



Publicly Accessible Penn Dissertations

---

2017

# Applied Dynamic Factor Modeling In Finance

Ross Maxwell Askanazi

University of Pennsylvania, [askross@sas.upenn.edu](mailto:askross@sas.upenn.edu)

Follow this and additional works at: <https://repository.upenn.edu/edissertations>



Part of the [Economics Commons](#)

---

## Recommended Citation

Askanazi, Ross Maxwell, "Applied Dynamic Factor Modeling In Finance" (2017). *Publicly Accessible Penn Dissertations*. 2918.  
<https://repository.upenn.edu/edissertations/2918>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2918>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Applied Dynamic Factor Modeling In Finance

## **Abstract**

In this dissertation, I study model misspecification in applications of dynamic factor models to finance. In Chapter 1, my co-author Jacob Warren and I examine factors for volatility of equities. Historical literature on the subject decomposes volatility into a factor component and an idiosyncratic remainder. Recent work has suggested that idiosyncratic volatility of US equities data has a factor structure, with the factor highly correlated with, and possibly precisely the market volatility. In this paper we attempt to characterize the underlying factor and find that it can be decomposed into a statistical (PCA) and structural (market volatility) factor. We also show that this feature is not unique to equities, appearing in diverse sets of financial data. Lastly, we find that this dual-factor approach is slightly dominated in forecasting environments by a single statistical factor, suggesting that accurate measurement of the factors provides a direction for future work. In Chapter 2, I explore the use of dynamic factor models in yield curve forecasting and an exploration of the spanning hypothesis – that is, whether all information necessary for forecasting yields is contained in the current yield curve. Only linear tests of the spanning hypothesis are typically conducted in the literature, and the results are subject to substantial disagreement. In this paper, I explore a key modern nonlinearity, namely the zero lower bound (ZLB). I first demonstrate in simulation that only very small nonlinearities in the measurement equation are necessary to break down the assumed linear spanning relationship. Because bond yields are determined by forward-looking behavior of investors, the effect of the ZLB affects spanning results as early as 1995. New nonlinear spanning tests are found to behave appropriately. Using the full set of yields instead of truncating to a small number of principal components is quantitatively important but does not eliminate the omitted nonlinearity effect.

## **Degree Type**

Dissertation

## **Degree Name**

Doctor of Philosophy (PhD)

## **Graduate Group**

Economics

## **First Advisor**

Francis X. Diebold

## **Keywords**

Bond yields, Factor Modeling, Time series econometrics, Volatility, Zero lower bound

## **Subject Categories**

Economics

APPLIED DYNAMIC FACTOR MODELING IN FINANCE

Ross Askanazi

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

---

Francis X. Diebold  
Professor of Economics

Graduate Group Chairperson

---

Jesús Fernández-Villaverde  
Professor of Economics

Dissertation Committee

Francis X. Diebold, Professor of Economics  
Xu Cheng, Associate Professor of Economics  
Francis J. DiTraglia, Assistant Professor of Economics

APPLIED DYNAMIC FACTOR MODELING IN FINANCE

© COPYRIGHT

2017

Ross Askanazi

This work is licensed under the  
Creative Commons Attribution  
NonCommercial-ShareAlike 3.0  
License

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

## ACKNOWLEDGEMENT

Completing a PhD is extremely difficult, and there are many people I would like to thank. I have made many mistakes on the path through graduate school, and without the wonderful people I've been surrounded by I would not have finished. First and foremost I must thank Professor Francis Diebold. From the first year in graduate school his passion for predictive modeling, forecasting, and applied work in finance has been contagious. Moreover, he has given me second, third, and fourth chances to prove myself, even when those chances were undeserved.

I am equally grateful to my other dissertation committee members, Professors Xu Cheng and Francis DiTraglia. They have both been extremely generous with their time, and their consistently insightful commentary has been helpful in refining the edges of this work into its final product. I would also like to thank Dr. Frank Schorfheide and Dr. Ben Connault – between coursework in econometrics, discussions in office hours, and helpful commentary in seminars they helped to shape my econometrics education. I am grateful to participants of our Econometrics Seminars and Econometrics Lunches; I am especially grateful to Dr. Glenn Rudebusch, whose work on the spanning hypothesis motivated and informed much of the second half of this dissertation.

The students of UPenn's economics department have been a wonderful resource. My frequent collaborators, Jacob Warren and Matt Cook, are extraordinary economists. They have been constant sources of innovative ideas, they have been sounding boards for my own terrible ideas, and they have helped to push those ideas into finished products. Most importantly, they made sure doing economics every day was fun. I learned extraordinary amounts from the students ahead of me: Minchul Shin, Molin Zhong, Laura Liu, and Lorenzo Braccini are brilliant minds, and were invaluable in my early attempts at research. Most importantly they are all experts in putting positive spins on the failures that are inevitable in the second and third years of graduate school. The econometrics students in year below mine, Paul

Sangrey and Minsu Chang, have kept me motivated and inspired me with their successes. I am excited for what the future holds for them. I thank my officemates Jacob Warren, Paul Sangrey, and Hanna Wang, who made coming to the office every day a pleasure. Kory Johnson and Raiden Hasegawa in the Statistics department have also been fantastic friends and helpful resources throughout the dissertation process.

I am more grateful than I can say to my mother, Dr. Karen Gil, who has inspired in me a lifelong love of learning and an appreciation for the value of putting one foot in front of the other every day. Her love and support was integral through every step of this process. I am also grateful to my father, Jeffrey Askanazi, and my brothers Evan and Cory Askanazi.

I have been lucky to have a wonderful family of friends in Philadelphia and abroad who have been constantly supportive. Stephen Marks, Adam Davis, Matt Frey, and Matt McErlean have been there for me as long as I can remember, and I hope to never know a time that they are not in my life. Jason Koski and Mike O'Reilly were perfect graduate school roommates and I miss them. Angela Patini was overwhelmingly supportive during the most daunting periods. I owe a special debt to the entire UPenn Climbing Team, and the greater Philadelphia climbing community. They were there when economics was fun, which was often, and they were there when economics was not fun, which was also often.

Finally, I would like to thank the National Science Foundation, whose Graduate Research Fellowship secured my acceptance into Penn's PhD program, and whose funding made this dissertation possible. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# ABSTRACT

## APPLIED DYNAMIC FACTOR MODELING IN FINANCE

Ross Askanazi

Francis X. Diebold

In this dissertation, I study model misspecification in applications of dynamic factor models to finance. In Chapter 1, my co-author Jacob Warren and I examine factors for volatility of equities. Historical literature on the subject decomposes volatility into a factor component and an idiosyncratic remainder. Recent work has suggested that idiosyncratic volatility of US equities data has a factor structure, with the factor highly correlated with, and possibly precisely the market volatility. In this paper we attempt to characterize the underlying factor and find that it can be decomposed into a statistical (PCA) and structural (market volatility) factor. We also show that this feature is not unique to equities, appearing in diverse sets of financial data. Lastly, we find that this dual-factor approach is slightly dominated in forecasting environments by a single statistical factor, suggesting that accurate measurement of the factors provides a direction for future work. In Chapter 2, I explore the use of dynamic factor models in yield curve forecasting and an exploration of the spanning hypothesis – that is, whether all information necessary for forecasting yields is contained in the current yield curve. Only linear tests of the spanning hypothesis are typically conducted in the literature, and the results are subject to substantial disagreement. In this paper, I explore a key modern nonlinearity, namely the zero lower bound (ZLB). I first demonstrate in simulation that only very small nonlinearities in the measurement equation are necessary to break down the assumed linear spanning relationship. Because bond yields are determined by forward-looking behavior of investors, the effect of the ZLB affects spanning results as early as 1995. New nonlinear spanning tests are found to behave appropriately. Using the full set of yields instead of truncating to a small number of principal components is quantitatively important but does not eliminate the omitted nonlinearity effect.

## TABLE OF CONTENTS

Acknowledgement . . . . .	iii
Abstract . . . . .	v
List of Tables . . . . .	vii
List of Illustrations . . . . .	viii
CHAPTER 1 : Factor Analysis For Volatility . . . . .	1
1.1 Introduction . . . . .	1
1.2 Modeling Procedure . . . . .	4
1.3 Equities Data . . . . .	11
1.4 Foreign-Exchange Rates . . . . .	26
1.5 Conditional Mean Misspecification . . . . .	30
1.6 Forecasting . . . . .	32
1.7 Conclusion . . . . .	38
1.8 Appendix . . . . .	39
CHAPTER 2 : The Spanning Puzzle and Nonlinearities at the Zero Lower Bound . . . . .	48
2.1 Introduction . . . . .	48
2.2 Yield Curve Modeling . . . . .	56
2.3 The Spanning Hypothesis . . . . .	65
2.4 Nonlinear Models and Spanning . . . . .	73
2.5 Simulation Results . . . . .	79
2.6 Empirics . . . . .	87
2.7 Conclusion . . . . .	94
2.8 Appendix . . . . .	97
Bibliography . . . . .	100



## LIST OF TABLES

TABLE 1 :	Statistical Tests Explained – Expected Outcomes . . . . .	24
TABLE 2 :	Statistical Tests for Equities . . . . .	25
TABLE 3 :	Statistical Tests for FX panel . . . . .	29
TABLE 4 :	Statistical Tests for Higher Powers of Market Return . . . . .	31
TABLE 5 :	Mean Square Error, Median Absolute Error of DOW 10 Rvariances .	34
TABLE 6 :	Mean Square Error, Median Absolute Error of S&P 100 Rvariances .	35
TABLE 7 :	Mean Square Error, Median Absolute Error of FX rate Rvariances .	36
TABLE 8 :	Simulation Results . . . . .	44
TABLE 9 :	DOW 10 Company List . . . . .	45
TABLE 10 :	S&P100 Company List . . . . .	45
TABLE 11 :	Forex List . . . . .	46
TABLE 12 :	Mean Square Error, Median Absolute Error of DOW 10 Rvariances (post 2009) . . . . .	46
TABLE 13 :	Mean Square Error, Median Absolute Error of S&P 100 Rvariances (post 2009) . . . . .	47
TABLE 14 :	Mean Square Error, Median Absolute Error of FX rate Rvariances (post 2009) . . . . .	47
TABLE 15 :	Breakdown of Spanning Literature . . . . .	52
TABLE 16 :	Linear Spanning Tests on a Nonlinear Model – Rejection Rates . . .	82
TABLE 17 :	Unspanned Risk – Monthly – Linear Predictive Regression . . . . .	92
TABLE 18 :	Unspanned Risk – Daily – Linear Predictive Regression . . . . .	92
TABLE 19 :	Unspanned Risk – Monthly – Nonlinear Predictive Regression . . . . .	93
TABLE 20 :	Unspanned Risk – Daily – Nonlinear Predictive Regression . . . . .	93

## LIST OF ILLUSTRATIONS

FIGURE 1 :	Principal Components Analysis . . . . .	9
FIGURE 2 :	Factor Structure in Equities Idiosyncratic Volatility . . . . .	17
FIGURE 3 :	Market Volatility and Idiosyncratic Volatility . . . . .	18
FIGURE 4 :	Market Volatility: Explained Variation . . . . .	19
FIGURE 5 :	First PC of Idiosyncratic Volatility . . . . .	19
FIGURE 6 :	Market Volatility and First PC of Idiosyncratic Volatility . . . . .	20
FIGURE 7 :	Market Volatility and PCA factor of Idiosyncratic Volatility . . . . .	21
FIGURE 8 :	FX Factor and Idiosyncratic Volatility . . . . .	27
FIGURE 9 :	Explained Variation in FX Idiosyncratic Volatility . . . . .	28
FIGURE 10 :	PCA Factor vs Market Volatility . . . . .	28
FIGURE 11 :	Equities Squared One-Step Prediction Errors . . . . .	33
FIGURE 12 :	FX Squared One-Step Prediction Errors . . . . .	37
FIGURE 13 :	Equities cumulative squared forecast errors – Post 2009 . . . . .	43
FIGURE 14 :	FX cumulative squared forecast errors . . . . .	43
FIGURE 15 :	Affine macroeconomic Term Structure Modeling . . . . .	55
FIGURE 16 :	US Government Bond Yields . . . . .	59
FIGURE 17 :	The ZLB Wedge Across Time . . . . .	63
FIGURE 18 :	The ZLB Wedge Across Maturities . . . . .	64
FIGURE 19 :	ZLB Wedge Invertibility . . . . .	74
FIGURE 20 :	Simulated Data . . . . .	80
FIGURE 21 :	ADS Index . . . . .	87
FIGURE 22 :	Residuals of Linear Predictive Regression . . . . .	96

# CHAPTER 1

## Factor Analysis For Volatility <sup>1</sup>

### 1.1 Introduction

As economists we find that large complex dynamics can usually be modeled as resulting from a small number of fundamental shocks. Factor models approach this formally:

$$y_t = \beta F_t + e_t, \quad E(F_t e_t) = 0,$$

where  $\dim(F_t) = k \ll \dim(y_t) = N$  and  $t = 1, \dots, T$ . One popular application of factor models (especially in finance) is for covariance matrix estimation. The factor model presents a useful decomposition, assuming factors and errors are orthogonal:

$$\Sigma^y = \beta \Sigma^F \beta' + \Sigma^e.$$

Here  $\Sigma^e$  is sparse, if not diagonal, and  $\Sigma^F$  is of small dimension, so  $\beta \Sigma^F \beta'$  is of low rank. This “low rank plus sparse” decomposition via factor models has facilitated tractable dynamic volatility: For  $\Sigma^y$  to be time-varying, at least one of  $\beta$ ,  $\Sigma^F$ , or  $\Sigma^e$  must be time-varying.

Over the years, there have been many variations to induce time-varying volatility in  $\Sigma^y$ . Most commonly (Diebold and Nerlove (1989), Jacquier et al. (1994), Adrian and Rosenberg (2008)),  $\Sigma^F$  is endowed with stochastic volatility, while other elements remain constant.

---

<sup>1</sup>This chapter is co-authored with Jacob Warren.

More recently though (Kim et al. (1998), Pitt and Shephard (1999), Aguilar and West (2000)), the diagonal elements of  $\Sigma^e$  were also allowed to time-vary, adding an additional layer of complexity.

However, recent empirical work has indicated that despite the factor model inducing orthogonal structure on the level equation, it ignores higher order dependence between the factor and idiosyncratic component. Specifically, Herskovic et al. (2014) find that idiosyncratic variances tend to (strongly) comove, and Barigozzi and Hallin (2016) further show that the comovement extends to the volatility of the level factor ( $\Sigma_t^F$ ) as well. Kalnina and Tewou (2017) and Duarte et al. (2014) are in the same vein. More specifically, let  $\sigma_t^e = \text{diag}(\Sigma_t^e)$ , then those papers suggest:

$$\log(\sigma_t^e) = AV_t + \varepsilon_t, \quad E(V_t \varepsilon_t) = 0, \quad \dim(V_t) \ll N,$$

where  $V_t$  is a factor for idiosyncratic volatility.

Our paper immediately builds off those recent contributions by using high-frequency based Realized Volatilities on two datasets of US Equities. In general, our findings support prior research: the panel of idiosyncratic volatilities has clear and strong factor structure, and the first principal component of the panel is highly correlated with market volatility.

The above literature is split on the nature of the factor for idiosyncratic volatility. While all agree that idiosyncratic volatility is dynamic and has factor structure, there is no consensus as to what precisely is the factor. Some use the market volatility as the factor, while others take a more statistical approach and merely use the first principal component. We attempt to provide clarity on that issue by accomplishing three main goals: First, we provide a framework for estimating the factor structure in idiosyncratic volatility using realized measures. Second, we attempt to answer (via a series of graphical tools and statistical tests) how exactly the factor for idiosyncratic volatility is related to market volatility. More specifically, we are interested in whether they are precisely the same, or if one supersedes

the other. Third, we demonstrate that the structure is a general feature of volatility, and not just limited to equities.

To accomplish the third goal, we extend this work to a panel of exchange rate volatilities in addition to the equities datasets. The same tractable dynamic volatility modeling has been used in forecasting exchange rate volatility (Diebold and Nerlove (1989)), and we explore the same questions of the nature of exchange rate idiosyncratic volatility. In contrast to equities, the correlation between the factor for idiosyncratic volatility and market volatility falls dramatically.

Despite that large difference, all datasets support the same general framework – namely that both the market volatility *and* an additional principal components factor is necessary for explaining cross-sectional variation. While on the one hand this presents a robust statistical fact, it is also troubling from an economic modeling perspective. Indeed, the question of *why* these statistical facts occur become all the more pronounced. Is there an economic theory that can support the phenomenon for both FX returns and equities? Or perhaps, is the framework a product of network effects, time-varying volatilities and financial markets? While we do not attempt to answer these questions in this paper, they provide the foundation for this and future work in the area.

The outline for the remainder of the paper is as follows: In Section 1.2, we outline the framework for estimating dynamic idiosyncratic volatility. In Section 1.3, we present the US equities data, and in subsections explore the outcomes of our model selection framework. In Section 1.4, we conduct the same set of exercises for foreign exchange rates. Section 1.5 explores robustness to the most obvious counterpoint to the proposed framework – namely that features of idiosyncratic volatility can simply be the result of conditional mean misspecification. Finally, Section 1.6 explores the implications of our findings for out-of-sample forecasting and Section 1.7 concludes. Post-conclusion, we provide simulation evidence that our battery of statistical tests perform and behave appropriately in our environment. This can be found in Section 1.8.1.

## 1.2 Modeling Procedure

### 1.2.1 Continuous Time Setup

For equities, we start with a continuous time price process that mimics the setup in Barndorff-Nielsen and Shephard (2004). Let  $S(t)$  be the price process of a security (or possibly a vector of securities), and  $X(t) = \log(S(t))$  be a semi-martingale, so

$$X(t) = \alpha(t) + m(t),$$

where  $\alpha(t)$  is the drift term and  $m(t)$  is a local martingale. For any sequence of partitions,  $t_0 = 0 < t_1 < t_2 \cdots < t_M = t$ , with  $\sup_j \{t_{j+1} - t_j\} \rightarrow 0$  for  $M \rightarrow \infty$ , we define the quadratic variation on day  $t$  as:

$$[X](t) = \text{plim}_{M \rightarrow \infty} \sum_{j=0}^{M-1} \{X(t_{j+1}) - X(t_j)\} \{X(t_{j+1}) - X(t_j)\}'.$$

In practice we only have a finite partition, so we construct the realized volatility as an estimator of the quadratic variation:

$$\widehat{[X]}(t) = RV_t = \sum_{j=0}^{M-1} \{X(t_{j+1}) - X(t_j)\} \{X(t_{j+1}) - X(t_j)\}'.$$

This is the standard definition of realized volatility, which has been well described and analyzed over the recent years (see, among others, Andersen et al. (2007)).

We further utilize two information sets, as in Sheppard and Xu (2014): a high frequency information set  $\mathcal{F}_t^{HF}$  and a low frequency information set  $\mathcal{F}_t^{LF}$ . The high frequency information set contains all the information of the low frequency information set, plus the intraday data necessary to construct the realized measure at date  $t$  (so that  $\mathcal{F}_t^{LF} \subset \mathcal{F}_t^{HF}$ ). We will subscript the high frequency information set by  $t_j$ ,  $j = 1, \dots, M_t$  for each date  $t$ .

Our primary objects of interest are as follows: We have returns  $r_t$ , factor loadings  $\beta_t$ , a level factor  $f_t$ , and idiosyncratic shocks  $v_t$  for the level equation. We posit the existence of a single factor structure at high frequency, so that the volatility of the factor is a scalar  $\sigma_t^f$ . The covariance of the idiosyncratic shocks is  $\Omega_{v_t}$ .

$$\begin{aligned}
r_{t_j} &= \beta_t f_{t_j} + v_{t_j} & t = 1, \dots, T & \quad j = 1, \dots, M_t, \\
f_{t_j} | \mathcal{F}_t^{HF} &\sim iidN(0, \sigma_t^f), \\
v_{t_j} | \mathcal{F}_t^{HF} &\sim iidN(0, \Omega_{v_t}).
\end{aligned}$$

Since the market factor and idiosyncratic error are continuous-time return sequences that are observed at distinct time partitions, we can compute their respective Realized Volatilities (assuming  $\beta$  is fixed and known intraday):

$$\begin{aligned}
RV_{f_t} &= \sum_{j=0}^{M-1} \{f_{t_{j+1}} - f_{t_j}\} \{f_{t_{j+1}} - f_{t_j}\}', \\
RV_{v_t} &= \sum_{j=0}^{M-1} \{v_{t_{j+1}} - v_{t_j}\} \{v_{t_{j+1}} - v_{t_j}\}'.
\end{aligned}$$

This factor structure at high frequencies time aggregates to a factor structure at the low (daily) frequency,

$$\begin{aligned}
r_t^{LF} &= \beta_t f_t^{LF} + v_t^{LF}, \\
f_t^{LF} | \mathcal{F}_t^{LF} &\sim N(0, \sigma_t^f), \\
v_t^{LF} | \mathcal{F}_t^{LF} &\sim iidN(0, \Omega_{v_t}).
\end{aligned}$$

From this point forward the  $LF$  superscript will be suppressed for brevity. We will at times use the notation  $X_t^{HF} = [X_{t_0}, X_{t_1}, \dots, X_{t_M}]'$  to represent the vector of high-frequency intraday observations of asset  $X$ .

### Factor Loadings

It remains to specify dynamics on the factor loadings as well. There is considerable debate on whether factor loadings actually have time-variation, and if so, at what frequency they should vary. There is also a debate about whether this time-variation has any broader implications for risk or returns. Braun et al. (1995) use bivariate EGARCH models to measure estimate conditional covariances of returns, but find only weak evidence of time-varying conditional (monthly) betas. Using an international panel, Ferson and Harvey (1993) find that nation-specific betas do time-vary with international risk factors, but that movements in the betas contribute only a small fraction to predicted variation in expected returns. Bali and Engle (2010) find substantial time-variation in betas with the market, and Bali et al. (2013) shows that the time-variation is meaningful for trading. Supporting this, Jagannathan and Wang (1996) allow betas to time-vary in a CAPM model, which is better able to explain cross-sectional returns. Lewellen and Nagel (2006) agree that betas time-vary, but disagree about their ability to explain cross-sectional returns. Sheppard and Xu (2014) combine realized measures with GARCH dynamics (HEAVY-GARCH) on factor models (including loadings) to great success. Most applicable to our setup, Andersen et al. (2006) compute realized betas, and find that they have much shorter memory than Realized Volatilities.

The debate about whether (and how much) betas vary over time is specifically important to our setup. Take for example, a toy model with time-varying betas:

$$y_t = \beta_t F_t + e_t,$$



but the econometrician instead estimates a model with constant betas:

$$y_t = \beta F_t + \bar{e}_t.$$

Then observe that the error term will include the time-variation in betas:

$$\bar{e}_t = (\beta_t - \beta)F_t + e_t.$$

This has large implications for the observed idiosyncratic covariance matrix from the misspecified regression:

$$\Sigma_{\bar{e}} = (\beta_t - \beta)\Sigma_F(\beta_t - \beta)' + \Sigma_e.$$

Thus, one could observe factor structure in the residual variances (and the factor would be highly correlated with factor volatility) simply due to misspecified dynamics in the factor loadings.

We therefore allow betas to time-vary at the daily level, but leave them fixed intraday. Mimicking the approach of Andersen et al. (2006), we use a realized beta setup:

$$R\beta_{i,t} = \frac{Cov(r_{it}^{HF}, f_t^{HF})}{Var(f_t^{HF})}$$

We allow dynamics on the factor loadings to follow independent autoregressions:

$$\Phi_{\beta_i}(L)\beta_{i,t} = \eta_{i,t}^\beta, \quad \eta_{i,t}^\beta \sim iidN(0, \sigma_i^\beta) \tag{1.2.1}$$

Dynamics on the factor loadings are given as independent univariate autoregressions, because having to estimate a vector autoregression of factor loadings defeats the purpose of employing

a factor structure in the first place, since there are  $N$  series of loadings.

It is important to note: while variation in realized betas has important implications for the cross-sectional and time-variation of asset realized volatility, modeling it greatly increases the number of parameters of the model (there are  $N \times k \times T$  realized betas). Therefore, for the purposes of *forecasting* asset realized volatility, it is not clear that allowing for variation in realized betas will improve outcomes. In fact we find that it is not – allowing for this variation increases mean squared forecast error. In our forecasting exercise, we therefore hold factor loadings constant, with the understanding that this may inflate the measured time-variation in idiosyncratic volatility. On balance, however, we find that this approach and a conservative interpretation of idiosyncratic volatility dynamics is more appropriate for forecasting.

### Factor Structure and PCA

In all empirical exercises, we use an observed factor for  $F_t$  in the level equation. This allows us to both ignore estimation error in  $F_t$ , and provides us with observed high frequency data for  $F_t$ , yielding realized measures of  $\sigma_{F_t}$  and  $\sigma_{e_t}$ .

In order to extract a statistical factor,  $V_t$ , we use principal components to extract a static factor for the idiosyncratic volatilities. Recall that for a panel  $\log(\sigma_e)$ , principal components extracts factors via the minimization problem

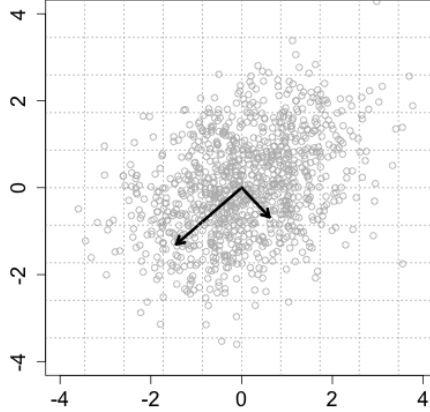
$$V(k) = \min_{\Lambda, F^k} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\log(\sigma_{e_{i,t}}) - \lambda_i^k V_t^k)^2,$$

subject to  $\Lambda^{k'} \Lambda^k / N = I_k$  or  $V^{k'} V^k / T = I_k$ .

Here  $k$  is the number of factors,  $V$  are the factors, and  $\lambda$  are the factor loadings. We can quickly see that  $\lambda$  and  $V$  are not separately identified. This is why we need the or-

Figure 1: Principal Components Analysis

A cloud of data. The black vectors represent the directions of greatest variation extracted by PCA. The length of each vector represents the variance in that direction.



thonormalization identifying constraint above. We can think of this minimization problem as extracting the directions of greatest variation. This can be visualized a la Figure 1.

Much like standard in-sample mean squared error (MSE) analysis, we can see from the above formula that  $V(k)$  is strictly decreasing in  $k$ . Therefore, optimizing  $V(k)$  is a poor choice for selecting the number of factors. As with MSE, this loss function can be augmented with a penalty function for the number of factors to create a consistent information criterion. We use the Bai and Ng (2002) information criterion to select the number of factors:

$$PC(k) = V(k) + kg(N, T)$$

Where  $N$  and  $T$  are the dimensions of the panel of interest, and  $g(\cdot)$  need only satisfy

$$g(N, T) \xrightarrow{N, T \rightarrow \infty} 0$$

$$\min(\sqrt{N}, \sqrt{T})g(N, T) \xrightarrow{N, T \rightarrow \infty} \infty$$

We use PCA as one of the options for generating a factor for volatility. Similar to most equity log-realized volatilities, the extracted PCA factor is approximately Gaussian and has long-memory. Since this factor is a linear combination of log volatilities, these features are to be expected.

## 1.3 Equities Data

The date ranges for the data analysis runs from January 2007 to November 2014. All low frequency (daily) returns were downloaded from the Center for Research in Security Prices (CRSP), while all high frequency data was downloaded from the Ticker and Quote (TAQ) dataset. We use high frequency data to construct realized measures from intraday returns, but use the low-frequency (daily) returns to make average realized volatility the same as the variance of returns.

We use two datasets: for low dimensional analysis, we use the DOW 10 and the SPY (a highly liquid ETF tracking the S&P 500) as an observed market factor. All companies in the DOW 10 are observed over the entire 2007–2014 trading period.

For high dimensional analysis, we use the S&P 100. Since companies enter and leave the index over the sample period, we keep the stocks in the index as of November 2014 that are traded across the entire 7 years. That leaves us with 90 assets. As with the DOW 10, we use the SPY as an observed market factor for this dataset. Lists of the DOW 10 and the stocks used in the S&P 100 (with sector designations) are presented in the Appendix.

### 1.3.1 Construction of Realized Measures

Continuing the discussion above, the Quadratic Variation of a log-price process is defined as

$$QV_t = \text{plim}_{M \rightarrow \infty} \sum_{j=0}^{M-1} \{X(t_{j+1}) - X(t_j)\} \{X(t_{j+1}) - X(t_j)\}'$$

The natural estimator of true quadratic variation truncates the number of intraday observations at some finite number. This estimator was introduced by Andersen et al. (2001) and Andersen et al. (2003) and it was shown to converge to  $QV_t$  as the number of observations goes to infinity by Barndorff-Nielsen and Shephard (2004).

Unfortunately that estimator is not robust to measurement error or jumps in the price

process, so many variations have been introduced in the subsequent years. In the presence of classical measurement error, the standard realized variance estimator is biased, and that bias depends on sample size. So as the sampling frequency increases, the estimator becomes worse and worse. To solve this issue, Ait-Sahalia et al. (2005a) propose a complex bias-corrected estimator, but also suggest that a subsampling approach can be nearly as good. Subsampling requires multiple intraday grids for the price process, where each sampling grid (say, 5 minutes) can be further subsampled at a higher frequency (say, 1 minute). Formally, let  $G^{(i)}$  be the partition of intraday returns at the  $i^{\text{th}}$  minute,  $G^{(i)} = \{t_i, t_{i+5}, t_{i+10}, \dots, t_{i+5(M-1)}\}$ , and associated estimate of realized volatility:  $[\widehat{X, X}]_t^{(i)} = \sum_{j \in G^{(i)}} \{X(t_{j+1}) - X(t_j)\} \{X(t_{j+1}) - X(t_j)\}'$ . Then the estimate for daily realized volatility is  $\widehat{RV}_t = \frac{1}{5} \sum_{i=1}^5 [\widehat{X, X}]_t^{(i)}$ . Liu et al. (2015) thoroughly investigate over 400 different estimators and find that 5 minute intervals (perhaps with 1-minute subsampling) is very hard to beat in terms of forecasting. Following their lead and the theoretical contributions of Ait-Sahalia et al. (2005a), that is the estimator we use. In our application,  $X$  is a vector of returns, which delivers a full realized covariance matrix:  $\widehat{RCov}_t$ .

To create our sampling time grid, we use the first observed return within minute  $j$  as  $X_{t_j}$  and fill in missing values with a return of 0. We also exclude the first and last 30 minutes of each trading day to avoid open and close effects.

Computing the daily realized betas in practice is a matter of simply taking components from the full Realized Covariance matrix described above:

$$\widehat{R}\beta_t = \frac{\widehat{RCov}(r_{it}^{HF}, f_t^{HF})}{\widehat{RV}(f_t^{HF})}.$$

Our method for computing realized measures is obviously not the only method of constructing a realized volatility – given the number of modeling choices including sampling rate, subsampling rate, functional form of the estimator (RV versus, say, a realized kernel), there are

hundreds of volatility estimators. Briefly, the realized kernel estimator of Barndorff-Nielsen et al. (2011) is an advanced method for these purposes, and has been further improved upon by Hautsch et al. (2012) and Hautsch et al. (2011) in an effort to construct more efficient estimators in high dimensions. Hautsch et al. (2011) finds that regularizing the kernel density estimator has significant implications for portfolio management. However, an additional branch of literature suggests that the marginal gains of more advanced estimators relative to the complexity required to calculate them is unclear. We refer to Liu et al. (2015), who show that complexity usually does not significantly increase accuracy.

### 1.3.2 Data Transformations

As a potential issue, we recognize that despite the theoretical and practical support for the Ait-Sahalia et al. (2005a) estimator, it does leave out significant trading information since it ignores possible overnight changes in returns. Since the low-frequency data is constructed using close-to-close returns, this lack of overnight information results in a nontrivial discrepancy between the high frequency realized measures and the low frequency realized measure, which is

$$\frac{1}{T} \sum_{t=1}^T r_t r_t'$$

We employ a simple scaling that matches the moments of realized measures of different frequencies, proposed in Sheppard and Xu (2014). Given

$$\begin{aligned} \bar{\Sigma} &= \frac{1}{T} \sum_{t=1}^T r_t r_t', \\ \bar{M} &= \frac{1}{T} \sum_{t=1}^T \widehat{RCov}_t, \\ \bar{\Gamma} &= \bar{\Sigma}^{1/2} \bar{M}^{-1/2}. \end{aligned}$$

Then define the scaled realized covariance:

$$\widetilde{RC}_t = \bar{\Gamma} \widehat{RCov}_t \bar{\Gamma}.$$

This yields

$$\frac{1}{T} \sum_{t=1}^T \widetilde{RC}_t = \frac{1}{T} \sum_{t=1}^T r_t r_t'.$$

As long as  $T$  is sufficiently larger than  $N$ , this transformation will be numerically stable. We apply the transformation to the entire Realized Covariance matrix, and then use the transformed values to construct realized betas. This means that although the moments for the full realized covariance match the low-frequency counterparts, the moments for realized betas do not. In practice we find that this overnight transformation does not impact the qualitative results, but in combination with improved intraday realized measures it is important to consider.

### 1.3.3 Estimation Procedure

Whether market volatility is precisely the factor for idiosyncratic volatility presents three possible DGPs, which in turn should influence theory and mechanisms explaining the phenomenon. There are three distinct cases for how the two can be related, and they lead to three separate models of interest that we must estimate:

1. The factor(s) for idiosyncratic volatility are precisely the volatilities of the market factor. This is the case employed in Kalnina and Tewou (2017). We call this FVOL MKT.
2. The factor(s) for idiosyncratic volatility are orthogonal to the volatilities of the market factors. We call this FVOL2.
3. The factor(s) for idiosyncratic volatility are separate from, though highly correlated with, the volatilities of the market factor. This case remains largely unexplored, though is related to work in Chen and Petkova (2012). We call this FVOL PCA.



While Duarte et al. (2014), Herskovic et al. (2014), Barigozzi and Hallin (2016), and Christoffersen et al. (2014) all utilize a statistical factor as their factor for idiosyncratic volatility, they do not comment on the relationship between Market Volatility and their statistical factor. It is therefore difficult to discern whether they support FVOL2 or FVOL PCA.

Case 1 would correspond to the following model:

$$\begin{aligned} r_t &= \beta F_t + e_t \\ \log(\sigma_{F_t}) &= \mu_F + \beta_F \log(\sigma_{F_{t-1}}) + u_t^F \\ \log(\sigma_{e_t^i}) &= \mu_i + \beta_i^e \log(\sigma_{F_t}) + u_t^i \end{aligned}$$

Case 2 would correspond to:

$$\begin{aligned} r_t &= \beta F_t + e_t \\ \log(\sigma_{F_t}) &= \mu_F + \beta_F \log(\sigma_{F_{t-1}}) + u_t^F \\ \log(\sigma_{e_t^i}) &= \mu_i + \beta_i^e \log(\sigma_{F_t}) + \gamma_i V_t + u_t^i \end{aligned}$$

Where  $V_t$  is an additional factor for volatility. The third case is if idiosyncratic volatility is orthogonal to market volatility,  $\beta_i = 0$ . Beginning with high frequency returns  $r_{t_j}$ , we proceed as follows.

- At each date  $t$ , we run the intraday regression

$$r_{t_j} = \beta_t f_{t_j} + v_{t_j}, \quad j = 1, \dots, M_t$$

- We construct the daily estimate of realized volatility for  $f_t$  and  $v_t$  according to Section

1.2.1. In practice, we compute the entire  $RCov$  for  $[r_t, f_t]$ , which is an  $(N+1) \times (N+1)$  matrix.

- We conduct the data transformations, namely the scaling transformation to adjust for overnight returns, according to Section 1.3.2.
- Decompose the adjusted  $RCov$  into market volatility,  $\sigma_t^f$  and idiosyncratic volatility,  $diag(\Omega_{v_t})$ .
- Finally, collect all elements of  $diag(\Omega_{v_t})$  into a  $T \times N$  panel.
- Analyze the panel according to the applicable model
  1. FVOL MKT – Single factor on idiosyncratic volatility, where the factor is market volatility.
  2. FVOL2 – Two factor model on idiosyncratic volatility, where the first factor is market volatility, and the second factor is extracted via PCA from the residuals.
  3. FVOL PCA – Single factor on idiosyncratic volatility, where the factor is extracted via PCA on the idiosyncratic volatility panel.

### 1.3.4 Equities: Factor structure in Idiosyncratic Volatility

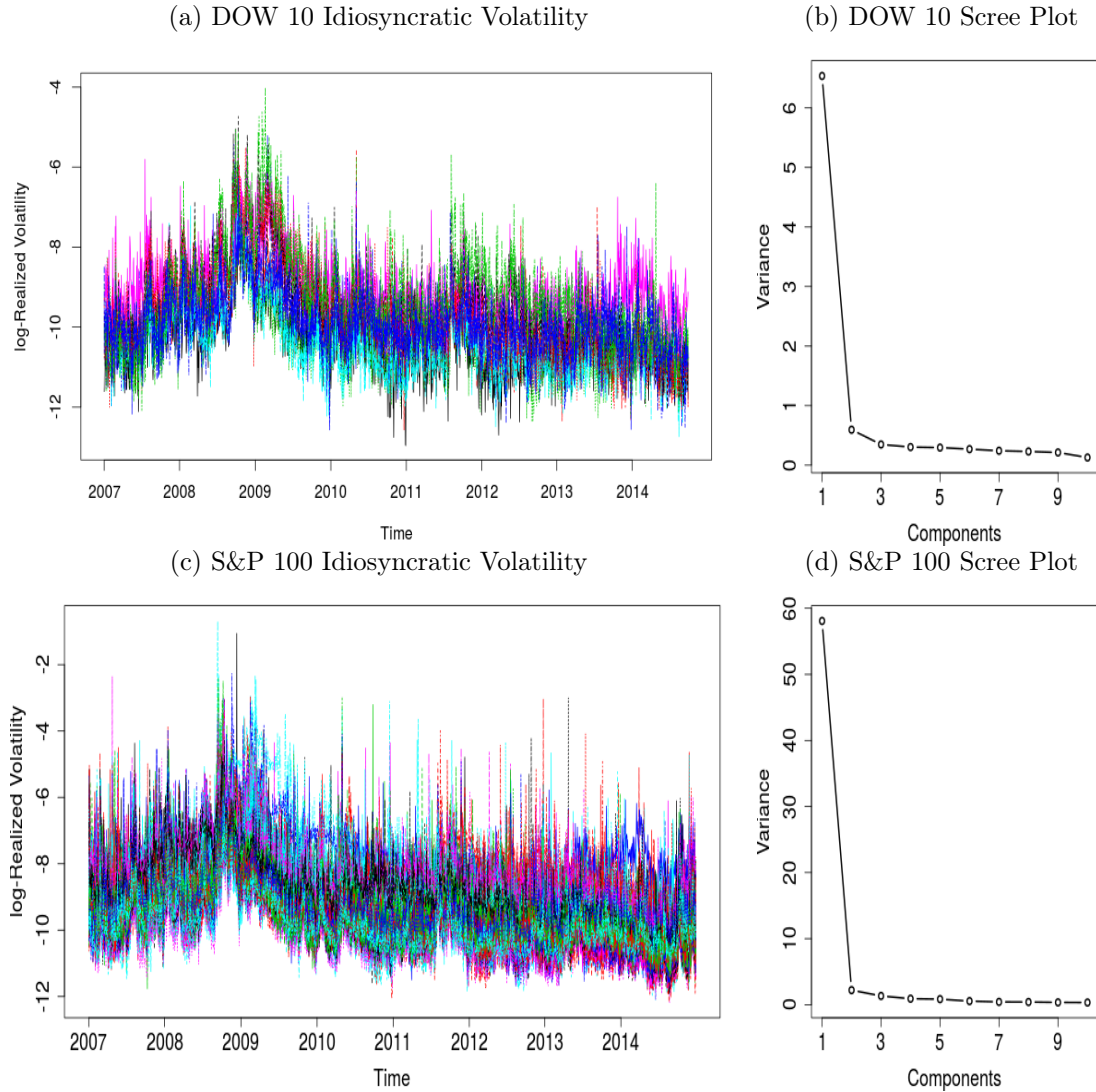
For both datasets, we start by verifying that idiosyncratic volatility is indeed dynamic and exhibits factor structure. We verify that it is dynamic by running univariate autoregressions with lag length chosen by AIC, all of which reject the null hypothesis of constant volatility with white noise. We verify factor structure by visual inspection of the panel and scree plots, which can be found in Figure 2.

### 1.3.5 Relationship between factor for volatility and factor volatility

Based on the figures, it is clear that there exists factor structure in idiosyncratic volatility. This is consistent with prior research in the field, as in Herskovic et al. (2014), Barigozzi and Hallin (2016), Kalnina and Tewou (2017) and Duarte et al. (2014). However, what is

Figure 2: Factor Structure in Equities Idiosyncratic Volatility

Figures 2a and 2c plot the log-realized volatilities of the DOW 10 and S&P 100 datasets from 2007–2015. Figures 2b and 2d plots corresponding scree plots (variances of the first 10 principal components)

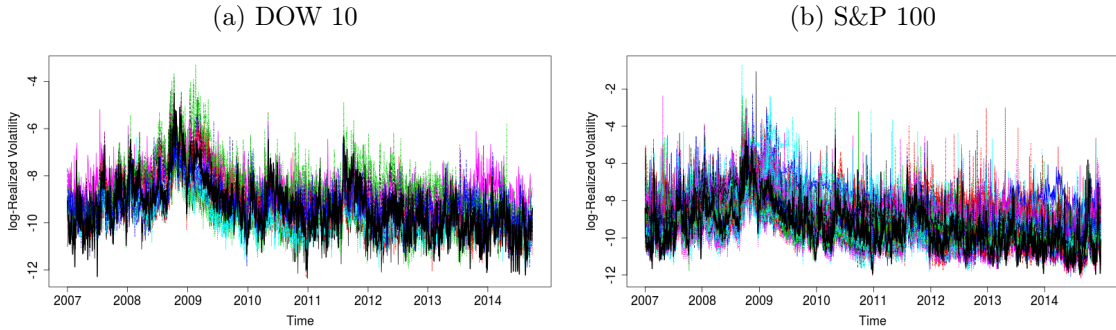


not clear from the above literature is the relationship between the factor for idiosyncratic volatility (i.e. the first principal component of the panel) and the volatility of the market factor. Kalnina and Tewou (2017) assume that they are the same, while Herskovic et al. (2014) and Barigozzi and Hallin (2016) do not.

In the following two sections, we argue that while the factor for idiosyncratic volatility and market volatility are highly correlated, they are not the same. We argue these facts based

Figure 3: Market Volatility and Idiosyncratic Volatility

The log-volatilities of the panel plotted against the SPY index volatility (in black) from 2007–2015.



on graphical analysis and a battery of statistical tests from the panel data literature.

### Graphical Analysis

We start by presenting the volatility of the market factor (SPY) overlaid on the plots of idiosyncratic volatility. The plots are in Figure 3. Taken together, the plots suggest that the market volatility explains amount of cross-sectional variation in the panel of idiosyncratic volatility. For the DOW 10, market volatility explains, on average 50% of cross sectional variation, while for the S&P100, it explains 55%. The distribution of explained variation across assets is in Figure 4. The explained variation is rather high for both panels, especially considering the naive modeling strategy would presume market volatility is unrelated to idiosyncratic volatility. These images heuristically support the methods in Kalnina and Tewou (2017).

Figure 4: Market Volatility: Explained Variation

The panel of  $R^2$  for each asset volatility in the panel regressed against SPY volatility. Approximately the same fraction of variation is explained by the SPY for each asset.

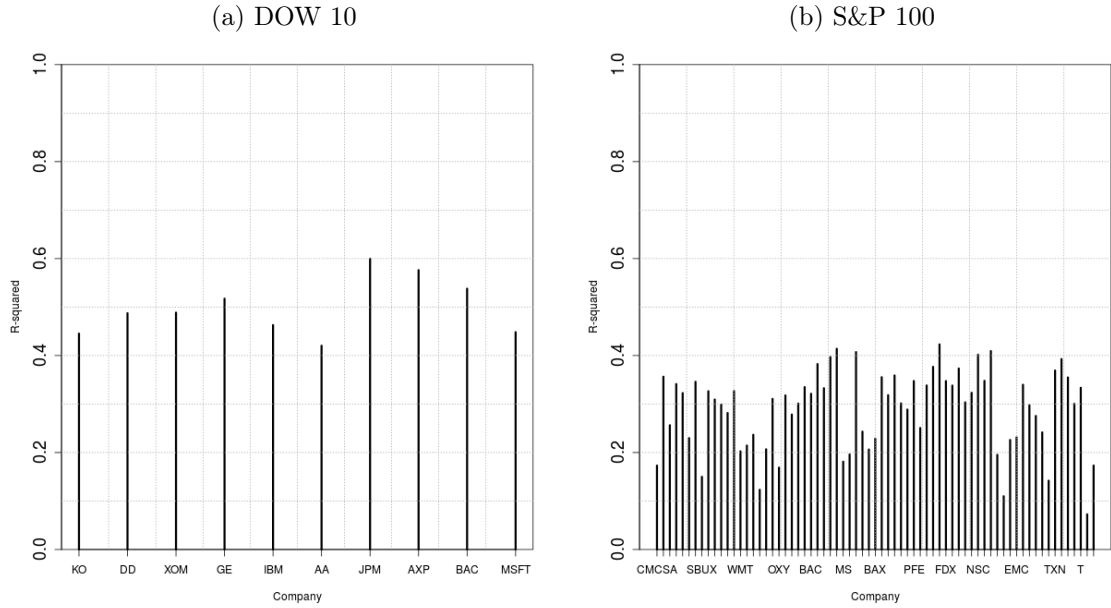


Figure 5: First PC of Idiosyncratic Volatility

The panel of  $R^2$  for each asset volatility in the panel regressed against first principal component of idiosyncratic volatility.

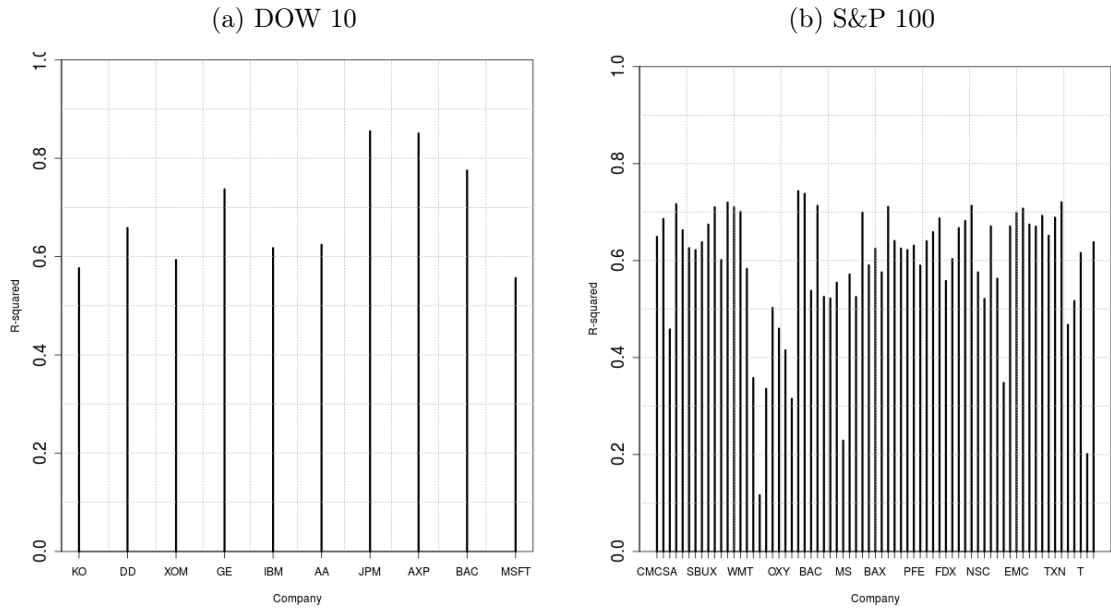
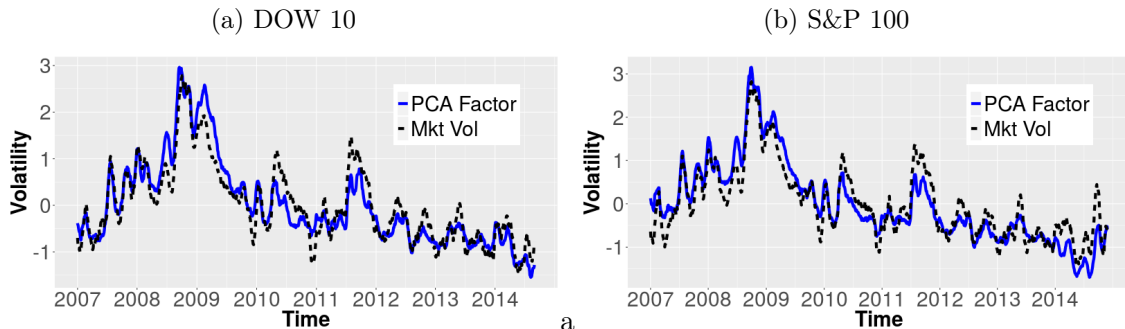


Figure 6: Market Volatility and First PC of Idiosyncratic Volatility

Each plot displays the 22-day rolling mean of the Market Volatility (black, dashed line) and the First PC of Idiosyncratic Volatility (blue, solid line) for that panel. Both volatilities have been centered and scaled to have mean 0 and variance 1.



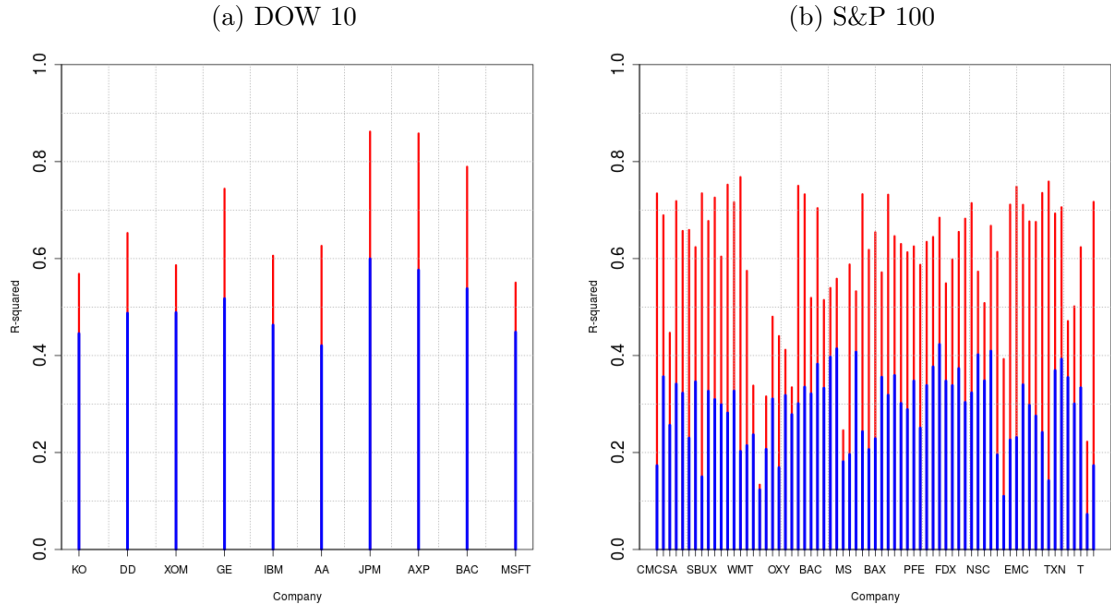
However, we also entertain the idea, as in Duarte et al. (2014), Herskovic et al. (2014), and others, that the factor for idiosyncratic volatility is a separate, PCA factor, that is possibly unrelated to market volatility. To support this, we present the distribution of explained variation, but this time with the first principal component of the panel of idiosyncratic volatilities replacing that of market volatility. These are in Figure 5. The average cross sectional  $R^2$  in the DOW 10 panel is 68%, while that in the S&P 100 panel is 76%. Unsurprisingly the first PC explains substantially more cross sectional variation than does market volatility. This supports Model 3.

Lastly, we also show that while the first PC explains more cross sectional variation than market volatility, the two are nonetheless highly correlated. In Figure 6 we plot the 22-day rolling average of the PCA factor and the Market log-Volatility. For both equities datasets, the correlation between the two (unsmoothed) is 0.85.

Despite this high correlation, we also consider whether both market volatility and the PC factor are important for explaining cross sectional variation in the panel. This is the Model 2 paradigm. We therefore plot the distribution of explained variations in Figure 7 with two factors – the first is the market volatility and the second is a PCA factor on residuals after regressing out market volatility. In this case, the average cross sectional  $R^2$  for DOW 10 is 68%, while that in the S&P 100 panel is 76%.

Figure 7: Market Volatility and PCA factor of Idiosyncratic Volatility

In blue, the panel of  $R^2$  for each asset volatility in the panel regressed against SPY volatility. In red is the increased in  $R^2$  from also regressing against first PCA.



One should note that the average explained variation for the two-factor paradigm is exactly the same as that for the principal components factor. Based on that observations, one might think that the market volatility plus a PCA factor merely spans the same space as the first PCA factor. Supporting this claim would be the fact that the canonical correlation between the first PCA factor and the two-factor model is almost exactly 1. Despite that, the two are not the same, insofar as it relates to explaining the panel of idiosyncratic volatility. Indeed, some assets are better explained by the two factor paradigm, and others are better explained by the principal components factor. Thus, while a linear combination of the two factors can nearly exactly generate the first PC factor, that linear combination is not optimal for explaining the panel.

Overall, graphical analysis supports the idea that both Model 2 or Model 3 are highly plausible. Despite the high correlation between Market Volatility and the first PC of Idiosyncratic Volatility, the PCA factor is able to explain a much larger share of overall variation.

## Statistical Tests

We propose a series of statistical tests for whether the factor for idiosyncratic volatility is the same, related or different from market volatility. We propose two versions of a likelihood ratio test, a test of factor structure from Onatski (2009), and a test for relating an observed factor to a PCA factor that is due to Bai and Ng (2006).

Using a normality assumption, we can use a likelihood ratio test for  $\beta_i^e = 0 \forall i$  in order to differentiate between cases 2 and 3. However, there are two LR tests necessary, since the construction of  $V_t$  via Principal Components will be different depending on whether the market volatility has been regressed out or not. As shown above, before regressing out the market volatility, the factor for idiosyncratic volatility is highly correlated with market volatility. As such, one would expect that if  $V_t$  is extracted from the entire panel of idiosyncratic volatility, then  $V_t$  might mainly include redundant information with  $\sigma_{f_t}$ . As such, we wish to test whether  $\sigma_{f_t}$  includes new information both before and after  $V_t$  has been extracted. In test LR-1 we construct  $V_t$  based on the residuals from first regressing out  $\sigma_{f_t}$ . In test LR-2, we construct  $V_t$  on the full panel, before regressing out  $\sigma_{f_t}$ . We expect, and find, that the test statistics for LR-1 are always substantially larger than those for LR-2. The LR test has asymptotic distribution as  $\chi_k^2$ , where  $k$  is the number of restrictions imposed. In all cases,  $k$  is the size of the cross sectional dimension.

In addition to a likelihood ratio test, we consider tests motivated by Bai and Ng (2006) and Onatski (2009). The former consists of using an observed factor  $G_t$  and PCA factor  $F_t$ , where the null hypothesis is that they are statistically the same. To deal with non-identification of the factor under rotation, the test statistics are constructed via canonical correlations as follows. Suppose we regress  $G_t$  against  $F_t$ , yielding  $\hat{G}_t$ . Then construct

$$\hat{\tau}_t = \frac{(\hat{G}_t - G_t)}{\widehat{\text{var}}(\hat{G}_t)^{(1/2)}}$$



In other words, this is the t-statistic for the null that  $G_t$  is spanned by  $F_t$ . Let  $\Phi_\alpha^\tau$  be the  $\alpha$  percentage point of the standard normal distribution. Then the statistics are

$$A = \frac{1}{T} \sum_{t=1}^T \mathbb{1}(|\hat{\tau}_t| > \Phi_\alpha^\tau)$$

$$M = \max |\hat{\tau}_t|$$

These exact tests have asymptotic distributions

$$A \rightarrow_p 2\alpha$$

$$M \text{ such that } P(M \leq x) \approx 2\Phi(x) - 1.$$

The rejection region for test  $M$  is found via simulation, as the  $(1 - \alpha)$  quantile of the maximum absolute value of standard normal vectors of length  $T$ .

They also propose approximate tests that are more heuristic. Consider regressing  $G_t$  against  $F_t$ . Then, under the null, the noise-to-signal ratio should be 0 and the  $R^2$  should be one. The heuristic tests say that the  $R^2$  should be “high,” and the noise-to-signal ratio should be “low.”

Lastly, we consider the test from Onatski (2009), which examines the number of factors in a panel. The exact test statistic is

$$R = \max_{k_0 < i \leq k_1} \frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}, \quad 0 \leq k_0 < k_1 \leq N - 2,$$

where  $\gamma_i$  is the  $i^{\text{th}}$  largest eigenvalue of the smoothed periodogram estimate of the spectral

Table 1: Statistical Tests Explained – Expected Outcomes

Test	Model 1	Model 2	Model 3
LR-1	Power, $\beta \neq 0$	Power, Reject	Power to correlated regressor, Reject
LR-2	Under-reject (correlated regressors)	Power, Reject	Under-reject (correlated regressors)
Onatski	Correct size	Power, Reject	Power, Reject more than 5%
A	0.1	N/A	N/A
M	$\sim 4$	N/A	N/A
NS	$\sim 0$	Moderately low	Moderately low
$R^2$	$\sim 1$	Moderately high	Moderately high

density matrix of data at a prespecified frequency. This test is valid for testing against a null of 0 factors. Therefore, after regressing out the market volatility, we test for the presence of factor structure, where the null hypothesis is no factor structure, and the alternative is anywhere from 1–3 factors. The test statistic has asymptotic Tracy-Widom distribution, whose critical values are tabulated in Onatski (2009).

For clarity, consider Table 1, where we provide the behavior of each test under the null hypotheses for Models 1–3 respectively.

The tests are statistically conclusive, and provide statistically significant estimates (except LR-2 for the DOW 10). All Bai and Ng (2006) easily reject the null that market volatility is the same as the PCA factor. The Onatski (2009) test supports the existence of at least one more factor after regressing out market volatility. The LR-1 test resoundingly rejects the null for both datasets, which supports the graphical evidence that the market volatility is a driver of the overall panel. The LR-2 null hypothesis is rejected for the S&P 100, but not for the DOW 10. This suggests that for the DOW 10 dataset the market volatility might be extraneous once the first PC is extracted, but that for the S&P 100 dataset, the market volatility still holds meaningful information for the cross-section even after extracting the first PC. All results can be found in Table 2.

The statistical tests therefore strongly support Model 2. Both market volatility *and* a

Table 2: Statistical Tests for Equities

Table with statistical tests for the two equities datasets (DOW 10 and S&P 100). LR-1 and LR-2 tests display likelihood ratio statistics for the null hypothesis that the coefficients on market volatility should be 0. LR-1 performs the test on the panel of idiosyncratic volatilities, while LR-2 performs the test on panel residuals after extracting the first Principal Component. Onatski is the test for factor structure described in Onatski (2009) where the null hypothesis is that there is no factor structure after regressing out the market volatility.  $A$  and  $M$  are exact tests from Bai and Ng (2006), while  $NS$  and  $R^2$  are approximate tests from the same paper. Note that  $A$  has no critical values, but the test statistic should converge to  $2\alpha$  for  $\alpha$  confidence level. \*\* denotes significant at 5%, \*\*\* denotes significant at 1%.

Test	DOW 10	S&P 100
LR - 1	36461***	471729***
LR - 2	13.72	135***
Onatski	12.13***	12.36***
$A$	0.50***	0.84***
$M$	14***	75***
$NS$	0.38	0.37
$R^2$	0.72	0.73
$CI(R^2)$	(0.70, 0.79)	(0.71, 0.75)

principal component factor are needed to explain the panel. The two are not the same, and neither makes the other extraneous.

## 1.4 Foreign-Exchange Rates

Next we move on to our analysis of Foreign Exchange rate returns. We consider a panel of 15 exchange rates from major economies (a full list can be found in the Appendix).

Our data consists of daily FX returns downloaded from FRED, confined to the post-Euro era, so our sample runs from January 1999 to October 2015. Since the returns are daily, we aggregate to monthly realized beta. For the market factor, we use an equal-weighted average of all the returns. This index is 99.9% correlated with the first principal component of returns. The order of our estimation procedure is exactly analogous to our equities data analysis, save that the frequencies are all shifted to be lower – high frequency exchange rate returns are now daily returns.

### 1.4.1 Comparison with Equities Results

We are interested if the structure in idiosyncratic volatility is confined merely to equities or also applies to other financial datasets. It turns out that many of the general features present in equities is also present in FX, though there are some important differences. We start with a graphical analysis of the data and then continue on to the statistical tests.

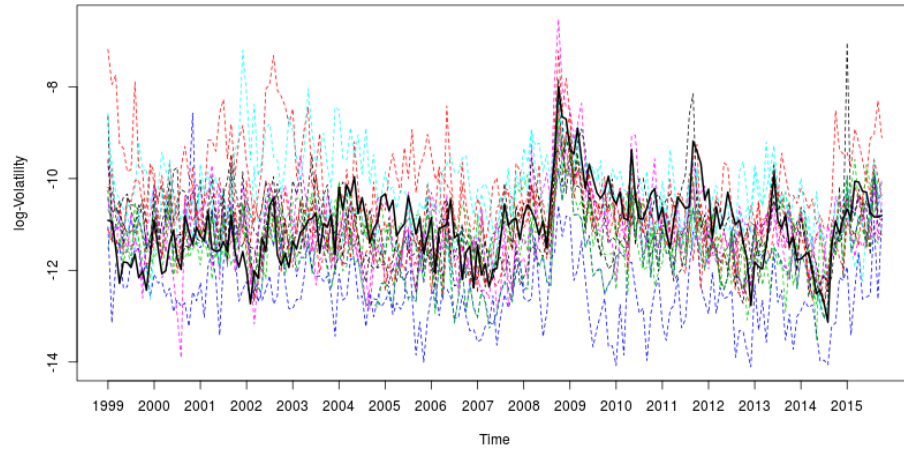
#### Graphical Analysis

The graphical analysis begins in Figure 8, where we present the panel of idiosyncratic volatility together with the market volatility. As in the case of equities, there are clear dynamics in idiosyncratic volatility, and they display factor structure. Moreover, the market volatility has dynamics consistent with the rest of the panel. In contrast to equities, the factor structure seems weaker here, as individual exchange rates frequently deviate from the rest of the panel.

The weaker factor structure is further supported by Figure 9. Whereas in equities the average  $R^2$  were 50% and 55% for DOW 10 and S&P 100, respectively, the market volatility only

Figure 8: FX Factor and Idiosyncratic Volatility

Panel of log exchange rate volatilities from 1999–2016 (AL, BZ, CA, DN, JP, KO, MX, NZ, NO, SI, SF, SZ, UK, EU). In black is the equal-weighted average of all returns (approximately the first PCA).



explains, on average, 18% of cross sectional variation. Additionally, the first PC explains only 47% of cross sectional variation, compared to 68% and 76% for the DOW 10 and S&P 100. Similar to equities, when we take two factors, the structure is familiar, though again, the levels are lower. Market volatility and a PC factor explain 50% of the cross sectional variation (as compared to 68% and 76% for DOW 10 and S&P 100).

Thus, in the case of FX returns, there are three major differences. First, the factor is much weaker. No matter which factor you use, the amount of cross-sectional variation is substantially lower. Second, the discrepancy between average explained variation from market volatility and the first PC is much larger. The first PC explains almost 30 percentage points more than cross-sectional variation of FX returns. Lastly, the two are much more dissimilar than their counterparts in the equities datasets. Indeed, the correlation between market volatility and the first PC of idiosyncratic volatility is 0.57, which is much lower than the 0.85 for both equities datasets. The 6-month rolling window of the PCA factor and the market volatility are plotted in Figure 10.

Thus, from graphical analysis, we immediately gain insight into similarities and differences between FX returns and equity returns. In the case of FX, market return does not do a good

Figure 9: Explained Variation in FX Idiosyncratic Volatility

Panel of  $R^2$  for each exchange rate when regressed against market volatility (the equal weighted average), the first PCA, and both. Despite high correlation of market volatility and first PCA, the first PCA has on average greater explanatory power. However, for a few assets, the gains from adding market volatility to the first PCA are also nontrivial (see BZ and CA).

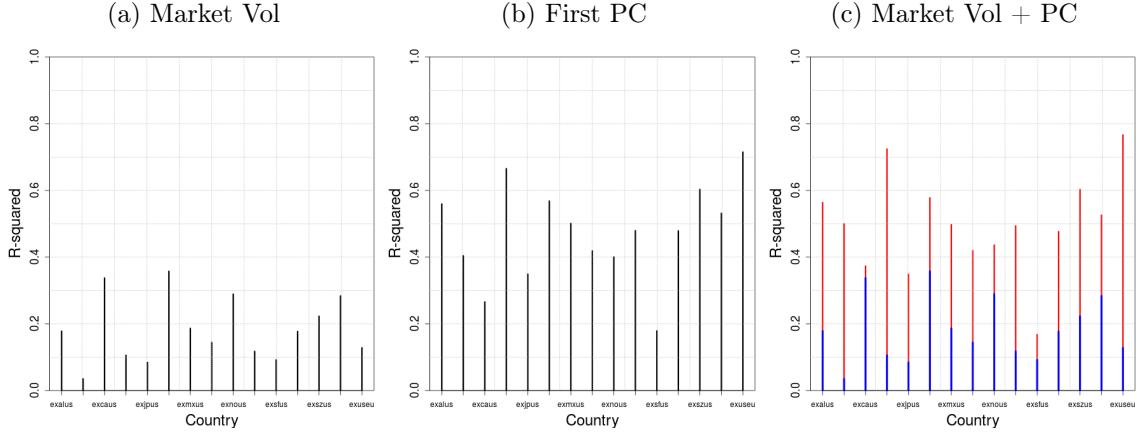
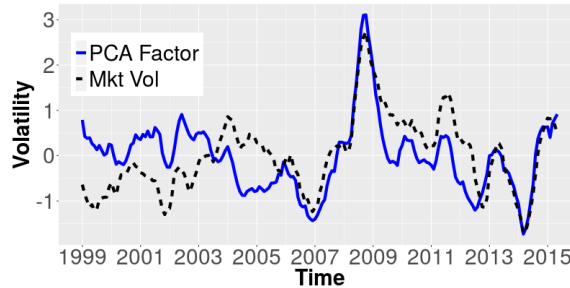


Figure 10: PCA Factor vs Market Volatility

6 month rolling window of volatility. Blue (solid) line displays the first PC of idiosyncratic volatility, while the black (dashed) line displays the market volatility. Both volatilities have been centered and scaled to have mean 0 and variance 1.



job explaining cross sectional variation, whereas the first PC does much better. Indeed, they are weakly correlated at 57%. Nonetheless, when the two are paired together, most cross sectional variation is explained. The average  $R^2$  from the two factor model is exactly the same as that of only the first PC, but similar to equities, the distribution of  $R^2$ s is not the same.

Table 3: Statistical Tests for FX panel

Table with statistical tests for the FX rate dataset. LR-1 and LR-2 tests display likelihood ratio statistics for the null hypothesis that the coefficients on market volatility should be 0. LR-1 performs the test on the panel of idiosyncratic volatilities, while LR-2 performs the test on panel residuals after extracting the first Principal Component. Onatski is the test for factor structure described in Onatski (2009) where the null hypothesis is that there is no factor structure after regressing out the market volatility.  $A$  and  $M$  are exact tests from Bai and Ng (2006), while  $NS$  and  $R^2$  are approximate tests from the same paper. Note that  $A$  has no critical values, but the test statistic should converge to  $2\alpha$  for  $\alpha$  confidence level. \*\* denotes significant at 5%, \*\*\* denotes significant at 1%.

Test	Forex
LR - 1	1221***
LR - 2	95***
Onatski	17.46***
$A$	0.76***
$M$	16***
$NS$	1.86
$R^2$	0.35
$CI(R^2)$	(0.24, 0.45)

### Statistical Tests

We run the same battery of statistical tests on the FX data as we did equities. Due to the graphical analysis above, we expect to easily reject the null that the PC factor is the same as market volatility (Bai and Ng (2006) tests) and that there is no factor structure once the market is taken into account (Onatski (2009) test). Somewhat surprisingly though, both LR tests also reject the null that market volatility should not be included at all. All results are in Table 3.

In conclusion, both datasets support the notion that there is factor structure in idiosyncratic volatility and that the panel of idiosyncratic volatility is best explained via two factors; one is the market factor and one is a PC factor.

## 1.5 Conditional Mean Misspecification

Conditional mean misspecification could also generate this observed factor structure. As a preliminary exercise, observe that if the true DGP is:

$$y_t = \beta_1 f_t + \beta_2 X_t + e_t$$
$$f_t \sim N(0, \sigma_{f,t}^2) \quad X_t \sim (0, \sigma_X^2)$$

Yet the estimated model is:

$$y_t = \bar{\beta}_1 f_t + \bar{e}_t$$

Then:

$$\mathbb{E}[\bar{\beta}_1] = \beta_1 + \beta_2 \frac{\text{Cov}(f_t, X_t)}{\mathbb{V}[f_t]} = \beta_1 \quad \text{if } \text{Cov}(f_t, X_t) = 0$$
$$\bar{e}_t = \beta_2(X_t) + e_t$$
$$\mathbb{V}_t[\bar{e}_t] = 2\sigma_X^2 \beta_2 \beta_2' + \mathbb{V}_t[e_t]$$

Even if  $\mathbb{V}_t[e_t] = c$ ,  $\mathbb{V}_t[\bar{e}_t]$  will be time-varying with factor structure. If  $X_t$  is a function of  $f_t$ , in particular suppose the conditional mean is a higher-order polynomial of  $f_t$ ,  $\mathbb{V}_t[\bar{e}_t]$  will also comove with market volatility! While this example is obviously contrived, it is important to point out that in the presence of *any* omitted factors from the level equation, there will be factor structure in idiosyncratic volatility. Indeed, Herskovic et al. (2014) fit a large factor model (5 principal components) to the level equation, but still found the same structure in idiosyncratic volatility. Since we are specifically interested in how the structure might effect the relationship with factor volatility, we run our intraday factor regression with four powers of the observed factor.



Table 4: Statistical Tests for Higher Powers of Market Return

Table with statistical tests for the two equities datasets (DOW 10 and S&P 100) where the factors are the first four powers of the observed market factor (SPY). LR-1 and LR-2 tests display likelihood ratio statistics for the null hypothesis that the coefficients on market volatility should be 0. LR-1 performs the test on the panel of idiosyncratic volatilities, while LR-2 performs the test on panel residuals after extracting the first Principal Component. Onatski is the test for factor structure described in Onatski (2009) where the null hypothesis is that there is no factor structure after regressing out the market volatility.  $A$  and  $M$  are exact tests from Bai and Ng (2006), while  $NS$  and  $R^2$  are approximate tests from the same paper. Note that  $A$  has no critical values, but the test statistic should converge to  $2\alpha$  for  $\alpha$  confidence level. While the Bai and Ng (2006) tests generate test statistics for each of the four powers, we only report the results for the first power (market volatility). \*\* denotes significant at 5%, \*\*\* denotes significant at 1%.

Test	DOW 10	S&P 100
LR - 1	33603.22***	395929***
LR - 2	41.1511	15436***
Onatski	8.75**	122***
$A$	0.87***	0.89***
$M$	64***	124***
$NS$	0.38	1.88
$R^2$	0.72	0.34
$CI(R^2)$	(0.70, 0.75)	(0.32, 0.36)

For the DOW10, the correlation between the market factor and the first PC of idiosyncratic volatility is 0.85, exactly the same as it was when we fit a single factor. For the S&P100, the correlation between market volatility and the first PC of idiosyncratic volatility drops from 0.85 to 0.56. While this drop is fairly large, our statistical tests, especially the LR tests, show that the market volatility is still a vital component of the panel. While the Bai and Ng (2006) test produces a statistic for all four observed factors (the volatilities of the powers of market returns), we only report statistics for the market volatility, as the results are nearly identical for all of them. The results of the statistical tests are in Table 4. A particularly striking result is the difference between the second likelihood ratio test at  $N = 10$  (the DOW10) and  $N = 100$  (the S&P100). This is likely owing to the blessing of dimensionality and improved inference of factor structure as  $N$  becomes large.

While the test statistics change, since we are now testing more restrictions (in the case of the LR tests), the overall picture is still the same. The LR tests are all resoundingly rejected, so the volatilities of powers of market returns cannot be excluded from the model.

## 1.6 Forecasting

In addition to assessing the relationship between the factor for idiosyncratic volatility and market volatility, we also explore what, if any, impact the factor for volatility has on volatility forecasting. In addition to the three models we presented in Section 1.3.5, we include two additional benchmark models:

1. BMK – The benchmark model where only the factor has time-varying volatility (constant idiosyncratic volatility). Jacquier et al. (1994) proposed a Stochastic Volatility version of this model, though they did not estimate it. Diebold and Nerlove (1989) proposed and estimated a similar model, where the factor volatility is an ARCH process.
2. AR – In addition to time-varying volatility in the factor, idiosyncratic volatility is also time-varying, but they vary as independent autoregressions. Kim et al. (1998) proposed this multivariate stochastic volatility model, though Pitt and Shephard (1999) and Aguilar and West (2000) independently (and with different MCMC techniques) actually produced estimation procedures.

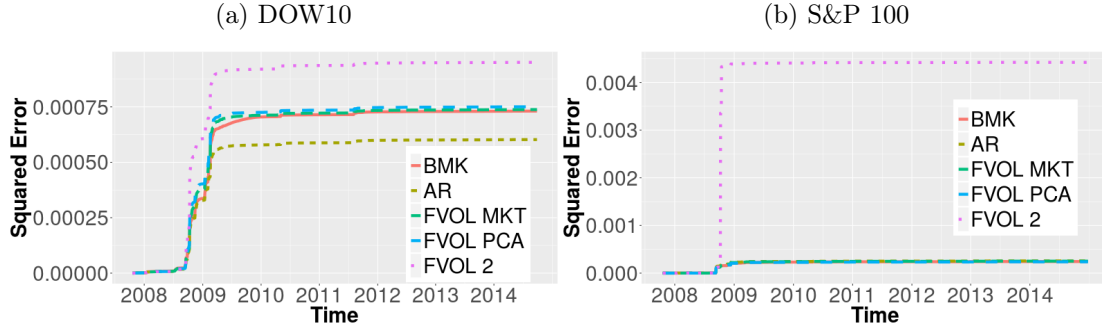
In order to estimate our three Factor for Idiosyncratic Volatility models, we proceed in one of the following ways:

- For model 1 of 3 (FVOL MKT), we regress each of the log-diagonal vector of  $\Omega_{v_t}$ ,  $\sigma_{e_t^i}$ , against  $\log(\sigma_t^f)$  to estimate  $\beta_i$ .
- For model 2 of 3 (FVOL2), regress  $\log(\sigma_{e_t^i})$  against  $\log(\sigma_{F_t})$ , and conduct PCA on the panel of residuals.
- For model 3 of 3 (FVOL PCA), we conduct PCA directly on the panel of log-idiosyncratic volatilities.  $\sigma_{e_t^i}$ . We then regress the residuals against  $\sigma_{F_t}$ .

For all datasets we focus on the forecast errors of the panel of variances. Correlations are modeled via loadings from the level regression, which are the same for all models. All models

Figure 11: Equities Squared One-Step Prediction Errors

Cumulative squared errors over time, 2007–2015, of DOW 10 and S&P100. Each date adds the average squared distance of true volatility to forecasted volatility over the panel. The models perform similarly outside the financial crisis 2008–2010, but there the discrepancies are large.



and datasets forecast poorly at the beginning of the financial crisis in 2008, so we report both average Mean Squared Error (MSE) and Median Absolute Error (MAE), where the mean/median is taken across time for each asset and then averaged across assets. We also plot the cumulative squared one-step ahead forecast errors, both for the whole sample and pre- and post-2008 (FX is also plotted pre-2008).

### 1.6.1 Equities

For both equities datasets, we use a 200 day rolling window estimation period. In each period we estimate each of the five competing models and forecast ahead 1–12 days. Due to the fact that there are some large outliers (even outside the financial crisis), we record both Average MSE and MAE. The DOW 10 forecasting results are presented in Table 5, while results for the S&P 100 dataset are in Table 6. One-step-ahead cumulative squared forecast errors for both datasets are plotted in Figure 11. To ensure the results are not solely driven by dynamics in the crisis, we also present (in the Appendix) tables of forecasting results and figures with squared forecast errors using forecasts only after January 2009. The DOW 10 forecasting results are in Table 12, while the S&P 100 results are in Table 13. Squared forecast errors for both datasets are plotted in Figure 13.

First focus on the DOW 10 dataset in Table 5. By average MSE, all FVOL models forecast

Table 5: Mean Square Error, Median Absolute Error of DOW 10 Rvariances

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. All models use a 200-day rolling window to estimate parameters, followed by forecasts for 1–12 days ahead.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	<b>0.82</b>	1.01	1.03	1.30	0.83	0.89	<b>0.82</b>	0.83
2	<b>0.86</b>	0.99	0.98	1.11	0.88	0.90	<b>0.83</b>	0.85
3	<b>0.90</b>	1.00	1.02	1.08	0.90	0.91	<b>0.85</b>	0.87
4	<b>0.92</b>	0.99	0.99	1.05	0.90	0.90	<b>0.85</b>	0.88
5	<b>0.93</b>	1.00	1.05	1.06	0.91	0.91	<b>0.85</b>	0.88
6	<b>0.94</b>	1.01	1.03	1.05	0.92	0.92	<b>0.86</b>	0.90
7	<b>0.94</b>	1.02	1.05	1.03	0.94	0.94	<b>0.87</b>	0.91
8	<b>0.95</b>	1.01	1.02	1.02	0.94	0.94	<b>0.86</b>	0.91
9	<b>0.94</b>	1.02	1.03	1.00	0.94	0.93	<b>0.86</b>	0.91
10	<b>0.94</b>	1.01	1.01	1.00	0.94	0.94	<b>0.87</b>	0.92
11	<b>0.95</b>	1.01	1.02	1.00	0.94	0.94	<b>0.88</b>	0.93
12	<b>0.95</b>	1.01	1.00	1.00	0.95	0.94	<b>0.88</b>	0.94

variances about as well, though FVOL 2 does slightly worse than the others at short horizons. In addition, the model of Pitt and Shephard (1999) (AR) does very well, clearly supporting the hypothesis that idiosyncratic variance is at least time-varying. Despite the FVOL models not performing particularly well, their worse performance is mainly centered around the financial crisis, specifically around late 2008. When we look at average MAE instead of MSE, we see that all models provide substantial forecasting improvements as compared to the benchmark model. The Pitt and Shephard (1999) (AR) model still performs about as well, but introducing some sort of factor on idiosyncratic volatility also performs comparably well with much fewer estimated parameters. Specifically, using a PCA factor to forecast idiosyncratic volatility works best at all horizons.

In the larger, S&P 100, sample, the results are qualitatively similar. Once again, all models perform very similarly when compared via average MSE. This time though, the AR model slightly underperforms the benchmark, the PCA factor slightly outperforms the benchmark,

Table 6: Mean Square Error, Median Absolute Error of S&P 100 Rvariances

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. All models use a 200-day rolling window to estimate parameters, followed by forecasts for 1–12 days ahead.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	1.06	1.04	<b>0.97</b>	18.18	0.70	0.79	0.70	<b>0.70</b>
2	1.17	1.05	<b>0.98</b>	27.19	0.76	0.83	<b>0.74</b>	0.74
3	1.18	1.07	<b>0.99</b>	9.77	0.79	0.84	<b>0.75</b>	0.77
4	1.34	1.09	<b>0.98</b>	15.11	0.81	0.85	<b>0.75</b>	0.77
5	1.16	1.08	<b>0.98</b>	9.49	0.83	0.86	<b>0.76</b>	0.79
6	1.29	1.04	<b>0.99</b>	7.58	0.85	0.87	<b>0.77</b>	0.81
7	1.22	1.07	<b>1.00</b>	10.56	0.86	0.87	<b>0.78</b>	0.82
8	1.18	1.05	<b>0.99</b>	4.61	0.87	0.89	<b>0.79</b>	0.84
9	1.24	1.11	<b>0.99</b>	3.97	0.88	0.89	<b>0.79</b>	0.84
10	1.28	1.12	<b>0.99</b>	3.23	0.89	0.90	<b>0.79</b>	0.84
11	1.26	1.13	<b>0.99</b>	2.49	0.89	0.91	<b>0.79</b>	0.86
12	1.24	1.16	<b>0.99</b>	1.36	0.90	0.91	<b>0.81</b>	0.87

and the FVOL 2 model performs substantially worse. The forecasting deficiencies are mainly due to the financial crisis, and by using average MAE, all FVOL models see large improvements over the benchmark model. The AR model performs very well, but this time both FVOL PCA and FVOL2 do even better. The FVOL MKT once again underperforms the other models, but still beats the benchmark.

Taken together, as the panel of volatilities grows in cross-sectional dimension, the improvements of using FVOL models increases. While using both the market volatility (model 1) and the PCA factor are each helpful, the PCA factor is better for forecasting. This reaffirms the traditional “Blessing of Dimensionality” in factor models – that when dimensions grow, there are increasingly large benefits to fitting factor models rather than attempting to model each series individually.

Table 7: Mean Square Error, Median Absolute Error of FX rate Rvariances

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. For all models, we use a 50-month rolling window where we estimate the model in every window and then forecast for 1–12 months ahead.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	0.98	<b>0.97</b>	1.22	2.47	0.83	0.93	<b>0.81</b>	0.83
2	1.13	1.01	<b>0.98</b>	1.18	0.86	0.96	<b>0.85</b>	0.88
3	1.16	1.02	<b>0.98</b>	1.69	0.91	1.00	<b>0.88</b>	0.92
4	1.10	1.03	1.17	2.47	0.94	0.98	<b>0.87</b>	0.94
5	1.08	1.03	37.84	2.40	0.96	1.00	<b>0.92</b>	0.96
6	1.11	1.06	4.44	15.41	0.99	1.02	<b>0.95</b>	1.01
7	1.06	1.11	1.42	25.39	0.97	0.99	<b>0.95</b>	1.00
8	1.03	1.03	4.20	1.53	0.98	1.01	<b>0.94</b>	1.03
9	1.02	1.05	1.60	1.21	0.97	1.00	<b>0.95</b>	0.99
10	1.03	1.04	1.32	1.19	0.97	1.01	<b>0.95</b>	1.01
11	1.03	1.02	1.01	1.06	0.95	0.97	<b>0.88</b>	0.96
12	1.04	1.02	1.00	1.02	0.98	0.99	<b>0.92</b>	0.97

### 1.6.2 Exchange Rates

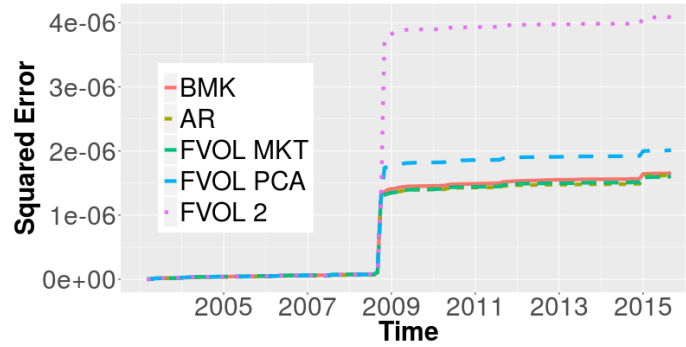
We use the same set of competing models to predict FX monthly volatilities, but this time use a rolling window of 50 months. We report both average MSE and MAE prediction error, as forecast errors are non-Gaussian. The table with forecasting performance is in Table 7 while the plot of squared prediction error is in Figure 12. We also include figures of squared prediction error for pre-August 2008 and post January 2009 in the Appendix (Figure 14), and forecasting results only post-2009 (Table 14).

When compared via MSE, most models do not make much of an improvement over the benchmark, if any at all. The FVOL MKT model performs slightly better at horizon 1, though worse at all other horizons. FVOL PCA performs best at horizon 2 and 3, but overall they both underperform the benchmark.

On the other hand, when compared via average MAE, the factor in idiosyncratic volatility

Figure 12: FX Squared One-Step Prediction Errors

Cumulative squared errors over time, 2007–2015, of the panel of exchange rate volatilities. Each date adds the average squared distance of true volatility to forecasted volatility over the panel. The models perform similarly outside the financial crisis 2008–2010, but there the discrepancies are large.



has a large impact on improving forecasts. All FVOL models perform much better (10–20%) than the benchmark, especially at short horizons. Similar to equities, the FVOL PCA model performs best at all horizons.

## 1.7 Conclusion

We have revisited the standard factor model, and its use in facilitating tractable dynamic volatility. We have shown that  $\Sigma_{e_t}$  is correlated with  $\Sigma_{F_t}$ , but that  $\Sigma_{F_t}$  alone is not sufficient for explaining time-variation in idiosyncratic volatility. This suggests that the classic decomposition is ultimately not an optimal approach to modeling time-varying volatility. Furthermore, one might conclude that if modeling panels of volatilities, and not covariances, is the practitioner's goal, then one should fit factor models to panels of volatilities directly. This result holds across a wide variety of asset classes and time frequencies.

We briefly explored the implications of these results for forecasting, but much remains to be done. In particular, do these hierarchical factor structures help in constructing density forecasts for returns? Are these risk factors for idiosyncratic volatility priced? Our preliminary evidence on both questions suggest negative results, but these results could be sensitive to the time horizon of the sample, the specific equity market, or even the industry.

The presence of this structure in both equities and FX data suggests it may be a more general feature of volatility. It remains to be argued why the nature of panels of volatility should lend themselves to such hierarchical structures, whether through network effects or an endogenous economic mechanism. Indeed, due to the fact that FX rates and equities are entirely different asset classes, the empirical phenomenon may be more of a statistical phenomenon (such as factor structure) than one that is driven by structural theory. It also remains to be shown whether this feature appears in other panels of volatilities, for example in the volatility of large macroeconomic panels. Finally, our framework here did not accurately account for measurement error in the panels of volatilities. Using frontier theory on the distribution of realized volatility estimators one can extend this work to account for measurement error, and this represents an avenue for future contributions.



## 1.8 Appendix

### 1.8.1 Simulation Appendix

In this section we confirm the appropriateness of our battery of statistical tests. There are several issues to consider that may warrant skepticism of their use in our environment: (1) Our observed factor volatility (market volatility) is actually observed with measurement error (as it is a realized measure), (2) our panel of interest itself is observed with measurement error (realized measures of idiosyncratic volatility), and (3) our models contain correlated regressors (as the market volatility factor is correlated with the first Principal Component of idiosyncratic volatility).

To assuage our concerns with all three issues, we conduct the following simulation. We generate output using Models 1, 2, and 3 as the data generating processes, for the cases of  $N = 10, 100, 200$ ,  $T = 500, 2000$ , and intraday observations of 100 and 1000. The log-market volatility is generated as an AR(1) process with AR parameter 0.9 and mean -9. The factor structure (whether Model 1, 2, or 3) is defined in terms of log-volatilities. All factor loadings (for all possible factors) are distributed as absolute value of normals with mean zero and standard deviation 0.5. In Model 2, the PCA factor is generated as the market (log) volatility plus classical measurement error with variance calibrated so that the PCA factor is 75% correlated with the market volatility. Intraday observations are taken as iid draws from a normal distribution with mean 0 and variance the true volatility. Realized volatilities are calculated as in Barndorff-Nielsen and Shephard (2004), the outer product of high-frequency returns. While we acknowledge that the high-frequency generation process is simplistic (and unrealistic), note that the most important object is the signal-to-noise ratio between true and realized volatility. With a more complex DGP, one should use a more sophisticated estimation procedure to maintain a similar amount of information. Factor loadings vary every day as iid noise centered around constant loadings.

We then conduct our battery of tests on each set of data generated for 1,000 simulations,

and determine if the tests have correct size and power for the respective data generating processes and null hypotheses. The results are very promising and presented in Table 8. Recall Table 1. We expect the LR-1 test to have appropriate power, rejecting the null in all cases<sup>2</sup>. In Model 3, LR-1 has appropriate power even against a correlated regressor, as we regress out  $\sigma_{F_t}$  first. By contrast, LR-2 will under-reject Models 1 and 3, as it is facing an alternative of correlated regressors (similar to a t-test in a simple regression setup). We see this in practice. The approximate Bai and Ng tests behave as expected. Notably, these tests are reasonably robust to measurement error both in the panel and in the observed factor for volatility: In the case of 100 intraday observations, the measurement error volatility in idiosyncratic volatility is 5% of the volatility in the panel, and the tests behave as expected.

### Measurement Error And Simulation Results

The exact Bai and Ng tests, as well as the Onatski tests, do not behave as desired in a high frequency simulation setting. We note in particular that when Model 1 is the null, both of these tests strongly over-reject. This suggests that our preference for Models 2 and 3 in the empirical results should potentially be taken with a grain of salt. In this section we explore the role that measurement error in the realized measures of market volatility and idiosyncratic volatility can play in explaining these results.

Recall the construction of realized measures, and suppose we are trying to select between Models 1, 2, and 3. Further suppose we measure idiosyncratic volatility accurately via a direct method. For example, with a high number of intraday observations, our measurement of realized beta will be accurate, so we may construct high frequency idiosyncratic returns directly, resulting in more classical measurement error in idiosyncratic volatilities. We still estimate the factor structure by estimating the regression

$$RIV_{it} = \mu_i + \beta_i RV_{f_t} + u_t^i \tag{1.8.1}$$

---

<sup>2</sup>Note that even in the case of Model 1, LR-1 should reject  $\beta = 0$  since the PCA factor should be the same as the market volatility.

If  $RV_{f_t} = \sigma_{F_t}^2 + \epsilon_{F_t}$ , then the parameter estimate  $\widehat{\beta}_i$  will be biased downward relative to the true regression coefficient between  $\sigma_{it}^2$  and  $\sigma_{F_t}^2$  due to attenuation bias from measurement error. The result is  $\widehat{u}_t^i$  will exhibit factor structure regardless of the nature of  $u_t^i$ . Note that in practice error in market volatility realized estimation will be correlated with error in idiosyncratic volatility realized estimation, which will reduce the magnitude of this problem – having correlated errors on LHS and RHS diminishes the impact of attenuation bias from RHS measurement error<sup>3</sup>.

Consider the results of our Onatski test: because the test statistic regresses out market volatility and examines the factor structure of the remaining residuals, it is subject to the above error. As a result, it over-rejects. We can find confirming evidence for this story by running the simulation using true market volatility in place of a realized estimate in the estimation of Equation 1.8.1. When we do this we find that Onatski rejects with the correct rate. We also consider alternative explanations of the phenomenon by running the simulation with different measurement error specifications – in particular we find that classical measurement error on the true market volatility still induces Onatski to over-reject. Thus the over-rejection is simply a matter of having positive measurement error at all, rather than depending on the exact nature of error in the realized estimator.

## 1.8.2 Data Lists

In this section, we present each of the three datasets used in the paper, as well as data descriptions. For the equity datasets we provide ticker and company name, and for the S&P 100 dataset we also present the sector. For the FX dataset we provide data label from FRED as well as the currencies.

---

<sup>3</sup>Consider regressing  $y$  against  $x$  when we observe  $\tilde{y} = y + \varepsilon$  and  $\tilde{x} = x + v$ . Then

$$\begin{aligned} y &= \beta x + u \\ \tilde{y} &= \beta \tilde{x} + u - \beta v + \varepsilon \end{aligned}$$

Thus a positive correlation between  $v$  and  $\varepsilon$  means the bias in  $\beta$  above is smaller than the bias in the case when  $\varepsilon = 0$ .

- Table 9 – List of companies used for DOW 10 analysis.
- Table 10 – List of companies (and sectors) used for S&P 100 analysis.
- Table 11 – List of currencies used for FX rate analysis.

Figure 13: Equities cumulative squared forecast errors – Post 2009

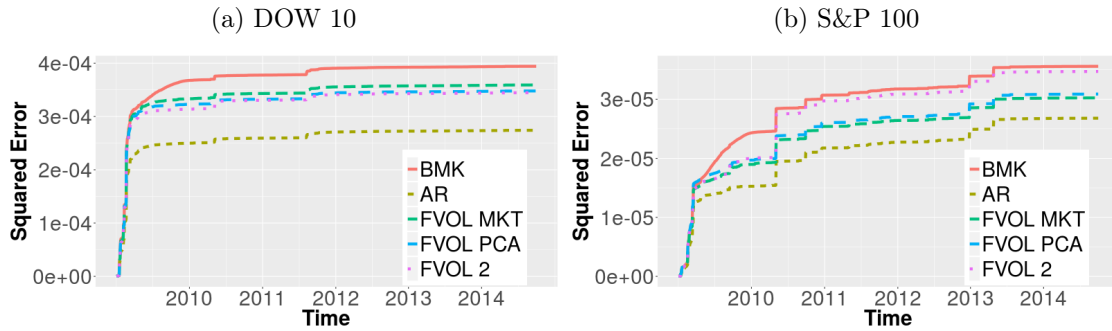
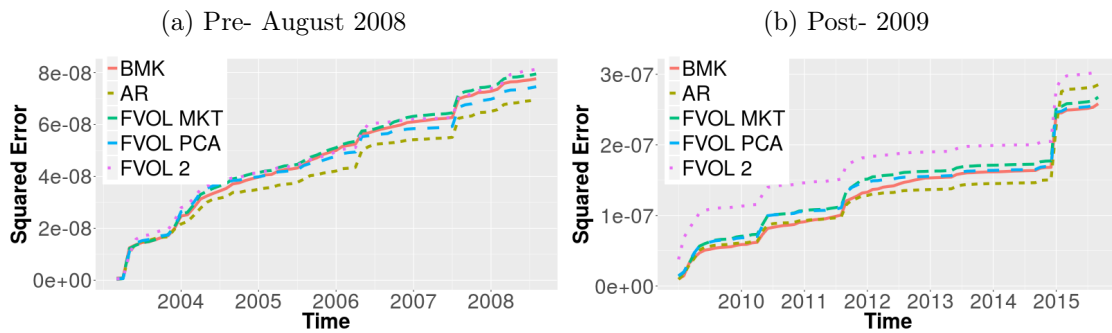


Figure 14: FX cumulative squared forecast errors



### 1.8.3 Forecasting Tables and Figures

In this Appendix we present extra tables and figures from the forecasting exercises.

- Table 12 – DOW 10 forecasting results (average MSE and MAE) using forecasts only after 2009.
- Table 13 – S&P 100 forecasting results (average MSE and MAE) using forecasts only after 2009.
- Table 14 – FX rate forecasting results (average MSE and MAE) using forecasts only after 2009.
- Figure 13 – Plot of squared forecast errors for both equity datasets, post 2009.
- Figure 14 – Plot of squared forecast errors for FX dataset, pre-2008 and post-2009.

Table 8: Simulation Results

Results of 1000 replications of each model. Columns are labelled M-1, M-2, and M-3 corresponding to Models 1, 2, and 3 respectively. Column and row segments are labelled based on corresponding dimensions N and T and number of intraday observations. Tests LR-1, LR-2 and Onatski report empirical size of 95% cutoff values. Tests A, M, NS and R2 (those from Bai and Ng (2006)) report average values across simulations.

		N = 10			N = 100			N = 200			
		M-1	M-2	M-3	M-1	M-2	M-3	M-1	M-2	M-3	
Intraday = 100	T = 500	LR-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		LR-2	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
		Onatski	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		A	0.64	0.95	0.91	0.88	0.97	0.97	0.91	0.98	0.98
		M	24.40	62.80	115.29	43.57	53.73	143.83	61.32	69.46	189.80
		NS	0.06	8.72	1.09	0.06	3.16	1.06	0.06	3.20	1.06
		R2	0.97	0.81	0.59	0.97	0.69	0.59	0.97	0.68	0.59
	T = 2000	LR-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		LR-2	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
		Onatski	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		A	0.62	0.94	0.91	0.86	0.97	0.96	0.90	0.97	0.97
		M	29.65	63.73	142.82	48.26	49.97	156.60	65.05	62.25	203.59
		NS	0.05	5.22	1.01	0.05	2.11	0.98	0.05	1.81	0.98
		R2	0.97	0.81	0.57	0.97	0.63	0.57	0.97	0.61	0.57
Intraday = 1000	T = 500	LR-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		LR-2	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
		Onatski	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		A	0.63	0.99	0.97	0.87	0.99	0.99	0.91	1.00	1.00
		M	23.07	55.09	348.38	41.32	49.87	422.75	58.08	62.42	573.11
		NS	0.01	7.23	0.97	0.01	3.38	0.97	0.01	2.76	0.98
		R2	1.00	0.84	0.61	1.00	0.70	0.61	1.00	0.70	0.61
	T = 2000	LR-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		LR-2	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
		Onatski	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		A	0.62	0.98	0.97	0.86	0.99	0.99	0.90	0.99	0.99
		M	27.92	54.86	443.58	45.58	45.09	467.36	62.37	58.50	614.67
		NS	0.00	5.16	0.92	0.00	1.95	0.90	0.00	1.81	0.90
		R2	1.00	0.83	0.59	1.00	0.65	0.59	1.00	0.62	0.60

Table 9: DOW 10 Company List

Ticker	Name
AA	Alcoa Inc
AXP	American Express
BAC	Bank of America
DD	Du Pont
GE	General Electric
IBM	International Business Machines
JPM	JPMorgan Chase
KO	Coca-Cola
MSFT	Microsoft
XOM	Exxon Mobil

Table 10: S&amp;P100 Company List

Ticker	Name	Sector	Ticker	Name	Sector
AAPL	Apple Inc.	Info Tech	IBM	Intl Business Machines Corp	Info Tech
ABT	Abbott Laboratories	Health Care	INTC	Intel Corp	Info Tech
ACN	Accenture plc	Info Tech	JNJ	Johnson & Johnson	Health Care
AIG	American International Group	Financials	JPM	JP Morgan Chase & Co	Financials
ALL	Allstate Corp	Financials	KO	Coca-Cola Co	Cons. Staples
AMGN	Amgen Inc	Health Care	LLY	"Lilly	Health Care
AMZN	Amazon.com Inc	Cons. Discret.	LMT	Lockheed Martin	Industrials
APA	Apache Corp	Energy	LOW	Lowe's Cos Inc	Cons. Discret.
APC	Anadarko Petroleum Corp	Energy	MA	Mastercard Inc A	Info Tech
BA	Boeing Co	Industrials	MCD	McDonald's Corp	Cons. Discret.
BAC	Bank of America Corp	Financials	MDT	Medtronic Inc	Health Care
BAX	Baxter Intl Inc	Health Care	MET	Metlife Inc	Financials
BHI	Baker Hughes Inc	Energy	MMM	3M Co	Industrials
BK	The Bank of New York Mellon Corp	Financials	MO	Altria Group Inc	Cons. Staples
BMJ	Bristol-Myers Squibb	Health Care	MON	Monsanto Co.	Materials
C	Citigroup Inc	Financials	MRK	Merck & Co Inc	Health Care
CAT	Caterpillar Inc	Industrials	MS	Morgan Stanley	Financials
CL	Colgate-Palmolive Co	Cons. Staples	MSFT	Microsoft Corp	Info Tech
CMCSA	Comcast Corp	Cons. Discret.	NKE	NIKE Inc B	Cons. Discret.
COF	Capital One Financial	Financials	NOV	National Oilwell Varco Inc	Energy
COP	ConocoPhillips	Energy	NSC	Norfolk Southern Corp	Industrials
COST	Costco Wholesale Corp	Cons. Staples	ORCL	Oracle Corp	Info Tech
CSCO	Cisco Systems Inc	Info Tech	OXY	Occidental Petroleum	Energy
CVS	CVS Caremark Corp.	Cons. Staples	PEP	PepsiCo Inc	Cons. Staples
CVX	Chevron Corp	Energy	PFE	Pfizer Inc	Health Care
DD	"DuPont	Materials	PG	Procter & Gamble	Cons. Staples
DIS	Walt Disney Co	Cons. Discret.	QCOM	QUALCOMM Inc	Info Tech
DOW	Dow Chemical	Materials	RTN	Raytheon Co	Industrials
DVN	Devon Energy Corp	Energy	SBUX	Starbucks Corp	Cons. Discret.
EBAY	eBay Inc.	Info Tech	SLB	Schlumberger Ltd	Energy
EMC	EMC Corp	Info Tech	SO	Southern Co	Utilities
EMR	Emerson Electric Co	Industrials	SPG	Simon Property Group	Financials
EXC	Exelon Corp	Utilities	T	AT&T Inc	Telecom Services
F	Ford Motor Co	Cons. Discret.	TGT	Target Corp	Cons. Discret.
FCX	Freeport McMoRan Copper & Gold	Materials	TWX	Time Warner Inc	Cons. Discret.
FDX	FedEx Corp	Industrials	TXN	Texas Instruments Inc	Info Tech
GD	General Dynamics	Industrials	UNH	Unitedhealth Group Inc	Health Care
GE	General Electric Co	Industrials	UNP	Union Pacific Corp	Industrials
GILD	Gilead Sciences Inc	Health Care	UPS	United Parcel Service Inc B	Industrials
GOOG	Google Inc	Info Tech	USB	US Bancorp	Financials
GS	Goldman Sachs Group Inc	Financials	UTX	United Technologies Corp	Industrials
HAL	Halliburton Co	Energy	VZ	Verizon Communications Inc	Telecom Services
HD	Home Depot Inc	Cons. Discret.	WFC	Wells Fargo & Co	Financials
HON	Honeywell Intl Inc	Industrials	WMT	Wal-Mart Stores	Cons. Staples
HPQ	Hewlett-Packard Co	Info Tech	XOM	Exxon Mobil Corp	Energy

Table 11: Forex List

FRED Label	Currency
exalus	Australia / US
exbzus	Brazil / US
excaus	Canada / US
exdnus	Denmark / US
exjpus	Japan / US
exkous	South Korea / US
exmxus	Mexico / US
exnzus	New Zealand / US
exnous	Norway / US
exsius	Singapore / US
exsfus	South Africa / US
exszus	Switzerland / US
exukus	UK / US
exeuus	EU / US

Table 12: Mean Square Error, Median Absolute Error of DOW 10 Rvariances (post 2009)

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. All models use a 200-day rolling window to estimate parameters, followed by forecasts for 1–12 days ahead. Table presents only forecast errors from predictions after 2009.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	<b>0.70</b>	0.91	0.88	0.87	0.83	0.88	<b>0.80</b>	0.81
2	<b>0.75</b>	0.93	0.93	0.89	0.87	0.90	<b>0.82</b>	0.83
3	<b>0.81</b>	0.92	0.94	0.89	0.89	0.90	<b>0.84</b>	0.86
4	<b>0.86</b>	0.93	0.94	0.90	0.90	0.90	<b>0.83</b>	0.86
5	<b>0.87</b>	0.92	0.94	0.90	0.92	0.92	<b>0.85</b>	0.88
6	<b>0.86</b>	0.93	0.94	0.91	0.92	0.92	<b>0.86</b>	0.90
7	<b>0.85</b>	0.92	0.93	0.90	0.93	0.93	<b>0.87</b>	0.90
8	<b>0.87</b>	0.93	0.93	0.91	0.93	0.93	<b>0.85</b>	0.90
9	<b>0.86</b>	0.94	0.94	0.91	0.94	0.93	<b>0.86</b>	0.90
10	<b>0.83</b>	0.93	0.93	0.91	0.94	0.94	<b>0.87</b>	0.90
11	<b>0.85</b>	0.94	0.94	0.92	0.94	0.95	<b>0.88</b>	0.92
12	<b>0.85</b>	0.94	0.93	0.93	0.95	0.93	<b>0.87</b>	0.93



Table 13: Mean Square Error, Median Absolute Error of S&P 100 Rvariances (post 2009)

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. All models use a 200-day rolling window to estimate parameters, followed by forecasts for 1–12 days ahead. Table presents only forecast errors from predictions after 2009.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	<b>0.75</b>	0.85	0.87	0.98	0.68	0.78	0.67	<b>0.65</b>
2	<b>0.83</b>	0.86	0.87	0.93	0.74	0.81	0.71	<b>0.70</b>
3	0.88	<b>0.87</b>	0.88	0.94	0.77	0.82	<b>0.72</b>	0.72
4	0.97	0.87	<b>0.86</b>	0.92	0.79	0.83	<b>0.72</b>	0.73
5	1.03	<b>0.85</b>	0.86	0.90	0.81	0.84	<b>0.74</b>	0.75
6	1.19	<b>0.85</b>	0.86	0.90	0.82	0.85	<b>0.75</b>	0.76
7	1.67	<b>0.86</b>	0.86	0.91	0.84	0.86	<b>0.76</b>	0.78
8	1.59	0.86	<b>0.86</b>	0.90	0.85	0.87	<b>0.77</b>	0.81
9	1.89	0.86	<b>0.85</b>	0.91	0.86	0.87	<b>0.77</b>	0.80
10	2.07	0.88	<b>0.86</b>	0.93	0.87	0.88	<b>0.77</b>	0.80
11	1.99	0.88	<b>0.85</b>	0.95	0.87	0.89	<b>0.78</b>	0.83
12	1.81	0.89	<b>0.86</b>	0.95	0.88	0.89	<b>0.79</b>	0.84

Table 14: Mean Square Error, Median Absolute Error of FX rate Rvariances (post 2009)

All values are relative to BMK forecasts. Bolded value in each row is the minimum, when better than BMK. BMK is benchmark, AR is with univariate autoregressive idiosyncratic volatility, FVOL MKT uses market volatility as a single idiosyncratic vol factor, FVOL PCA uses a single principal component as an idiosyncratic vol factor, FVOL 2 uses both. For all models, we use a 50-month rolling window where we estimate the model in every window and then forecast for 1–12 months ahead. Table presents only forecast errors from predictions after 2009.

$h$	Average MSE				Average MAE			
	AR	FVOL MKT	FVOL PCA	FVOL2	AR	FVOL MKT	FVOL PCA	FVOL2
1	1.10	1.04	1.00	1.19	0.75	0.83	<b>0.72</b>	0.75
2	1.06	1.15	1.00	1.29	0.81	0.91	<b>0.77</b>	0.83
3	1.12	1.19	<b>0.96</b>	1.73	0.92	0.99	<b>0.83</b>	0.94
4	1.72	1.19	<b>0.94</b>	3.10	0.93	1.03	<b>0.86</b>	0.97
5	1.52	1.23	211.40	3.37	0.98	1.04	<b>0.89</b>	0.99
6	1.76	1.35	16.09	8.90	1.01	1.06	<b>0.94</b>	1.01
7	1.36	1.47	3.33	14.53	0.98	1.03	<b>0.91</b>	1.02
8	1.23	1.21	23.83	4.25	0.99	1.05	<b>0.94</b>	1.05
9	1.12	1.30	5.26	2.28	0.96	1.02	<b>0.96</b>	1.03
10	1.17	1.22	3.33	2.07	0.96	0.98	<b>0.92</b>	1.03
11	1.21	1.08	<b>0.99</b>	1.36	0.96	0.97	<b>0.85</b>	0.98
12	1.23	1.09	<b>0.94</b>	1.07	0.95	0.98	<b>0.86</b>	0.95

## CHAPTER 2

# The Spanning Puzzle and Nonlinearities at the Zero Lower Bound

### 2.1 Introduction

The relationship between the yield curve and macroeconomic fundamentals is a modern but rich area of research. Historically, the literature has emphasized different fundamental qualities of the yield curve, such as whether yields must be arbitrage-free, whether the yield curve must be expectations hypothesis-driven, and whether the yield curve is driven by latent or observed factors. Many of these questions that are still actively under discussion. A recent deep divide between the macroeconomic literature and the financial literature is the spanning hypothesis.

On the one hand, simple statistical and financial arguments suggests that the current yield curve must span any factor that is a determinant of yields. The statistical argument notes that most term structure models assume instantaneous risk-free interest rates are affine in risk factors. This implies bond yields are affine in risk factors. Since yield-specific risk factors are approximately the first principal components of yields (and are thus linear combinations of yields), the measurement equation can be inverted. In the inverted measurement equation, macroeconomic risk factors are solely a linear function of yields. This is what the literature typically refers to as “spanning”: macroeconomic risk factors can be deduced a linear combinations of yields from a sufficiently rich portfolio. The financial argument is concerned with the efficiency of markets. If macroeconomic factors inform the distribution of future yields, agents would trade on this information until it was reflected in today’s prices. This suggests there is no extraneous information regarding future yields contained in

macroeconomic fundamentals after controlling for the entire contemporaneous yield curve. This is a key result for simplifying estimation in most macroeconomic term structure models. For examples of macroeconomic literature describing the role of spanning results in estimation, see Ang and Piazzesi (2000), Ang et al. (2006), as well as Rudebusch and Wu (2008) and Bikbov and Chernov (2010). A short literature reviewed by Duffee (2013) attempts to demonstrate directly the invertibility of the yield curve measurement equation.

On the other hand, much of the financial econometric and predictive modeling/forecasting literature finds evidence of unspanned macroeconomic risk. Here unspanned macroeconomic risk means that macroeconomic factors improve forecasts of yields (or returns) even once the contemporaneous yield curve is controlled for. There is a large body of literature finding direct evidence against the spanning relationship. First, Cooper and Priestley (2009) and Ludvigson and Ng (2009) demonstrate that macroeconomic factors provide new information on contemporaneous risk premiums. Stock and Watson (2003) shows that the future value of macroeconomic factors capable of explaining cross-sectional variation of the yield curve is not exclusively determined by bond yields; this is a corollary to unspanned risk denoted “unspanned macroeconomic forecasts.” Joslin et al. (2014) provides rigorous tests of macroeconomic term structure model (MTSM) restrictions implied by the spanning relationship. This work also provides a more conceptual statistical argument against spanning, namely that the number of factors needed to span the yield curve is far fewer than the number of macroeconomic factors believed to be correlated with the yield curve<sup>4</sup>. There is also a large body of work that ties macroeconomic fundamentals to the yield curve for the purposes of forecasting, but does not directly tackle the spanning question. Nevertheless, the role of macroeconomic information in forecasting is suggestive of a failure of the spanning relationship. See, for example, Evans and Marshall (1998), Dewachter and Lyrio (2006), and most recently Diebold et al. (2006).

---

<sup>4</sup>This conceptual argument invites a formal test via canonical correlations of the yield curve and a large macroeconomic panel. A relevant modern framework for doing so is Andreou et al. (2016). Such a test is beyond the scope of the framework here, but as of this paper no such test has been conducted.

Bauer and Rudebusch (2017) is the most recent significant attempt to reconcile this literature. Bauer and Rudebusch (2017) argues that failing to account for measurement error can induce spurious findings of unspanned risk<sup>5</sup>. The measurement error argument states that even if the spanning hypothesis holds vis a vis true yields, observed yields contain measurement errors that can induce spuriously observed unspanned risk. That work was motivated by the disagreement between the theory and evidence of term structure modeling, employing the term “spanning puzzle”. On a similar note, Bauer and Hamilton (Forthcoming) show that statistical tests of the spanning hypothesis suffer from biased standard errors owing to highly serially correlated regressors, thus leading to over-rejection. Correcting for this bias calls into question the findings of numerous studies that had previously rejected spanning.

This paper is motivated by the findings of Bauer and Rudebusch (2017) and the ensuing discussion of the subject undertaken by Rebonato (2016). In particular, Rebonato (2016) suggests that spanning of one form or another must hold, as bond markets are highly efficient.

Thus one of two possibilities must be true:

1. Spanning tests will find unspanned risk no matter how many *linear* combinations of yields are used, in which case affine models are inadequate<sup>6</sup>.
2. If spanning is to hold in affine models, sufficiently many yields must be used as regressors. This is in contrast to the existing literature which typically employs 3 or 5 principal components in forecasting.

However, the two explanations are actually closely linked, because small misspecification in the model or data can induce large spurious rejection of spanning; rotating the data to principal components (PCs) and only using the first 3 or five PCs magnifies the effects of small misspecification. This paper will explore this link by means of a single key nonlinearity of interest in modern times: namely the zero lower bound (ZLB) wedge. If the true data

---

<sup>5</sup>Several arguments are presented, but the measurement error result is the relevant finding for this framework.

<sup>6</sup>By the same token, Gurkaynak and Wright (2012) posits the question of whether affine models can accommodate existing discrepancies in the spanning literature as an important research question.

generating process is nonlinear, then the classical approaches to testing the spanning hypothesis (which are all linear tests and procedures) can spuriously reject, even if the degree of nonlinearity is small. This is relevant in modern times, where the ZLB makes yields a nonlinear function of risk factors. In fact, the ZLB induces a nonlinear wedge between risk factors and actual yields in the presence of a *positive probability* of negative future short rates. The probability of a future ZLB regime has been positive since as early as 1993, and as a result nonlinearities in the yield curve have been present at least as long. This was first observed by Bomfim (2003), who notes the small probability of a ZLB regime approaching in 2000–2003. This argument is consistent with that of Cochrane and Piazzesi (2005), who found smaller degrees of unspanned risk given a sample that ended in 2003 as opposed to ending the sample in 2007. Notably, these nonlinearities first appear in the long end of the yield curve (since longer maturities will be where positive probabilities of the ZLB binding first appear). This is in contrast to much of the literature that emphasizes the role of nonlinearities at short maturities, and in particular emphasizes that the ZLB regime is not relevant until post-2007. Since the spanning literature frequently emphasizes time spans that run up until 2007, the ZLB can still be a relevant omitted nonlinearity. In particular, 10-year and 20-year yields are nontrivially influenced by the ZLB as early as 2002 and 1995, respectively.

This is an important first nonlinearity to consider for two reasons: (1) the new nonlinear terms in the model are spanned if and only if yields are spanned, and (2) updated appropriate updated spanning tests can be easily derived. A thorough examination would require exploring the potentially endless world of nonlinear models, where determining what *spanning* even means is no longer obvious – See Collin-Dufresne and Goldstein (2002), Shin and Zhong (2017), and Christensen et al. (2014), who present differing methods and results on whether dynamic yield volatility is spanned. Another leading alternative is the quadratic term structure model of Ahn et al. (2002), advanced by Li and Zhao (2006). Even simpler omitted nonlinearities are relevant: Cochrane (2015) observes the potential that unspanned risks attributed to inflation are potentially “proxying (and poorly) for detrending [and sea-

Table 15: Breakdown of Spanning Literature

	Linear DGP	Nonlinear DGP
Linear Model	Already Treated	<b>*Open*</b>
Nonlinear Model	Robust	Model-Dependent

sonal effects] of the level and slope factors”. This paper does not undertake a comprehensive exploration of remaining nonlinearities and their role in explaining remaining unspanned macroeconomic risk, leaving this task for a future work.

One can consider this exploration via Table 15. The single omitted nonlinearity considered is the ZLB wedge. There are a number of techniques available for modifying an affine model to respect the ZLB, but the leading method is the shadow rate model of Black (1995). This paradigm posits the existence of an unconstrained instantaneous risk-free rate (or *shadow rate*), and real-world instantaneous risk-free rate is the supremum of the shadow rate and zero. The difference between shadow yields and actual yields is deduced as the option price of holding currency. The most popular exposition of this model and the necessary estimation techniques for it is Christensen and Rudebusch (2013)<sup>7</sup>. Approximations based on Krippner (2012) and Krippner (2013) have led shadow-rate models to the forefront of ZLB yield curve modeling.

By contrast, exploring the second possibility from Rebonato (2016) is relatively simple. By employing higher frequency data, increasing the number of available yield observations, one can extend standard 3-factor spanning tests to 5-factor spanning tests without loss of power. Moreover, these tests can actually be expanded to fully general spanning tests using all observed yields. To proceed on this front, this paper will compare the effect of the omitted nonlinearity mechanism in both a monthly frequency and a daily frequency model, wherein the daily frequency model will be able to have an order of magnitude more observations. Though the full infill asymptotic theory of how this model should behave relative to a monthly frequency model will not be explored here, this methodology represents a first pass

---

<sup>7</sup>Shadow-rate models were in partial use immediately after Black (1995), but difficulties in analytical solutions and estimation largely precluded their use.

at estimating the relative effects of omitted nonlinearities against the effects of having failed to use a sufficient number of yields.

The paper proceeds as follows. Sections 2.2 and 2.3 first discuss the relevant yield curve modeling techniques, along with classical approaches for testing the spanning hypothesis. The models considered here rely on no-arbitrage restrictions. The other leading specification is the popular forecasting Dynamic Nelson-Siegel (DNS) model. In fact, the two can often be combined, and the resulting Arbitrage-Free Nelson-Siegel (AFNS, Christensen et al. (2011)) has been successful in the empirical MTSM literature. However, once the DNS model is augmented with unspanned observable macroeconomic risk factors, there is little exploration in the literature on whether the arbitrage-free restrictions are helpful for model fit/forecasting, or if they are even correctly specified restrictions. See Figure 15 to visualize the nesting of these models and why the no-arbitrage restriction strategy is preferable in this framework.

Section 2.4 discusses the statistical properties of our main hypothesis: that when the true data generating process is nonlinear and satisfies spanning, linear spanning tests can spuriously reject the spanning hypothesis. Section 2.5 demonstrates the main hypotheses in simulation, and explores the ability of corrections to linear spanning tests to accurately recover macroeconomic spanning. It is shown that the degrees of nonlinearity caused by the ZLB pre-2007 are sufficient to cause substantial over-rejection in simulation, but that the numerical importance of failing to use sufficiently many yields is generally greater. The omitted nonlinearity effect is compounded by the use of PCs instead of all yields, and greatly compounded by the use of only three PCs.

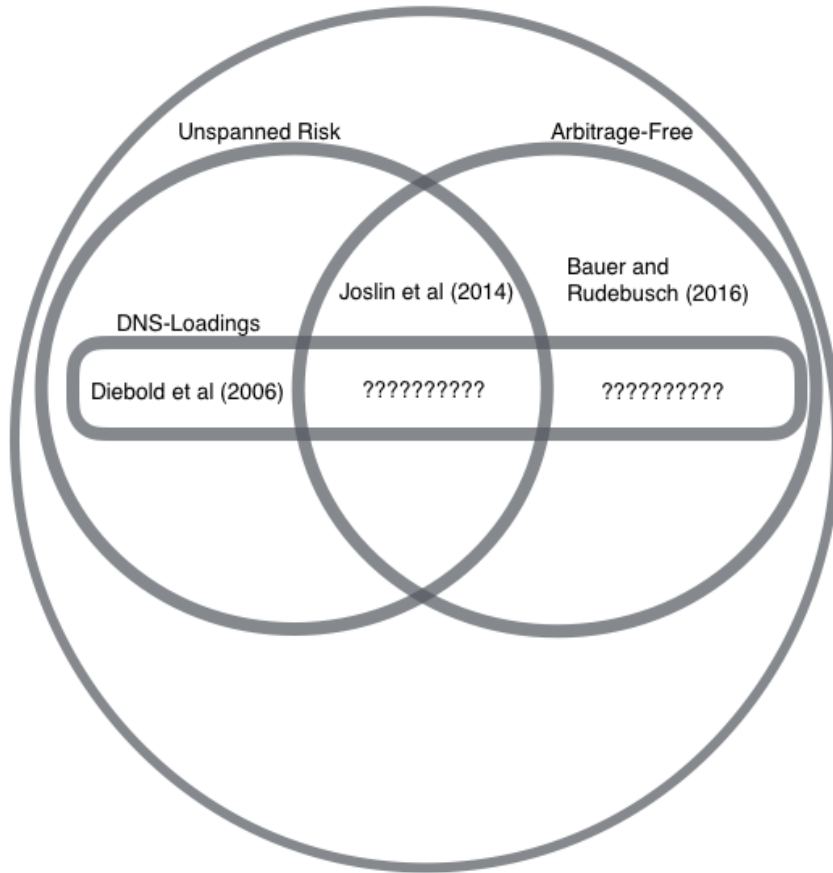
Section 2.6 applies the models to the data. This section finds that standard linear spanning tests indeed reject the spanning hypothesis. After adjusting for nonlinearities induced by the ZLB in the pre-2007 period, we find results broadly consistent with the simulation results. The omitted ZLB wedge mechanism indeed explains a nontrivial portion of the observed unspanned risk, though the effect in practice is not nearly as strong at the long end of the yield curve as one might expect based on the shape of the ZLB wedge. The linear spanning

tests understandably find enormous unspanned risk in the post-ZLB period. Once corrections are made to account for the ZLB, however, one finds that there is less unspanned risk at all maturities post-ZLB as compared to pre-ZLB. The effects of omitting data (i.e., using PCs instead of a full set of yields) was found to be comparable in magnitude to the effect of omitting the ZLB nonlinearity. However, there is additional evidence that the use of PCs instead of a full set of yields as predictors compounds the effect of omitting the ZLB nonlinearity. Section 2.7 summarizes the conclusions of this work and offers recommendations for future work.



Figure 15: Affine macroeconomic Term Structure Modeling

This figure represents the nesting of macroeconomic term structure models together with the major literature upon which this paper is based. The blank central piece of the diagram represents the open question in the literature of whether a term structure model augmented with macroeconomic factors can be satisfy no-arbitrage while still matching the data well. This question is separate from the debate on spanning, and suggests that for the purposes of this work arbitrage-free models might be preferable to DNS models.



## 2.2 Yield Curve Modeling

### 2.2.1 Yield Curve Notation

Here introductory notation is established. All bond yield models must begin with a risk-free rate  $r_t$  and risk factors  $X_t$ , which may consist of yield-specific risk factors  $Z_t$  and/or macroeconomic risk factors  $M_t$ . Affine models assume:

$$r_t = \delta_0 + \delta_1 X_t = \delta_0 + \delta_1^Z Z_t + \delta_1^M M_t \quad (2.2.1)$$

In the Gaussian class of yield curve models, real-world transition dynamics on  $X_t = [Z_t, M_t]$  are given by a Gaussian vector autoregression (VAR):

$$X_t = \mu + \phi X_{t-1} + \Sigma_t$$

The distinction between real-world and risk-neutral dynamics will be elaborated upon further in Section 2.2.2. One can then introduce bond yields  $y_t(\tau)$ , prices  $P_t(\tau)$ , and forward rates  $f_t(\tau)$ , each a function of maturity  $\tau$ , via the following classical system of equations, described in McCulloch (1971), that governs the relationship between the three:

$$\begin{aligned} P_t(\tau) &= e^{-\tau y_t(\tau)} \\ f_t(\tau) &= -\frac{P_t'(\tau)}{P_t(\tau)} \\ y_t(\tau) &= \frac{1}{\tau} \int_0^\tau f_t(u) du \end{aligned}$$

These are the discount curve, forward rate curve, and yield curve respectively. A thorough discussion of how to construct an actual data set of bond yields can be found in Diebold

et al. (2006), who draws on the work of McCulloch (1975) and Vasicek and Fong (1982). Given a set of yields, the broad problem is how to estimate the curve  $y_t(\tau)$ , an infinite dimensional time-varying object. Under the affine assumption for interest rates, Duffie and Kan (1996) shows that bond yields are affine in risk factors:

$$y_t(\tau) = A(\tau) + B(\tau)X_t = A(\tau) + B^Z(\tau)Z_t + B^M(\tau)M_t \quad (2.2.2)$$

It follows that forward rates are also affine in risk factors, with notation<sup>8</sup>:

$$f_t(\tau) = A'(\tau) + B'^Z(\tau)Z_t + B'^M(\tau)M_t$$

The most widely used class of term structure models assumes Gaussian VAR transition dynamics for risk factors:

$$\begin{pmatrix} Z_t \\ M_t \end{pmatrix} = \begin{pmatrix} \phi_{ZZ} & \phi_{ZM} \\ \phi_{MZ} & \phi_{MM} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ M_{t-1} \end{pmatrix} + \Sigma_t, \quad \Sigma_t \sim N(0, \Omega) \quad (2.2.3)$$

With transition dynamics on risk factors specified, the task of modeling  $y_t(\tau)$  now falls to the determination of appropriate functions  $A(\tau)$ ,  $B(\tau)$ .

### 2.2.2 No-Arbitrage Modeling

A canonical arbitrage-free affine macroeconomic term structure model proceeds as follows.

As before, the short-rate is affine in the risk factors:

$$r_t = \rho_0 + \rho_1^Z Z_t + \rho_1^M M_t$$

---

<sup>8</sup>Once the ZLB is introduced, it will occasionally be useful to work with forward rates in addition to or instead of yields. In particular, see Section 2.4.2, where employing nonlinear predictive regression tests is made easier through the use of forward rates. Noting that forward rates are also affine in risk factors implies that working with forward rates adds no computational burden.

Importantly, to appropriately specify no-arbitrage restrictions, it is necessary to separately specify the real-world ( $\mathbb{P}$ ) and the risk-neutral ( $\mathbb{Q}$ ) dynamics separately. Thus:

$$\begin{pmatrix} Z_t \\ M_t \end{pmatrix} = \mu^{\mathbb{P}} + \phi^{\mathbb{P}} \begin{pmatrix} Z_{t-1} \\ M_{t-1} \end{pmatrix} + \Sigma_t^{\mathbb{P}}, \quad \Sigma_t^{\mathbb{P}} \sim N(0, \Omega^{\mathbb{P}}) \quad (2.2.4)$$

And separately:

$$\begin{pmatrix} Z_t \\ M_t \end{pmatrix} = \mu^{\mathbb{Q}} + \phi^{\mathbb{Q}} \begin{pmatrix} Z_{t-1} \\ M_{t-1} \end{pmatrix} + \Sigma_t^{\mathbb{Q}}, \quad \Sigma_t^{\mathbb{Q}} \sim N(0, \Omega^{\mathbb{Q}}) \quad (2.2.5)$$

Following Joslin et al. (2013), who builds upon the work of Duffie and Kan (1996), this set of assumptions is sufficient to guarantee that model-implied yields are arbitrage free in addition to being affine in risk factors.

Implicitly,  $Y_t$ ,  $A$ , and  $B^Z, B^M$  are all functions of the maturity  $\tau$ . So far, nothing has changed from the standard model except for having restricted factor loadings. However, with specifications on both the risk-neutral and physical dynamics of the risk factors, Duffie and Kan (1996) show that the following difference equations yield  $A(\tau)$  and  $B(\tau)$  that satisfy no-arbitrage conditions:

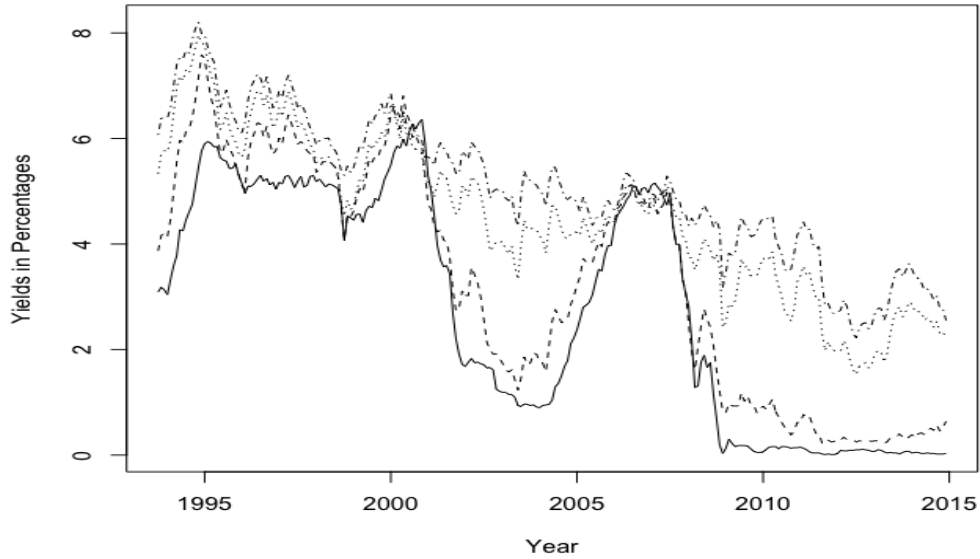
$$(\tau + 1)A(\tau + 1) = \tau A(\tau) + \mu^{\mathbb{Q}} \tau B(\tau) - \frac{1}{2}(\tau B(\tau))\Omega(\tau B(\tau)) + \rho_0 \quad (2.2.6a)$$

$$(\tau + 1)B(\tau + 1) = \phi^{\mathbb{Q}} \tau B(\tau) + \rho_1 \quad (2.2.6b)$$

These restrictions facilitate estimation, as now the full set of factor loadings (potentially of large dimension) are pinned down by the few remaining parameters in Equations 2.2.6a and 2.2.6b.

Figure 16: US Government Bond Yields

Plotted are market yields on US Treasury securities at constant maturities of 3m, 1yr, 10yr, and 20yr from December 1993–June 2015 at a monthly frequency.



### 2.2.3 Zero Lower Bound and the Shadow Rate

Consider the modern US yield curve and compare its general history with the recent ZLB period in Figure 16. The model described above, in normal times, posits that bond yields are affine in risk factors, and that risk factor transition dynamics are linear and conditionally Gaussian. Following 2007, when bond yields are close to the ZLB and facing that constraint, a wholly affine and conditionally Gaussian state space system performs poorly. While there are a number of paradigms for remedying this, shadow rate models have proven far and away the most widely accepted. The standard thought exercise to motivate this is to consider the price of a bond with maturity  $\tau$  in two alternative worlds: the real world where currency exists, and a “shadow” world where currency does not exist. As a result, the instantaneous risk-free rate  $r_t$  should always satisfy:

$$r_t = \max\{0, s_t\}, \quad (2.2.7)$$

where  $s_t$  is the instantaneous risk-free rate in the *shadow* world, henceforth the shadow rate. Bond prices in the two worlds are similarly related like so:

$$\begin{aligned} P_t(\tau) &= \min[1, \tilde{P}_t(\tau)] \\ &= \tilde{P}_t(\tau) - \max[\tilde{P}_t(\tau) - 1, 0] = \tilde{P}_t(\tau) - C_t^A(\tau, \tau, 1) \end{aligned}$$

The real world bond price must equal the shadow bond price, unless the price of the shadow bond is above 1. The term  $\max[\tilde{P}_t(\tau) - 1, 0]$  is observed to be an American call option on the shadow bond with the same maturity and a strike price of 1, hence the notation  $C_t^A(\tau, \tau, 1)$ .

Krippner (2012) and Krippner (2013) observe that the American call option is extraordinarily difficult to analyze analytically because of its early exercise option. It is therefore difficult to estimate in practice. Instead it can be approximated by a European option. For arbitrary maturities, Krippner (2012) introduces this approximating bond and prices it thusly:

$$P_t^a(\tau + \delta) = \tilde{P}_t(\tau + \delta) - C_t^E(\tau, \tau + \delta, 1)$$

As discussed in Christensen and Rudebusch (2016), Krippner (2012) deduces that, while the approximating bond price might not be identical to the price of the real world bond of interest, the corresponding forward rate converges as  $\delta \rightarrow 0$ . In short:

$$f_t(\tau) = \lim_{\delta \rightarrow 0} \left[ -\frac{\partial}{\partial \delta} P_t^a(\tau + \delta) \right]$$

And as a result:

$$f_t(\tau) = \tilde{f}_t(\tau) + \lim_{\delta \rightarrow 0} \left[ -\frac{\partial C_t^E(\tau, \tau + \delta, 1)}{\partial \delta \tilde{P}_t(\tau + \delta)} \right] \quad (2.2.8)$$

In the Gaussian class of models this can be written more explicitly as:

$$f_t(\tau) = \tilde{f}_t(\tau) \Phi \left( \frac{\tilde{f}_t(\tau)}{\Psi(\tau)} \right) + \Psi(\tau) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\tilde{f}_t(\tau)}{\Psi(\tau)} \right]^2 \right) \quad (2.2.9)$$

The resulting approximating yield curve in the shadow rate model is expressed:

$$y_t(\tau) = \frac{1}{\tau} \int_t^{t+\tau} \left[ \tilde{f}_t(s) \Phi \left( \frac{\tilde{f}_t(s)}{\Psi(s)} \right) + \Psi(s) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\tilde{f}_t(s)}{\Psi(s)} \right]^2 \right) \right] ds \quad (2.2.10)$$

The variable  $\Psi(s)$  gives the conditional variance of shadow bond option prices. In the Gaussian class, it can be shown (Christensen and Rudebusch (2013), or Bauer and Rudebusch (2016)) that its value is determined according to:

$$\begin{aligned} \Psi(1) &= \rho_1' \Omega \rho_1 \\ \Psi(\tau) &= \rho_1' \left[ \sum_{i=0}^{\tau-1} \phi^i \Omega (\phi^i)' \right] \rho_1, \tau > 1 \end{aligned}$$

One can explore the nature of shadow forward rates and shadow yields within this framework.

Notice that if:

$$\tilde{y}_t(\tau) = \frac{1}{\tau} \int_t^{t+\tau} \tilde{f}_t(s) ds$$

Then:

$$y_t(\tau) = \tilde{y}_t(\tau) + \frac{1}{\tau} \int_t^{t+\tau} \left[ \tilde{f}_t(s) \left( \Phi \left( \frac{\tilde{f}_t(s)}{\Psi(s)} \right) - 1 \right) + \Psi(s) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\tilde{f}_t(s)}{\Psi(s)} \right]^2 \right) \right] ds \quad (2.2.11)$$

Define:

$$ZLB_t(\tau) = \frac{1}{\tau} \int_t^{t+\tau} \left[ \tilde{f}_t(s) \left( \Phi \left( \frac{\tilde{f}_t(s)}{\Psi(s)} \right) - 1 \right) + \Psi(s) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\tilde{f}_t(s)}{\Psi(s)} \right]^2 \right) \right] ds.$$

This term is commonly denoted as the ZLB wedge. Conceptually this term represents an approximation to the option value from holding currency. Since investors will always hold cash instead of buying a bond with negative yield, this option value of currency drives the difference between shadow yields and model-fitted yields. Thus:

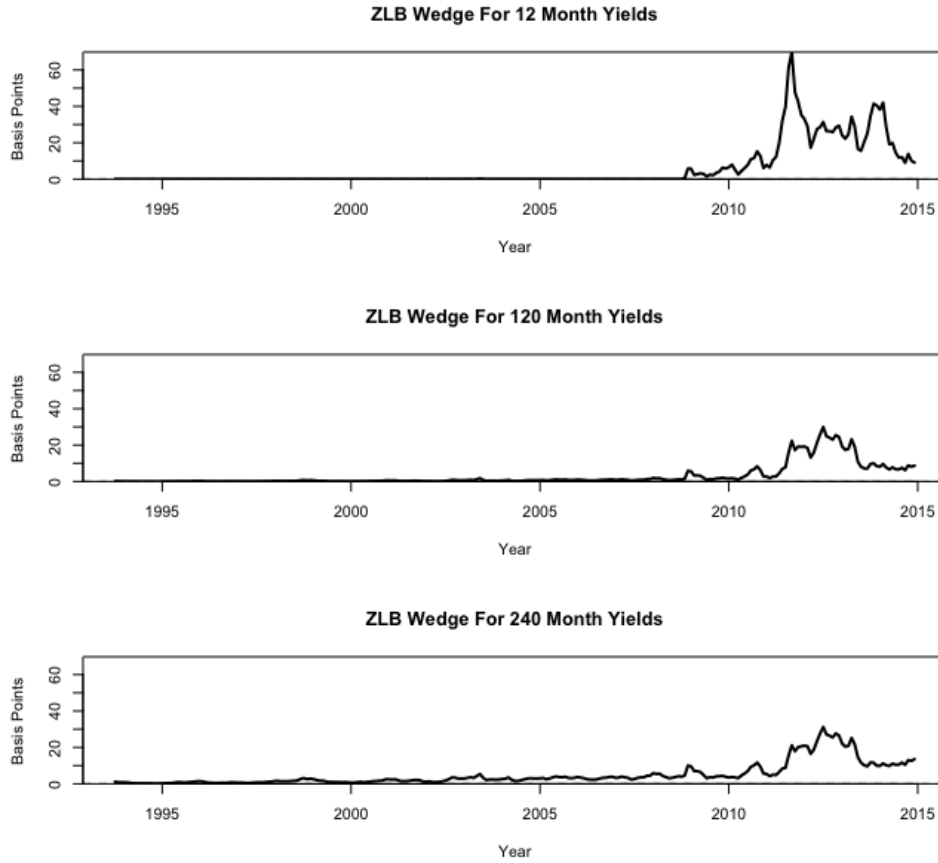
$$y_t(\tau) = \tilde{y}_t(\tau) + ZLB_t(\tau) \quad (2.2.12)$$

The cross-sectional and time-varying nature of the function  $ZLB_t(\tau)$  is of central importance to this study. The ZLB literature has classically emphasized the nature of this function in the period post-2007. However, as discussed in the introduction, investors placed positive probability on future short rates being negative as early as 2000–2002, and we can see this in the ZLB wedge in Figure 17. Although these nonlinearities are small, their statistical significance should not be underestimated. We can also see that between 2000 and 2016, the impact of the ZLB on yields fluctuated dramatically (see Figure 18). Because the yield curve was downward trending, but not yet at the ZLB, the shape of the yield curve in the early 2000s represents the expectation that the ZLB constraint would begin to bind at some



Figure 17: The ZLB Wedge Across Time

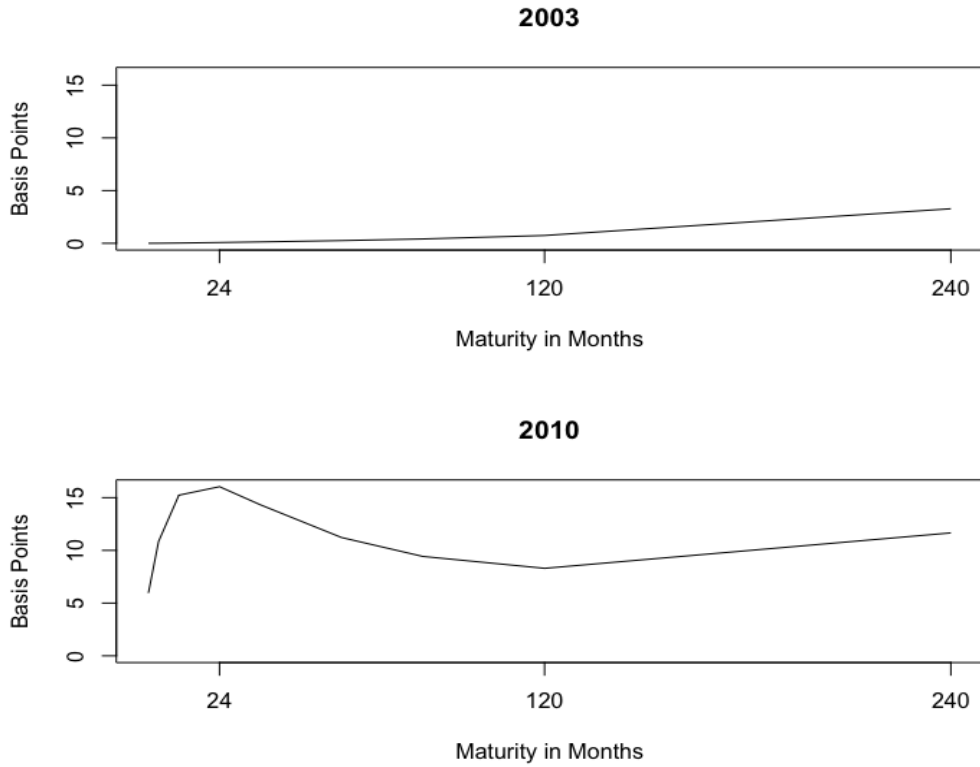
In each plot is the ZLB wedge from the model of Bauer and Rudebusch (2016). This is a Gaussian MTSM with inflation and output gap as macroeconomic factors evaluated at a monthly frequency. Each plot gives the wedge for a particular maturity from December 1993 to June 2015.



point in the future, albeit not the immediate future. Thus the long end of the yield curve is affected more sharply than the short end. This terminology now facilitates a discussion of the nature of the spanning hypothesis at the ZLB.

Figure 18: The ZLB Wedge Across Maturities

In each plot is the ZLB wedge from the model of Bauer and Rudebusch (2016). In the upper plot is the ZLB wedge across maturities from a representative month in the year 2003. In the lower plot is the ZLB wedge across maturities from a representative month in 2010.



## 2.3 The Spanning Hypothesis

Under the assumptions of Section 2.2.1, bond yields are affine in risk factors  $X_t$ :

$$y_t = A + BX_t = A + B^Z Z'_t + B^M M_t \quad (2.3.1)$$

This equation gives rise to the spanning puzzle. Suppose  $B^M$  is a nonsingular matrix, in which case one can invert this equation to arrive at  $M_t$  as a linear function of yield curve information. The standard argument in the literature is as follows: consider the case that  $Z_t$  are not just dynamic latent factors, but are instead the first three principal components of yields<sup>9</sup>. Then if Equation 2.3.1 holds with nonsingular  $B^M$ , for some portfolio of bond yields  $\mathbb{P}_t$ :

$$M_t = \alpha + \beta \mathbb{P}_t$$

Therefore, in a regression of  $M_t$  on the full set  $y_t$  should have an  $R^2$  of one. The non-statistical formulation of this argument has an efficient markets basis: if macroeconomic information available today were helpful for forecasting the yield curve, bond traders would be aware of this information and trade on it, meaning that this macroeconomic information would be present in today's yield curve.

To test this in an arbitrage-free affine model, there are two necessary and sufficient parameter restrictions for this class of models to have unspanned macroeconomic risk:

$$\rho_1^M = 0 \quad (2.3.2)$$

---

<sup>9</sup>Much of the yield curve forecasting literature, recently Diebold and Li (2006) and Joslin et al. (2014), agree that the dynamic factors are extremely highly correlated with the first three principal components.

$$\phi_{PM}^{\mathbb{Q}} = 0 \tag{2.3.3}$$

Unfortunately, this calculation shows that the increased parsimony in the model does equate with increased power to test the restrictions implied by the existence of unspanned risk. Bauer and Rudebusch (2017) notes that under the null hypothesis the risk-neutral dynamics of the macroeconomic risk factors are not identified. Following Hansen (1996), unidentified parameters under the null hypothesis prevent the likelihood ratio test statistic from achieving its standard  $\chi^2$  limiting distribution. The solution Bauer and Rudebusch (2017) present is to instead conduct a likelihood ratio test on a fully flexible, unrestricted dynamic factor model wherein the null hypothesis supposes only that the factor loadings on macroeconomic risk factors are zero. Since this fully flexible model nests the unrestricted (i.e., allowing for spanning) arbitrage-free model, this test will be conservative for the actual desired test statistic. The likelihood ratio test statistic will not be conducted here as a short argument shows it to be redundant in the Gaussian class of models: if an observed risk factor is correlated with yields, it must be correlated with shadow yields. This is because yields are the sum of shadow yields and the ZLB wedge, and it is impossible in the Gaussian class of models to be correlated with one and not the other. If one rejects the unspanned risk restrictions on an affine model, one will invariably have to reject them on a shadow rate model too. Therefore, the results of considering the likelihood ratio tests here are redundant to those considered in Bauer and Rudebusch (2017), in which restrictions implied by the existence of unspanned risk are strongly rejected.

The spanning puzzle literature often cites two inauspicious corollaries. The first is that after controlling for an adequate portfolio of bond yields, macroeconomic factors are uninformative about their own future values. This corollary is the *unspanned macroeconomic forecasts* puzzle, following Bauer and Rudebusch (2017). That the cross section of relevant macroeconomic factors can be fully explained by contemporaneous bond yields is denoted the *unspanned macroeconomic variation* element of the puzzle, again following Bauer and

Rudebusch (2017).

Though these corollaries are interesting phenomenon in their own right, the question of most practical importance is the existence of *unspanned macroeconomic risk*. This question is asking whether all information necessary for forecasting yields is contained in today's yield curve. Formally, this is a question of the predictive regression:

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 M_t + e_t, \quad (2.3.4)$$

and whether or not  $\beta_2 = 0$ . The *unspanned macroeconomic variation* and *unspanned macroeconomic forecasts* corollaries are examined in practice via two additional regressions:

$$M_t = \gamma_0 + \gamma_1 y_t + \varepsilon_t \quad (2.3.5a)$$

$$M_{t+1} = \lambda_0 + \lambda_1 y_t + \lambda_2 M_t + \epsilon_t \quad (2.3.5b)$$

These regression tests appear in Joslin et al. (2014), Duffee (2011), and Bauer and Rudebusch (2017). The (informal) test statistics are that the  $R^2$  of the first regression should be near one, and that  $\lambda_2$  in the second regression should be zero. However, there is some disagreement over whether the corollary regressions are necessary. In particular, there is an intuitive discussion of the following in Bauer and Rudebusch (2017): One need not expect regression 2.3.5a to have an  $R^2$  of 1, rather the residual from this regression should be expected to have no predictive power. The predictive regression 2.3.4 tests whether the residual from 2.3.5a has predictive power for yields, and 2.3.5b tests whether the residual from 2.3.5a has predictive power for macroeconomic variables. The previous discussion is formalized in the following theorem:

**Theorem 1.** *The following statements are true:*

1. *Unspanned macroeconomic risk implies unspanned macroeconomic variation.*
2. *Unspanned macroeconomic variation does not imply unspanned macroeconomic risk.*
3. *Outside of knife-edge cases, unspanned macroeconomic risk implies unspanned macroeconomic forecasts.*
4. *Unspanned macroeconomic forecasts do not imply unspanned macroeconomic risk.*

*Proof.* In the predictive regression

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 M_t + e_t,$$

finding  $\beta_2 \neq 0$  is equivalent to the following: Consider the residual from a regression of  $M_t$  on  $y_t$ ,

$$M_t = \hat{\gamma} y_t + \widehat{OM}_t. \tag{2.3.6}$$

Then unspanned risk implies that the orthogonal component  $\widehat{OM}_t$  has predictive power for  $y_{t+h}$ . That is, the following regression could equivalently be considered:

$$y_{t+h} = \tilde{\beta}_0 + \tilde{\beta}_1 y_t + \tilde{\beta}_2 \widehat{OM}_t + e_t \tag{2.3.7}$$

Thus, unspanned macroeconomic risk implies the existence of an orthogonal component in the projection of  $M_t$  on yields, and that this orthogonal component has predictive power. Thus  $\tilde{\beta}_2 \neq 0$ . That this orthogonal component must exist implies there is unspanned macroeconomic variation. This proves statement 1. However, the existence of this orthogonal component does not necessarily imply  $\tilde{\beta}_2 \neq 0$ , thus proving statement 2.

Considering regression 2.3.6 again. In the regression:

$$M_{t+h} = \tilde{\lambda}_0 + \tilde{\lambda}_1 y_t + \tilde{\lambda}_2 \widehat{OM}_t + \epsilon_t, \quad (2.3.8)$$

unspanned macroeconomic forecasts are defined as  $\tilde{\lambda}_2 \neq 0$ . Combining 2.3.6 and 2.3.4 yields:

$$M_{t+h} = \hat{\gamma}\tilde{\beta}_0 + \hat{\gamma}\tilde{\beta}y_t + \hat{\gamma}\tilde{\beta}_2\widehat{OM}_t + \widehat{OM}_{t+h}$$

Thus, unless the coefficient in a regression of  $\widehat{OM}_{t+h}$  on  $\widehat{OM}_t$  is exactly  $-\hat{\gamma}\tilde{\beta}$ , it will be the case that  $\tilde{\beta} \neq 0 \implies \tilde{\lambda}_2 \neq 0$ . This proves statement 3.

Finally, the left hand side of 2.3.8 is a sum of  $\hat{\gamma}y_{t+h}$  and  $\widehat{OM}_{t+h}$ . As can clearly be seen, unspanned macroeconomic forecasts by themselves do not indicate whether  $\widehat{OM}_t$  has predictive power for  $y_{t+h}$ , in which case there is also unspanned macroeconomic risk, or  $\widehat{OM}_{t+h}$  only, in which case there is not. This proves statement 4.

□

It follows that although the literature treats tests for unspanned macroeconomic variation and unspanned macroeconomic forecasts as robustness checks for unspanned macroeconomic risk, they are in fact three separate phenomena. Moreover, while finding the absence of unspanned variation or unspanned macroeconomic forecasts would have implications for unspanned macroeconomic risk, it is very unlikely that bond yields have a fixed perfectly linear deterministic relationship with any collection of contemporaneous macroeconomic factors. Therefore, one should always expect to find unspanned macroeconomic variation. Finally, one should observe that the existence of unspanned macroeconomic variation is all that is necessary to induce error in the measurement equation for yields<sup>10</sup>. Therefore, again

---

<sup>10</sup>This follows from a standard attenuation bias argument. If:

$$\begin{aligned} M_t &= \tilde{M}_t + OM_t \\ y_t &= \beta\tilde{M}_t, \end{aligned}$$

the existence of error in all term structure measurement equations is not necessarily related to whether unspanned *risk* exists, only unspanned *macroeconomic variation*.

As a result, the focus should be solely on regression 2.3.4 to test the spanning hypothesis. To summarize, the primary testing of unspanned risk is  $\beta_2 = 0$  in the predictive regressions 2.3.4 and 2.3.9. This restriction holds under the null of **no unspanned risk**. One may proceed by an  $F$ -test of this restriction, or by examining differences in adjusted  $R$  squared's between restricted and unrestricted regressions.

### 2.3.1 PCA-Based Spanning Tests

It is frequently observed that small amounts of model misspecification can induce substantial spuriously observed unspanned risk. This is an unintuitive result. Examining regression 2.3.4, intuition suggests that as estimation error or model misspecification of yields smoothly vanishes, spurious rejections should similarly smoothly vanish. However, in practice, we do not see this occurring.

However, the exact specification provided in Equation 2.3.4 is also rarely examined in practice. These predictive regression are typically constrained further. In particular, since macroeconomic term structure models are usually estimated on a monthly basis, there are rarely enough observations to estimate the large matrix  $\beta_1$  (which will typically have 50 – 100 parameters). Furthermore, most models suggest regressing against the first several principal components, denoted  $\mathcal{P}_t$ , is sufficient, and so regressing against the full yet of  $y_t$ 's on the right hand side is unnecessary. Common term structure models posit that only three factors are sufficient to explain the majority of the cross section of the yield curve. In the Nelson-Siegel literature, for example, factors have explicit shape interpretations: level, slope, and curvature factors<sup>11</sup>. This gives rise to the test:

$$y_{t+h} = \beta_0 + \beta_1 \mathcal{P}_t + \beta_2 M_t + e_t \tag{2.3.9}$$

---

but  $M_t$  is used in a term structure model, the measurement equation will have some error.

<sup>11</sup>See Diebold et al. (2005) Diebold and Li (2006), Christensen et al. (2011).



The two corollary tests, when constrained to PCs, yield

$$M_t = \gamma_0 + \gamma_1 \mathcal{P}_t + \varepsilon_t \tag{2.3.10a}$$

$$M_{t+1} = \lambda_0 + \lambda_1 \mathcal{P}_t + \lambda_2 M_t + \epsilon_t \tag{2.3.10b}$$

However, there is a problem with this restriction, namely that these tests are highly sensitive to small model misspecification. This is primarily a result of the approximate factor structure of yields. The higher-order principal components (i.e., 4th and 5th) only explain a very small portion of cross-sectional variation, thus making them very difficult to estimate. However, including all relevant principal components is crucial for ensuring the accuracy of tests of the spanning hypothesis. See Bauer and Rudebusch (2017), who shows that while the fourth and fifth component are small, removing them from the regression leads to spurious rejection of spanning. Similarly, Duffee (2011) demonstrated the potential for factors to have statistically imperceptible effects on the cross section of yields but strong forecasting power. Therefore, a small misspecification can result in a large error in the measurement of the space spanned by the first five yield PCs, and the subsequent frequent inappropriate rejection rates of spanning. Consider the following regressions:

$$y_t = \tilde{y}_t + \varepsilon_{1,t} \tag{2.3.11a}$$

$$M_t = \mathcal{P}_t + \varepsilon_{2,t} \tag{2.3.11b}$$

$$M_t = \tilde{\mathcal{P}}_t + \varepsilon_{3,t} \tag{2.3.11c}$$

The hope is that the behavior of the bottom two regressions should be nearly identical when the  $R^2$  of the first regression is almost 1. That said, this will not be the case. To test this, we simulated a shadow rate model with two macroeconomic factors and three latent, yield-

specific factors. Shadow yields and yields are almost identical, so that regression 2.3.11a has an average  $R^2 = 0.95$ . However, the average  $R^2$  of regression 2.3.11c is 0.99, while the average  $R^2$  of regression 2.3.11b is 0.8.

This theoretical discrepancy between PC-based tests and tests that use the full set of yields drives a substantial portion of the conversation on spurious unspanned risk. Spanning in a predictive regression will hold if, and only if, sufficiently many yields are used on the right hand side, and every explanation for spurious unspanned risk has its effect amplified by using PCs instead of a full set of yields. This drives the effort in simulation and in the empirical study to compare the effects of omitted ZLB nonlinearities under both PC-based tests and fully general tests.

## 2.4 Nonlinear Models and Spanning

Statistically, the argument thus far has hinged on the presence of linearity. If the true relationship is nonlinear, as represented by:

$$y_t = g(A + B^Z Z_t' + B^M M_t)$$

for some nonlinear but invertible  $g(\cdot)$ , the argument no longer holds. In other words, the linear regression of  $M_t$  on  $y_t$  may have a very low  $R^2$  – suggesting unspanned macroeconomic variation - even though in fact  $M_t$  can be constructed entirely from information contained in  $y_t$ .

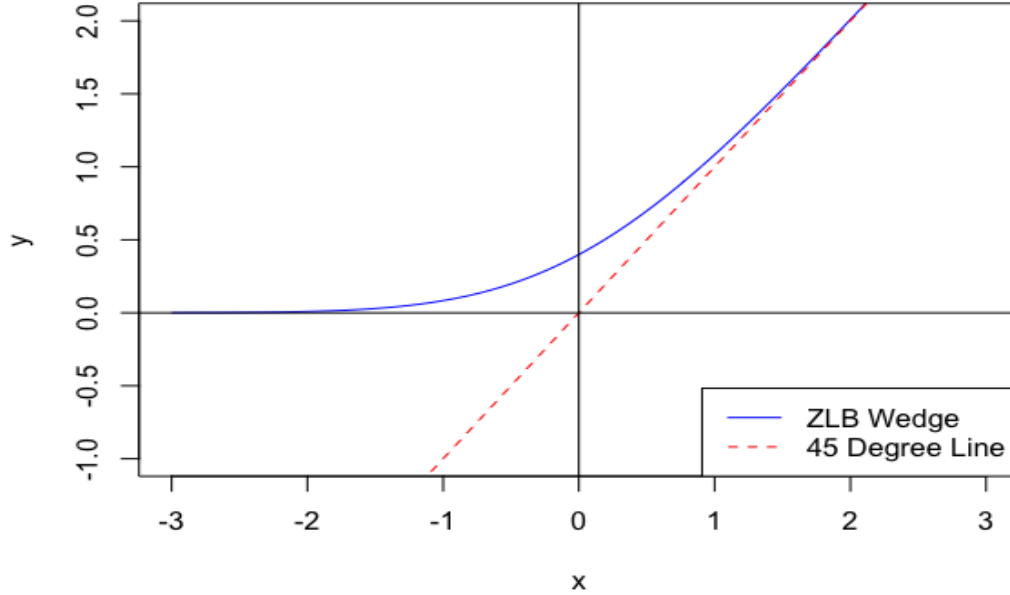
### 2.4.1 Linear Vs Nonlinear Tests: Shadow Rate

The central question of the spanning hypothesis remains: given appropriate conditioning on the current yield curve, do macroeconomic risk factors offer additional predictive ability? In the shadow rate model, observed yields themselves are no longer affine in risk factors, and are thus unlikely to reliably linearly span those factors. This includes macroeconomic factors. However, under the modeling assumptions of Section 2.2.3, there are nonlinear transformations of the yield curve that produce objects that linearly span risk factors. These are shadow forward rates and shadow yields,  $\tilde{f}_t(\tau)$  and  $\tilde{y}_t(\tau)$ .

Consider Equation 2.2.9. Following Wu and Xia (2016), one can observe this equation gives observed forward rates as a monotonic function of shadow forward rates. Therefore, conditional on parameter estimates that govern  $\Psi(\tau)$  the equation can always be inverted. Figure 19 demonstrates two facts: (1) that this transformation from shadow yields to actual yields induces a wedge for positive shadow short rates; and (2) this function is monotonic and invertible, save for extreme negative values of the shadow short rate. This holds even when the shadow short rate is below zero, as the inversion relies solely on parametric assumptions

Figure 19: ZLB Wedge Invertibility

This figure illustrates the function that maps a shadow yield to its observed yield.



made that drive the dynamics of shadow bond prices<sup>12</sup>.

The proposal here, therefore, is to use shadow yields as a predictor in the classical spanning tests<sup>13</sup>. Using yields will overestimate the degree of unspanned macroeconomic variation, leading to spurious estimates of unspanned risk. Formally:

**Theorem 2.** *Suppose  $y_t$  is generated from the spanned model of Section 2.2.3, and suppose*

<sup>12</sup>Shadow yields follow by integration. This integration step can introduce another source of estimation uncertainty: shadow forward rates are already affine in risk factors, meaning that it should not be necessary to integrate to shadow yields. However, in spite of this the focus will be on shadow yields for consistency with existing spanning literature.

<sup>13</sup>One could proceed further by augmenting the linear form of the spanning tests to allow for nonlinear regression terms, and a procedure for doing so is outlined in Section 2.4.2. The effect of this correction is small in practice. One might expect that the correction allows for more flexibility in the predictive ability of both yields and macroeconomic factors, and these will have opposing effects on the magnitude of observed unspanned risk. Because the effect is small relative to the necessary additional exposition for the method, the results here emphasize standard tests. Moreover, this is consistent with the results of Bauer and Rudebusch (2017), in which measurement errors of excess returns as a forecast target are not considered. On both counts, empirically the most important task is accurately measuring the space spanned by contemporaneous yields.

shadow yields are known. Then consider the regressions:

$$y_{t+h} = \Omega_0 + \Omega_1 \mathcal{P}_t + \Omega_2 M_t + \epsilon_t \quad (2.4.1a)$$

$$y_{t+h} = \bar{\Omega}_0 + \bar{\Omega}_1 \tilde{\mathcal{P}}_t + \bar{\Omega}_2 M_t + \bar{\epsilon}_t, \quad (2.4.1b)$$

where  $\mathcal{P}_t$  are PCs of  $y_t$  and  $\tilde{\mathcal{P}}_t$  are PCs of  $\tilde{y}_t$ . Then  $\mathbb{E}[\Omega_2] \neq 0$ , and  $\lim_{T \rightarrow \infty} \Omega_2 \neq 0$ , and  $\mathbb{E}[\bar{\Omega}_2] = 0$ , and  $\lim_{T \rightarrow \infty} \bar{\Omega}_2 = 0$ .

In the Gaussian class of models,  $M_t$  is orthogonal to  $\bar{\epsilon}_t$  under the null.  $\mathbb{E}[\bar{\Omega}_2] = 0$  by an application of Frisch-Waugh-Lovell, and that it is consistently estimable follows from standard ordinary least squares (OLS) results. The first statements of the theorem then follow from the fact that  $M_t$  is positively correlated with  $\epsilon_t$ .

When conducting the fully general spanning test 2.3.4, the following can be deduced from Theorem 4 of Pagan (1984):

**Theorem 3.** *Suppose  $y_t$  is generated from the spanned model of Section 2.2.3, and consider the following regressions:*

$$y_{t+h} = \Omega_0 + \Omega_1 y_t + \Omega_2 M_t + \epsilon_t \quad (2.4.2a)$$

$$y_{t+h} = \bar{\Omega}_0 + \bar{\Omega}_1 \hat{y}_t + \bar{\Omega}_2 M_t + \bar{\epsilon}_t \quad (2.4.2b)$$

Suppose regressions 2.4.2 are conducted as the second stage of a two-step estimator, where  $\hat{y}_t$  is estimated from the data and then PCs are computed in the first stage. Then  $\bar{\Omega}_2$  is consistently estimable by OLS and  $\lim_{T \rightarrow \infty} \bar{\Omega}_2 = 0$ .

## 2.4.2 Nonlinear Predictive Regression

The asymmetries of the ZLB imply that the residuals in the linear regressions will have nonstandard distributions. Consider actual yields at date  $t + h$ :

$$\begin{aligned}
 y_{t+h} &= \tilde{y}_{t+h}(\tau) + ZLB_{t+h}(\tau) \\
 &= A(\tau) + B(\tau)X_{t+h} + ZLB_{t+h} \\
 &= A + B^Z Z'_{t+h} + B^M M_{t+h}
 \end{aligned}$$

Projecting  $X_{t+h}$  back to date  $t$  via Gaussian transition dynamics yields:

$$y_{t+h} = \underbrace{A(\tau) + \bar{B}(\tau)X_t}_{\text{Linearly spanned by shadow yields under null}} + \underbrace{ZLB_{t+h}}_{\text{Never linearly spanned by shadow yields}} + \varepsilon_t \quad (2.4.3)$$

Thus, even if we use the correct nonlinear transformation of yields in a linear predictive regression, the residual from this regression will include the date  $t + h$  ZLB wedge. However, this nonlinear predictive regression consists of a nonlinear transformation of the sum of (1) shadow yields, obtained by inverting Equation 2.2.12, (2) macroeconomic risk factors, and (3) a Gaussian shock. The nonlinear predictive regression is thus a nonlinear transformation of the RHS of the linear predictive regression. In particular, if one only needs correct inference on a test of the null restriction that the coefficient on  $M_t$  is zero, this approach is redundant: Since the residuals in this nonlinear predictive regression are Gaussian, under the null of spanning the coefficient on  $M_t$  will behave identically in a correct nonlinear regression as in the linear regression above estimated by least squares.

Should inference on overall predictability of yields be necessary, as is the case when the null

is correctly rejected, the magnitude of unspanned risk will not be appropriately measured by a linear regression. Notice that conditional on rejecting the null, using the incorrect linear regression could in theory either overstate or understate the degree of unspanned risk; the effect depends on whether the coefficient on  $M_t$  has the same or opposite sign as its correlation with the ZLB wedge term.

The following method for overcoming this arises as a direct extension of methods for censored regressions in Rigobon and Stoker (2007) and Cameron and Trivedi (2005). Consider again Figure 19: Unlike an actual censoring environment, the nonlinear transformation induced by a ZLB wedge is invertible, and so the approach is much easier than correcting for censoring. Equation 2.4.3, rather than having to be estimated by MLE, can be inverted before projecting the right hand side back to date  $t$  to arrive at:

$$\tilde{f}_{t+h} = \bar{\Omega}_0 + \bar{\Omega}_1 \hat{y}_t + \bar{\Omega}_2 M_t + \bar{\epsilon}_t \quad (2.4.4)$$

Controlling for the nonlinear effect of  $M_t$  on yields can be accomplished by running Regression 2.4.2b but with shadow forward rates as the regressand – this is precisely Regression 2.4.4 – and then applying the ZLB transformation to arrive at an appropriate forecast for yields. The full procedure is thus:

- Run predictive regression 2.4.4.
- Construct a distribution of forecasts  $\hat{f}_{i,t+h}$  by simulation. Either by bootstrapping  $\bar{\epsilon}_t$ , or estimating the distribution of  $\bar{\epsilon}_t$  and taking deviates. Both in theory and empirically the distribution of  $\bar{\epsilon}_t$  is approximately Gaussian, and so the distribution is easy to simulate.

- Construct a distribution of forecasts  $\hat{y}_{i,t+h}$  via

$$\hat{y}_{i,t+h} = \frac{1}{\tau} \int_{t+h}^{t+h+\tau} \left[ \hat{f}_{i,t+h}(s) \Phi \left( \frac{\hat{f}_{i,t+h}(s)}{\Psi(s)} \right) + \Psi(s) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\hat{f}_{i,t+h}(s)}{\Psi(s)} \right]^2 \right) \right] ds$$

- Take  $\hat{y}_{t+h} = \frac{1}{M} \sum_i \hat{y}_{i,t+h}$  to be the fitted value for yields, where  $M$  is the number of simulated draws taken of  $\hat{f}_{i,t+h}$ .
- Calculate adjusted  $R^2$  from residuals  $y_{t+h} - \hat{y}_{t+h}$ .

The combined features of (1) Gaussian transition dynamics on state variables (and therefore Gaussian transition dynamics for shadow objects); and (2) invertible relationships between shadow objects and observable objects makes for simple nonlinear predictive regressions/nonlinear forecasting.

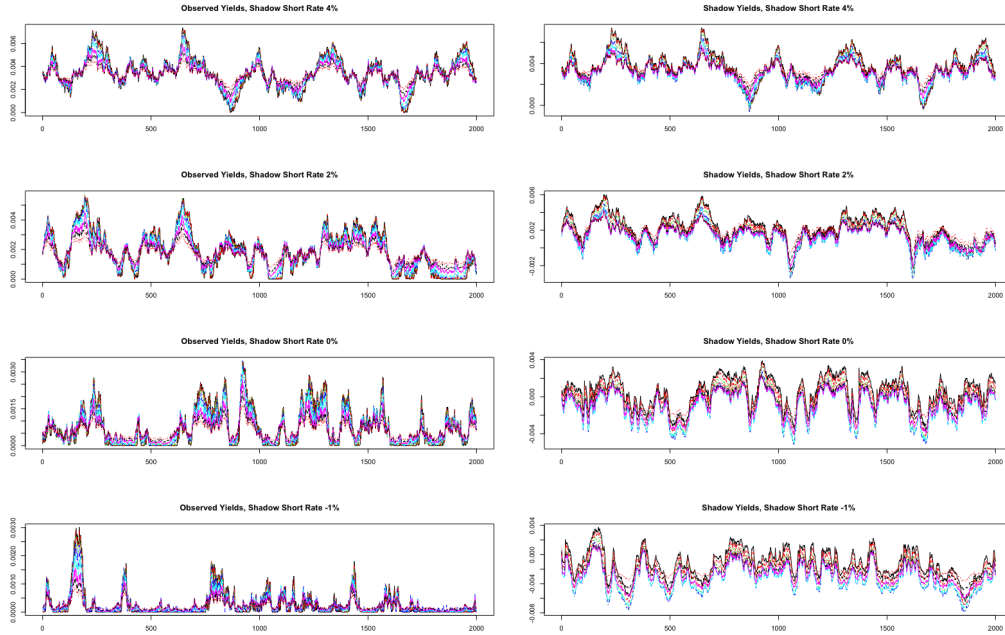


## 2.5 Simulation Results

This section discusses a set of simulations with two goals in mind: (1) Testing the degree of nonlinearity necessary in the bond yield measurement equation necessary to break linear spanning relationships; and (2) testing for the validity of our nonlinear spanning test procedure. A secondary goal is to compare the contributions of nonlinearity to spurious rejection to other leading sources of model misspecification: namely (1) measurement error, (2) the omission of higher order PCs, and (3) the general use of PCs over the full set of yields. This simulation only considers the linear predictive regressions. The central question of interest is whether incorrect methods can spuriously reject the null, which only requires the linear predictive regressions to test. Measuring the size of unspanned risk is an empirical matter. Moreover, since conducting the nonlinear predictive regressions themselves require a simulation, doing so repeatedly in a simulation environment is a tremendous computational burden relative to the gains in understanding.

In the model of Section 2.2.3, nonlinearity can be generated whenever there is a positive probability of future short rates being negative. As a motivating exercise, consider the model of Bauer and Rudebusch (2016), a shadow-rate model augmented with macroeconomic risk factors. Classically, the ZLB literature has emphasized the role of the ZLB nonlinearity in short-maturity yields, because these are the yields most obviously constrained by the boundary. However, as the yield curve approached the ZLB, the longest maturity yields were actually affected first. Since the relevant variable is when *positive probability* of negative short rates appear over the duration of the bond's maturity, this probability first becomes positive at a maturity of 10 years before it becomes positive at a maturity of 1 year. This is explicitly demonstrated in Figure 17. This is relevant for the spanning literature, because the standard window to examine ends in 2007, and it is thought that the ZLB has no impact in this window. However, we demonstrate that nonlinearities are already emerging at the long end of the yield curve during this window.

Figure 20: Simulated Data



For the purposes of this simulation exercise, we generated a long dataset from the arbitrage-free shadow rate model on the assumption that nonlinear spanning holds. We then conducted a number of spanning tests to examine the rate of rejection in favor of unspanned risk. The following tests were thus considered in population:

- Predictive regressions 2.4.2, with yields regressed against lagged:
  - 5 principal components of yields and macroeconomic risk factors.
  - 5 principal components of shadow yields and macroeconomic risk factors.
- The same tests but using three principal components instead of five.
- The same tests but using full yields and shadow yields instead of principal components.
- The same tests but generating observed yields as having classical measurement error with a standard deviation of 6 basis points (bp).

The main hypothesis is that in cases where macroeconomic risk factors can be deduced

from yield curve information, albeit only with a nonlinear relationship (i.e., via shadow yields), linear spanning tests will incorrectly suggest the presence of significant unspanned macroeconomic risks. Furthermore, modified nonlinear spanning tests that employ shadow yields suggest that it may be possible to recover appropriate spanning relationships.

The data is generated from the model given in Section 2.2.3 with dimensions of  $N = 8$  maturities ranging from three months to 10 years. The macroeconomic risk factors used are inflation and the unemployment gap, as there is a common view that these macroeconomic conditions are correlated with yield curve dynamics (see in particular the above discussion on the meaning of the level and slope factors). Note that the paths of the macroeconomic risk factors are simulated as well, but the choice of macroeconomic factors determines the parameters that will govern their dynamics. This exercise is conducted for several different degrees of nonlinearity, and the outcomes of the spanning tests are given below (along with their expected outcomes). Nonlinearities are controlled via the unconditional mean and starting value of the shadow level factor. Implicitly, this controls the distance of the shadow short rate from zero (since the two are affine functions of each other), and thus the probability of negative future short rates. This probability varies for yields of different maturities, and so the  $R^2$  of the regression of yields on contemporaneous risk factors varies across maturities. When the process is close to the ZLB, there will be stronger nonlinearities, and the linear  $R^2$  will be lower. Refer to Figure 20 for the divergence between observed and shadow yields across the different levels of the unconditional shadow short rate.

The qualitative nature of the phenomenon can be observed in Table 16. Small degrees of nonlinearity induce high rates of rejection of the null when shadow yields are omitted. The tests achieve correct size once shadow yields are used as regressors. The only irregularity in this table is that the rejection rates are not all 1; the predictive regressions are being run in population and in the regressions ignoring shadow yields the null hypothesis is in fact false. The reduction in power is likely a product of the high correlation between regressors when the null hypothesis is close to being satisfied, as the model is only weakly identified.

Table 16: Linear Spanning Tests on a Nonlinear Model – Rejection Rates

This table presents the results of spanning tests applied to nonlinear models in population, with the set of models indexed by degree of nonlinearity. The metric is difference in rejection rates of the restriction that coefficients on macroeconomic factors are zero for the predictive regression that tests for unspanned macroeconomic risk. The first column gives the unconditional mean of the shadow short rate. The following three panels are divided into five columns. The first gives the approximate  $R^2$  of the measurement equation for the given yield. The remaining columns are differences in rejection rates of the restriction on macroeconomic coefficients in each spanning test: Column 2 is for the nonlinear spanning test with 5 PCs applied to yields with  $6bps$  of error. Column 3 is for the linear test with 5 PCs. Columns 4 and 5 are for the nonlinear test with 3 PCs and 5 PCs, respectively. The top panel tests regression 2.3.9, using principal components as regressors, while the bottom uses the full dataset, testing the fully general regression 2.3.4 (hence columns 4 and 5 of each lower panel are combined). The tests are conducted at horizon  $h = 12$  periods ahead.

Shadow Short Rate		Risks By Maturities														
		3m					2yr					10yr				
		$R^2$	ME	L	NL(3)	NL(5)	$R^2$	ME	L	NL(3)	NL(5)	$R^2$	ME	L	NL(3)	NL(5)
PCs	4%	0.99	0.07	0.55	0.07	0.05	0.99	0.07	0.36	0.06	0.06	0.99	0.07	0.41	0.10	0.06
	2%	0.97	0.24	0.94	0.08	0.07	0.97	0.07	0.59	0.07	0.06	0.98	0.06	0.67	0.08	0.07
	0%	0.75	0.25	0.96	0.12	0.06	0.84	0.07	0.94	0.09	0.06	0.92	0.12	0.99	0.07	0.06
	-1%	0.54	0.14	0.93	0.10	0.07	0.71	0.08	0.94	0.09	0.08	0.86	0.10	0.99	0.07	0.07
Full	4%	0.99	0.10	0.07		0.04	0.99	0.05	0.08		0.04	0.99	0.06	0.07		0.05
	2%	0.97	0.25	0.14		0.06	0.97	0.08	0.05		0.05	0.98	0.06	0.06		0.04
	0%	0.75	0.23	0.35		0.07	0.84	0.07	0.23		0.06	0.92	0.11	0.39		0.06
	-1%	0.54	0.14	0.33		0.07	0.71	0.07	0.33		0.07	0.86	0.09	0.69		0.06

The top row of the table demonstrates in simulation our intuition that the nonlinearity induced by the ZLB can affect magnitudes of unspanned risk even prior to 2007. The magnitude of the effect on short-maturity yields increases as the ZLB approaches. Though this result may be surprising given that overall variability of short-maturity yields falls close to the ZLB, the intuition follows from the fact that these are long-horizon forecasts. Consider the bottom two plots in Figure 20: there are long periods where yields are constant at zero, followed by regimes where yields become positive (when the shadow short rate has wandered sufficiently far above its long run mean). For example, in the linear test forecasts for the highly volatile period at approximately  $t = 800$  are constructed from the period preceding, in which yields are almost perfectly flat at zero. By contrast, shadow yields have their full affine space from which to construct forecasts.

Consistent with the existing literature, the strongest mechanism for inducing spurious unspanned risk is the omission of the 4th and 5th factors. A notable irregularity is the non-monotonicity of the effect of excluding 4th and 5th PCs as the ZLB approaches for both medium and long maturity yields, and no intuitive explanation seems obvious. This simulation also facilitates a test of the explanation that the unstable projection onto principal components is to blame for the large errors even in the case of small nonlinearities (or small measurement error). Since very large datasets are produced, there is no need to conduct the spanning test by only regressing against principal components – in fact predictive regression 2.3.4 can be conducted. This predictive regression should find that for very small degrees of nonlinearity, or for very small degrees of measurement error, model-implied yields and shadow yields should be approximately collinear, and results should converge. The over-rejection of the null is cut dramatically for medium maturity bonds when controlling for the full set of yields, but the improvement for the long end of the yield curve remains relatively robust. This suggests that while the projection of yields onto principal components before conducting spanning tests may be problematic, and while spanning tests at higher frequencies (with enough observations to conduct the fully general tests) may offer different evidence on the spanning hypothesis, it is unlikely to be the full story.

This simulation also facilitates an exploration of the hypothesis of Bauer and Rudebusch (2017), namely that the most plausible explanation reconciling the theory and empirics of the spanning problem is that bond yields are measured with error. Consider a measurement equation for yields where instead of observing yields  $y_t$  we observe yields with some small measurement error,  $y_t^* = y_t + \varepsilon_t$ . Then the model for  $y_t^*$  is

$$y_t^* = A + BX_t + \varepsilon_t = A + B^Z Z_t' + B^M M_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (2.5.1)$$

So long as  $\sigma^2 > 0$ , it is possible that  $B^M \neq 0$  and still the above equation cannot be inverted to find  $M_t$  as affine in observed bond yields. Therefore, macroeconomic factors can be helpful in forecasting bond yields, fail to be affine in *observed* bond yields, and be affine in true bond yields. The paper finds that small degrees of measurement error can induce substantial spurious unspanned risk, because small degrees of measurement error can substantially bias the estimates of higher principal components, inducing spurious unspanned risk.

However, the measurement error explanation is potentially unsatisfying: Bauer and Rudebusch (2017) even offers the caveat that sufficient measurement error renders any two models indistinguishable. The grievance may even be more specific – the spanning hypothesis states that if yields are affine in macroeconomic risk factors then those factors cannot improve predictive regressions, and adding measurement error is specifically breaking the assumption that observed yields are affine in risk factors. Without a meaningful economic distinction between observed yields and model-implied (or “actual”) yields, imposing measurement error is an imposition that spanning does not hold.

This intuition facilitates comparison between the two explanations: Bauer and Rudebusch (2017) notes that a measurement error with a standard deviation of 6bp is sufficient to explain spurious evidence for unspanned risks in a model where spanning holds. It is confirmed that 6bps of measurement error is sufficient to generate over-rejection of the null of unspanned risk, as seen from the column labeled *ME*. However, the measurement error

explanation and the nonlinear misspecification story do not have identical strengths at all maturities. For several maturities, the imposition of measurement error induces no spuriously observed unspanned risk, while for others the effect is far greater than the nonlinear mechanism. This is especially true far away from the ZLB. In some sense we should expect this, as the measurement error imposes uniform bias on all yields, while the ZLB influences different maturities at different times. The measurement error explanation is more contingent on the omission of higher order principal components, as the effect is weaker when the full set of yields is used. There is no obvious intuition for this, but a potential explanation is that since the same degree of measurement error is applied to all yields, but longer-maturity yields experience more observed unspanned risk than short-maturity yields, applying the same magnitude of measurement error to long-maturity yields as short-maturity ones further biases the principal components. This creates large observed unspanned risk that would not be present if only long-maturity yields were observed without error. However, both mechanisms induce over-rejection.

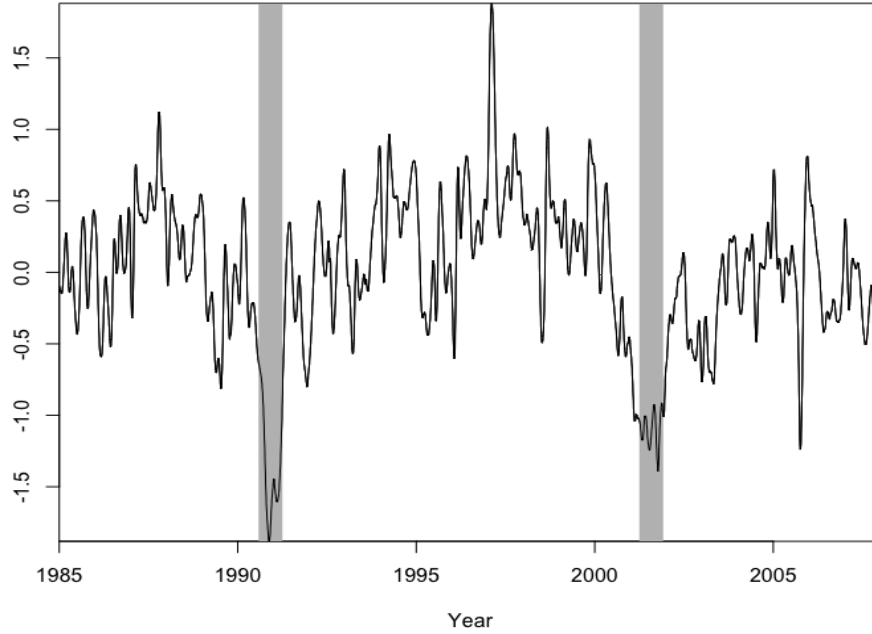
There are some potential issues with this simulation study. The first is that it is conducted in population, and so empirical studies with smaller samples may not exhibit precisely the same phenomenon, or exhibit the same numerical importance. In smaller samples, there are direct small-sample issues in the predictive regression tests as considered by Bauer and Hamilton (Forthcoming). There will also be estimation error in shadow yields, as conducting the inversion to estimate them requires estimating the variance of the option value of the shadow bond – this increases the importance of distinguishing whether the ZLB wedge is observable by agents. This estimation error also leads to a nontrivial generated regressor problem that makes conducting the appropriate F-tests difficult, though of course differences in  $R^2$  of the predictive regression are still computable. Following Bauer and Rudebusch (2017) and Cochrane (2015), the qualitative nature of the phenomenon is likely to be preserved in spite of these issues. However, a full treatment of the small-sample theory in this environment a la Bauer and Hamilton (Forthcoming) is left for future work. The second is that even though this data is generated from the same parameters frequently found

in the data, it is more obviously stationary – meanwhile, the actual data is often asserted to be detrending and thus nonstationary, and in particular Cochrane (2015) points to this detrending as a potential source of spurious unspanned risk. A similar simulation study that includes strong trend and cyclical components in the level factor and inflation yields qualitatively and quantitatively similar results.



Figure 21: ADS Index

This is the Aruoba-Diebold-Scotti (ADS) Index, available at daily frequency from the Philadelphia Federal Reserve. Dates of recessions shaded in grey are taken from Rudebusch et al. (2007).



## 2.6 Empirics

### 2.6.1 Data and Estimation

The key questions raised in the introduction section of this paper were (1) whether the effect of omitting an obvious nonlinearity is stronger than the effect of failing to use a full set of yields; and (2) how strong is the interaction between these effects, following the argument that rotating to PCs and truncating higher order components will generally magnify the effects of model misspecification. The simulations described here have demonstrated that the spurious rejections can be large, and that appropriate tests are available. The size of the effect in practice is an empirical question, and so this section now turns to data.

Yields are market yields on US Treasury securities at a constant maturity of 3–240 months, at

both monthly and daily frequencies. The use of the daily dataset facilitates the fully general spanning tests, in which a larger number of observations allows more accurate use of the full yield curve for regression instead of their principal components. Yields for all maturities at both monthly and daily frequency are available directly from the Federal Reserve<sup>14</sup>. As discussed in the simulation results, the macroeconomic factors used at the monthly frequency are inflation (measured by annual percent change in core consumer price index (CPI)) and unemployment gap. This is the same dataset used to explore monetary policy expectations in Bauer and Rudebusch (2016). A key additional macroeconomic factor used elsewhere, in particular by Diebold et al. (2006), is the federal funds rate. However, as argued in Rudebusch (2009), the federal funds rate is largely determined by inflation and real activity, and thus its inclusion on top of a price index and the unemployment gap would be redundant.

The key macroeconomic risk factor used is the Aruoba-Diebold-Scotti activity index. This is a high-frequency measure, available at the daily frequency, of US economy real activity<sup>15</sup>. For a thorough review of the ADS construction and results stemming from its use see Aruoba et al. (2009). There are several estimates of daily inflation indices, such as Knotek and Zaman (2015) (who provide a procedure for daily nowcasts of the current quarter core inflation) and Watanabe and Watanabe (2014) (who construct a daily price index using point-of-sale scanner data in Japan). However, the high frequency inflation indices literature is inconclusive with respect to reliability, as compared to the use of high frequency real activity indices. The model brought to the daily data in this paper will therefore only consider a real activity index, and thus have only a single macroeconomic factor. There is precedence for this: models with a sole macroeconomic risk factor include Ang et al. (2006) and Jardet et al. (2013), both of whom rely only on a growth factor. Real activity has a stronger

---

<sup>14</sup>See <http://www.federalreserve.gov/releases/h15/data.htm>. For a discussion on methodologies of constructing yield curve data, see Gurkaynak et al. (2007).

<sup>15</sup>This index is available from the Philadelphia Federal Reserve at <https://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index>. The index is an extracted daily business activity factor built from a mixed frequency model of many indicators. However, none of these indicators are at a daily frequency, so new information is only incorporated into the ADS index every 2–4 days. Between these periods, movements are model-driven.

link to yield curve dynamics than growth, meaning that this model is motivated by their parsimony and that the ADS index should prove a more relevant variable than any daily growth indicator<sup>16</sup>.

The sample is from 1993–2016. Beginning in 1993 facilitates the inclusion of 20-year maturity yields, as US Treasury yields of 20-year maturity are only available during that window<sup>17</sup>. Estimation is conducted by Extended Kalman Filter, following the belief that there are relevant nonlinearities throughout the full sample period. Details are in Section 2.8.1.

## 2.6.2 Empirical Spanning Tests

Tables 17 through 20 contain the results of the empirical spanning tests. The general features of the linear spanning tests are consistent with the literature. There is less unspanned risk found in the full sample of 1993–2016 than in 1993–2007 and 2007–2016 separately; this is consistent with the view that the transition to the ZLB was accompanied by a regime change.

There are several new features to point out. First, the negative result: including shadow yields in the predictive regressions, in a very small number of cases, actually *increases* the degree of observed unspanned risk. This result is quite noticeable in the linear predictive regressions, and mostly vanishes when nonlinear predictive regressions are considered. Where it appears in the nonlinear predictive regressions, it is mostly at the short end of the yield curve in the post-ZLB regime; this is arguably because adjusted  $R^2$  are penalizing for additional regressors, and since the short end of the yield curve is strongly restricted most predictive gains come from regressing just against a constant. Broadly, this is a result in favor of the measurement error explanation for spurious unspanned risk posited by Bauer and Rudebusch (2017): Since shadow yields are estimated from a model, if the data are

---

<sup>16</sup>Bauer and Rudebusch (2017) finds a connection between growth measures and curvature of yields – however, given the high level of noise in measuring both, they are skeptical of this connection and consider it likely to be spurious overfitting. In contrast, Moller (2014) provides new evidence that the relationship between growth and curvature is rather robust.

<sup>17</sup>Little is lost by starting this late, as there was certainly a break in the dynamics of the yield curve in the late 1980s, as documented by Joslin et al. (2014).

observed with error, the model will be improperly estimated, and thus estimated shadow yields can have substantial error. This results in greater observed unspanned risk.

However, the empirical results are largely supportive of the main hypothesis – the inclusion of the shadow yields reduces the gains from adding macroeconomic factors to the predictive regression. This is most strongly seen when considering the nonlinear predictive regressions. We may conclude that the correct functional form of the regression is more important for testing spanning than is using the correct set of regressors. Many pieces of the literature are interested in predicting returns instead of yields. Returns are a further nonlinear transformation of yields, and so even when yields are affine in risk factors there is little reason to expect linear predictive regressions to correctly measure unspanned risk. In some cases, once shadow yields are conditioned on, adding macroeconomic factors actually reduces the adjusted  $R^2$  (i.e., the negative entries in the tables), though this only occurs in the cases of using the full set of yields. Finally, the proposition that the effect of the ZLB would show up most strongly at the long end of the yield curve pre-ZLB ends up being quite dubious – though the ZLB is larger at the long end of the yield curve than the short end, the extra fraction of variation explained by the ZLB wedge is small.

Using the full set of yields instead of 5 PCs reduces observed unspanned risk, but the effect is not as large as the effect of moving from 3 PCs to 5 PCs. This is to be expected: The sixth through ninth PCs are vanishingly small. While the suggestion that PCs with small cross-sectional importance may be crucial for forecasting is believable for the fourth and fifth PCs, the argument can only be stretched so far before it strains credulity. In terms of magnitudes of effects, it is a very mixed bag whether the omitted nonlinearity or the omitted highest order PCs (i.e. using the full set of yields) have a stronger effect in inducing spurious unspanned risk. However, it is clear that the interaction of the two is quite relevant: examining the 93–07 rows of Table 18 and 20, when the nonlinearity is weak its ability to generate large spurious unspanned risk hinges on the use of PCs. Conditioning on a full set of yields makes the effect of the weak nonlinearity on spurious unspanned risk appropriately

weak. A concise summary of the results would be that approximately 10% of unspanned risk from 1993–2007 is attributable to the ZLB wedge, while upwards of 40% – 50% are attributable to the use of PCs instead of the full set of yields. In addition, comparing the pre-2007 and post-2007 periods, both models suggest that unspanned risk has fallen, but this hinges strongly on the use of all yields instead of PCs.

Table 17: Unspanned Risk – Monthly – Linear Predictive Regression

This table provides the results of spanning tests for yields applied to a model at monthly frequency, using the unemployment gap and inflation as the macroeconomic risk factors. The full sample of data is from 1993–2016, and each row presents the tests conducted on a subsample with the years given in the far left column. The following panels are broken into four columns: The first column gives the linear  $R^2$  of the measurement equation for the bond of the given maturity. The remaining columns provide the differences in adjusted  $R^2$  for the linear spanning test, the nonlinear spanning test with three PCs, and the nonlinear test with five PCs respectively. The top panel conducts this series of tests using PC-based tests, and the lower panel conducts this series using the fully general tests (hence columns 3 and 4 of each lower panel are combined). The results are conducted at horizon  $h = 12$  periods (1-year) ahead.

		Risks By Maturities															
		3m				2yr				10yr				20yr			
		$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)
PCs	93–16	0.99	0.0033	0.0037	0.0031	0.99	-0.0003	0.0048	-0.0004	0.99	0.0068	0.0127	0.0093	0.99	0.0140	0.0208	0.0186
	93–07	0.99	0.0724	0.0511	0.0745	0.99	0.1012	0.0471	0.0989	0.99	0.1432	0.0516	0.1333	0.98	0.1344	0.0424	0.1220
	07–16	0.64	0.1355	0.2919	0.1185	0.96	0.3403	0.3695	0.3822	0.99	0.0955	0.0988	0.1094	0.98	0.0108	0.0110	0.0127
Full	93–16	0.99	0.0030	0.0007		0.99	-0.0015	-0.0016		0.99	0.0061	0.0070		0.99	0.0138		0.0133
	93–07	0.99	0.0433	0.0408		0.99	0.0524	0.0435		0.99	0.0970	0.0640		0.98	0.1000		0.0601
	07–16	0.64	0.1036	0.0009		0.96	0.1840	-0.0048		0.99	0.0418	-0.0050		0.98	0.0020		-0.0025

Table 18: Unspanned Risk – Daily – Linear Predictive Regression

This table provides the results of spanning tests for yields applied to a model at daily frequency, using the ADS index as a single macroeconomic risk factor. The full sample of data is from 1993–2016, and each row presents the tests conducted on a subsample with the year given in the far left column. The following panels are broken into four columns: The first column gives the linear  $R^2$  of the measurement equation for the bond of the given maturity. The remaining columns provide the differences in adjusted  $R^2$  for the linear spanning test, the nonlinear spanning test with three PCs, and the nonlinear test with five PCs respectively. The top panel conducts this series of tests using PC-based tests, and the lower panel conducts this series using the fully general tests (hence columns 3 and 4 of each lower panel are combined). The results are conducted at horizon  $h = 365$  periods (1-year) ahead.

		Risks By Maturities															
		3m				2yr				10yr				20yr			
		$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)
PCs	93–16	0.99	0.0476	0.0433	0.0422	0.99	0.0168	0.0138	0.0132	0.99	-0.0000	-0.0000	-0.0000	0.99	0.0023	0.0025	0.0024
	93–07	0.99	0.0728	0.1675	0.0693	0.99	0.0524	0.1423	0.0499	0.99	0.0140	0.0567	0.0131	0.99	0.0054	0.0260	0.0049
	07–16	0.87	-0.0002	0.0061	0.0011	0.99	0.0943	0.1148	0.0740	0.99	0.0241	0.1509	0.0192	0.99	0.0089	0.1139	0.0076
Full	93–16	0.99	0.0574	0.0580		0.99	0.0314	0.0295		0.99	0.0035	0.0028		0.99	0.0000		-0.0000
	93–07	0.99	0.0484	0.0484		0.99	0.0299	0.0299		0.99	0.0061	0.0061		0.99	0.0020		0.0020
	07–16	0.87	-0.0001	0.0007		0.99	0.0354	0.0319		0.99	0.0088	0.0184		0.99	0.0033		0.0116

Table 19: Unspanned Risk – Monthly – Nonlinear Predictive Regression

This table provides the results of spanning tests for yields applied to a model at monthly frequency, using the unemployment gap and inflation as the macroeconomic risk factors. The full sample of data is from 1993–2016, and each row presents the tests conducted on a subsample with the years given in the far left column. The following panels are broken into four columns: The first column gives the linear  $R^2$  of the measurement equation for the bond of the given maturity. The remaining columns provide the differences in adjusted  $R^2$  for the linear spanning test, the nonlinear spanning test with 3 PCs, and the nonlinear test with 5 PCs respectively. The top panel conducts this series of tests using PC-based tests, and the lower panel conducts this series using the fully general tests (hence columns 3 and 4 of each lower panel are combined). The results are conducted at horizon  $h = 12$  periods (1 year) ahead.

		Risks By Maturities															
		3m				2yr				10yr				20yr			
		$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)
PCs	93–16	0.99	0.0052	-0.0007	0.0002	0.99	0.0015	0.0021	-0.0010	0.99	0.0087	0.0122	0.0089	0.99	0.0157	0.0209	0.0169
	93–07	0.99	0.0771	0.0450	0.0677	0.99	0.1069	0.0449	0.0988	0.99	0.1493	0.0521	0.1476	0.98	0.1398	0.0424	0.1405
	07–16	0.66	0.1497	-0.2240	-0.1904	0.97	0.3761	0.3310	0.3663	0.99	0.1167	0.0756	0.1016	0.98	0.0259	-0.0496	-0.0375
Full	93–16	0.99	0.0030		0.0035	0.99	-0.0015		-0.0020	0.99	0.0061		0.0048	0.99	0.0139		0.0119
	93–07	0.99	0.0435		0.0424	0.99	0.0527		0.0514	0.99	0.0976		0.0914	0.98	0.1006		0.0959
	07–16	0.66	0.1056		-0.0775	0.97	0.1878		0.1758	0.99	0.0424		0.0345	0.98	0.0019		-0.0430

Table 20: Unspanned Risk – Daily – Nonlinear Predictive Regression

This table provides the results of spanning tests for yields applied to a model at daily frequency, using the ADS index as a single macroeconomic risk factor. The full sample of data is from 1993–2016, and each row presents the tests conducted on a subsample with the year given in the far left column. The following panels are broken into four columns: The first column gives the linear  $R^2$  of the measurement equation for the bond of the given maturity. The remaining columns provide the differences in adjusted  $R^2$  for the linear spanning test, the nonlinear spanning test with 3 PCs, and the nonlinear test with 5 PCs respectively. The top panel conducts this series of tests using PC-based tests, and the lower panel conducts this series using the fully general tests (hence columns 3 and 4 of each lower panel are combined). The results are conducted at horizon  $h = 365$  periods (1 year) ahead.

		Risks By Maturities															
		3m				2yr				10yr				20yr			
		$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)	$R^2$	L	NL(3)	NL(5)
PCs	93–16	0.99	0.0417	0.0434	0.0369	0.99	0.0142	0.0143	0.0117	0.99	0.0000	0.0011	0.0005	0.99	0.0019	0.0027	0.0023
	93–07	0.99	0.0626	0.1324	0.0603	0.99	0.0477	0.1166	0.0459	0.99	0.0172	0.0515	0.0163	0.99	0.0082	0.0270	0.0078
	07–16	0.87	0.0022	-0.0026	-0.0819	0.99	0.1090	0.1325	0.0915	0.99	0.0398	0.1789	0.0314	0.99	0.0178	0.1415	0.0131
Full	93–16	0.99	0.0466		0.0439	0.99	0.0251		0.0219	0.99	0.0027		0.0027	0.99	0.0000		0.0004
	93–07	0.99	0.0445		0.0423	0.99	0.0292		0.0272	0.99	0.0090		0.0080	0.99	0.0045		0.0039
	07–16	0.87	-0.0000		-0.0049	0.99	0.0459		0.0464	0.99	0.0195		0.0202	0.99	0.0093		0.0095

## 2.7 Conclusion

The results here are a further exploration of the spanning hypothesis, motivated by and building from the existing literature's disagreement on the validity of spanning. This disagreement is most starkly highlighted by four works: Joslin et al. (2014), Bauer and Rudebusch (2017), Bauer and Hamilton (Forthcoming), and Cochrane (2015) exhibit substantial disagreement on the qualitative nature of spanning and, even where they agree on whether unspanned risks exist, disagree on the economic significance. Weighing in on this disagreement, Rebonato (2016) posits one of two explanations:

1. Affine models are wholly unsatisfactory, and exploring spanning through affine models will always yield irreconcilable disagreements.
2. Spanning holds in affine models, but can only be observed if enough principal components (or indeed, a full set of sufficiently many yields) are used.

To weigh in on these possibilities, this chapter first examined the significance of omitting the nonlinear wedge induced by the ZLB. In contrast to much of the existing ZLB literature, forward-looking behavior in bond yields means that a nontrivial ZLB wedge can be observed as early as 1993–2007. Omitting this nonlinearity induces spuriously observed unspanned risk, much like the measurement error explanations of Bauer and Rudebusch (2017) or the small-sample bias explanations of Bauer and Hamilton (Forthcoming). This omission is corrected by extracting shadow yields, which can be included as regressors in the predictive regression tests of spanning. Updating the functional form of the predictive regression to respect the ZLB further reduces observed unspanned risk. A simulation study shows the effect of this omission can be large and that corrected tests behave appropriately.

Second, by using higher frequency data, predictive regression tests of the spanning hypothesis can use a higher-dimensional set of yields as regressors – in fact the full set of observed yields is used. This was difficult to do prior to technical innovations such as the ADS index. By comparing typical linear tests against proper tests that account for the nonlinearity of



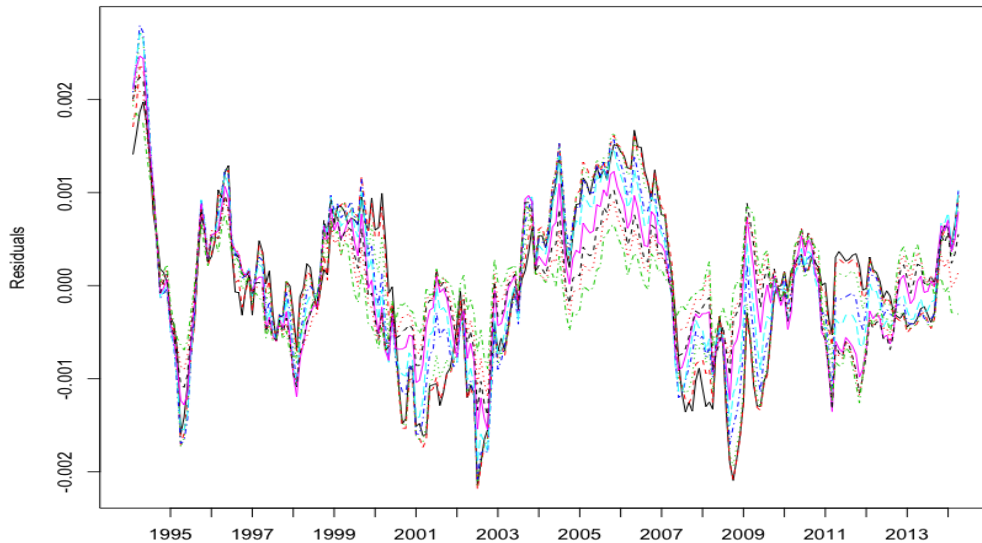
the ZLB wedge, and examining the economic significance in small and large datasets, the results here can weigh in on the explanations posited by Rebonato (2016). Combining these techniques, the work here can be considered a first step in the comparison of the effects of omitted nonlinearities and insufficient data vis a vis generating spurious unspanned risk.

The results suggest that while omitting the ZLB nonlinearities are significant and can account for approximately 10% of observed unspanned risk, using additional yield data is more important, accounting for approximately 40% – 50% of observed unspanned risk.. This is consistent with the more recent empirical developments that emphasize the role of higher order principal components, namely that spanned factors with economically imperceptible but statistically significant cross-sectional effects can be crucial for forecasting.

### **2.7.1 Directions for Future Work**

There are several obvious next steps for the work presented here. The nonlinear predictive regression corrections here were more important than estimating the space spanned by regressors. That is, using shadow yields as regressors corrected for less observed unspanned risk than using a correct functional form of the regression. In our case the nonlinear predictive regression was quite simple to construct. Overall, even both mechanisms combined were insufficient to account for even a majority of observed unspanned risk, let alone all of it. However, much of the literature considers the prediction of excess returns instead of yields directly. Since returns are a further transformation of the data, it is unclear if using a nonlinear predictive regression would be even more crucial in that setting. In addition, and more importantly: the ZLB wedge is an obvious omitted nonlinearity for which it is simple to adjust spanning tests, but the set of nonlinear yield curve models is large and as such remaining effects could be large. As stressed in Cochrane (2015), there are even quite simple omitted model dynamics. Consider Figure 22, a plot of the residuals from a linear predictive regression spanning test: There are obvious business cycle dynamics that could be easily accounted for without the incorporation of macroeconomic factors (say, by deseasonalizing).

Figure 22: Residuals of Linear Predictive Regression



The theory of the daily frequency model and its relation to spurious unspanned risk has much work to be done. In particular, one needs an estimate of how much remaining unspanned risk at the daily frequency could still be attributable to small-sample issues, despite the large number of observations. The strong business cycle component of yields means infill asymptotics and calendar-time asymptotics will behave quite differently; a daily-frequency model still only observes 3-4 business cycles over 20 years. The daily frequency model also offers a framework for measuring efficiency of bond markets: Higher frequency data results in less measured unspanned risk, even when using the same number of regressors. This suggests the results at a monthly frequency are partly caused by failing to factor in the short amount of time it takes for agents to trade away macroeconomic information into the yield curve<sup>18</sup>. Finally, the results suggest that additional methods of disciplining a predictive regression should be promising, such as shrinkage or selection.

---

<sup>18</sup>Exploring this requires a structural model, as reduced-form simulations attempting to demonstrate the effect will beg the question. Consider a simulation in which intra-month yields load on previous day's macroeconomic information, and that macroeconomic information is highly autocorrelated. This will result in observations of substantial unspanned risk if the process is sampled at a monthly frequency and little to no unspanned risk if the process is sampled daily.

## 2.8 Appendix

### 2.8.1 Estimation Appendix

The use of the Kalman Filter for estimation in regular times is well-documented (e.g., Christensen et al. (2011) or Diebold and Li (2006)). Our use of the Kalman Filter will be standard, save for small adjustments owing to the fact that the measurement equation contains both latent and observed states. In normal times one might simply regress yields against both Nelson-Siegel polynomials and observed macroeconomic factors, then run a VAR on level/slope/curvature and macroeconomic risk factors. However, this method is not possible when augmenting to a nonlinear model, and so the Extended Kalman filter must be used. The second half of this section presents the necessary adjustments to the filter for estimation in the ZLB state.

Recall our system:

$$y_t = A(\tau) + B^Z(\tau)Z_t + B^M(\tau)M_t + \varepsilon_t \quad (2.8.1a)$$

$$\begin{pmatrix} Z_t \\ M_t \end{pmatrix} = \begin{pmatrix} \mu_Z \\ \mu_M \end{pmatrix} + \begin{pmatrix} \phi_{ZZ} & \phi_{ZM} \\ \phi_{MZ} & \phi_{MM} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ M_{t-1} \end{pmatrix} + \Sigma_t \quad (2.8.1b)$$

$$\begin{pmatrix} \varepsilon_t \\ \Sigma_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & \Omega \end{pmatrix} \right] \quad (2.8.1c)$$

Rewriting the system in appropriate state space representation – let  $M_t$  be true macroeco-

nomic variables, and  $M_t^*$  be their observed value. Then:

$$\begin{pmatrix} y_t \\ M_t^* \end{pmatrix} = \begin{pmatrix} A(\tau) + B^Z(\tau)Z_t + B^M(\tau)M_t + \varepsilon_t \\ M_t + \varepsilon_t^M \end{pmatrix} \quad (2.8.2a)$$

$$(2.8.2b)$$

$$\begin{pmatrix} Z_t \\ M_t \end{pmatrix} = \begin{pmatrix} \mu_Z \\ \mu_M \end{pmatrix} + \begin{pmatrix} \phi_{ZZ} & \phi_{ZM} \\ \phi_{MZ} & \phi_{MM} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ M_{t-1} \end{pmatrix} + \Sigma_t \quad (2.8.2c)$$

$$(2.8.2d)$$

$$\begin{pmatrix} \varepsilon_t \\ \varepsilon_t^M \\ \Sigma_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Omega \end{pmatrix} \right] \quad (2.8.2e)$$

With the inclusion of the auxiliary measurement equation for macroeconomic factors, in particular a measurement equation with zero error variance, the state space representation now facilitates a completely standard Kalman Filter. For clarity on the use of the regular Kalman Filter and the Extended Kalman Filter and its history in estimating yield models, see Christensen and Rudebusch (2016).

The Extended Kalman Filter requires only a numerical adjustment to estimate the nonlinearity in the measurement equation. For the shadow rate model, the measurement equation post-ZLB is:

$$y_t(\tau) = \frac{1}{\tau} \int_t^{t+\tau} \left[ \tilde{f}_t(s) \Phi \left( \frac{\tilde{f}_t(s)}{\Psi(s)} \right) + \Psi(s) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{\tilde{f}_t(s)}{\Psi(s)} \right]^2 \right) \right] ds,$$

where shadow forward rates  $\tilde{f}_t(\tau)$  are also affine in risk factors by model assumption. The standard Kalman Filter will not achieve optimal estimates of the filtered states, and therefore, cannot be used in any maximum likelihood procedure. The Extended Kalman Filter considers arbitrary measurement and transition equations of the form:

$$\begin{aligned} y_t &= f(\alpha_t, \theta_1) + \varepsilon_t \\ \alpha_t &= g(\alpha_{t-1}, \theta_2) + \Sigma_t \end{aligned}$$

The filter approximates the nonlinear system through a first-order linear approximation when updating covariance estimates. Because yields are affine in the state variables, an affine approximation to the measurement equation is needed instead of a linear one. The transition dynamics are still linear. Define the matrices  $A_t(\theta_1)$ ,  $B_t(\theta_1)$  as

$$\begin{aligned} A_t(\theta_1) &= f(\alpha_{t|t-1}, \theta_1) - \frac{\partial f(\alpha_t, \theta_1)}{\partial \alpha_t} \Big|_{\alpha_t = \alpha_{t|t-1}} \alpha_{t|t-1} \\ B_t(\theta_1) &= \frac{\partial f(\alpha_t, \theta_1)}{\partial \alpha_t} \Big|_{\alpha_t = \alpha_{t|t-1}} \alpha_{t|t-1} \end{aligned}$$

This yields the approximation:

$$y_t = A_t(\theta) + B_t(\theta)\alpha_t + \varepsilon_t \tag{2.8.3}$$

The filter then proceeds exactly as in the standard Kalman Filter, substituting  $f(\cdot)$  in to the predictive step and update step, and substituting  $A_t(\theta_1)$ ,  $B_t(\theta_1)$  in the remaining equations.

## BIBLIOGRAPHY

- Adrian, Tobias and Joshua Rosenberg (2008), “Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk,” *The Journal of Finance*, 63, 2997–3030.
- Aguilar, Omar and Mike West (2000), “Bayesian Dynamic Factor Models and Portfolio Allocation,” *Journal of Business & Economic Statistics*, 18, 338–357.
- Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant (2002), “Quadratic Term Structure Models: Theory and Evidence,” *Review of Financial Studies*, 15, 243–288.
- Ait-Sahalia, Yacine, Per Mykland, and Lan Zhang (2005a), “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise,” *Review of Financial Studies*, 18, 351–416.
- Andersen, Torben, T. Bollerslev, and F.X. Diebold (2007), “Parametric and Nonparametric Volatility Measurement,” in “Handbook of Financial Econometrics,” (edited by Ait-Sahalia, Y. and L.P. Hansen), Amsterdam; North Holland.
- Andersen, Torben, Tim Bollerslev, Francis Diebold, and Paul Labys (2001), “The Distribution of Realized Exchange Rate Volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- Andersen, Torben, Tim Bollerslev, Francis X. Diebold, and Paul Labys (2003), “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Jin Wu (2006), “Realized Beta: Persistence and Predictability,” *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R.F. Engle and C.W.J. Granger*, B, 1–40.
- Andreou, Elena, Patrick Gagliardini, Eric Ghysels, and Mirco Rubin (2016), “Is Industrial Production Still the Dominant Factor for the US Economy?” Swiss Finance Institute Research Paper Series 16-11, Swiss Finance Institute.
- Ang, Andrew and Monika Piazzesi (2000), “A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables,” *Journal of Monetary Economics*, 50, 745–787.
- Ang, Andrew, Monika Piazzesi, and Min Wei (2006), “What Does The Yield Curve Tell Us about GDP Growth?” *Journal of Econometrics*, 131, 359–403.
- Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti (2009), “Real-Time Measurement of Business Conditions,” *Journal of Business and Economic Statistics*, 27, 417–427.
- Bai, Jushan and Serena Ng (2002), “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221.

- Bai, Jushan and Serena Ng (2006), “Evaluating Latent and Observed Factors in Macroeconomics and Finance,” *Journal of Econometrics*, 507–537.
- Bali, Turan G. and Robert F. Engle (2010), “The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations,” *Journal of Monetary Economics*, 57, 377–390.
- Bali, Turan G., Robert F. Engle, and Yi Tang (2013), “Dynamic Conditional Beta is Alive and Well in the Cross-Section of Daily Stock Returns,” Koc University-TUSIAD Economic Research Forum Working Papers 1305, Koc University-TUSIAD Economic Research Forum.
- Barigozzi, Matteo and Marc Hallin (2016), “Generalized Dynamic Factor Models and Volatilities: Recovering the Market Volatility Shocks,” *The Econometrics Journal*, 19, C33–C60.
- Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard (2011), “Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading,” *Journal of Econometrics*, 162, 149–169.
- Barndorff-Nielsen, Ole E. and Neil Shephard (2004), “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics,” *Econometrica*, 72, 885–925.
- Bauer, Michael D. and James D. Hamilton (Forthcoming), “Robust Bond Risk Premia,” *Review of Financial Studies*.
- Bauer, Michael D. and Glenn D. Rudebusch (2016), “Monetary Policy Expectations at the Zero Lower Bound,” *Journal of Money, Credit and Banking*, 48, 1439–1465.
- Bauer, Michael D. and Glenn D. Rudebusch (2017), “Resolving the Spanning Puzzle in Macro-Finance Term Structure Models,” *Review of Finance*, 21, 511–553.
- Bikbov, Ruslan and Mikhail Chernov (2010), “No-Arbitrage Macroeconomic Determinants of the Yield Curve,” *Journal of Econometrics*, 159, 166–182.
- Black, Fischer (1995), “Interest Rates as Options,” *The Journal of Finance*, 50, 1371–1376.
- Bomfim, Antulio N. (2003), “Interest Rates as Options: Assessing the Markets’ View of the Liquidity Trap,” Finance and Economics Discussion Series 2003-45, Board of Governors of the Federal Reserve System (U.S.).
- Braun, Phillip A., Daniel B. Nelson, and Alain M. Sunier (1995), “Good News, Bad News, Volatility, and Betas,” *The Journal of Finance*, 50, 1575–1603.
- Cameron, A. Colin and Pravin K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press.

- Chen, Zhanhui and Ralitsa Petkova (2012), “Does Idiosyncratic Volatility Proxy for Risk Exposure?” *The Review of Financial Studies*, 30.
- Christensen, Jens H. E. and Glenn D. Rudebusch (2013), “Estimating Shadow-Rate Term Structure Models with Near-Zero Yields,” *Journal of Financial Econometrics*, 13, 226.
- Christensen, Jens H. E. and Glenn D. Rudebusch (2016), “Modeling Yields at the Zero Lower Bound: Are Shadow Rates the Solution?” in “Dynamic Factor Models (Advances in Econometrics, Volume 35),” (edited by Eric Hillebrand, Siem Jan Koopman), chap. 3, 75–125, Emerald Group Publishing Limited.
- Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch (2011), “The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models,” *Journal of Econometrics*, 164, 4–20.
- Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch (2014), “Can Spanned Term Structure Factors Drive Stochastic Yield Volatility?” Working Paper Series 2014-3, Federal Reserve Bank of San Francisco.
- Christoffersen, Peter, Asger Lunde, and Kasper Olesen (2014), “Factor Structure in Commodity Futures Return and Volatility,” *CREATES Research Papers*.
- Cochrane, J. H. (2015), “Comments on ‘Robust Risk Premia’ By Michael Bauer and Jim Hamilton,” Working paper.
- Cochrane, John H. and Monika Piazzesi (2005), “Bond Risk Premia,” *American Economic Review*, 95, 138–160.
- Collin-Dufresne, Pierre and Robert S. Goldstein (2002), “Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility,” *Journal of Finance*, 57, 1685–1730.
- Cooper, Ilan and Richard Priestley (2009), “Time-Varying Risk Premiums and the Output Gap,” *Review of Financial Studies*, 22, 2801.
- Dewachter, Hans and Marco Lyrio (2006), “Macro Factors and the Term Structure of Interest Rates,” *Journal of Money, Credit and Banking*, 38, 119–140.
- Diebold, Francis and Marc Nerlove (1989), “The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model,” *Journal of Applied Econometrics*, 4, 1–21.
- Diebold, Francis X. and Canlin Li (2006), “Forecasting the Term Structure of Government Bond Yields,” *Journal of Econometrics*, 130, 337–364.
- Diebold, Francis X., Monika Piazzesi, and Glenn D. Rudebusch (2005), “Modeling Bond Yields in Finance and Macroeconomics,” *American Economic Review*, 95, 415–420.



- Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba (2006), “The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach,” *Journal of Econometrics*, 131, 309–338.
- Duarte, Jefferson, Avraham Kamara, Stephan Siegel, and Celine Sun (2014), “The Systematic Risk of Idiosyncratic Volatility,” Available at SSRN: <https://ssrn.com/abstract=1905731> or <http://dx.doi.org/10.2139/ssrn.1905731>.
- Duffee, Gregory (2013), “Forecasting Interest Rates,” in “Handbook of Economic Forecasting,” (edited by Graham Elliot, Allan Timmermann), 2, chap. 7, 385–426, Elsevier.
- Duffee, Gregory R. (2011), “Information In (and Not In) the Term Structure,” *The Review of Financial Studies*, 24, 2895–2934.
- Duffie, Darrell and Rui Kan (1996), “A Yield-Factor Model of Interest Rates,” *Mathematical Finance*, 6, 379–406.
- Evans, Charles L. and David A. Marshall (1998), “Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory,” *Carnegie-Rochester Conference Series on Public Policy*, 49, 53–111.
- Ferson, Wayne E and Campbell R Harvey (1993), “The Risk and Predictability of International Equity Returns,” *Review of Financial Studies*, 6, 527–566.
- Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007), “The U.S. Treasury Yield Curve: 1961 to the Present,” *Journal of Monetary Economics*, 54, 2291–2304.
- Gurkaynak, Refet S. and Jonathan H. Wright (2012), “Macroeconomics and the Term Structure,” *Journal of Economic Literature*, 50, 331–367.
- Hansen, Bruce E. (1996), “Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis,” *Econometrica*, 64, 413–430.
- Hautsch, Nikolaus, Lada Kuj, and Peter Malec (2011), “The Merit of High-Frequency Data in Portfolio Allocation,” CFS Working Paper Series.
- Hautsch, Nikolaus, Lada M. Kyj, and Roel C. A. Oomen (2012), “A blocking and regularization approach to high-dimensional realized covariance estimation,” *Journal of Applied Econometrics*, 27, 625–645.
- Herskovic, Bernard, Bryan T. Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh (2014), “The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications,” Working Paper 20076, National Bureau of Economic Research.
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (1994), “Bayesian Analysis of Stochastic Volatility Models,” *Journal of Business & Economic Statistics*, 12, 371–389.

- Jagannathan, Ravi and Zhenyu Wang (1996), “The Conditional CAPM and the Cross-Section of Expected Returns,” *The Journal of Finance*, 51, 3–53.
- Jardet, Caroline, Alain Monfort, and Fulvio Pegoraro (2013), “No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth,” *Journal of Banking and Finance*, 37, 389–402.
- Joslin, Scott, Anh Le, and Kenneth J. Singleton (2013), “Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs,” *Journal of Financial Economics*, 109, 604–622.
- Joslin, Scott, Marcel Priebisch, and Kenneth J. Singleton (2014), “Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks,” *Journal of Finance*, 69, 1197–1233.
- Kalnina, Ilze and Kokouvi Tewou (2017), “Cross-sectional Dependence in Idiosyncratic Volatility,” Cahiers de recherche 08-2015, Centre interuniversitaire de recherche en économie quantitative, CIREQ.
- Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *The Review of Economic Studies*, 65, 361–393.
- Knotek, Edward S. and Saeed Zaman (2015), “Nowcasting U.S. Headline and Core Inflation,” Working Paper 1403, Federal Reserve Bank of Cleveland.
- Krippner, Leo (2012), “Modifying Gaussian Term Structure Models When Interest Rates Are Near the Zero Lower Bound,” Reserve Bank of New Zealand Discussion Paper Series DP2012/02, Reserve Bank of New Zealand.
- Krippner, Leo (2013), “A Tractable Framework for Zero Lower Bound Gaussian Term Structure Models,” Reserve Bank of New Zealand Discussion Paper Series DP2013/02, Reserve Bank of New Zealand.
- Lewellen, Jonathan and Stefan Nagel (2006), “The Conditional CAPM Does Not Explain Asset-pricing Anomalies,” *Journal of Financial Economics*, 82, 289–314.
- Li, Haitao and Feng Zhao (2006), “Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives,” *Journal of Finance*, 61, 341–378.
- Liu, Lily, Andrew Patton, and Kevin Sheppard (2015), “Does Anything Beat 5-Minute RV?” *Journal of Econometrics*, 187, 293–311.
- Ludvigson, Sydney C. and Serena Ng (2009), “Macro Factors in Bond Risk Premia,” *Review of Financial Studies*, 22, 5027.
- McCulloch, J Huston (1971), “Measuring the Term Structure of Interest Rates,” *Journal of Business*, 44, 19–31.

- McCulloch, J. Huston (1975), “The Tax-Adjusted Yield Curve,” *Journal of Finance*, 30, 811–830.
- Moller, Stig V. (2014), “GDP Growth and the Yield Curvature,” *Finance Research Letters*, 11, 1–7.
- Onatski, Alexei (2009), “Testing Hypotheses About the Number of Factors in Large Factor Models,” *Econometrica*, 77, 1447–1479.
- Pagan, Adrian (1984), “Econometric Issues in the Analysis of Regressions with Generated Regressors,” *International Economic Review*, 25, 221–247.
- Pitt, Michael K. and Neil Shephard (1999), “Time varying covariances: a factor stochastic volatility approach,” in “Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting,” (edited by J.M. Bernardo, A.P. Dawid, J.O. Berger and A.F.M Smith), 547–570, Oxford University Press, Oxford.
- Rebonato, Riccardo (2016), *Bond Pricing and Yield Curve Modelling: A Structural Approach*, Cambridge University Press.
- Rigobon, Roberto and Thomas M. Stoker (2007), “Estimation With Censored Regressors: Basic Issues,” *International Economic Review*, 48, 1441–1467.
- Rudebusch, Glenn D. (2009), “The Fed’s Monetary Policy Response to the Current Crisis,” *FRBSF Economic Letter*.
- Rudebusch, Glenn D., Brian P. Sack, and Eric T. Swanson (2007), “Macroeconomic Implications of Changes in the Term Premium,” *Federal Reserve Bank of St. Louis Review*, 89, 241–269.
- Rudebusch, Glenn D. and Tao Wu (2008), “A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy,” *Economic Journal*, 118, 906–926.
- Sheppard, Kevin and Wen Xu (2014), “Factor High-Frequency Based Volatility (HEAVY) Models,” Available at SSRN: <https://ssrn.com/abstract=2442230> or <http://dx.doi.org/10.2139/ssrn.2442230>.
- Shin, Minchul and Molin Zhong (2017), “Does Realized Volatility Help Bond Yield Density Prediction?” *International Journal of Forecasting*, 33, 373–389.
- Stock, James H. and Mark Watson (2003), “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, 41, 788–829.
- Vasicek, Oldrich A. and H. Gifford Fong (1982), “Term Structure Modeling Using Exponential Splines,” *Journal of Finance*, 37, 339–348.
- Watanabe, Kota and Tsutomu Watanabe (2014), “Estimating Daily Inflation Using Scanner

Data: A Progress Report,” University of Tokyo Price Project Working Paper Series 020, University of Tokyo, Graduate School of Economics.

Wu, Jing Cynthia and Fan Dora Xia (2016), “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound,” *Journal of Money, Credit and Banking*, 48, 253–291.