\ matrix\ by\ row\text{-}stacked\ 1\ and\ \nu_t.$$

K is up to 129 and their interactions in the paper. If *K* > *T*, one can use Lasso.

Our model is called 'affine' because the price of risk (λ_t) is a linear function of risks (ν_t). ν_t is a *K*-dimensional vector of state variables (e.g., macro-financial variables, big data).

The value of entering two offsetting forward contracts is zero, i.e., *E*(*M* · *dF*) = 0. Hence, the definition of pricing kernel implies:

$$E(M_{t+dt} \cdot dF_{n,t}/F_{n,t}) = 0.$$
$$\iff E(\exp(-\lambda_t' \lambda_t/2dt - \lambda_t' dz_{\nu,t} - df_{n,t}\Delta)) = 1.$$

Applying Ito's lemma produces:

$$1 = E\left(1 - \lambda_t'\lambda_t/2 \cdot dt - \lambda_t' dz_{\nu,t} - df_{n,t}\Delta + \left(\lambda_t'\lambda_t/2 \cdot dt + \lambda_t' dz_{\nu,t} + df_{n,t}\Delta\right)^2/2\right)$$

$$\iff E\left(df_{n,t} \cdot \Delta\right) = E\left(-\lambda_t'\lambda_t/2 \cdot dt - \lambda_t' dz_{\nu,t} + \left(\lambda_t'\lambda_t/2 \cdot dt + \lambda_t' dz_{\nu,t} + df_{n,t}\Delta\right)^2/2\right)$$

$$\iff E\left(df_{n,t} \cdot \Delta\right) = E\left(-\lambda_t' dz_{\nu,t} + \left(df_{n,t}\right)^2 \Delta^2/2 + \lambda_t' dz_{\nu,t} \cdot df_{n,t}\Delta\right)$$

To use notations as follows:

$$\sigma_{f,n,t}^2 dt \equiv (df_{n,t})^2.$$
$$\vec{\beta}_{n,t} dt \equiv cov(dz_{\nu,t}, df_{n,t})/\sigma_{f,n,t}^2, \; ie, \; regression \; coefficients.$$

The regression coefficients are similar to those at the first cross-sectional stage of Fama-MacBeth regression (Fama & MacBeth, 1973). Thus, our model becomes:

$$\implies E\left(df_{n,t} \cdot \Delta\right) = E\left(-\lambda_t' dz_{\nu,t} + \sigma_{f,n,t}^2 \cdot dt\Delta^2/2 + \lambda_t' \cdot \vec{\beta_{n,t}} \cdot \sigma_{f,n,t}^2 \Delta dt\right)$$

$$\iff E\left(df_{n,t} \cdot \Delta\right) = E\left[-\lambda_t' dz_{\nu,t} + \sigma_{f,n,t}^2 \left(\Delta/2 + \lambda_t' \cdot \vec{\beta_{n,t}}\right)\Delta dt\right]$$

Then, the empirical design becomes similar to the second time-series stage of Fama-MacBeth regression as:

$$df_{n,t} \cdot \Delta = -\lambda_t' dz_{\nu,t} + \sigma_{f,n,t}^2 \left(\Delta/2 + \lambda_t' \cdot \vec{\beta_{n,t}}\right)\Delta dt + \sigma_n \Delta dz_{n,t}$$

We can forecast a yield curve iteratively by adding $E[df_{n,t}]$ to $f_{n,t}$ because $\sigma_n \Delta dz_{n,t}$ is a cross-sectional measurement error. This is in line with the usual affine model approach in which each yield equation with measurement errors is specified as:

$$\gamma_t^{(\tau)} = A(\tau) + B(\tau)^T x_t + \epsilon_t^{(\tau)}.$$

Matching the volatility terms produces:

$$\left(df_{n,t} \cdot \Delta\right)^2 = \left(-\lambda_t' dz_{\nu,t} + \sigma_{f,n,t}^2 \left(\Delta/2 + \lambda_t' \cdot \vec{\beta_{n,t}}\right)\Delta dt + \sigma_n \Delta dz_{n,t}\right)^2$$

$$\iff \sigma_{f,n,t}^2 dt \cdot \Delta^2 = \lambda_t'\lambda_t dt + \left(\sigma_n \Delta\right)^2 dt - 2\sigma_n \Delta \lambda_t' dt$$

$$\iff \sigma_{f,n,t}^2 \Delta^2 = \lambda_t'\lambda_t + \sigma_n^2 \Delta^2$$

$$\iff \sigma_{f,n,t}^2 = (\lambda_0 + \lambda_1 \nu_t)' (\lambda_0 + \lambda_1 \nu_t) + \sigma_n^2$$

Then, our empirical design becomes:

$$df_{n,t} = -(\lambda_0 + \lambda_1 \nu_t)' dz_{\nu,t} + \kappa_{n,t} \Delta dt + \sigma_n dz_{n,t},$$

where

$$\kappa_{n,t} = \sigma_{f,n,t}^2 \left(1/2 + (\lambda_0 + \lambda_1 \nu_t)' \cdot \vec{\beta_{n,t}} \right)$$

$$\sigma_{f,n,t}^2 = (\lambda_0 + \lambda_1 \nu_t)'(\lambda_0 + \lambda_1 \nu_t) + \sigma_n^2.$$

The model does not have to be linear. For example, any nonlinear models can generate *dz* terms and be matched with *df* terms to estimate the model parameters.

Appendix II. Five-step procedure to balance the panel of macro-financial variables

Step 1. Use the *tcode* from McCracken and Ng (2015) to transform the data.

Step 2. Normalize the outcome from the first step because "observations that are missing are initialized to the unconditional mean based on the non-missing values (which is zero since the data are demeaned and standardized) so that the panel is re-balanced (McCracken and NG, 2015)."

Step 3. Use the generated panel data to obtain factors and loadings before rewriting the missing values with estimates of the lambda times factor.

Step 4. Use the standard deviation and mean estimates obtained in the process of normalization in Step 2 to inverse the normalization to revert to the original data form.

Step 5. Repeat Step 2 to 4 until missing values do not change.

Appendix III. List of macro-financial variables

The table lists all 127 macro-financial variables along with the variable names, descriptions, and *tcodes*, following Ludvigson and Ng (2009). The *tcode* column denotes the following data transformation for a series x:

(1) No transformation
(2) Δx_t
(3) $\Delta^2 x_t$
(4) $\log(x_t)$
(5) $\Delta \log(x_t)$
(6) $\Delta^2 \log(x_t)$
(7) $\Delta(x_t/x_{t-1} - 1.0)$

Category	FRED	Description	tcode
Output and Income	IPDCONGD	IP: Durable Consumer Goods	5
	IPFUELS	IP: Fuels	5
	IPBUSEQ	IP: Business Equipment	5
	IPDMAT	IP: Durable Materials	5
	IPNCONGD	IP: Nondurable Consumer Goods	5
	IPFPNSS	IP: Final Products and Nonindustrial Supplies	5
	IPNMAT	IP: Nondurable Materials	5
	IPCONGD	IP: Consumer Goods	5
	IPMAT	IP: Materials	5
	IPFINAL	IP: Final Products (Market Group)	5
	INDPRO	IP Index	5
	RPI	Real Personal Income	5
	IPB51222S	IP: Residential Utilities	5
	IPMANSICS	IP: Manufacturing (SIC)	5
	W875RX1	Real personal income ex transfer receipts	5
	CUMFNS	Capacity Utilization: Manufacturing	2
Labor Market	UNRATE	Civilian Unemployment Rate	2
	DMANEMP	All Employees: Durable goods	5
	USCONS	All Employees: Construction	5
	AWHMAN	Avg Weekly Hours : Manufacturing	1
	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	5
	USTPU	All Employees: Trade, Transportation & Utilities	5
	PAYEMS	All Employees: Total nonfarm	5
	HWIURATIO	Ratio of Help Wanted/No. Unemployed	2
	CES3000000008	Avg Hourly Earnings : Manufacturing	6
	CES2000000008	Avg Hourly Earnings : Construction	6
	CLF16OV	Civilian Labor Force	5
	NDMANEMP	All Employees: Nondurable goods	5
	CES0600000007	Avg Weekly Hours : Goods-Producing	1
	CE16OV	Civilian Employment	5
	SRVPRD	All Employees: Service-Providing Industries	5
	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	5
	UEMPMEAN	Average Duration of Unemployment (Weeks)	2

	MANEMP	All Employees: Manufacturing	5
	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	5
	CLAIMSx	Initial Claims	5
	UEMP15T26	Civilians Unemployed for 15-26 Weeks	5
	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	5
	USFIRE	All Employees: Financial Activities	5
	USGOOD	All Employees: Goods-Producing Industries	5
	USGOVT	All Employees: Government	5
	USTRADE	All Employees: Retail Trade	5
	CES0600000008	Avg Hourly Earnings : Goods-Producing	6
	USWTRADE	All Employees: Wholesale Trade	5
	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	2
	CES1021000001	All Employees: Mining and Logging: Mining	5
	HWI	Help-Wanted Index for United States	2
Consumption and Orders	HOUSTMW	Housing Starts, Midwest	4
	HOUSTNE	Housing Starts, Northeast	4
	PERMITS	New Private Housing Permits, South (SAAR)	4
	PERMITW	New Private Housing Permits, West (SAAR)	4
	HOUST	Housing Starts: Total New Privately Owned	4
	PERMIT	New Private Housing Permits (SAAR)	4
	HOUSTW	Housing Starts, West	4
	PERMITMW	New Private Housing Permits, Midwest (SAAR)	4
	PERMITNE	New Private Housing Permits, Northeast (SAAR)	4
	HOUSTS	Housing Starts, South	4
Orders and Inventories	UMCSENTx	Consumer Sentiment Index	2
	DPCERA3M086SBEA	Real personal consumption expenditures	5
	RETAILx	Retail and Food Services Sales	5
	AMDMUOx	Unfilled Orders for Durable Goods	5
	BUSINVx	Total Business Inventories	5

	ISRATIOx	Total Business: Inventories to Sales Ratio	2
	ANDENOx	New Orders for Nondefense Capital Goods	5
	ACOGNO	New Orders for Consumer Goods	5
	CMRMTSPLx	Real Manu. and Trade Industries Sales	5
	AMDMNOx	New Orders for Durable Goods	5
Money and Credit	M1SL	M1 Money Stock	6
	DTCTHFNM	Total Consumer Loans and Leases Outstanding	6
	M2REAL	Real M2 Money Stock	5
	INVEST	Securities in Bank Credit at All Commercial Banks	6
	REALLN	Real Estate Loans at All Commercial Banks	6
	M2SL	M2 Money Stock	6
	NONBORRES	Reserves Of Depository Institutions	7
	TOTRESNS	Total Reserves of Depository Institutions	6
	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	6
	BUSLOANS	Commercial and Industrial Loans	6
	NONREVSL	Total Nonrevolving Credit	6
	CONSPI	Nonrevolving consumer credit to Personal Income	2
Interest rate and Exchange Rates	AAA	Moody’s Seasoned Aaa Corporate Bond Yield	2
	EXJPUSx	Japan / U.S. Foreign Exchange Rate	5
	T10YFFM	10-Year Treasury C Minus FEDFUNDS	1
	TB3MS	3-Month Treasury Bill	2
	GS1	1-Year Treasury Rate	2
	BAAFFM	Moody’s Baa Corporate Bond Minus FEDFUNDS	1
	EXCAUSx	Canada / U.S. Foreign Exchange Rate	5
	BAA	Moody’s Seasoned Baa Corporate Bond Yield	2
	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	5
	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	1
	CP3Mx	3-Month AA Financial Commercial Paper Rate	2

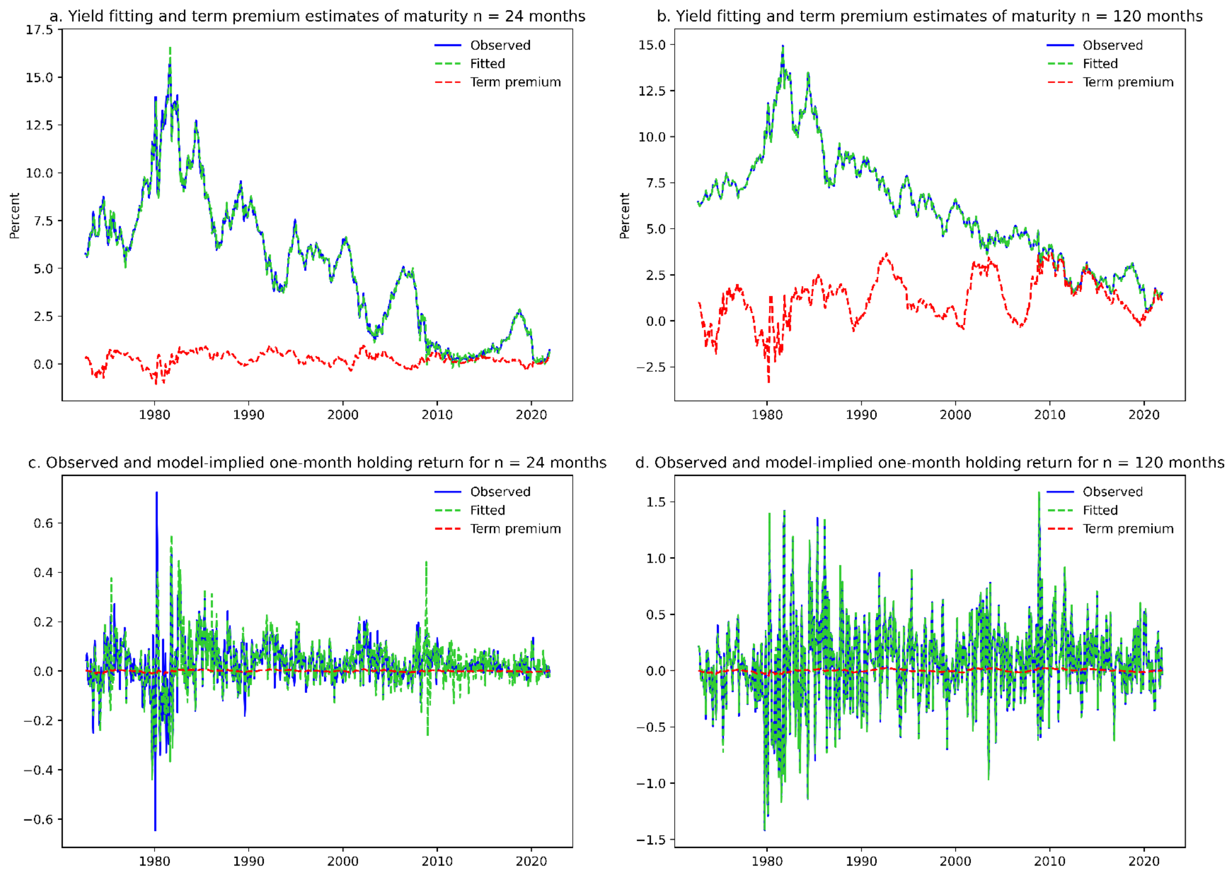
	GS5	5-Year Treasury Rate	2
	T1YFFM	1-Year Treasury C Minus FEDFUNDS	1
	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	1
	FEDFUNDS	Effective Federal Funds Rate	2
	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	1
	GS10	10-Year Treasury Rate	2
	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	1
	TB6MS	6-Month Treasury Bill	2
	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	5
	T5YFFM	5-Year Treasury C Minus FEDFUNDS	1
Prices	CUSR0000SA0L2	CPI : All items less shelter	6
	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	6
	CPIMEDSL	CPI : Medical Care	6
	WPSFD49207	Producer Price Index by Commodity: Final Demand: Finished Goods	6
	WPSID62	Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand	6
	CPIAUCSL	CPI : All Items	6
	CPIAPPSL	CPI : Apparel	6
	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	6
	CUSR0000SA0L5	CPI : All items less medical care	6
	WPSID61	Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand	6
	CUSR0000SAS	CPI : Services	6
	OILPRICEx	Crude Oil, spliced WTI and Cushing	6
	CUSR0000SAD	CPI : Durables	6
	CPITRNSL	CPI : Transportation	6
	PCEPI	Personal Cons. Expend: Chain Index	6
	DSERRG3M086SBEA	Personal Cons. Exp: Services	6
	CPIULFSL	CPI : All Items Less Food	6
	WPSFD49502	Producer Price Index by Commodity: Final Demand: Personal Consumption Goods	6
	PPICMM	PPI: Metals and metal products	6
	CUSR0000SAC	CPI : Commodities	6

Stock Market	S&P: indust	S&P’s Common Stock Price Index: Industrials	5
	S&P div yield	S&P’s Composite Common Stock: Dividend Yield	2
	VIXCLSx	CBOE S&P 100 Volatility Index: VXO	1
	S&P 500	S&P’s Common Stock Price Index: Composite	5
	S&P PE ratio	S&P’s Composite Common Stock: Price-Earnings Ratio	5
Others	TWEXAFEGSMTHx	Nominal Major Currencies U.S. Dollar Index (Goods Only)	5
	BOGMBASE	St. Louis Adjusted Monetary Base	6

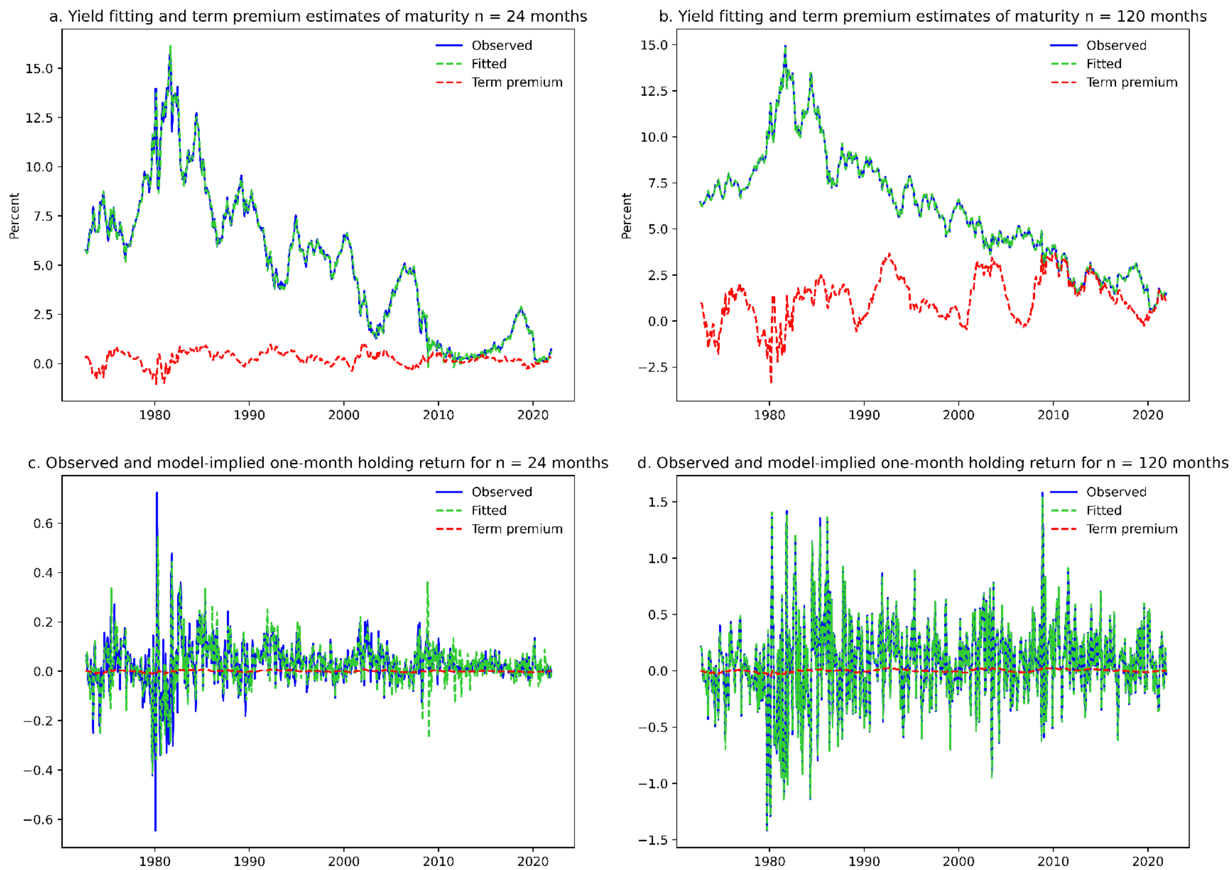
Appendix IV. Time-series fitting and term premia of model-implied yields using four and five yield factors

This figure plots the yield fitting and term premium estimates, as well as predictability for one-month holding period excess returns of zero-coupon yield curve data for Treasuries with two- and ten-year maturities, as observed and implied by the ACM model and our proposed model using four yield factors (Panel A) and five yield factors (Panel B). The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. For both panels, solid lines represent observed yields and returns, dashed green lines represent model-implied yields and returns and dashed red lines represent the model-implied term premia.

Panel A. Time-series fit and term premia of model-implied yields using four yield factors



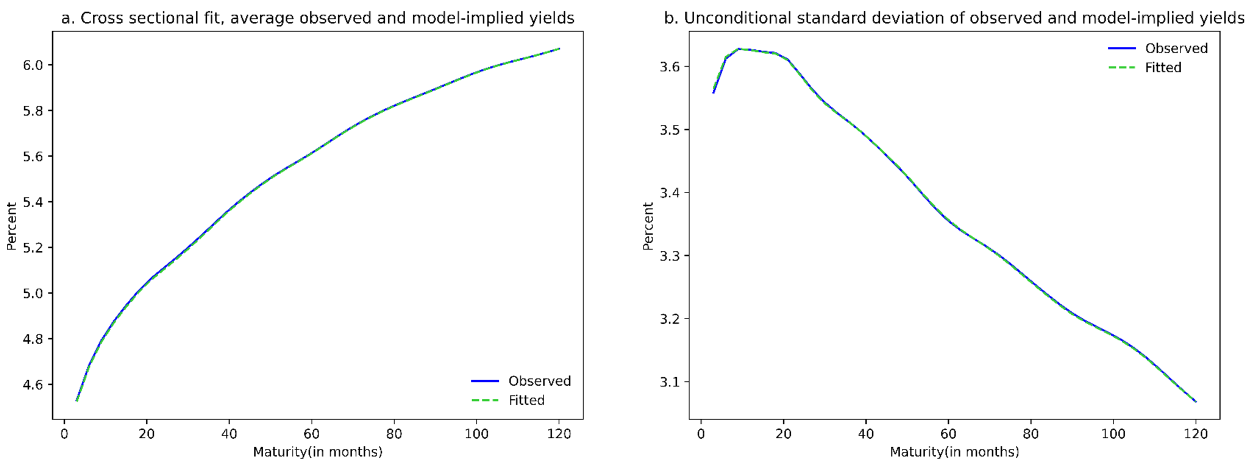
Panel B. Time-series fit and term premia of model-implied yields using five yield factors



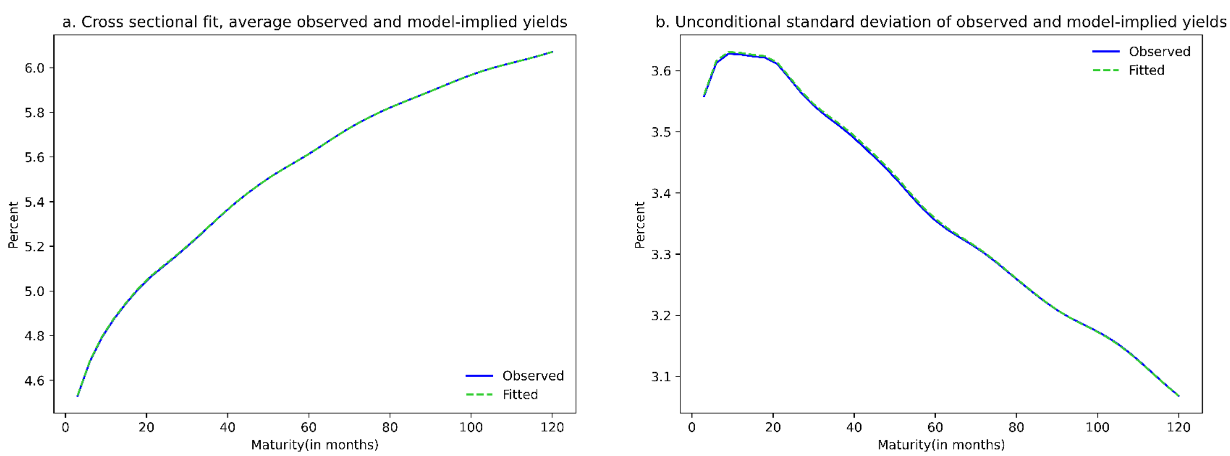
Appendix V. Cross-sectional fit of model-implied yields using four and five yield factors

The figures plot the cross-sectional fit of the yields generated by using our proposed model with four (Panel A) and five factors (Panel B). The sample data are constructed using Liu and Wu's (2021) yield curve data. The sample period is from 1972:8 to 2021:12. For both panels, the graph on the left-hand side plots the unconditional means while the graph on the right-hand side plots the unconditional standard deviations of the observed and model-implied yields.

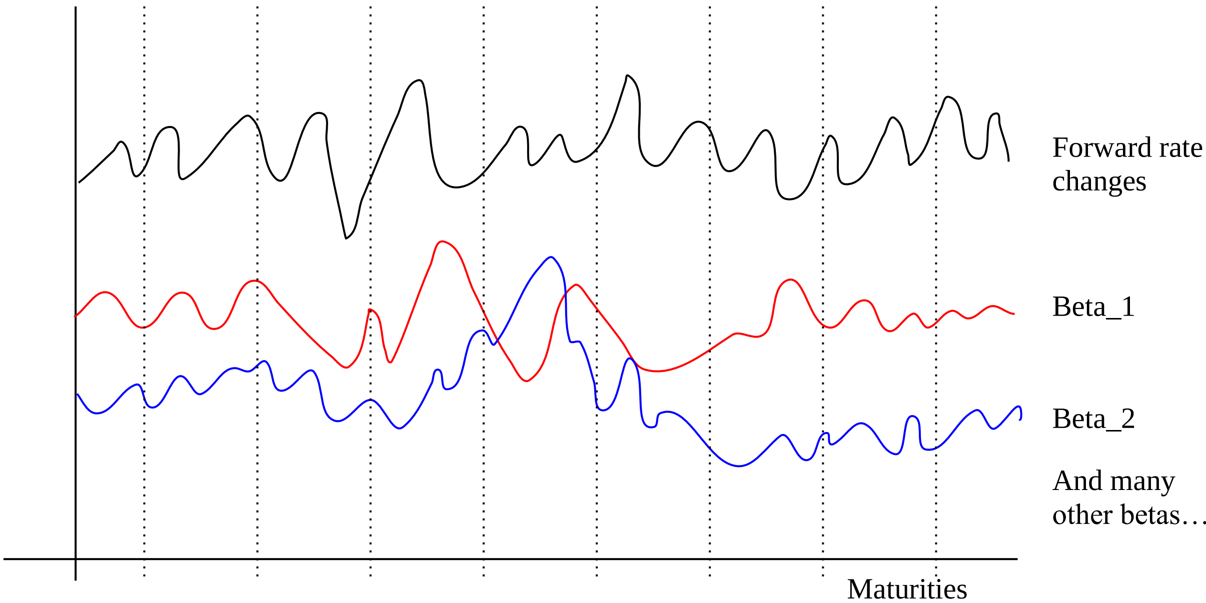
Panel A. Cross-sectional fit of model-implied yields using four yield factors



Panel B. Cross-sectional fit of model-implied yields using five yield factors



Appendix VI. Intuition for data augmentation



The above figure shows the snapshot for three waves at time *t*. The black wave plots the relation between maturities and forward rate changes. The red wave plots the movement of Beta_1, the sensitivity of forward rate changes to macro-financial variable_1. The blue wave plots the movement of Beta_2, the sensitivity of forward rate changes to macro-financial variable_2. Let us assume that the series of Beta_*i* (*i* = 1 .. *N*) exist, which the above figure does not show due to the space limit. The vertical dotted lines are the maturities for which forward rate data are available (i.e., observables maturities).

Our goal is to cross-sectionally fit the forward rate changes with the waves of betas or factor loadings. Our simple functional analysis is as follows.

☐ Suppose the number of macro-financial variables (betas) is small. First, regress the forward rate changes on betas using the data observed at the vertical dotted lines (observable maturities). Second, compute the time-series average of the regression coefficients (Fama-MacBeth approach).

☐ Suppose the number of macroeconomic variables is large. First, connect observed forward rates at time *t* to create a continuous wave. This is our "data augmentation". For instance, we simply connect the dots with line segments or other generative algorithms. Similarly, connect observed forward rates at time *t*-1.

Second, generate a wave of forward rate changes by subtracting the second forward-rate wave from the first forward-rate wave. Iterate this process for all *t*.

Third, select many maturities (points on the X-axis) whether data are available or not ("augmented maturities"). At each *t* and at each augmented maturity, identify the augmented changes of forward rates.

Fourth, regress the "augmented forward rate changes" on macro-financial variables one by one. This creates "augmented betas". This paper regresses the augmented changes on the principal components of macro-financial variables to alleviate the concern about univariate regressions (e.g., missing variables). This step creates *waves of betas*. The larger the number of macro-financial variables included, the larger the number of beta waves (e.g., 129 and their interactions in our model).

Fifth, cross-sectionally regress the augmented forward rate changes on the augmented betas at augmented plus observed maturities. We experiment with Lasso regression at this step to identify irrelevant betas. If the size of the data is too small for the cross-sectional regressions, return to the third step to create more artificial data at more augmented maturities.

Sixth, compute the time-series average of the (Lasso) regression coefficients as Fama-MacBeth regression to identify which macro-financial variables matter to explain yield-curve dynamics.

Alternatively, undo PCA at the fourth step and recover the (Lasso) regression coefficients for each macro-financial variable at each t. Next, average the coefficients in the time series for Fama-MacBeth regression.

Appendix VII. Nonlinear relationship between macro-financial variables and bond returns

No.	Variable Name	Description	t-value
1	pc3_IPB51222S	(3rd Yield Factor) * (IP: Residential Utilities)	3.182
2	CMRMTSPLx_CES1021000001	(Real Manu. and Trade Industries Sales) * (All Employees: Mining and Logging: Mining)	3.114
3	HOUSTNE_PERMITNE	(Housing Starts, Northeast) * (New Private Housing Permits, Northeast (SAAR))	3.108
4	HOUSTNE_HOUSTW	(Housing Starts, Northeast) * (Housing Starts, West)	3.029
5	pc3_IPNCONGD	(3rd Yield Factor) * (IP: Nondurable Consumer Goods)	2.970
6	RETAILx_CES1021000001	(Retail and Food Services Sales) * (All Employees: Mining and Logging: Mining)	2.935
7	HWIURATIO_EXCAUSx	(Ratio of Help Wanted/No. Unemployed) * (Canada / U.S. Foreign Exchange Rate)	-2.899
8	HOUSTNE_PERMITW	(Housing Starts, Northeast) * (New Private Housing Permits, West (SAAR))	2.894
9	CES1021000001_PPICMM	(All Employees: Mining and Logging: Mining) * (PPI: Metals and metal products)	2.854
10	FEDFUNDS_BAA	(Effective Federal Funds Rate) * (Moody’s Seasoned Baa Corporate Bond Yield)	-2.835
11	HOUSTW_PERMITNE	(Housing Starts, West) * (New Private Housing Permits, Northeast (SAAR))	2.812
12	CE16OV_GS1	(Civilian Employment) * (1-Year Treasury Rate)	2.771
13	CMRMTSPLx_T10YFFM	(Real Manu. and Trade Industries Sales) * (10-Year Treasury C Minus FEDFUNDS)	-2.761
14	BUSINVx_CES0600000008	(Total Business Inventories) * (Avg Hourly Earnings : Goods-Producing)	-2.757
15	PERMITS_S&P div yield	(New Private Housing Permits, South (SAAR)) * (S&P’s Composite Common Stock: Dividend Yield)	2.739
16	UEMP15T26_SRVPRD	(Civilians Unemployed for 15-26 Weeks) * (All Employees: Service-Providing Industries)	-2.736
17	AWHMAN_S&P div yield	(Avg Weekly Hours : Manufacturing) * (S&P’s Composite Common Stock: Dividend Yield)	2.735

18	CMRMTSPLx_T5YFFM	(Real Manu. and Trade Industries Sales) * (5-Year Treasury C Minus FEDFUNDS)	-2.734
19	pc5_TB3MS	(5th Yield Factor) * (3-Month Treasury Bill)	-2.716
20	BUSLOANS_EXJPUSx	(Commercial and Industrial Loans) * (Japan / U.S. Foreign Exchange Rate)	2.711
21	CMRMTSPLx_AAAFFM	(Real Manu. and Trade Industries Sales) * (Moody's Aaa Corporate Bond Minus FEDFUNDS)	-2.709
22	UEMP27OV_WPSID62	(Civilians Unemployed for 27 Weeks and Over) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	2.708
23	pc2_EXJPUSx	(2nd Yield Factor) * (Japan / U.S. Foreign Exchange Rate)	-2.705
24	NONREVSL_EXCAUSx	(Total Nonrevolving Credit) * (Canada / U.S. Foreign Exchange Rate)	-2.702
25	CES0600000007_S&P div yield	(Avg Weekly Hours : Goods-Producing) * (S&P's Composite Common Stock: Dividend Yield)	2.701
26	USFIRE_BUSLOANS	(All Employees: Financial Activities) * (Commercial and Industrial Loans)	-2.700
27	PERMIT_S&P div yield	(New Private Housing Permits (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.682
28	USGOVT_WPSID62	(All Employees: Government) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	2.682
29	HOUSTS_S&P div yield	(Housing Starts, South) * (S&P's Composite Common Stock: Dividend Yield)	2.680
30	ISRATIOx_EXUSUKx	(Total Business: Inventories to Sales Ratio) * (U.S. / U.K. Foreign Exchange Rate)	-2.672
31	TB3SMFFM_CUSR0000SA0L5	(3-Month Treasury C Minus FEDFUNDS) * (CPI : All items less medical care)	-2.667
32	PERMITW_S&P div yield	(New Private Housing Permits, West (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.667
33	HOUSTNE_PERMITMW	(Housing Starts, Northeast) * (New Private Housing Permits, Midwest (SAAR))	2.663
34	CE16OV_TB6MS	(Civilian Employment) * (6-Month Treasury Bill)	2.663
35	UEMP15T26_PAYEMS	(Civilians Unemployed for 15-26 Weeks) * (All Employees: Total nonfarm)	-2.661

36	PERMITMW_S&P div yield	(New Private Housing Permits, Midwest (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.643
37	HOUSTNE_PERMIT	(Housing Starts, Northeast) * (New Private Housing Permits (SAAR))	2.623
38	CES1021000001_ISRATIOx	(All Employees: Mining and Logging: Mining) * (Total Business: Inventories to Sales Ratio)	-2.622
39	CE16OV_UEMP15OV	(Civilian Employment) * (Civilians Unemployed - 15 Weeks & Over)	-2.616
40	PPICMM_VIXCLSx	(PPI: Metals and metal products) * (CBOE S&P 100 Volatility Index: VXO)	-2.603
41	HOUST_S&P div yield	(Housing Starts: Total New Privately Owned) * (S&P's Composite Common Stock: Dividend Yield)	2.603
42	HOUSTW_S&P div yield	(Housing Starts, West) * (S&P's Composite Common Stock: Dividend Yield)	2.598
43	T10YFFM_EXUSUKx	(10-Year Treasury C Minus FEDFUNDS) * (U.S. / U.K. Foreign Exchange Rate)	-2.595
44	TB3SMFFM_OILPRICEx	(3-Month Treasury C Minus FEDFUNDS) * (Crude Oil, spliced WTI and Cushing)	-2.591
45	BUSINVx_TOTRESNS	(Total Business Inventories) * (Total Reserves of Depository Institutions)	-2.583
46	ACOGNO_DSERRG3M086SBEA	(New Orders for Consumer Goods) * (Personal Cons. Exp: Services)	-2.580
47	PERMITNE_PERMITW	(New Private Housing Permits, Northeast (SAAR)) * (New Private Housing Permits, West (SAAR))	2.573
48	HOUSTNE_S&P div yield	(Housing Starts, Northeast) * (S&P's Composite Common Stock: Dividend Yield)	2.569
49	IPDMAT_CES3000000008	(IP: Durable Materials) * (Avg Hourly Earnings : Manufacturing)	2.558
50	IPDCONGD_USGOVT	(IP: Durable Consumer Goods) * (All Employees: Government)	2.551
51	EXUSUKx_PPICMM	(U.S. / U.K. Foreign Exchange Rate) * (PPI: Metals and metal products)	-2.551
52	HOUSTMW_S&P div yield	(Housing Starts, Midwest) * (S&P's Composite Common Stock: Dividend Yield)	2.550
53	CMRMTSPLx_BAAFFM	(Real Manu. and Trade Industries Sales) * (Moody's Baa Corporate Bond Minus FEDFUNDS)	-2.548
54	UMCSENTx_DTCTHFNM	(Consumer Sentiment Index) * (Total Consumer Loans and Leases Outstanding)	-2.538
55	UEMP15T26_BUSLOANS	(Civilians Unemployed for 15-26 Weeks) * (Commercial and Industrial Loans)	2.538

56	pc2_IPB51222S	(2nd Yield Factor) * (IP: Residential Utilities)	-2.530
57	TB6SMFFM_CUSR0000SA0L5	(6-Month Treasury C Minus FEDFUNDS) * (CPI : All items less medical care)	-2.519
58	HOUST_HOUSTNE	(Housing Starts: Total New Privately Owned) * (Housing Starts, Northeast)	2.508
59	UEMP27OV_M2SL	(Civilians Unemployed for 27 Weeks and Over) * (M2 Money Stock)	2.499
60	REALLN_WPSID61	(Real Estate Loans at All Commercial Banks) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand)	2.499
61	HOUSTNE_HOUSTMW	(Housing Starts, Northeast) * (Housing Starts, Midwest)	2.498
62	DDURRG3M086SBEA_CES0600000008	(Personal Cons. Exp: Durable goods) * (Avg Hourly Earnings : Goods-Producing)	-2.490
63	UEMP5TO14_WPSFD49207	(Civilians Unemployed for 5-14 Weeks) * (Producer Price Index by Commodity: Final Demand: Finished Goods)	-2.489
64	HOUSTNE	Housing Starts, Northeast	2.488
65	USGOVT_DDURRG3M086SBEA	(All Employees: Government) * (Personal Cons. Exp: Durable goods)	2.481
66	IPBUSEQ_USCONS	(IP: Business Equipment) * (All Employees: Construction)	-2.469
67	EXUSUKx_WPSFD49207	(U.S. / U.K. Foreign Exchange Rate) * (Producer Price Index by Commodity: Final Demand: Finished Goods)	-2.465
68	HOUST_PERMITNE	(Housing Starts: Total New Privately Owned) * (New Private Housing Permits, Northeast (SAAR))	2.458
69	W875RX1_EXCAUSx	(Real personal income ex transfer receipts) * (Canada / U.S. Foreign Exchange Rate)	-2.456
70	IPFPNSS_USCONS	(IP: Final Products and Nonindustrial Supplies) * (All Employees: Construction)	-2.455
71	IPFINAL_USCONS	(IP: Final Products (Market Group)) * (All Employees: Construction)	-2.451
72	IPCONGD_USCONS	(IP: Consumer Goods) * (All Employees: Construction)	-2.450
73	PERMITNE_S&P div yield	(New Private Housing Permits, Northeast (SAAR)) * (S&P's Composite Common Stock: Dividend Yield)	2.447
74	UNRATE_BOGMBASE	(Civilian Unemployment Rate) * (St. Louis Adjusted Monetary Base)	-2.443
75	UEMPLT5_BUSINVx	(Civilians Unemployed - Less Than 5 Weeks) * (Total Business Inventories)	-2.438

76	CES1021000001_USCONS	(All Employees: Mining and Logging: Mining) * (All Employees: Construction)	2.432
77	PERMITNE	New Private Housing Permits, Northeast (SAAR)	2.425
78	AAAFFM_EXUSUKx	(Moody’s Aaa Corporate Bond Minus FEDFUNDS) * (U.S. / U.K. Foreign Exchange Rate)	-2.409
79	UNRATE_UEMP15T26	(Civilian Unemployment Rate) * (Civilians Unemployed for 15-26 Weeks)	2.406
80	UEMP27OV_CPIAUCSL	(Civilians Unemployed for 27 Weeks and Over) * (CPI : All Items)	2.401
81	pc5_CUSR0000SAS	(5th Yield Factor) * (CPI : Services)	2.400
82	UEMP5TO14_CES1021000001	(Civilians Unemployed for 5-14 Weeks) * (All Employees: Mining and Logging: Mining)	-2.400
83	CPITRNSL_CPIMEDSL	(CPI : Transportation) * (CPI : Medical Care)	-2.388
84	BOGMBASE_CONSPI	(St. Louis Adjusted Monetary Base) * (Nonrevolving consumer credit to Personal Income)	2.387
85	IPB51222S_S&P: indust	(IP: Residential Utilities) * (S&P’s Common Stock Price Index: Industrials)	-2.387
86	PERMITNE_PERMITMW	(New Private Housing Permits, Northeast (SAAR)) * (New Private Housing Permits, Midwest (SAAR))	2.385
87	CE16OV_UEMP15T26	(Civilian Employment) * (Civilians Unemployed for 15-26 Weeks)	-2.383
88	IPDMAT_UEMP15OV	(IP: Durable Materials) * (Civilians Unemployed - 15 Weeks & Over)	-2.374
89	UEMP15OV_BUSLOANS	(Civilians Unemployed - 15 Weeks & Over) * (Commercial and Industrial Loans)	2.368
90	BOGMBASE_CES0600000008	(St. Louis Adjusted Monetary Base) * (Avg Hourly Earnings : Goods-Producing)	2.359
91	FEDFUNDS_COMPAPFFx	(Effective Federal Funds Rate) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.356
92	IPMANSICS_UEMP15OV	(IP: Manufacturing (SIC)) * (Civilians Unemployed - 15 Weeks & Over)	-2.352
93	UEMP15OV_SRVPRD	(Civilians Unemployed - 15 Weeks & Over) * (All Employees: Service-Providing Industries)	-2.347
94	UNRATE_CUSR0000SA0L5	(Civilian Unemployment Rate) * (CPI : All items less medical care)	2.344
95	ISRATIOx_EXSZUSx	(Total Business: Inventories to Sales Ratio) * (Switzerland / U.S. Foreign Exchange Rate)	2.339
96	DMANEMP_CES3000000008	(All Employees: Durable goods) * (Avg Hourly Earnings : Manufacturing)	2.339

97	ISRATIOx_GS5	(Total Business: Inventories to Sales Ratio) * (5-Year Treasury Rate)	-2.338
98	CLAIMSx_EXJPUSx	(Initial Claims) * (Japan / U.S. Foreign Exchange Rate)	-2.336
99	S&P 500_T10YFFM	(S&P's Common Stock Price Index: Composite) * (10-Year Treasury C Minus FEDFUNDS)	-2.331
100	EXUSUKx_WPSFD49502	(U.S. / U.K. Foreign Exchange Rate) * (Producer Price Index by Commodity: Final Demand: Personal Consumption Goods)	-2.324
101	CES1021000001_USTPU	(All Employees: Mining and Logging: Mining) * (All Employees: Trade, Transportation & Utilities)	2.323
102	S&P 500_BAAFFM	(S&P's Common Stock Price Index: Composite) * (Moody's Baa Corporate Bond Minus FEDFUNDS)	-2.320
103	BUSINVx_CES2000000008	(Total Business Inventories) * (Avg Hourly Earnings : Construction)	-2.319
104	HOUSTMW_PERMITNE	(Housing Starts, Midwest) * (New Private Housing Permits, Northeast (SAAR))	2.316
105	UEMP15OV_PAYEMS	(Civilians Unemployed - 15 Weeks & Over) * (All Employees: Total nonfarm)	-2.310
106	S&P PE ratio_CES3000000008	(S&P's Composite Common Stock: Price-Earnings Ratio) * (Avg Hourly Earnings : Manufacturing)	2.308
107	TB6SMFFM_CUSR0000SAS	(6-Month Treasury C Minus FEDFUNDS) * (CPI : Services)	-2.307
108	BAAFFM_EXUSUKx	(Moody's Baa Corporate Bond Minus FEDFUNDS) * (U.S. / U.K. Foreign Exchange Rate)	-2.306
109	NONBORRES_T10YFFM	(Reserves Of Depository Institutions) * (10-Year Treasury C Minus FEDFUNDS)	-2.304
110	T5YFFM_EXUSUKx	(5-Year Treasury C Minus FEDFUNDS) * (U.S. / U.K. Foreign Exchange Rate)	-2.300
111	pc2_IPNCONGD	(2nd Yield Factor) * (IP: Nondurable Consumer Goods)	-2.300
112	REALLN_DTCTHFNM	(Real Estate Loans at All Commercial Banks) * (Total Consumer Loans and Leases Outstanding)	-2.296
113	UEMP5TO14_CES0600000008	(Civilians Unemployed for 5-14 Weeks) * (Avg Hourly Earnings : Goods-Producing)	-2.294
114	PERMIT_PERMITNE	(New Private Housing Permits (SAAR)) * (New Private Housing Permits, Northeast (SAAR))	2.287

115	S&P div yield_BAAFFM	(S&P's Composite Common Stock: Dividend Yield) * (Moody's Baa Corporate Bond Minus FEDFUNDS)	2.285
116	CES1021000001_EXUSUKx	(All Employees: Mining and Logging: Mining) * (U.S. / U.K. Foreign Exchange Rate)	-2.282
117	UEMP15OV_S&P PE ratio	(Civilians Unemployed - 15 Weeks & Over) * (S&P's Composite Common Stock: Price-Earnings Ratio)	-2.282
118	T10YFFM_EXJPUSx	(10-Year Treasury C Minus FEDFUNDS) * (Japan / U.S. Foreign Exchange Rate)	2.276
119	UEMP27OV_CUSR0000SA0L2	(Civilians Unemployed for 27 Weeks and Over) * (CPI : All items less shelter)	2.272
120	TB6SMFFM_OILPRICEx	(6-Month Treasury C Minus FEDFUNDS) * (Crude Oil, spliced WTI and Cushing)	-2.270
121	IPNCONGD_USCONS	(IP: Nondurable Consumer Goods) * (All Employees: Construction)	-2.267
122	DPCERA3M086SBEA_S&P div yield	(Real personal consumption expenditures) * (S&P's Composite Common Stock: Dividend Yield)	-2.264
123	INDPRO_UEMP15OV	(IP Index) * (Civilians Unemployed - 15 Weeks & Over)	-2.252
124	AWHMAN_HOUSTNE	(Avg Weekly Hours : Manufacturing) * (Housing Starts, Northeast)	2.240
125	CES0600000007_HOUSTNE	(Avg Weekly Hours : Goods-Producing) * (Housing Starts, Northeast)	2.239
126	IPNCONGD_BAAFFM	(IP: Nondurable Consumer Goods) * (Moody's Baa Corporate Bond Minus FEDFUNDS)	2.239
127	UNRATE_PCEPI	(Civilian Unemployment Rate) * (Personal Cons. Expend: Chain Index)	2.238
128	HWIURATIO_CES1021000001	(Ratio of Help Wanted/No. Unemployed) * (All Employees: Mining and Logging: Mining)	2.238
129	UEMP15T26_USTRADE	(Civilians Unemployed for 15-26 Weeks) * (All Employees: Retail Trade)	-2.235
130	UEMP27OV_CUSR0000SA0L5	(Civilians Unemployed for 27 Weeks and Over) * (CPI : All items less medical care)	2.234
131	W875RX1_UEMP15OV	(Real personal income ex transfer receipts) * (Civilians Unemployed - 15 Weeks & Over)	-2.229
132	NONREVSL_CONSPI	(Total Nonrevolving Credit) * (Nonrevolving consumer credit to Personal Income)	-2.229
133	M2SL_M2REAL	(M2 Money Stock) * (Real M2 Money Stock)	2.228

134	DPCERA3M086SBEA_CES1021000001	(Real personal consumption expenditures) * (All Employees: Mining and Logging: Mining)	2.222
135	UNRATE_WPSID62	(Civilian Unemployment Rate) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	2.219
136	CUMFNS_CES3000000008	(Capacity Utilization: Manufacturing) * (Avg Hourly Earnings : Manufacturing)	2.219
137	NONBORRES_UMCSENTx	(Reserves Of Depository Institutions) * (Consumer Sentiment Index)	2.219
138	TB6SMFFM_CPIULFSL	(6-Month Treasury C Minus FEDFUNDS) * (CPI : All Items Less Food)	-2.203
139	CMRMTSPLx_EXSZUSx	(Real Manu. and Trade Industries Sales) * (Switzerland / U.S. Foreign Exchange Rate)	-2.202
140	DPCERA3M086SBEA_S&P 500	(Real personal consumption expenditures) * (S&P's Common Stock Price Index: Composite)	2.201
141	CLF16OV_TB6MS	(Civilian Labor Force) * (6-Month Treasury Bill)	2.198
142	USFIRE_CPIMEDSL	(All Employees: Financial Activities) * (CPI : Medical Care)	-2.197
143	IPDMAT_NONREVSL	(IP: Durable Materials) * (Total Nonrevolving Credit)	-2.193
144	DPCERA3M086SBEA_HWIURATIO	(Real personal consumption expenditures) * (Ratio of Help Wanted/No. Unemployed)	2.188
145	IPDMAT_HWIURATIO	(IP: Durable Materials) * (Ratio of Help Wanted/No. Unemployed)	2.187
146	CES0600000007_PERMITNE	(Avg Weekly Hours : Goods-Producing) * (New Private Housing Permits, Northeast (SAAR))	2.185
147	M2REAL_DDURRG3M086SBEA	(Real M2 Money Stock) * (Personal Cons. Exp: Durable goods)	-2.184
148	USGOOD_CES1021000001	(All Employees: Goods-Producing Industries) * (All Employees: Mining and Logging: Mining)	2.183
149	IPMANSICS_CES3000000008	(IP: Manufacturing (SIC)) * (Avg Hourly Earnings : Manufacturing)	2.177
150	BOGMBASE_CES2000000008	(St. Louis Adjusted Monetary Base) * (Avg Hourly Earnings : Construction)	2.164
151	S&P div yield_PPICMM	(S&P's Composite Common Stock: Dividend Yield) * (PPI: Metals and metal products)	-2.159
152	FEDFUNDS_DTCOLNVHFNM	(Effective Federal Funds Rate) * (Consumer Motor Vehicle Loans Outstanding)	2.156
153	TB3SMFFM_CPIULFSL	(3-Month Treasury C Minus FEDFUNDS) * (CPI : All Items Less Food)	-2.154

154	MANEMP_CES3000000008	(All Employees: Manufacturing) * (Avg Hourly Earnings : Manufacturing)	2.154
155	RPI_BOGMBASE	(Real Personal Income) * (St. Louis Adjusted Monetary Base)	-2.153
156	CMRMTSPLx_CUSR0000SA0L5	(Real Manu. and Trade Industries Sales) * (CPI : All items less medical care)	-2.151
157	GS1_CPIMEDSL	(1-Year Treasury Rate) * (CPI : Medical Care)	2.150
158	pc3_T1YFFM	(3rd Yield Factor) * (1-Year Treasury C Minus FEDFUNDS)	-2.148
159	IPB51222S_BOGMBASE	(IP: Residential Utilities) * (St. Louis Adjusted Monetary Base)	2.148
160	TB6SMFFM_CPIMEDSL	(6-Month Treasury C Minus FEDFUNDS) * (CPI : Medical Care)	2.147
161	CES1021000001_USGOVT	(All Employees: Mining and Logging: Mining) * (All Employees: Government)	2.143
162	M1SL_EXSZUSx	(M1 Money Stock) * (Switzerland / U.S. Foreign Exchange Rate)	-2.135
163	ISRATIOx_TWEXAFEGSMTHx	(Total Business: Inventories to Sales Ratio) * (Nominal Major Currencies U.S. Dollar Index (Goods Only))	2.121
164	S&P: indust_DSERRG3M086SBEA	(S&P’s Common Stock Price Index: Industrials) * (Personal Cons. Exp: Services)	-2.120
165	IPFPNSS_EXCAUSx	(IP: Final Products and Nonindustrial Supplies) * (Canada / U.S. Foreign Exchange Rate)	-2.119
166	DDURRG3M086SBEA_INVEST	(Personal Cons. Exp: Durable goods) * (Securities in Bank Credit at All Commercial Banks)	2.115
167	IPBUSEQ_HWI	(IP: Business Equipment) * (Help-Wanted Index for United States)	-2.114
168	UEMPLT5_GS10	(Civilians Unemployed - Less Than 5 Weeks) * (10-Year Treasury Rate)	-2.113
169	GS5_CPIAPPSL	(5-Year Treasury Rate) * (CPI : Apparel)	2.112
170	USWTRADE_VIXCLSx	(All Employees: Wholesale Trade) * (CBOE S&P 100 Volatility Index: VXO)	2.110
171	AWHMAN_PERMITNE	(Avg Weekly Hours : Manufacturing) * (New Private Housing Permits, Northeast (SAAR))	2.108
172	UEMPLT5_AAA	(Civilians Unemployed - Less Than 5 Weeks) * (Moody’s Seasoned Aaa Corporate Bond Yield)	-2.106
173	HOUSTW_PERMITW	(Housing Starts, West) * (New Private Housing Permits, West (SAAR))	2.105
174	CPIAUCSL_CUSR0000SAS	(CPI : All Items) * (CPI : Services)	2.104
175	pc4_GS1	(4th Yield Factor) * (1-Year Treasury Rate)	-2.101
176	CMRMTSPLx_NONREVSL	(Real Manu. and Trade Industries Sales) * (Total Nonrevolving Credit)	-2.097

177	CUMFNS_USCONS	(Capacity Utilization: Manufacturing) * (All Employees: Construction)	-2.092
178	BAAFFM_PPICMM	(Moody's Baa Corporate Bond Minus FEDFUNDS) * (PPI: Metals and metal products)	-2.089
179	PERMITNE_COMPAPFFx	(New Private Housing Permits, Northeast (SAAR)) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.085
180	HWI_CES0600000008	(Help-Wanted Index for United States) * (Avg Hourly Earnings : Goods-Producing)	-2.084
181	UNRATE_CPIAUCSL	(Civilian Unemployment Rate) * (CPI : All Items)	2.084
182	pc4_CES0600000007	(4th Yield Factor) * (Avg Weekly Hours : Goods-Producing)	2.083
183	S&P 500_DSERRG3M086SBEA	(S&P's Common Stock Price Index: Composite) * (Personal Cons. Exp: Services)	-2.082
184	CUMFNS_UEMP15OV	(Capacity Utilization: Manufacturing) * (Civilians Unemployed - 15 Weeks & Over)	-2.082
185	CES0600000007_COMPAPFFx	(Avg Weekly Hours : Goods-Producing) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.076
186	USWTRADE_NONBORRES	(All Employees: Wholesale Trade) * (Reserves Of Depository Institutions)	2.075
187	TB3SMFFM_CPIMEDSL	(3-Month Treasury C Minus FEDFUNDS) * (CPI : Medical Care)	2.074
188	IPNCONGD_USGOOD	(IP: Nondurable Consumer Goods) * (All Employees: Goods-Producing Industries)	-2.074
189	IPNCONGD_AAAFFM	(IP: Nondurable Consumer Goods) * (Moody's Aaa Corporate Bond Minus FEDFUNDS)	2.073
190	COMPAPFFx_WPSID61	(3-Month Commercial Paper Minus FEDFUNDS) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand)	2.072
191	PERMITW_COMPAPFFx	(New Private Housing Permits, West (SAAR)) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.071
192	UEMP27OV_PCEPI	(Civilians Unemployed for 27 Weeks and Over) * (Personal Cons. Expend: Chain Index)	2.071
193	S&P 500_AAAFFM	(S&P's Common Stock Price Index: Composite) * (Moody's Aaa Corporate Bond Minus FEDFUNDS)	-2.070
194	IPCONGD_PCEPI	(IP: Consumer Goods) * (Personal Cons. Expend: Chain Index)	-2.070

195	M2REAL_AAA	(Real M2 Money Stock) * (Moody's Seasoned Aaa Corporate Bond Yield)	-2.068
196	HOUSTMW_CPIMEDSL	(Housing Starts, Midwest) * (CPI : Medical Care)	-2.068
197	USFIRE_EXUSUKx	(All Employees: Financial Activities) * (U.S. / U.K. Foreign Exchange Rate)	-2.067
198	UEMPLT5_AMDMUOx	(Civilians Unemployed - Less Than 5 Weeks) * (Unfilled Orders for Durable Goods)	-2.065
199	IPMAT_UEMPMEAN	(IP: Materials) * (Average Duration of Unemployment (Weeks))	-2.064
200	PERMIT_COMPAPFFx	(New Private Housing Permits (SAAR)) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.064
201	PERMITS_COMPAPFFx	(New Private Housing Permits, South (SAAR)) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.064
202	DPCERA3M086SBEA_S&P: indust	(Real personal consumption expenditures) * (S&P's Common Stock Price Index: Industrials)	2.064
203	UNRATE_CPIULFSL	(Civilian Unemployment Rate) * (CPI : All Items Less Food)	2.063
204	IPBUSEQ_CES3000000008	(IP: Business Equipment) * (Avg Hourly Earnings : Manufacturing)	2.060
205	CES1021000001_USTRADE	(All Employees: Mining and Logging: Mining) * (All Employees: Retail Trade)	2.059
206	UNRATE_CUSR0000SA0L2	(Civilian Unemployment Rate) * (CPI : All items less shelter)	2.059
207	S&P div yield_WPSID61	(S&P's Composite Common Stock: Dividend Yield) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand)	-2.056
208	pc4_AWOTMAN	(4th Yield Factor) * (Avg Weekly Overtime Hours : Manufacturing)	-2.054
209	UEMPMEAN_S&P div yield	(Average Duration of Unemployment (Weeks)) * (S&P's Composite Common Stock: Dividend Yield)	-2.053
210	IPNMAT_CPIMEDSL	(IP: Nondurable Materials) * (CPI : Medical Care)	2.050
211	PERMITS_S&P 500	(New Private Housing Permits, South (SAAR)) * (S&P's Common Stock Price Index: Composite)	-2.048
212	AWHMAN_COMPAPFFx	(Avg Weekly Hours : Manufacturing) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.046
213	IPFPNSS_TB6MS	(IP: Final Products and Nonindustrial Supplies) * (6-Month Treasury Bill)	2.045

214	FEDFUNDS_AAA	(Effective Federal Funds Rate) * (Moody's Seasoned Aaa Corporate Bond Yield)	-2.045
215	UNRATE_DTCTHFNM	(Civilian Unemployment Rate) * (Total Consumer Loans and Leases Outstanding)	-2.044
216	T1YFFM_CUSR0000SAS	(1-Year Treasury C Minus FEDFUNDS) * (CPI : Services)	-2.043
217	AMDMUOx_EXSZUSx	(Unfilled Orders for Durable Goods) * (Switzerland / U.S. Foreign Exchange Rate)	-2.042
218	HOUSTNE_COMPAPFFx	(Housing Starts, Northeast) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.040
219	USTRADE_ANDENOx	(All Employees: Retail Trade) * (New Orders for Nondefense Capital Goods)	-2.039
220	UEMP5TO14_DTCTHFNM	(Civilians Unemployed for 5-14 Weeks) * (Total Consumer Loans and Leases Outstanding)	-2.039
221	UNRATE_S&P PE ratio	(Civilian Unemployment Rate) * (S&P's Composite Common Stock: Price-Earnings Ratio)	-2.038
222	IPMANSICS_UEMP15T26	(IP: Manufacturing (SIC)) * (Civilians Unemployed for 15-26 Weeks)	-2.034
223	USGOOD_CES3000000008	(All Employees: Goods-Producing Industries) * (Avg Hourly Earnings : Manufacturing)	2.034
224	W875RX1_HWIURATIO	(Real personal income ex transfer receipts) * (Ratio of Help Wanted/No. Unemployed)	2.033
225	HOUST_COMPAPFFx	(Housing Starts: Total New Privately Owned) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.032
226	pc5_MANEMP	(5th Yield Factor) * (All Employees: Manufacturing)	-2.032
227	GS10_CPIAPPSL	(10-Year Treasury Rate) * (CPI : Apparel)	2.031
228	UEMP15OV_USTRADE	(Civilians Unemployed - 15 Weeks & Over) * (All Employees: Retail Trade)	-2.029
229	HOUSTS_COMPAPFFx	(Housing Starts, South) * (3-Month Commercial Paper Minus FEDFUNDS)	-2.029
230	IPDCONGD_WPSID61	(IP: Durable Consumer Goods) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand)	-2.028
231	pc4_T10YFFM	(4th Yield Factor) * (10-Year Treasury C Minus FEDFUNDS)	2.028
232	IPBUSEQ_CPIAPPSL	(IP: Business Equipment) * (CPI : Apparel)	-2.025

233	HWIURATIO_S&P: indust	(Ratio of Help Wanted/No. Unemployed) * (S&P's Common Stock Price Index: Industrials)	2.025
234	pc4_PERMITS	(4th Yield Factor) * (New Private Housing Permits, South (SAAR))	2.023
235	S&P div yield_CES3000000008	(S&P's Composite Common Stock: Dividend Yield) * (Avg Hourly Earnings : Manufacturing)	-2.023
236	pc4_T5YFFM	(4th Yield Factor) * (5-Year Treasury C Minus FEDFUNDS)	2.021
237	IPBUSEQ_DTCOLNVHFNM	(IP: Business Equipment) * (Consumer Motor Vehicle Loans Outstanding)	2.020
238	PERMITW_S&P 500	(New Private Housing Permits, West (SAAR)) * (S&P's Common Stock Price Index: Composite)	-2.020
239	pc2_M1SL	(2nd Yield Factor) * (M1 Money Stock)	-2.019
240	pc5_UEMP27OV	(5th Yield Factor) * (Civilians Unemployed for 27 Weeks and Over)	2.019
241	pc2_TWEXAFEGSMTHx	(2nd Yield Factor) * (Nominal Major Currencies U.S. Dollar Index (Goods Only))	-2.017
242	IPMANSICS_DTCOLNVHFNM	(IP: Manufacturing (SIC)) * (Consumer Motor Vehicle Loans Outstanding)	2.017
243	CLAIMSx_CES1021000001	(Initial Claims) * (All Employees: Mining and Logging: Mining)	-2.017
244	HWIURATIO_GS5	(Ratio of Help Wanted/No. Unemployed) * (5-Year Treasury Rate)	2.016
245	PERMITNE_M2REAL	(New Private Housing Permits, Northeast (SAAR)) * (Real M2 Money Stock)	2.016
246	PERMITNE_CPIMEDSL	(New Private Housing Permits, Northeast (SAAR)) * (CPI : Medical Care)	-2.011
247	pc1_T1YFFM	(1st Yield Factor) * (1-Year Treasury C Minus FEDFUNDS)	-2.010
248	S&P div yield_CPIMEDSL	(S&P's Composite Common Stock: Dividend Yield) * (CPI : Medical Care)	-2.009
249	IPFINAL_TB6MS	(IP: Final Products (Market Group)) * (6-Month Treasury Bill)	2.006
250	PERMIT_S&P 500	(New Private Housing Permits (SAAR)) * (S&P's Common Stock Price Index: Composite)	-2.004
251	AAA_CUSR0000SAC	(Moody's Seasoned Aaa Corporate Bond Yield) * (CPI : Commodities)	2.003
252	UEMPLT5_UEMP5TO14	(Civilians Unemployed - Less Than 5 Weeks) * (Civilians Unemployed for 5-14 Weeks)	2.002
253	pc1_CES0600000008	(1st Yield Factor) * (Avg Hourly Earnings : Goods-Producing)	-2.002
254	IPFUELS_S&P div yield	(IP: Fuels) * (S&P's Composite Common Stock: Dividend Yield)	-2.000
255	USFIRE_CUSR0000SA0L5	(All Employees: Financial Activities) * (CPI : All items less medical care)	1.996

256	AWHMAN_S&P 500	(Avg Weekly Hours : Manufacturing) * (S&P's Common Stock Price Index: Composite)	-1.996
257	IPNMAT_DDURRG3M086SBEA	(IP: Nondurable Materials) * (Personal Cons. Exp: Durable goods)	-1.994
258	BAAFFM_EXJPUSx	(Moody's Baa Corporate Bond Minus FEDFUNDS) * (Japan / U.S. Foreign Exchange Rate)	1.989
259	RETAILx_HWIURATIO	(Retail and Food Services Sales) * (Ratio of Help Wanted/No. Unemployed)	1.989
260	UEMP5TO14_WPSFD49502	(Civilians Unemployed for 5-14 Weeks) * (Producer Price Index by Commodity: Final Demand: Personal Consumption Goods)	-1.987
261	HOUSTW_COMPAPFFx	(Housing Starts, West) * (3-Month Commercial Paper Minus FEDFUNDS)	-1.987
262	M1SL_WPSID62	(M1 Money Stock) * (Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand)	-1.986
263	CUSR0000SAC_PCEPI	(CPI : Commodities) * (Personal Cons. Expend: Chain Index)	-1.984
264	pc4_BAAFFM	(4th Yield Factor) * (Moody's Baa Corporate Bond Minus FEDFUNDS)	1.984
265	WPSID61_DSERRG3M086SBEA	(Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand) * (Personal Cons. Exp: Services)	-1.984
266	IPB51222S_S&P 500	(IP: Residential Utilities) * (S&P's Common Stock Price Index: Composite)	-1.981
267	S&P: indust_CP3Mx	(S&P's Common Stock Price Index: Industrials) * (3-Month AA Financial Commercial Paper Rate)	-1.981
268	CES0600000007_S&P 500	(Avg Weekly Hours : Goods-Producing) * (S&P's Common Stock Price Index: Composite)	-1.980
269	IPCONGD_TB6MS	(IP: Consumer Goods) * (6-Month Treasury Bill)	1.979
270	S&P PE ratio_CP3Mx	(S&P's Composite Common Stock: Price-Earnings Ratio) * (3-Month AA Financial Commercial Paper Rate)	-1.978
271	IPFUELS_S&P 500	(IP: Fuels) * (S&P's Common Stock Price Index: Composite)	1.978
272	IPNMAT_ANDENOx	(IP: Nondurable Materials) * (New Orders for Nondefense Capital Goods)	-1.977
273	HOUST_S&P 500	(Housing Starts: Total New Privately Owned) * (S&P's Common Stock Price Index: Composite)	-1.976
274	T5YFFM_EXJPUSx	(5-Year Treasury C Minus FEDFUNDS) * (Japan / U.S. Foreign Exchange Rate)	1.976

275	M1SL_CUSR0000SAD	(M1 Money Stock) * (CPI : Durables)	-1.976
276	ISRATIOx_AAA	(Total Business: Inventories to Sales Ratio) * (Moody’s Seasoned Aaa Corporate Bond Yield)	-1.971
277	PERMITMW_COMPAPFFx	(New Private Housing Permits, Midwest (SAAR)) * (3-Month Commercial Paper Minus FEDFUNDS)	-1.970
278	USCONS_EXUSUKx	(All Employees: Construction) * (U.S. / U.K. Foreign Exchange Rate)	-1.969
279	USFIRE_AMDMNOx	(All Employees: Financial Activities) * (New Orders for Durable Goods)	1.966
280	CUMFNS_UEMP15T26	(Capacity Utilization: Manufacturing) * (Civilians Unemployed for 15-26 Weeks)	-1.965
281	IPFPNSS_CES1021000001	(IP: Final Products and Nonindustrial Supplies) * (All Employees: Mining and Logging: Mining)	1.962
282	S&P PE ratio_DSERRG3M086SBEA	(S&P’s Composite Common Stock: Price-Earnings Ratio) * (Personal Cons. Exp: Services)	-1.962
283	CLF16OV_GS1	(Civilian Labor Force) * (1-Year Treasury Rate)	1.961
284	HOUSTW_CPIMEDSL	(Housing Starts, West) * (CPI : Medical Care)	-1.961