Federal Reserve Bank of New York
Staff Reports

Nonlinearity and Flight to Safety in the Risk-Return Trade-Off for Stocks and Bonds

Tobias Adrian
Richard Crump
Erik Vogt

Staff Report No. 723
April 2015
Revised November 2017

Nonlinearity and Flight to Safety in the Risk-Return Trade-Off for Stocks and Bonds

Tobias Adrian, Richard Crump, and Erik Vogt

Federal Reserve Bank of New York Staff Reports, no. 723
April 2015; revised November 2017
JEL classification: G01, G12, G17

Abstract

We document a highly significant, strongly nonlinear dependence of stock and bond returns on past equity market volatility as measured by the VIX. We propose a new estimator for the shape of the nonlinear forecasting relationship that exploits additional variation in the cross section of returns. The nonlinearities are mirror images for stocks and bonds, revealing flight to safety: expected returns increase for stocks when volatility increases from moderate to high levels, while they decline for Treasury securities. These findings provide support for dynamic asset pricing theories where the price of risk is a nonlinear function of market volatility.

Key words: flight to safety, risk-return trade-off, dynamic asset pricing, volatility, nonparametric estimation and inference, intermediary asset pricing, asset management

---

Adrian (corresponding author): International Monetary Fund (tadrian@imf.org). Crump: Federal Reserve Bank of New York. Vogt contributed to this paper while working at the Federal Reserve Bank of New York. The authors thank Michael Bauer, Tim Bollerslev, Andrea Buffa, John Campbell, Itamar Drechsler, Robert Engle, Eric Ghysels, Arvind Krishnamurthy, Ivan Shaliastovich, Kenneth Singleton, Allan Timmermann, Peter Van Tassel, Jonathan Wright, and three anonymous referees, as well as seminar and conference participants at Boston University, the Federal Reserve Banks of New York and San Francisco, the NYU Stern Volatility Institute, the NBER 2015 Summer Institute, and the 2016 AFA Annual Meeting for helpful comments and suggestions. Daniel Stackman, Rui Yu, and Oliver Kim provided outstanding research assistance.

1 Introduction

Investor flight-to-safety is pervasive in times of elevated risk (Longstaff (2004), Beber et al. (2009), Baele et al. (2013)). Economic theories of investor flight-to-safety predict highly nonlinear asset pricing relationships (Vayanos (2004), Weill (2007), Caballero and Krishnamurthy (2008), Brunnermeier and Pedersen (2009)). Such nonlinear pricing relationships are difficult to document empirically as the particular shape of the nonlinearity is model specific, and inference of nonlinear relationships presents econometric challenges.

In this paper, we document an economically and statistically strong nonlinear risk-return tradeoff by estimating the relationship between stock market volatility as measured by the VIX and future returns. The nonlinear risk-return tradeoff features evidence of flight-to-safety from stocks to bonds in times of elevated stock market volatility consistent with the above cited theories. The VIX strongly forecasts stock and bond returns up to 24 months into the future when the nonlinearity is accounted for, in sharp contrast to the insignificant linear relationship.

The nature of the nonlinearity in the risk-return tradeoffs for stocks and bonds are virtually mirror images, as can be seen in Figure 1 on the next page, estimated from a large cross-section of stocks and bonds. Both stock and bond returns have been normalized by their unconditional standard deviation in order to allow plotting them on the same scale. There are three notable regions that characterize the nature of the nonlinear risk-return tradeoff, defined by the VIX median of 18 and the VIX 99.3rd-percentile of 50. When the VIX is below its median of 18, both stocks and bonds exhibit a risk return tradeoff that is relatively insensitive to changes in the VIX. In the intermediate 18-50 percent range of the VIX, the nonlinearity is very pronounced: as the VIX increases above its unconditional median, expected Treasury returns tend to fall, while expected stock returns rise. This finding is consistent with a flight-to-safety from stocks to bonds, raising expected returns to stocks and compressing expected returns to bonds. For levels of the VIX above 50, which has only occurred in the aftermath of the Lehman failure, this logic reverses, and a further increase in the VIX is associated with lower stock and higher bond returns. The latter finding for very high values of the VIX likely reflects the fact that severe financial crises are followed by abysmal stock returns and aggressive interest rate cuts, due to a collapse in real activity, thus reflecting changes in cash flow expectations (see Campbell et al. (2013)).

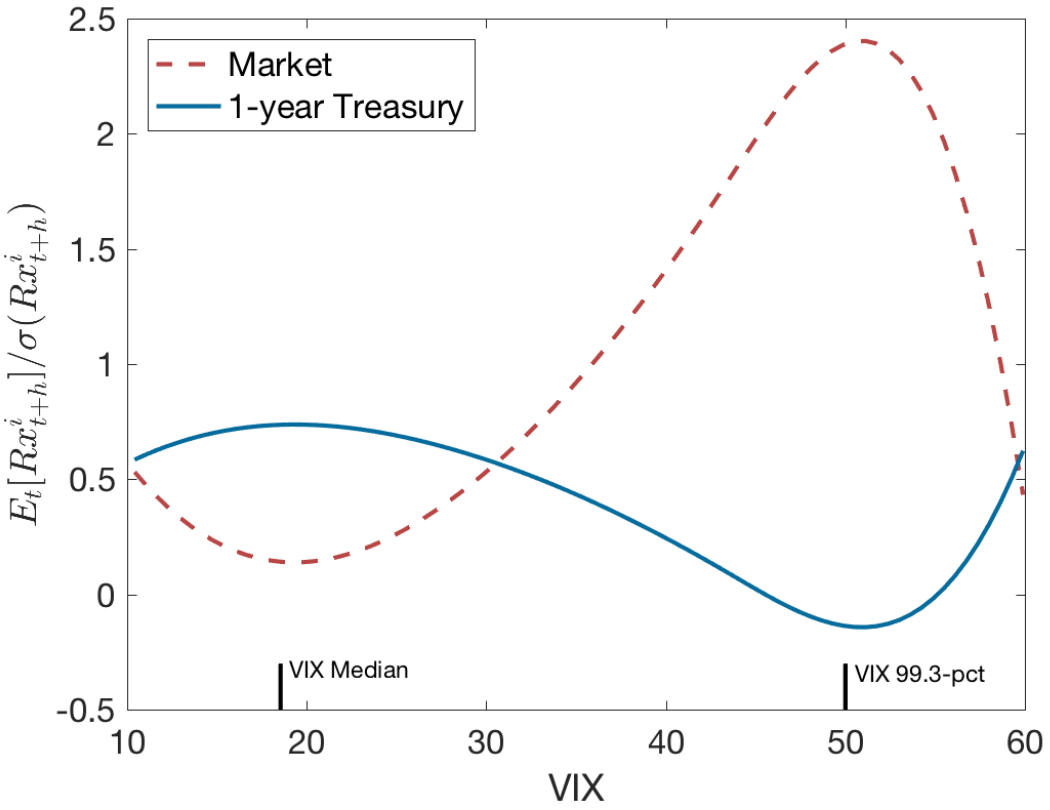


Figure 1: This figure shows the relationship between the six month cumulative equity market return (CRSP value-weighted US equity market portfolio) and the six month lag of the VIX in red, as well as the relationship between the six month cumulative 1-year Treasury return (CRSP 1-year constant maturity Treasury portfolio) and the six month lag of the VIX in blue. Both nonlinear relationships are estimated using reduced rank sieve regressions on a large cross-section of stocks and bonds. The y-axis is expressed as a ratio of returns to the full sample standard deviation. The x-axis shows the VIX.

What is most notable is that a linear regression using the VIX does not forecast stock or bond returns significantly at any horizon. Nonlinear regressions, on the other hand, do forecast stock and bond returns with very high statistical significance and reveal the striking mirror image property of Figure 1. We study the nature of the nonlinearity and mirror image property in a variety of ways, using kernel regressions, polynomial regressions, as well as nonparametric sieve regressions. In all cases and on subsamples, we find pronounced nonlinearity within risky assets and reversed nonlinearities for safe assets, in terms of both statistical and economic magnitudes.

In order to estimate the shape of the nonlinearity in a robust way, we propose a novel

way to nonparametrically estimate the shape using a reduced-rank sieve regression on a large cross-section of stock and bond returns. We specify a nonlinear forecasting function $\phi_h(v)$ according to the following set of equations

$$Rx^i_{t+h} = a^i_h + b^i_h \cdot \phi_h(vix_t) + \varepsilon^i_{t+h}, \qquad i = 1, \ldots, n,$$

where h denotes the forecasting horizon and i refers to the individual stock and bond portfolios and Rx are excess returns. The nonlinearity of the function $\phi_h(v)$ is highly significant, and its forecasting power is very strong. Importantly, when we estimate $\phi_h(v)$ separately for stocks and bonds, we obtain statistically indistinguishable functions (up to an affine transformation).

A major advantage of estimating $\phi_h(v)$ from a large panel of stock and bond returns is that it exploits additional cross-sectional variation unavailable in the univariate regressions that are typical in the return forecasting literature. The algebra for the estimator can be described intuitively in two stages. In the first stage, returns to each asset are regressed in the time series on lagged sieve expansions of the VIX. In the second stage, the rank of the matrix of forecasting coefficients is reduced using an eigenvalue decomposition, and only a rank one approximation is retained (see Adrian et al. (2015) for a related derivation). This is a dimensionality reduction that is optimal when errors are conditionally Gaussian and the number of regressors are fixed. The resulting factor $\phi_h(v)$ is a nonlinear function of volatility and is the best common predictor for the whole cross section of stock and bond returns.

In addition to the new estimator, we also introduce asymptotically valid inference procedures for four hypotheses of interest, which may be implemented using standard critical values. The first is a joint test of significance for whether the whole cross-section of test assets loads on ϕ_h. The second tests the null that ϕ_h does not predict excess returns for a specific asset, while allowing it to predict for other assets. The third test is a comparison of whether the function $\phi_h(\cdot)$ is different from zero at any fixed value $\bar{v}$, generating pointwise confidence intervals for the unknown function. The fourth is a specification test for the rank-one restriction that the same nonlinear function of volatility ϕ_h drives expected stock and bond returns. To conduct inference when estimating multi-horizon returns with overlapping data we extend the "reverse regression" approach of Hodrick (1992) to our reduced-rank, nonparametric setting.

Our finding that the VIX forecasts stock and bond returns in a nonlinear fashion is robust to the inclusion of standard predictor variables such as the dividend yield, the BAA/10-year Treasury default spread, the 10-year/3-month Treasury term spread, and the variance risk premium. Furthermore, we show that the nonlinear relationship is highly significant for the 1990-2007 sample which excludes the 2008-09 financial crisis. Importantly, the shape of the nonlinearity in the 1990-2007 and the 1990-2014 sample resemble each other closely, even though the tail events in those samples are distinct. We also verify that Treasury returns are forecasted only by a nonlinear function of the VIX, not the Treasury implied volatility as measured by the MOVE. The latter result suggests that pricing of risk is proxied by the VIX as a common forecasting variable for stocks and bonds. These findings thus point towards joint dynamic asset pricing of stocks and bonds, as explored in linear settings by Mamaysky (2002), Bekaert et al. (2010), Lettau and Wachter (2011), Ang and Ulrich (2012), Koijen et al. (2013), and Adrian et al. (2015).

Our main result that expected returns to stocks, Treasury bonds, and credit returns are forecast by a common nonlinear function $\phi_h(v)$ suggests that this function is a price of risk variable in a dynamic asset pricing model. The sieve reduced rank regression estimator restricts expected returns of each asset i to be an affine transformation of $\phi_h(v)$ with intercept a_h^i and slope b_h^i. Asset pricing theories predict these coefficients to be determined by risk factor loadings. We take this prediction to the data, estimating the beta representation of a dynamic asset pricing kernel that features the market return, the one year Treasury return, and innovations to $\phi_h(v)$ as cross-sectional pricing factors, and $\phi_h(v)$ as price of risk variable. We show that this asset pricing model performs well in pricing the cross-section of stock, bond, and credit portfolios, and that there is a tight cross-sectional relationship between the forecasting slopes b_h^i and the risk factor loadings.

The dynamic asset pricing results indicate that the pricing of risk over time is related to the level of volatility in a nonlinear fashion. A number of alternative theories are compatible with such a finding, including 1) flight-to-safety theories due to redemption constraints on asset managers, 2) macro-finance models with financial intermediaries, and 3) representative agent models with a) habit formation and b) long-run risks.

Among asset management pricing theories, our findings are particularly in line with the theory of Vayanos (2004), where asset managers are subject to funding constraints that (endogenously) depend on the level of market volatility. When volatility increases, the likelihood

of redemptions rises, leading to a decline in the risk appetite of the asset managers. Increases in volatility generate flight-to-safety as managers attempt to mitigate the impact of higher volatility on redemption risk by allocating more to relatively safe assets. In equilibrium, the dependence of risk appetite on volatility generates expected returns with features that are qualitatively similar to our estimated function $\phi_h(v)$. Furthermore, the Vayanos (2004) theory gives rise to a dynamic asset pricing kernel that would predict that the forecasting slope b_h^i is cross-sectionally related to risk factor loadings, as explained earlier.

We present direct evidence in favor of the flight-to-safety mechanism related to asset managers by estimating the shape of global mutual fund flows' dependence on the VIX. The shape of the function resembles the shape of the return forecasting function $\phi_h(v)$ closely for the range of the VIX from 18 to 50. Furthermore, the signs of the loadings on the flow function exhibits evidence of flight-to-safety.

Our findings are also closely linked to intermediary asset pricing theories. In Adrian and Boyarchenko (2012), intermediaries are subject to value at risk (VaR) constraints that directly link intermediaries' risk taking ability to the level of volatility. Prices of risk are a nonlinear function of intermediary leverage, which has a one-to-one relationship to the level of volatility. A similar nonlinear risk-return tradeoff is also present in the theories of He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014). We also discuss the extent to which our findings are compatible with the habit formation model of Campbell and Cochrane (1999) and the long-run risk model of Bansal and Yaron (2004).

The remainder of the paper is organized as follows. Section 2 presents evidence of the nonlinearity in the risk-return tradeoff using polynomial, spline, and kernel regressions for stocks and bonds. Section 3 develops our novel sieve reduced rank regression estimator and tests of the stability of the nonlinear shape across asset classes and over time. Section 4 provides an economic analysis of the flight-to-safety feature in the nonlinear risk-return tradeoff, establishing a link to dynamic asset pricing models and relating our findings to theories of flight-to-safety. Furthermore, we discuss the theoretical literature in light of our findings in detail. Section 5 concludes. Additional results are available in a Supplementary Appendix (hereafter, Appendix).

2 Motivating Evidence on the Nonlinear Risk-Return Trade-off

This section presents initial evidence from univariate predictive regressions of stock and bond returns on VIX polynomials, motivating the presence of nonlinearities in expected returns (subsection 2.1). We find evidence of a mirror image property: the shape of the nonlinearity of bonds mirrors inversely the shape of the nonlinearity of stocks. We document that this mirror image property not only holds for polynomial regressions, but also for nonparametric estimators such as kernel regressions or sieve regressions (subsection 2.2). This motivates us to develop a panel estimation method for the shape of the nonlinearity that allows each asset return to be an affine function of a common nonlinear function of market volatility. The theory and our main empirical evidence using this new estimator are presented in Section 3.

2.1 Suggestive Univariate Evidence from VIX Polynomials

To demonstrate the gains that can be obtained by allowing for nonlinearities, we estimate the linear regression

$$Rx^i_{t+h} = a^i_h + b^i_h\,(vix_t) + \varepsilon^i_{t+h}, \tag{2.1}$$

the polynomial regression

$$Rx^i_{t+h} = a^i_h + b^i_h(vix_t) + c^i_h\,(vix_t)^2 + d^i_h\,(vix_t)^3 + \varepsilon^i_{t+h}, \tag{2.2}$$

$$Rx^i_{t+h} = a^i_h + b^i_h(move_t) + c^i_h\,(move_t)^2 + d^i_h\,(move_t)^3 + \varepsilon^i_{t+h}, \tag{2.3}$$

and the augmented polynomial regressions

$$\begin{aligned} Rx^i_{t+h} = {} & a^i_h + b^i_h(vix_t) + c^i_h\,(vix_t)^2 + d^i_h\,(vix_t)^3 \\ & + bm^i_h(move_t) + cm^i_h\,(move_t)^2 + dm^i_h\,(move_t)^3 + \mathbf{f}^{i\prime}\,\mathbf{z}_t + \varepsilon^i_{t+h} \end{aligned} \tag{2.4}$$

separately for i representing equity market or Treasury excess returns. Here, $\mathbf{z}_t$ is a vector of predictors, and $Rx^i_{t+h} = (12/h)[(r^i_{t+1} - r^f_t) + \cdots + (r^i_{t+h} - r^f_{t+h-1})]$ denotes the continuously compounded h-month holding period return of asset i in excess of the one-month riskfree

rate r_t^f (at an annual rate). For comparison, we include both equity market option-implied volatility (VIX) as well as its Treasury counterpart (MOVE).

Table 1 reports t-statistics for the coefficients of regressions (2.1) through (2.4), as well as p-values for the joint hypothesis test under the null of no predictability (H_0: $Rx_{t+h}^i = a_h^i + \varepsilon_{t+h}^i$). While the top panel of Table 1 shows results where Rx_{t+h}^i represents excess returns on 1-year maturity US Treasuries for forecasting horizons $h = 6$, 12, and 18 months, the bottom panel reports analogous results for excess returns on the CRSP value-weighted US equity market portfolio. Since our sample represents monthly observations from 1990:1 to 2014:9, we follow Ang and Bekaert (2007) and compute standard errors using the Hodrick (1992) correction for multihorizon overlapping observations.[1]

The most striking features of Table 1 are the predictive gains obtained by simply augmenting the VIX with squares and cubes of itself. For 1-year Treasuries, the t-statistic on the VIX coefficient jumps from 1.91 in the linear regression to 4.13 when squares and cubes of VIX are included at the $h = 6$ month forecasting horizon, from 1.86 to 3.60 at the $h = 12$ month horizon, and from 1.13 to 3.21 at the $h = 18$ month horizon. Moreover, the coefficients on the squares and cubes of the VIX are themselves highly statistically significant – an effect that persists even with the inclusion of standard forecasting variables like the BAA/10-year Treasury default spread (DEF), the variance risk premium (VRP) following Bollerslev et al. (2009), the 10-year/3-month Treasury Term Spread (TERM), and the (log) dividend yield (DY). The p-values also suggest strong evidence for the joint predictive content of the VIX polynomial.

Similar gains are obtained for equity market excess returns. While we can confirm the findings of Bekaert and Hoerova (2014) and Bollerslev et al. (2013) that the VIX itself does not (linearly) forecast excess stock market returns, we document marked improvements in predictability when polynomials of the VIX are included: p-values for the joint test of no predictability drop from 0.316 for the linear regression case to 0.007 for the VIX polynomial case at the $h = 6$ month horizon, from 0.460 to 0.032 at the $h = 12$ month horizon, and from 0.439 to 0.088 at the $h = 18$ month horizon. More formally, a direct test of linearity (the joint null hypothesis that the coefficients on the VIX squared and cube terms are zero) is

[1] Ang and Bekaert (2007) strongly argue in favor of Hodrick (1992) standard errors in return predictability regressions with overlapping observations, as these exhibit substantially better size control than Newey and West (1987) standard errors in the same setting. This argument is confirmed in the simulation evidence of Wei and Wright (2013) as well as the simulations we present in the Appendix.

strongly rejected in favor of higher order polynomial terms.[2] For the short forecast horizon $h = 6$, we note further that evidence for the VIX polynomial's predictability remains even after the inclusion of other forecasting variables, while for longer horizons, the predictive content of VIX polynomials appears to subside.

The MOVE is an analogous portfolio of yield curve weighted options written on Treasury futures. To the extent that some form of segmentation between Treasury and equity markets could give rise to separate pricing kernels for bonds and stocks, one may surmise that excess returns in either market reflect compensation for exposure to different types of volatility or uncertainty risk. Somewhat surprisingly, however, Table 1 shows that this is not the case. Whereas the VIX polynomials exhibit t-statistics at times above five, t-statistics on MOVE polynomials coefficients are struggling to exceed one. The p-values show that regressions on the MOVE cannot be statistically distinguished from regressions on a constant.

A final noteworthy feature of Table 1 are the signs on the constant and coefficients of the VIX polynomials for Treasuries compared to equities. While the coefficients on the VIX, VIX^2, and VIX^3 alternate as $(\hat{b}_h^i > 0)$, $(\hat{c}_h^i < 0)$, and $(\hat{d}_h^i > 0)$ for $i =$ Treasuries, they are exactly the opposite for equities across all forecasting horizons. The same is true for the intercepts: while the intercepts in the Treasury regressions are all negative when VIX polynomials are included, the intercepts for the equity regressions are all positive. In contrast, the linear VIX specification appears to make no signed distinction between Treasury and equity market excess returns and is instead reporting a statistically insignificant relationship between the VIX and future excess returns across all horizons.

The findings are also robust to a range of forecast horizons: Figure 2 plots p-values by h ranging from 1 to 24 months for both the linear regression (2.1) (blue line) and polynomial regression (2.2) (red line). Several noteworthy features emerge from the figure. First, the predictive gains that result from allowing for VIX nonlinearities, as measured by the distance between the blue and red lines, are substantial for all horizons h, Treasury maturities, and equity returns. In particular, the polynomial specification dominates the linear one across all Treasury and equity excess returns for horizons $h = 3, \ldots, 24$. Second, the null of lack of predictability for the polynomial specification is strongly rejected at the 5% level (the

[2]Strictly speaking, the Hodrick (1992) standard errors are asymptotically valid under the null of no predictability. However, it is straightforward to show that they extend to hypotheses of weak (local-to-zero) predictability. See the Appendix for further discussion.

red line falls below the dashed line) for short-maturity Treasuries and for a wide range of forecast horizons h. Third, as Treasury maturities lengthen, VIX polynomial predictability begins to wane as the red line gradually shifts upward and becomes insignificant for 10-year Treasuries. Fourth, for equity market returns, the null of no predictability is rejected at the 5% level for forecast horizons $h = 3, \ldots, 15$ months.

As a robustness check, we examine to what extent the VIX's predictive results are driven by the 2008 financial crisis. Figure 3 repeats the exercise of Figure 2 with a sample spanning only 1990:1 to 2007:7. On this pre-crisis sample, 1-year Treasuries are still strongly predicted by VIX polynomials across horizons $h = 3, \ldots, 24$ while outperforming the linear specification in all panels. As Treasury maturities increase, VIX polynomial predictability appears stronger than even in the full sample. 10-year Treasuries, in particular, are showing signs of predictability at longer horizons, which contrasts with the result on the sample ending in 2014. On the other hand, equity market return predictability diminishes slightly and rejects the null of no predictability only at longer horizons. In relative terms, however, we again note that the gains from allowing the VIX to nonlinearly predict returns are substantial compared to the linear specification across both Treasuries and equities.

We will present more details on the economic interpretation of the shape of the nonlinear forecasting relationships in Section 4. However, we can characterize the broad findings in the following way. When the VIX is below its median of 18, both stocks and bonds are relatively insensitive to changes in the VIX. In the intermediate 18-50 percent range of the VIX, the nonlinearity is very pronounced: as the VIX increases above its unconditional median, expected Treasury returns tend to fall, while expected stock returns rise. This finding is consistent with a flight-to-safety from stocks to bonds, raising expected returns to stocks and compressing expected returns to bonds. For levels of the VIX above 50 which has only occurred during the 2008 crisis, this logic reverses, and a further increase in the VIX is associated with lower stock and higher bond returns. The latter finding for very high values of the VIX likely reflects the fact that severe financial crises are followed by abysmal stock returns and aggressive interest rate cuts, due to a collapse in real activity, thus reflecting changes in cash flow expectations (see Campbell et al. (2013)). We discuss the economic interpretation in more detail in Section 4.

Our findings are consistent with economic theories that suggest a risk-return tradeoff in the pricing of risky assets (e.g. Sharpe (1964), Merton (1973), Ross (1976)). An unexpected

increase in riskiness should be associated with a contemporaneous drop in the asset price and an increase in expected returns. While the first half of this logic is readily verified—asset returns and volatility changes tend to be strongly negatively correlated contemporaneously—the latter prediction has been much harder to prove. Our results so far indicate that there is a strongly significant positive risk return tradeoff in the data, once one allows for nonlinearity. Previous studies that have documented a positive risk return tradeoff in the time series have typically relied on the use of mixed frequency data (see Ghysels et al. (2005)), cross-sectional approaches (see Guo and Whitelaw (2006), Bali and Engle (2010)), or very long historical data (see Lundblad (2007)). A simple regression of asset returns on lagged measures of risk such as the VIX or realized volatility typically do not yield any statistically significant relationship for the risk-return tradeoff (e.g. Bekaert and Hoerova (2014) and Bollerslev et al. (2013)). In contrast, we show that there is a strong nonlinear relationship between stock and bond returns and lagged equity market volatility.[3]

2.2 Motivation of Sieve Reduced Rank Regressions

The preceding results showed that polynomials – rather than linear functions – of the VIX have important predictive power for future excess stock and bond returns. But instead of accepting a cubic VIX polynomial as the true data generating process for excess returns, we conjecture that the polynomials provide an approximation to some general nonlinear relationship between equity implied volatility and future excess stock and bond returns. To test this conjecture, we nonparametrically estimate the relationship between the VIX and future excess stock and bond returns via the method of sieves, which facilitates intuitive comparisons to polynomial regressions. To motivate our nonparametric sieve estimation framework, fix asset i and forecast horizon h and consider

$$Rx^i_{t+h} = \phi^i_h(v_t) + \varepsilon^i_{t+h}, \tag{2.5}$$

where $v_t = vix_t$. Equation (2.5) effectively replaces the polynomial $(a^i_h + b^i_h v_t + c^i_h v_t^2 + d^i_h v_t^3)$ from before with an unknown function $\phi^i_h(v_t)$.

[3] Ghysels et al. (2014) presents a similar result using a regime switching approach. One regime features high volatility with a negative risk-return relation, whereas the risk-return relation is positive in the second regime.

To estimate the function $\phi_h^i(\cdot)$ nonparametrically, we assume that $\phi_h^i \in \Phi$, where Φ is a general function space of sufficiently smooth functions. In practice, estimation over the entire function space Φ is challenging because it is infinite dimensional. In settings like these, the method of sieves (e.g., Chen (2007)) proceeds instead by estimation on a sequence of m-dimensional approximating spaces $\{\Phi_m\}_{m=1}^{\infty}$. We say that $\{\Phi_m\}_{m=1}^{\infty}$ is a valid sieve for Φ if it is nested (i.e. $\Phi_m \subset \Phi_{m+1} \subset \cdots \subset \Phi$) and eventually becomes dense in Φ (i.e. $\cup_{m=1}^{\infty}\Phi_m$ is dense in Φ). Letting $m = m_T \to \infty$ slowly as the sample size $T \to \infty$, the idea then is that the spaces Φ_{m_T} grow and increasingly resemble Φ, so that the least squares solution

$$\hat{\phi}_{m_T,h}^i \equiv \arg\min_{\phi \in \Phi_{m_T}} \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(Rx_{t+h}^i - \phi(v_t)\right)^2 \tag{2.6}$$

converges to the true unknown function $\phi_h^i \in \Phi$ in (2.5) in some suitable sense.

For our choice of Φ_m we use the space spanned by linear combinations of m B-splines of the VIX. Thus, any element $\phi_m \in \Phi_m$ may be written as $\phi_m(v) = \sum_{j=1}^{m}\gamma_j \cdot B_j(v)$ where $\gamma_j \in \mathbb{R}$ for $j = 1, \ldots m$, v is a value in the support of the VIX, and B_j is the j^{th} B-spline (see the Appendix for further details). B-splines have a number of appealing features such as well-established approximation properties and substantial analytical tractability. This is because for fixed m, the solution to the least squares problem (2.6) is simply the OLS estimator on B-spline coefficients $\gamma^h = \left(\gamma_1^h, \ldots, \gamma_m^h\right)'$:

$$\hat{\gamma}^h = \left(X_m X_m'\right)^{-1} X_m \mathrm{Rx}^i, \tag{2.7}$$

where $\mathrm{Rx}^i = \left(Rx_{1+h}^i, \ldots, Rx_{T+h}^i\right)'$ and X_m is the $(m \times T)$ matrix of predictors with j^{th} row equal to $[B_j(\mathrm{VIX}_1), \ldots, B_j(\mathrm{VIX}_T)]$. Therefore, for fixed m, the solution to (2.6) becomes

$$\hat{\phi}_{m,h}^i(v) = \sum\nolimits_{j=1}^{m}\hat{\gamma}_j^h \cdot B_j(v). \tag{2.8}$$

Equation (2.8) makes clear that the simple polynomial specification introduced in the previous subsection may be thought of as an alternative nonparametric estimate of $\phi_h^i(\cdot)$ using polynomial basis functions, v^j, instead of B-splines $B_j(v)$. However, this approach was informal in the sense that the choice of the maximum degree of polynomial was not made with relation to the sample size. Instead, we think of the number of basis functions $m = m_T$ as

growing to infinity at some appropriate rate that depends on the sample size T.[4]

The top half of Figure 4 shows various estimates for $\phi_h^i(v)$, where $h = 6$ and i refers to either 1-year Treasury excess returns (blue line) or equity market excess returns (red line) over the full sample period from 1990:1 to 2014:9. In the left graph, we show the cross-validated sieve B-spline estimates $\hat{\phi}^i_{m_T,h}(v)$ (equation (2.8)), whereas the middle graph shows the functional form implied by the simple polynomial specification of the previous section. The estimated functional forms in both the left and middle graphs are very similar, implying that the cubic polynomial choice in the previous section provided a reasonable first pass at investigating the nonlinear relationship. As a further robustness check, the right panel shows the estimated function based on a nonparametric kernel regression, which gives a qualitatively similar impression of $\phi_h^i(v)$ for stocks and bonds.

Figure 4 also demonstrates another noteworthy empirical regularity. If we compare $\hat{\phi}^i_h(v)$ using either equity returns or bond returns as test assets it appears that they are related by a simple scale and reflection transformation. This could already be deduced from the alternating coefficient signs from the polynomial regressions, and is now additionally confirmed with two alternative nonparametric estimators. Moreover, the bottom panel of Figure 4 shows that the mirror image relationship between $\phi_h^{\text{Treasuries}}(v)$ and $\phi_h^{\text{Stocks}}(v)$ existed prior to the 2008 financial crisis and is therefore not an artifact of a few extreme observations. Instead, the crisis is merely helpful in identifying $\phi_h^i(v)$ for large v.

We interpret this finding as strongly suggestive that equity market and Treasury excess returns load on a *common* $\phi_h(v)$ function, up to location, scale, and reflection transformations. In this case, we show next that $\phi_h(\cdot)$ could then be estimated jointly across assets rather than estimating univariate regressions equation by equation, as was done above. This has the benefit of allowing Treasury returns across multiple maturities as well as the equity market excess returns to jointly inform the estimate of the common $\phi_h(\cdot)$, thereby exploiting information in the cross-section of asset returns.

[4]In particular, it can be shown that m behaves very much like a bandwidth parameter in that it is chosen to optimally trade off notions of bias and variance: heuristically, if m is too small, Φ_m is too small relative to Φ, which causes bias, and if m is too big, it results in overfitting. In the remainder of the paper, we follow the existing literature in sieve estimation and choose m_T by cross-validation (see, e.g., Li and Racine (2007)). We use for our results a mean-square forecast error (MSFE) cross validation procedure. Full details are provided in the Appendix.

3 Main Results

We start this section by introducing our main panel estimation method, which exploits the common nonlinearity revealed in expected stock and bond returns (subsection 3.1). We label this panel estimation method sieve reduced rank regressions. This method exploits cross-sectional variation in excess returns to estimate the shape of the nonlinearity. We use the sieve reduced rank regressions to document that the nature of the nonlinearity is reversed when the excess return to be predicted is the equity market versus Treasuries, pointing towards flight-to-safety from stocks to bonds as equity market volatility rises above its unconditional median (subsection 3.2). We also document the robustness of the predictive relationships across forecasting horizons and Treasury maturities. Strikingly, we show that the shape of the nonlinearity is statistically indistinguishable whether it is extracted from only bonds or only stocks. We also present results for broader cross sections, including industry sorted portfolios, maturity sorted Treasury returns, and credit returns (subsection 3.3). We then provide robustness checks in the form of out-of-sample forecasting performance (subsection 3.4).

3.1 Derivation of Sieve Reduced Rank Regressions

In this subsection, we formalize the intuition of a common volatility function $\phi_h(v)$ by introducing a reduced-rank, sieve-based procedure which produces a nonparametric estimate of $\phi_h(v)$ under only weak assumptions. The novelty of our approach is that we use cross-sectional information across assets to better inform our estimate of this function. We label our approach "sieve reduced-rank regression" (SRR regression) as it combines the cross-sectional restrictions implied by a reduced-rank assumption with the flexibility of a nonparametric sieve estimator. We will see that the estimator is conveniently available in closed form and hypothesis tests rely on standard critical values.

Suppose we observe excess returns on $i = 1, \ldots, n$ assets that follow

$$Rx^i_{t+h} = a^i_h + b^i_h \cdot \phi_h\left(v_t\right) + \varepsilon^i_{t+h}. \tag{3.1}$$

Here, a^i_h and b^i_h are asset-specific shift and scale parameters, $\phi_h\left(\cdot\right)$ is the same for all assets, and $v_t = vix_t$. This specification can be compared with equation (2.5), which held that

$Rx^i_{t+h} = \phi^i_h(v_t) + \varepsilon^i_{t+h}$. Thus in the univariate regressions from the previous section, $\phi^i_h(v_t)$ was estimated separately for each asset i, with no cross-asset restrictions imposed. In contrast, the specification (3.1) implies that the same function $\phi_h(v_t)$ forecasts returns across assets, which amounts to the restriction $\phi^i_h(v_t) = a^i_h + b^i_h \cdot \phi_h(v_t)$. We will provide a formal test of this restriction later in this section.

If we take the same approach as in the univariate specification we can rewrite this equation as

$$Rx^i_{t+h} = a^i_h + b^i_h(\gamma_h' X_{m,t}) + f^i_h z_t + \tilde{\varepsilon}^i_{t+h} \tag{3.2}$$

where

$$\tilde{\varepsilon}^i_{t+h} = \varepsilon^i_{t+h} + b^i_h \cdot (\phi_h(v_t) - \gamma_h' X_{m,t}). \tag{3.3}$$

$\tilde{\varepsilon}^i_{t+h}$ in equation (3.3) is composed of two terms. The first term is the error term from the original regression equation. The second term represents the approximation error of the true nonlinear function and the best approximation from the space Φ_m. As m grows with the sample size this approximation error vanishes in the appropriate sense. Finally, the z_t terms allow us to consider additional predictors.

If we stack equation (3.2) across n assets we obtain

$$\mathrm{Rx}_{t+h} = \mathbf{a}_h + \mathbf{A}_h X_{m,t} + \mathbf{F}_h Z_t + \tilde{\varepsilon}_{t+h}, \qquad \mathbf{A}_h = \mathbf{b}_h \gamma_h' \tag{3.4}$$

where $\mathbf{a}_h = (a^1_h, \ldots, a^n_h)'$, $b_h = (b^1_h, \ldots, b^n_h)'$, $\mathrm{Rx}_{t+h} = \left(Rx^i_{t+h}, \ldots, Rx^n_{t+h}\right)'$ and $\tilde{\varepsilon}_{t+h} = \left(\tilde{\varepsilon}^1_{t+h}, \ldots, \tilde{\varepsilon}^n_{t+h}\right)'$. For any fixed m, equation (3.4) is a reduced-rank regression where $\mathbf{A}_h$ is assumed to be of rank one.[5] The parameters $(\mathbf{a}_h', \mathbf{b}_h', \gamma_h', \mathbf{F}_h')'$ may be estimated in closed form. However, in order to separately identify $\mathbf{a}_h$ and $\mathbf{b}_h$ additional restrictions must be imposed. In our empirical analysis we impose the normalization $\phi_h(0) = 0$ and $b^1_h = b^{\mathrm{MKT}}_h = 1$. The first restriction allows us to identify the constant term for each asset, while the second implies that the market return is our reference asset.

To describe the estimation procedure, let $\hat{\mathbf{a}}_{h,\mathrm{ols}}$ $(n \times 1)$, $\hat{\mathbf{A}}_{h,\mathrm{ols}}$ $(n \times m)$ and $\hat{\mathbf{F}}_{h,\mathrm{ols}}$ $(n \times p)$ be the stacked OLS estimates and W a symmetric, positive-definite weight matrix. In our

[5] See Reinsel and Velu (1998) for a general introduction. Examples of parametric reduced-rank regressions are systems-based cointegration analysis (see e.g. Johansen (1995)), beta representations of dynamic asset pricing models (see e.g. Adrian et al. (2013, 2015)), and bond return forecasting (see e.g. Cochrane and Piazzesi (2008)).

empirical application, we set W to a diagonal matrix that scales excess returns by the inverse of their standard deviation to avoid overweighting high-variance assets in the estimation. Then,

$$\hat{\mathbf{b}}_h = \frac{\tilde{\mathbf{b}}_h}{\tilde{b}_h^1}, \qquad \hat{\gamma}_h = \tilde{\gamma}_h \cdot \tilde{b}_h^1, \qquad \left[\hat{\mathbf{a}}_h \;\vdots\; \hat{\mathbf{F}}_h\right] = \left[\hat{\mathbf{a}}_{h,\text{ols}} \;\vdots\; \hat{\mathbf{F}}_{h,\text{ols}}\right] + \left(\hat{\mathbf{A}}_{h,\text{ols}} - \tilde{\mathbf{b}}_h \tilde{\gamma}_h'\right) X_m Z' \left(ZZ'\right)^{-1},$$

where $Z = (Z_1, Z_2, \ldots, Z_T)$, $\tilde{\mathbf{b}}^h = W^{1/2}L$, $\tilde{\gamma}^h = \hat{\mathbf{A}}_{h,\text{ols}}' W^{-1} \hat{\mathbf{b}}_h$ and L is the eigenvector associated with the maximum eigenvalue of the matrix $W^{-1/2}\hat{\mathbf{A}}_{h,\text{ols}} \left(X_m M_Z X_m'\right) \hat{\mathbf{A}}_{h,\text{ols}}' W^{-1/2}$ where $M_Z = I_T - Z\left(ZZ'\right)^{-1} Z'$. If it were the case that $\tilde{\varepsilon}_{t+h} \sim_{iid} \mathcal{N}(0, W)$ and m was fixed, then $\left(\hat{\mathbf{a}}_h', \hat{\mathbf{b}}_h', \hat{\gamma}_h', \text{vec}\left(\hat{\mathbf{F}}_h\right)'\right)'$ would be the maximum likelihood estimates of $\left(\mathbf{a}_h', \mathbf{b}_h', \gamma_h', \text{vec}\left(\mathbf{F}_h\right)'\right)'$.

In this paper the first three hypotheses of interest are:

$$\begin{array}{ll} \mathbb{H}_{1,0} : \forall v \in \mathcal{V},\ \phi_h(v) = 0 & \mathbb{H}_{1,A} : \exists v \in \mathcal{V} \text{ s.t. } \phi_h(v) \neq 0 \\ \mathbb{H}_{2,0} : \forall v \in \mathcal{V},\ b_h^j \phi_h(v) = 0 & \mathbb{H}_{2,A} : \exists v \in \mathcal{V} \text{ s.t. } b_h^j\ \phi_h(v) \neq 0 \\ \mathbb{H}_{3,0} : \phi_h\left(\bar{v}\right) = 0 & \mathbb{H}_{3,A} : \phi_h\left(\bar{v}\right) \neq 0 \end{array} \tag{3.5}$$

The first hypothesis is a joint test of significance for whether the whole cross-section of test assets jointly loads on ϕ_h. For example, using equation (3.1), under $\mathbb{H}_{1,0}$, returns would be invariant to the level of the VIX (i.e., would be characterized by a horizontal line). If we were in the parametric case then this hypothesis could be written more simply as a null hypothesis that $\mathbf{A}_h = \mathbf{0}$. Instead, since we are in a nonparametric setting, we must formulate the hypothesis in terms of the unknown function $\phi_h(\cdot)$. The second hypothesis tests the null that ϕ_h does not predict excess returns Rx_{t+h}^j of asset j, while allowing it to predict another asset $i \neq j$. This test replaces t-tests on the loadings b_h^j because the scale of $\mathbf{b}_h$ cannot be determined separately from the scale of ϕ_h, which prompted our normalization $b_h^1 = 1$. This means, in particular, that a test of $b_h^1 = 0$ cannot be conducted, motivating our test on the product $b_h^j\ \phi_h$. In finite samples when the number of sieve expansion terms is fixed at some m, we show below that $\mathbb{H}_{1,0}$ and $\mathbb{H}_{2,0}$ may be tested using the standard χ^2 test.[6] This represents an additional convenient aspect of the sieve-based nonparametric procedure, since it allows us to test hypotheses about predictability in effectively the same

[6]The use of a χ^2 test is a small sample correction. See Crump et al. (2008) and the references therein.

way as a parametric joint test of significance. Finally, the third hypothesis is a comparison of whether the function $\phi_h(\cdot)$ is different from zero at a fixed value $\bar{v}$. By inverting a test of this hypothesis for different values of $\bar{v}$ we are able to construct pointwise confidence intervals for the unknown function.

In the following, we characterize the limiting distributions of the test statistics associated with each of these three hypotheses.

Theorem 1. *Under regularity conditions given in the Appendix*

$$
\begin{aligned}
&(i) \quad \hat{\mathcal{T}}_1 \to_{d,\mathbb{H}_{1,0}} \quad \mathcal{N}(0,1), \qquad \hat{\mathcal{T}}_1 = \left(\text{vec}\left(\hat{\mathbf{A}}\right)' \hat{\text{V}}_1^{-1} \text{vec}\left(\hat{\mathbf{A}}\right) - mn\right) \Big/ \sqrt{2mn}, \\
&(ii) \quad \hat{\mathcal{T}}_2 \to_{d,\mathbb{H}_{2,0}} \quad \mathcal{N}(0,1), \qquad \hat{\mathcal{T}}_2 = \left(\left(\hat{b}^j\right)^2 \hat{\gamma}' \hat{\text{V}}_2^{+} \hat{\gamma} - 1\right) \Big/ \sqrt{2}, \\
&(iii) \quad \hat{\mathcal{T}}_3 \to_{d,\mathbb{H}_{3,0}} \quad \mathcal{N}(0,1), \qquad \hat{\mathcal{T}}_3 = \left(\hat{\phi}_{h,m}(\bar{v}) - \phi_h(\bar{v})\right) \Big/ \hat{\text{V}}_3,
\end{aligned}
$$

as $T \to \infty$, where $\hat{\mathbf{A}}$, $\hat{b}^j$ and $\hat{\gamma}$ are obtained from the reverse regression, $\hat{\text{V}}_1$, $\hat{\text{V}}_2$, and $\hat{\text{V}}_3$ are defined in the Appendix, V^+ is the Moore-Penrose generalized inverse of V and $\to_{d,\mathbb{H}_0}$ signifies convergence in distribution under the hypothesis $\mathbb{H}_0$.

Theorem 1 provides the appropriate limiting distributions to implement the main asset pricing tests for the paper. To conduct inference when estimating multi-horizon returns with overlapping data we extend the "reverse regression" approach of Hodrick (1992) to our reduced-rank, nonparametric setting (see the Appendix for further details). The test statistics $\hat{\mathcal{T}}_1$ and $\hat{\mathcal{T}}_2$ are based on the estimated parameters from the reverse regression. They represent joint tests of zero coefficients on the basis functions and are akin to a standard F-test in a linear regression setting. The key difference here is that the number of basis functions, m, are growing with the sample size and so a balance must be achieved to attain the distributional approximation. In particular, the number of basis functions must grow fast enough to achieve the nonparametric approximation to the true, unknown parameters including ϕ_h, but also slow enough so that the approximation of the test statistics to a standard normal random variable is maintained.[7] A similar approach was taken by Crump

[7] In Section A.2 of the Appendix we provide results from a series of Monte Carlo simulations. These simulations show that our new inference procedures control empirical size very well in cases with modest sample sizes and data-generating processes which mimic the properties commonly encountered in financial time series.

et al. (2008) for the case of *i.i.d.* data to test for heterogeneity in treatment effects. Here we introduce tests for our reduced-rank, nonparametric setting which are also valid under more general forms of time-series dependence in the data. This framework is more appropriate for finance and macroeconomic applications.

We are also interested in the following hypothesis which allows us to test the cross-sectional restrictions across assets which are a key feature of our specification:

$$\begin{aligned}\mathbb{H}_{4,0} &: \exists\, b_h,\ \phi_h(\cdot)\ \text{ s.t. } \mathbb{E}_t\left[\mathrm{Rx}_{t+h}\right] = b_h\phi_h(v_t)\\ \mathbb{H}_{4,A} &: \nexists\, b_h,\ \phi_h(\cdot)\ \text{ s.t. } \mathbb{E}_t\left[\mathrm{Rx}_{t+h}\right] = b_h\phi_h(v_t)\,.\end{aligned} \tag{3.6}$$

Under the null hypothesis, $\mathbb{H}_{4,0}$, there exists a single, common function ϕ_h which drives the time variation in conditional expected returns. Under the alternative hypothesis this is not the case. For example, the alternative hypothesis would include the case where the data feature asset-specific, possibly nonlinear functions of the VIX.

The following theorem allows us to formally test $\mathbb{H}_{4,0}$.

Theorem 2. *Under regularity conditions given in the Appendix*

$$\hat{\mathcal{T}}_4 \rightarrow_{d,\mathbb{H}_{4,0}} \quad \mathcal{N}(0,1)\,, \qquad \hat{\mathcal{T}}_4 = \left(\left(\hat{\mathbf{A}} - \hat{b}\hat{\gamma}'\right)' \hat{V}_4^{+} \left(\hat{\mathbf{A}} - \hat{b}\hat{\gamma}'\right) - s\right) \Big/ \sqrt{2s}$$

as $T \to \infty$, where $\hat{\mathbf{A}}$, $\hat{b}$ and $\hat{\gamma}$ are obtained from the reverse regression, $\hat{\mathrm{V}}_4$ is defined in the Appendix, V^{+} is the Moore-Penrose generalized inverse of V, $s = (m-1)(n-1)$, and $\rightarrow_{d,\mathbb{H}_{4,0}}$ signifies convergence in distribution under the hypothesis $\mathbb{H}_{4,0}$.

The test statistic $\hat{\mathcal{T}}_4$ compares the unrestricted estimate, $\hat{\mathbf{A}}$, to the estimate under the cross-sectional restrictions which impose a common function across assets, $\hat{b}\hat{\gamma}'$.[8] Under a strictly parametric specification for conditional expected returns this can be interpreted as a Lagrange multiplier (LM) test statistic. Alternatively, it can also be interpreted as a minimum distance test statistic as it compares restricted and unrestricted coefficient estimates (for more details see Remark 1 in Appendix A.5). The proofs of Theorems 1 and 2 utilize technical results for sieve estimators in time-series settings from Chen and Christensen (2015) along with convergence rates for the multivariate central limit theorem for dependent

[8] We thank an anonymous referee for suggesting this additional test.

data from Bulinskii and Shashkin (2004).

3.2 Estimation of Sieve Reduced Rank Regressions

Our main empirical findings using the SRR regressions of equation (3.1) for the market return and the maturity sorted bond returns are presented in Table 2. As we saw in the univariate VIX polynomial regressions, substantial improvements are gained when allowing the cross-section of market returns and maturity sorted bond returns to depend on the VIX nonlinearly: Whereas the VIX does not linearly forecast excess returns in panel (1), the nonlinear forecasting relationship for stocks and bonds is highly significant in panel (2). Moreover, panel (3) shows that the nonlinear forecasting factor is robust to the inclusion of common predictor variables (the default spread DEF, the variance risk premium VRP, the term spread TERM, and the log dividend yield DY). Furthermore, the significant predictability is present in the 1990-2007 period which excludes the financial crisis (Table 3).

Examining Table 2 in more detail, we see that the market return is most strongly predicted at the six month horizon at the one percent level. Per construction, the coefficient on the market return is 1. Overall, the strongest predictability appears for shorter-maturity bonds, as the one year and two year bond return is highly significantly predicted at the one percent level for the 6, 12, and 18 month horizons. Interestingly, the significance is unchanged when even the variance risk premium (a volatility measure constructed from the VIX) is included, suggesting that the nonlinear forecasting factor is unrelated to VRP. Longer maturity Treasuries such as the twenty year or the thirty year bond return tend to be somewhat less significant at longer horizons. The sign on all of the Treasury variables is negative whereas the market return is positive, revealing a mirror image relationship that is statistically strongest for liquid short-maturity Treasuries. While individual coefficient significance was tested by $\mathbb{H}_{2,0}$, joint significance for the function $\phi_h(vix_t)$ in the cross-section of excess returns is tested with $\mathbb{H}_{1,0}$.[9] Again the joint test provides strong justification for nonlinearities $\phi_h(vix_t)$ across all forecasting horizons, whereas the linear VIX specification cannot be statistically distinguished from specifications featuring only the constant a_h^i. As an additional test of $\mathbb{H}_{1,0}$ the last row in each panel reports p-values based on a block bootstrapped distribution for the test statistic $\hat{\mathcal{T}}_1$ (see Section A.4 in the Appendix for details on

[9]To improve the finite-sample properties of the test statistic $\hat{\mathcal{T}}_1$, in our empirical implementation we construct the estimated variance matrix under the null hypothesis, $\mathbb{H}_{1,0}$.

the resampling approach). The bootstrap-based p-values produce very similar conclusions as inference based on asymptotic results.

For the pre-crisis period 1990-2007 presented in Table 3, the equity market is significant at least at the 10% level at all horizons. Treasury returns are again very significant, and result in stronger rejections for shorter maturity Treasuries. In particular, the shorter maturity Treasury returns are significant at the 1% level across all specifications and forecasting horizons for both the pre-crisis and full samples, which include the specifications with common predictor variables. Furthermore, test $\mathbb{H}_{1,0}$ confirms that $\phi_h(vix_t)$ is a strong predictor of excess returns jointly across all test assets and horizons. We highlight again the gains obtained by allowing excess returns to nonlinearly depend on the VIX.[10] Most importantly, the mirror image property between stock and bond returns is revealed in the pre-crisis period as suggested by the coefficient signs, although we are careful to point out that for specification (3), coefficient signs are difficult to interpret when ϕ_h interacts with the control predictors. In sum, we find that the nonlinear sieve reduced rank regressions reveal the mirror image property, which is not manifested for the linear VIX regressions.

We interpret the mirror image property as evidence of flight-to-safety, since it reveals that the required return for holding stocks and bonds is intimately linked to the same function of aggregate volatility $\phi_h(vix_t)$. An alternative interpretation is that the VIX nonlinearly forecasts stock and bond returns independently and that our SRR regression imposes a link that is not supported by the data. We show, however, that this is not the case. Specifically, Table 4 reports the outcome of the test of a common nonlinear function $\phi_h(vix_t)$ driving expected stock and bond returns (see Theorem 2). For all three of our horizons ($h = 6, 12,$ or 18) and in both the full sample and pre-crisis sample, we fail to reject the null hypothesis of a single common function at the 10% level. Thus, in our sample we do not find sufficient evidence in favor of the alternative hypothesis and, in unison with the other results presented in this section, we find strong supporting evidence for a common nonlinear dependence of

[10] A related advantage of the nonlinear transformation of the VIX is that it serves to modulate the persistence properties of the VIX and, consequently, better align the time series properties of the predictor with returns. Marmer (2008) shows that this can lead to improvements when estimating forecasting relationships as compared to a linear specification. More generally, this is related to the literature on nonlinear transformations and memory properties of a time series (see, for example, Granger (1995), Park and Phillips (1999, 2001), de Jong and Wang (2005), Pötscher (2004), or Berenguer-Rico and Gonzalo (2014)). Separately, in Section A.3 of the Appendix we show that our results are robust to using the VIX transformed by the natural logarithm.

stock and bond returns on the VIX.

These results fall within the vast literature on asset return forecasting. Seminal papers include Campbell and Shiller (1988a,b), Lettau and Ludvigson (2001), Cochrane and Piazzesi (2005), Ang and Bekaert (2007), and are surveyed by Cochrane (2011). The majority of this literature focuses on forecasting returns using financial ratios or yields. While much of that literature employs linear forecasting relationships, some do model nonlinearities. Lettau and Van Nieuwerburgh (2008) present a regime shifting model for stock return forecasting. Pesaran et al. (2006) present forecasting relationships for US Treasury bonds subject to stochastic breakpoint processes. Rossi and Timmermann (2010) document a nonlinear risk-return tradeoff in equities using boosted regression trees. To the best of our knowledge, no paper has estimated a common nonlinear forecasting relationship across different asset classes. In a related literature, Martin (2017) shows that expected equity market returns are bounded below by risk-neutral expected variance and shows that the latter can be measured by the SVIX, a portfolio of S&P 500 index options that is closely related to the VIX. Our results support the link between expected equity market returns and option-implied volatility, but additionally show that Treasury returns have an opposite reaction, revealing flight-to-safety. The flight-to-safety response is relatively less explored and is not an immediate consequence of the theory in Martin (2017).

3.3 Evidence Using Broader Cross-Sections of Assets

While our results so far have focused on the aggregate stock return and maturity sorted Treasury bond portfolios, we now estimate the SRR regression on a broader set of test assets in order to improve our understanding of the cross-section of returns. We use the 12 industry sorted stock portfolios from Kenneth French's website[11] and the industry and rating sorted investment grade credit returns from Barclays. We also continue to include the maturity sorted Treasury bond portfolios.

Figure 5 displays the results of hypothesis tests $\mathbb{H}_{2,0}$. The height of bar j represents the point estimate of $\hat{b}_h^j$ for $h = 6$. For each $j = 1, \ldots, n$, the color of the bar denotes the significance of the associated $\hat{b}_h^j$ coefficient based on the results in Theorem 1. The figure shows that the majority of stock and Treasury portfolios load significantly on $\phi_h(vix_t)$.

[11] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Manufacturing, known to be highly procyclical, has a strong positive exposure while the only equity portfolios that appear invariant to the volatility factor are non-durables, energy, utilities, and healthcare, which are known to represent inelastic sectors. Most strikingly, the only assets with negative exposures to the volatility factor $\hat{\phi}_h(v)$ are Treasury portfolios and AAA corporate bonds, which is consistent with a flight to safety interpretation. Furthermore, the results confirm our previously reported findings, which showed the strongest predictive content for nonlinear VIX functions at shorter maturity Treasuries. For $h = 6$, however, the corporate bond loadings are not generally statistically significant.

Next, Figure 6 shows $\hat{E}_t[Rx^i_{t+h}] = \hat{a}^i_h + \hat{b}^i_h \hat{\phi}_h(v)$ for each of the $i = 1, \ldots, 26$ portfolios and horizon $h = 6$ months. The dashed lines in blue represent assets with a negative $\hat{b}^i_h$ exposure to $\hat{\phi}_h(v)$, while the solid lines in red denote positive $\hat{b}^i_h$ exposures. To differentiate our reference asset, the black line denotes the market excess return estimate, whereas the gray dashed line represents the short-maturity Treasury return. We note again that all dashed assets (blue and gray) are Treasury returns and the AAA corporate bond return and are distinguished by their negative loadings on $\hat{\phi}_h(v)$.

Finally, we examine the plausibility of the affine structure in (3.1) by estimating $Rx^i_{t+h} = a^i_h + b^i_h \phi_h(v_t) + \varepsilon^i_{t+h}$ by SRR regression for i ranging over equities and separately for i ranging over bonds. That is, we estimate sieve reduced rank regressions, where the left-hand side variables are the excess return on the equity market and 11 industry portfolios, which yields an estimate of $\hat{\phi}^{\text{Stocks}}_h(\cdot)$.[12] Next, we repeat the regression, but where i includes only the seven maturity-sorted Treasury portfolios, yielding an estimate of $\hat{\phi}^{\text{Treas}}_h(\cdot)$. Under the common $\phi_h(v)$ assumption implicit in (3.1), $\phi^{\text{Stocks}}_h(\cdot)$ and $\phi^{\text{Treas}}_h(\cdot)$ should be equivalent in the population when identified separately from stocks and bonds, up to a location and scale parameter. We find the location and scale parameters by regressing $\hat{\phi}^{\text{Stocks}}_h(v)$ on $\hat{\phi}^{\text{Treas}}_h(v)$ for a range of v in the support of the VIX. The result of this exercise is plotted in Figure 7 which shows $\hat{\phi}^{\text{Stocks}}_h(v)$ along with the location- and scale-shifted $\hat{\phi}^{\text{Treas}}_h(v)$. The figure clearly supports our conjecture of a common $\phi_h(v)$ function and underscores the mirror image property in expected stock and bond returns.

[12] The 12th industry portfolio is omitted because the market portfolio is included in the regression.

3.4 Robustness of Estimates

The preceding tables strongly suggest that expected excess returns for the market and Treasury returns are driven by a common nonlinear function of the VIX, i.e. $E_t[Rx^i_{t+h}] = a^i_h + b^i_h \phi_h(vix_t)$. One potential concern is that these in-sample results are an artifact of overfitting arising from the flexibility of our nonparametric approach. If this were the case, we would expect out-of-sample forecasting performance of our SRR regression to be particularly poor relative to more parsimonious models. To this end, we study the out-of-sample properties relative to a model with no predictability in returns (e.g., Welch and Goyal (2008)).[13]

We begin by splitting our monthly sample of excess returns $\text{Rx}_t = (Rx^{MKT}_t, Rx^{cmt1}_t, \ldots, Rx^{cmt30}_t)'$ and the VIX for $t = 1, \ldots, T$ into an initial in-sample estimation period $t = 1, \ldots, t^*$, and an out-of-sample forecasting period $t = (t^* + 1), \ldots, T$. We consider three different in-sample cutoffs corresponding to $(t^*/T) = 0.4$, 0.5, and 0.6. We next implement the SRR regression estimator in exactly the same way as in the full-sample analysis: at each point $t = (t^* + 1), \ldots, T$, we choose the model specification via cross-validation using only information up to and including time t and produce a forecast of the subsequent h-period holding period return. We then repeat this exercise for each successive out-of-sample forecast and store the resulting forecast errors.

Table 5 shows two sets of relative performance measures: the log ratio of root mean-square forecast errors (RMSFEs) from the SRR regression model (in the numerator) and the mean-only model (in the denominator) along with the log ratio of mean absolute forecast errors (MAFEs). Positive numbers therefore mean that the SRR regression model produces larger out-of-sample forecast errors on average, whereas negative numbers mean the SRR model produces smaller forecast errors. The table shows three panels corresponding to $h = 6, 12$, and 18 month forecast horizons. The two performance measures, three sample splits, eight assets, and three forecast horizons result in 144 forecast error comparisons.

Reassuringly, the SRR model outperforms the mean-only model over 60% of the time across horizons and sample splits. Perhaps unsurprisingly, given the modest number of time series observations we have available, the SRR regression generally has its strongest performance for the $(t^*/T) = 0.6$ sample split. For example, if we measure performance

[13] In Appendix A.1 we report forecast error comparisons relative to a linear model in the VIX, along with tests of equal predictive accuracy of Clark and West (2007). These results are consistent with the findings in this section.

by average absolute forecast errors, the SRR regression model outperforms the mean-only model at least 60% at the time and as high as 100% of the time across the three horizons. The key takeaway from Table 5 is that the SRR regression provides reasonable forecasts in real time which should alleviate concerns about the risk of overfitting from our flexible, nonparametric model. We can therefore focus on the full-sample results where we have more power to discriminate between competing hypotheses.

4 Economics of Flight-to-Safety

We now turn to the economic interpretation of the nonlinear risk-return tradeoff. We first establish the link of the SRR regression model to dynamic asset pricing theories (subsection 4.1). We empirically show that the cross-sectional dispersion of the forecasting slopes b_h^i from the SRR regression are related to risk factor loadings on the market return, the one year Treasury return, and the nonlinear volatility function $\phi_h(v_t)$. This is evidence in favor of a dynamic pricing kernel where $\phi_h(v_t)$ is a price of risk variable.

We next turn to asset pricing theories that can give rise to time varying effective risk aversion as a nonlinear function of market volatility. The first types of theories that we discuss feature flight-to-safety (subsection 4.2) that leads to time varying pricing of risk because asset managers are subject to withdrawal after poor performance. We also verify empirically that global mutual fund flows exhibit the nonlinear (contemporaneous) relationship to the VIX as the expected returns do. We then review intermediary asset pricing theories that link the pricing of risk to the level of volatility, for example due to the VaR constraint of Adrian and Boyarchenko (2012) (subsection 4.3). We also discuss the extent to which consumption based theories might explain the shape of $\phi_h(v_t)$, focusing on habit formation and long-run risks (subsection 4.4). Finally, we show in subsection 4.5 that the VIX also forecasts macroeconomic activity in a nonlinear fashion when the crisis period is included.

4.1 Dynamic Asset Pricing

Equation (3.1) shows that expected returns are affine functions of $\phi_h(v_t)$ with intercept a^i and slope b^i. Asset pricing theory suggests that these intercepts and slopes are cross-sectionally related to risk factor loadings (see Sharpe (1964), Merton (1973), Ross (1976)). In particular,

an equilibrium pricing kernel with affine prices of risk, as for example presented by Adrian et al. (2015), would suggest that

$$E_t[Rx^i_{t+h}] \quad = \quad \alpha^i_h + \beta^i_h \left(\lambda_0 + \lambda_1 \phi_h(v_t) + \Lambda_2 x_t\right). \tag{4.1}$$

In this expression, β^i_h denotes a $(1 \times K)$ vector of risk factor loadings, λ_0 is the $(K \times 1)$ vector of constants for the prices of risk, λ_1 is the $(K \times 1)$ vector defining how prices of risk vary as a function of $\phi_h(v_t)$, and Λ_2 is the $(K \times p)$ matrix mapping that defines how the price of risk depends on p additional risk factors x_t. The expression also allows for a pricing error α^i_h, representing deviations from no-arbitrage due to trading frictions. Equation (4.1) is the beta representation of expected returns when the pricing kernel is of an essentially affine form (see Duffee (2002)).

Equation (4.1) has time series and cross-sectional predictions, which in turn can be linked to alternative theories of time varying pricing of risk. We start with an investigation of the cross-sectional predictions. In comparison to the SRR regression model of equation (3.1), the asset pricing theories of equation (4.1) put the following constraints on the intercept and slope:

$$a^i_h \quad = \quad \alpha^i_h + \beta^i_h \lambda_0 \tag{4.2}$$

$$b^i_h \quad = \quad \beta^i_h \lambda_1. \tag{4.3}$$

In order to show that the cross-sectional dispersion of a^i_h and slope b^i_h is compatible with such asset pricing restrictions, we proceed in two steps. We first estimate the unrestricted panel forecasting relationship $Rx^i_{t+h} = a^i_h + b^i_h \phi_h(v_t) + \varepsilon^i_{t+h}$ by sieve reduced rank regression, yielding a cross-section of parametric estimates of a^i_h and b^i_h and a single nonparametric estimate of $\phi_h(v_t)$. Here, $i = 1, \ldots, n$ ranges over the CRSP market excess return, maturity-sorted Treasury excess returns, industry-sorted portfolio excess returns, and ratings and industry sorted corporate bond excess returns. Next, we estimate prices of risk as well as risk factor exposures according to the dynamic asset pricing restrictions in (4.1). That is, we estimate jointly $Rx^i_{t+h} = (\alpha^i_h + \beta^i_h \lambda_0) + \beta^i_h \lambda_1 \phi_h(v_t) + \beta^i u_{t+h} + \varepsilon^i_{t+h}$, where the estimate of $\phi_h(v_t)$ is taken as given from the unrestricted first step regression, and where $u_{t+h} \equiv Y_{t+h} - E_t[Y_{t+h}]$ represents vector autoregression innovations to the risk factors Y_t consisting of the market

return, the one-year Treasury return, and the nonlinear volatility factor $\phi_h(v_t)$. Results of the regressions are given in Table 6, while Figure 8 shows the cross-sectional relationships between a_h^i and $\alpha_h^i + \beta_h^i \lambda_0$ and between b_h^i and $\beta_h^i \lambda_1$ for $h = 1$.[14]

Table 6 shows that the industry portfolios and credit portfolios are highly significantly exposed to equity market risk, while few if any of the Treasury returns load significantly on the market return. The Treasury and corporate bond returns are significantly exposed to the one year Treasury return and also to innovations to the nonlinear volatility factor, although to a lesser extent. Importantly, the prices of risk λ_1 show that the market return commands a significant positive risk premium, while the Treasury return and nonlinear volatility factor each command a negative risk premium, which is consistent with a flight-to-safety interpretation. Taken together, the product $\beta_h^i \lambda_1$, representing loadings on the forecasting factor $\phi_h(v_t)$, has negative signs for all bond returns and positive signs for all equity returns.

Figure 8 shows the cross-sectional relationship between the forecasting intercept a_h^i and slope b_h^i and the risk factor exposures. The top left panel shows that the forecasting slope b_h^i is strongly related to the risk factor loadings β_h^i and prices of risk λ_1. Correspondingly, when deviations from arbitrage α_h^i are permitted, the top right panel supports the view that the dynamic asset pricing model that restricts slope coefficients b_h^i to be $\beta_h^i \lambda_1$ results in correct predictions about unconditional excess returns in the cross-section. These predictions are also captured in the unrestricted forecasting regressions (bottom right panel). However, under no-arbitrage in frictionless markets α_h^i is forced to zero, the pricing performance deteriorates in the bottom left panel.[15] Taken together, the results from Table 6 and Figure 8 strongly suggest the interpretation of $\phi_h(v_t)$ as a price of risk variable in a dynamic asset pricing model with frictions.

Equation (4.1) also has the time series prediction that $\phi_h(v_t)$ might be related to alternative proxies for the time variation in the pricing of risk. There is a large literature documenting the extent to which asset prices are predictable. An obvious question is to what extent our estimated return predictor $\phi_h(v_t)$ is related to other forecasting variables that have been shown to be significant. To do so, we plot eight commonly used variables

[14] We focus on $h = 1$ to simplify estimation of the dynamic asset pricing model.

[15] The deterioration in pricing performance is attributable to our choice of industry sorted portfolios as test assets for the equities, as those are well known to generate pricing errors relative to risk based explanations. Size or book to market sorted portfolios would improve cross-sectional pricing performance.

together with $\phi_h(v_t)$ in the time series for $h = 6$ (see Figure 9).[16] In particular, we show the slope of the Treasury yield curve as measured by TERM, the Cochrane and Piazzesi (2005) CP factor, the BAA-AAA credit spread (DEF), the dividend yield (DY) of the S&P500, the CAY factor by Lettau and Ludvigson (2001), the variance risk premium (VRP) of Bollerslev et al. (2009), the illiquidity factor of Kyle and Obizhaeva (2016), and the macroeconomic uncertainty series of Jurado et al. (2015). It is apparent from Figure 9 that the relationships of $\phi_6(v_t)$ with all eight variables is very weak. The only variables that bear some resemblance are the credit spread DEF, the variance risk premium VRP, and the macroeconomic uncertainty measure, all of which increase during the financial crisis in tandem with $\phi_6(v_t)$. However, outside of crisis periods, neither DEF nor the macroeconomic uncertainty variable co-move strongly with $\phi_6(v_t)$. The latter observation is supported in the original paper by Jurado et al. (2015), who found that their macroeconomic uncertainty variable is not subsumed by other measures of uncertainty, including measures of stock market volatility.

The relationship between VRP and $\phi_6(v_t)$ merits further discussion. In particular, as a measure of option-implied volatility, the VIX is a function of both expected future equity realized variance and the variance risk premium. Hence we expect the VRP to bear some resemblance to the VIX, and in fact, their correlation in our data is a sizeable 62%. Figure 9, however, compares the VRP with $\phi_6(v_t)$, a nonlinear function of the VIX. In our data, the VRP and this nonlinear function $\phi_6(v_t)$ have a much lower correlation of 15%. In unreported results, we estimated the sieve reduced rank regressions on the VRP itself and found evidence against a nonlinear relationship between VRP and future stock and bond returns. Finally, we note that our earlier SRR regression forecasting results showed that DEF and VRP do not impact the significance of $\phi_h(v_t)$ markedly.

4.2 Theories and Evidence of Flight-to-Safety

The theoretical literature has proposed a number of distinct mechanisms generating flight-to-safety. Vayanos (2004) studies equilibrium asset pricing where volatility is stochastic, and assets are invested by fund managers. Assets' illiquidity arises as trading is subject to

[16] We provide a visual representation of the alternative forecasting variables, since four of the eight plotted series were previously shown to be unrelated to the predictive content of $\phi_h(v_t)$ (Tables 1, 2, and 3), and since CAY is only available quarterly. In separate results (not included for brevity), we also included the CP factor as a predictive control and found that it did not affect our results.

fixed transactions costs. Fund managers are subject to withdrawals when fund performance is poor, generating a preference for liquidity that is a time varying function of volatility. Vayanos (2004) derives equilibrium expected returns of the following form[17]

$$E_t\left[Rx^i_{t+1}\right] = \alpha^i\left(v_t\right) + A\left(v_t\right) Cov_t\left(Rx^i_{t+1}, Rx^M_{t+1}\right) + Z\left(v_t\right) Cov_t\left(Rx^i_{t+1}, v_{t+1}\right). \tag{4.4}$$

Expected returns are thus functions of their covariance with the market return and with volatility, where the impact of these covariances on expected returns depends on the endogenously time varying effective risk aversion $A\left(v_t\right)$ and the endogenously time varying volatility risk premium $Z\left(v_t\right)$. Our estimated function $\phi\left(v_t\right)$ corresponds to $A\left(v_t\right) \cdot v_t$ and our estimated exposures on $\phi\left(v_t\right)$ are related to the covariance of asset returns with the risk factors, including market risk and interest rate risk. The solution for $A\left(v_t\right)$ is highly nonlinear, and must be computed numerically. Pricing errors α^i_t are related to trading costs across assets times the withdrawal likelihood, and are also a function of v_t, but they are unrelated to risk factor loadings. We note, however, that the downward sloping $\phi\left(v_t\right)$ for the VIX below its median and above its 99th percentile is not an immediate implication of the theory. We will discuss possible explanations below.

In Vayanos (2004), $A\left(v_t\right)$ is convex. Furthermore, our estimated $\phi\left(v_t\right)$ has to be compared to $A\left(v_t\right) \cdot v_t$, making it more convex. Our estimate of $\phi\left(v_t\right)$ is convex when the VIX is in the intermediate range. Within Vayanos' theory, the convexity arises because risk premia are affected by fund managers' concern with withdrawals. Withdrawals are costly to the managers because the managers' fee is reduced, and holding a riskier portfolio makes withdrawals more likely by increasing the probability that performance falls below the threshold. When volatility is low, managers are not concerned with withdrawals and hence the component of the risk premium that corresponds to withdrawals is very small and almost insensitive to volatility. That component starts increasing rapidly, however, when volatility increases, leading the managers' effective risk aversion to increase with volatility.

In the model of Vayanos (2004), conditional betas are constant, and stock-bond correlations are constant. This is an equilibrium outcome that is consistent with our assumption of having constant betas in the dynamic asset pricing model. Comparative statics in the

[17] We are changing notation from continuous to discrete time to make it consistent with the rest of the notation in this paper.

Vayanos model do suggest that changes in the contemporaneous correlation will be related to the magnitude of expected returns. A more general model could allow for time varying betas, both theoretically and empirically, but such an approach would substantially complicate estimation and inference, and go beyond the current scope of the paper.

We next investigate direct evidence of flight-to-safety by analyzing global mutual fund flows from the Investment Company Institute (ICI) (Table 7). The table reports a contemporaneous SRR regression of the mutual fund flows on the VIX. The regression is contemporaneous, as flows drive expected returns contemporaneously with volatility. The table shows that US equity, world equity, and hybrid funds have strongly negative loadings on the fund flow function $\phi^{FF}(v_t)$ while government bond funds exhibit strongly positive loadings. To interpret the meaning of these opposite signs, the top left panel of Figure 10 plots SRR-estimated fund flows $\hat{F}lows_t^i = \hat{a}^i + \hat{b}^i\hat{\phi}^{FF}(v_t)$ for government bond fund flows and the three significant equity fund flows, where v ranges over the empirical support of the VIX. The plot shows that as the VIX rises above 22, government bond funds experience inflows. As the VIX rises a few points higher, the risky equity funds experience outflows, presenting direct evidence of flight-to-safety at a slightly higher VIX threshold than we previously identified from excess returns data. Table 7 also shows that the flight-to-safety pattern in loading signs is present in the mutual fund flows whether the crisis is included or not. It is interesting to note that money market funds are not found to be a safe asset in these regressions, consistent with the fact that investors ran on money market funds.

The bottom panel of Figure 10 corroborates the SRR regression findings in the time series. The plot shows the relationship between the VIX and aggregated flight-to-safety flows, defined as the sum of equity fund outflows and contemporaneous government bond fund inflows.[18] Our earlier risk premium-based estimates of $\phi(v_t)$ indicated the expected returns to stocks and bonds begin to diverge when the VIX rises above its median of 18. The plot correspondingly shows that in states of the world where the VIX exceeds its median threshold, flight-to-safety flows strongly co-move with the VIX, with a correlation of 68%.

The top right panel of Figure 10 compares (an affine transformation of) the shape of the nonlinearity from the fund flows, $\phi^{FF}(v_t)$, to the $\phi(v_t)$ function estimated from excess returns. While the mutual fund flows line up with the $\phi(v_t)$ function in the intermediate range of the VIX between 18 and 50, the $\phi(v_t)$ function is downward sloping outside of

[18]Note that Figure 10 plots the VIX in excess of its median.

that range. For the lower end of the range, the theory by Buffa et al. (2014) might offer an explanation. The upper part of the range, when stock returns depend negatively on volatility when the VIX is above 50 is not consistent with Vayanos (2004). In our interpretation, that negative dependence of future returns on the VIX is due to changes in cash flow news during the depth of the financial crisis, as argued by Campbell et al. (2013).[19]

Buffa et al. (2014) augment the Vayanos (2004) model with competition among fund managers. Because of agency frictions, investors make managers' fees more sensitive to performance and benchmark performance against a market index. This makes managers unwilling to deviate from the index and exacerbates price distortions. Because trading against overvaluation exposes managers to greater risk of deviating from the index than trading against undervaluation, agency frictions bias the aggregate market upwards. They can also generate a negative relationship between risk and return because they raise the volatility of overvalued assets.

Caballero and Krishnamurthy (2008) study a very different but complementary mechanism for flight-to-safety, based on Knightian uncertainty. In their model, agents faced with Knightian uncertainty consider the worst case among the scenarios over which they are uncertain. When the aggregate quantity of liquidity is limited, Knightian agents grow concerned with liquidity shortages and they therefore sell risky financial claims in favor of safe and uncontingent claims, i.e. there is flight-to-safety. Even though the flight-to-safety seems prudent from individuals' point of views, it is collectively costly for the macroeconomy because scarce liquidity goes wasted. To the extent that a high level of the VIX might trigger Knightian agents to a flight-to-safety, or a high level of the VIX is correlated with an increase in uncertainty, the predictions of Caballero and Krishnamurthy (2008) broadly support our empirical finding that riskier securites tend to load positively on $\phi(v_t)$, while safe securities load negatively on $\phi(v_t)$.

[19]Note that our interpretation of the economic driver of the nonlinear risk return tradeoff differs from the interpretation by Ghysels et al. (2014), who argue that their high volatility regime corresponds to the flight-to-safety regime. In contrast, our evidence presented in this paper is that flight-to-safety is most relevant for the range of the VIX between 18 and 50, which occurs 49.3 percent of the time. For the VIX above 50 (occurring .7 percent of the sample), cash flow news, not discount rate news are the likely driver of returns.

4.3 Intermediary Asset Pricing Theories

Intermediary asset pricing theories model the impact of intermediary balance sheet frictions on the pricing of risk and real activity within dynamic general equilibrium models of the macroeconomy. The strand of literature was pioneered by He and Krishnamurthy (2013), who model an intermediary sector whose ability to raise external equity capital depends on past performance, similar to Vayanos (2004).

Adrian and Boyarchenko (2012) expand on those intermediary asset pricing theories by introducing intermediary leverage as an additional state variable. The intermediary leverage state variable arises endogenously, as intermediaries are subject to value at risk (VaR) constraints. Furthermore, intermediaries face liquidity shocks in addition to productivity shocks, while the models of He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014) only feature a single shock.

The VaR constraint directly links aggregate volatility to the leverage of intermediaries. Increases in volatility, which arise endogenously, tighten the leverage constraints on intermediaries, increasing their effective risk aversion, as well as equilibrium pricing of risk. Booms correspond to periods when volatility is endogenously low, pricing of risk is compressed, and intermediary leverage is elevated. When adverse shocks hit, either to productivity or to liquidity, volatility rises endogenously, tightening balance sheet constraints and leading to a widening in the pricing of risk.

In order to gauge the relationship of our pricing of risk function $\phi(v_t)$ to intermediary asset pricing theories, we estimate the relationship between the cross-section of VaRs over time and the VIX, using SRR regressions. We obtain the VaRs from Bloomberg for Bank of America, Citigroup, Goldman Sachs, JP Morgan, and Morgan Stanley. VaRs are expressed in dollar terms. We use the aforementioned five banks as those institutions are the main US banking organizations with trading operations that have reported data continuously since 2004. The VaR data is, unfortunately, only available at a quarterly frequency, so that return forecasting regressions have very few observations.[20] Instead, we present in Figure 11 the results of regressing the panel of VaRs contemporaneously on the VIX, using SRR regressions. The result of this is shown in the lower panel of Figure 11, while the upper

[20]We did perform SRR regressions of the stock and bond portfolios on the six month lag of the summed VaRs and found significant forecasting ability using the SRR regressions. We also uncovered the mirror image property.

panel presents the sum of VaRs together with the VIX. We can see that there is a tight association between the VIX and the VaRs, and that the SRR regression is suggestive of a slightly concave relationship.

The empirical results of Figure 11 support the assumption of Adrian and Boyarchenko (2012) that intermediary balance sheet constraints are related to market volatility due to risk management constraints. Furthermore, our earlier finding that expected returns are systematically related to the VIX are compatible with the notion that constraints on intermediary balance sheets matter for the pricing of risk. Of course, this evidence is only suggestive, and more rigorous analysis would require the calibration of intermediary asset pricing models, or an identification strategy for exogenous variation in dealer balance sheet capacity. We leave such research for future work.

4.4 Time Varying Pricing of Risk in Consumption-Based Asset Pricing

Besides theories of flight-to-safety and intermediary asset pricing theories, consumption-based asset pricing models can also give rise to variation in the pricing of risk. Among the workhorse consumption-based asset pricing models, we consider habit formation and long-run risk models with recursive preferences and compare them with the SRR-implied time varying price of risk.

In habit formation theories, utility depends not just on the level of current consumption, but rather on current consumption in excess of previously experienced consumption. Asset pricing implications of habit formation were pioneered by Abel (1990), Constantinides (1990), and Sundaresan (1989). Our discussion will focus on the theory by Campbell and Cochrane (1999, 2000), which has synthesized earlier work, and proven successful in explaining asset pricing puzzles.

The key state variable in the Campbell and Cochrane (1999, 2000) setup is the surplus consumption ratio s_t, which is a slow moving, mean reverting function of past shocks to aggregate consumption. The pricing of risk is a function of the surplus consumption ratio. Campbell and Cochrane (1999, 2000) present parameter calibrations that have been shown to be able to explain asset pricing puzzles, both in the cross-section and in the time series.

In order to gauge the plausibility of our estimated $\phi(v_t)$ function within the Campbell

and Cochrane (1999, 2000) asset pricing context, we consider the conditional Sharpe ratio generated by that pricing kernel, $\frac{E_t\left[R^*_{t+1}\right]}{\sqrt{Var_t\left[R^*_{t+1}\right]}} = \sqrt{Var_t\left(\ln M_{t+1}\right)} = \gamma\left(1+\lambda\left(s_t\right)\right)\sigma$, where R^*_{t+1} corresponds to the returns of the maximum Sharpe ratio portfolio. Because the maximum Sharpe ratio represents the conditional variance of the (log) stochastic discount factor, it provides an analytically tractable way to study the habit model's price of risk. Given the SRR-estimated $\phi\left(v_t\right)$ interpretation as a price of risk, the time-variation in the two objects can be meaningfully compared.

In long-run risk (LRR) models, preferences are recursive and consumption and dividends contain a small, persistent expected growth rate component, as well as stochastic consumption volatility. Bansal and Yaron (2004) initially showed that such models capture salient features of the equity premium, the risk-free rate, and market volatility. By analogy to the habit model, we examine the LRR-implied maximum Sharpe ratio $\frac{E_t\left[R^*_{t+1}\right]}{\sqrt{Var_t\left[R^*_{t+1}\right]}} = \sqrt{Var_t\left(\ln M_{t+1}\right)} = \sqrt{a_0 + a_1\sigma_t^2}$, which is driven by conditional consumption volatility σ_t, and use the parameter calibrations for a_0 and a_1 from the empirical LRR implementation in Bansal et al. (2012). σ_t is estimated as a 4-quarter moving average of squared consumption AR(1) innovations that have been projected on lagged innovations and the dividend yield.[21]

The top panel of Figure 12 shows the Campbell-Cochrane implied maximum conditional Sharpe ratio together with the (mean shifted) SRR-implied price of risk $\phi\left(v_t\right)$ over time. The bottom panel shows the analogous figure for the long-run risk maximum Sharpe ratio. There is some positive correlation among the two consumption-based price of risk measures, and the SRR regression implied price of risk $\phi\left(v_t\right)$. For example, during the financial crisis, when $\phi\left(v_t\right)$ increases sharply, the Sharpe ratio from the habit model and the LRR model also increases unusually strongly. However, while $\phi\left(v_t\right)$ reverts back to lower levels after 2009, the habit Sharpe ratio remains high through 2015. The LRR Sharpe ratio also stays elevated through 2011. Another striking divergence occurs during the late 1990s and early 2000s, when the tech boom in the stock market was associated with high volatility, but a low level of the habit and LRR Sharpe ratios. Hence overall, the correlations between the $\phi\left(v_t\right)$ and the consumption-based Sharpe ratios (0.07 for habit and 0.21 for LRR) are fairly low.

Our takeaway from this finding is that time variation of expected returns derived from

[21] We thank Amir Yaron for providing us with this data.

the SRR regressions on the VIX captures distinct economic mechanism when compared to the time variation of expected returns induced by habit formation. While habit formation and LRR pricing kernels are very tightly linked to the growth of aggregate consumption, expected returns derived from the VIX are likely to capture funding constraints on financial institutions such as major banking organizations and fund managers, as we argued above. Hence the two sources of time variation in expected returns and Sharpe ratios capture complementary economic forces.

4.5 Macroeconomic Consequences

Our final investigation concerns macroeconomic aggregates. A strand of macroeconomics has developed concerning the impact of uncertainty shocks on real economic activity. The model of Bloom (2009), for example, provides a mechanism in which time-varying uncertainty about productivity and demand (i.e. "business conditions") affects firm investment and hiring decisions. Interestingly, his usage of non-convex adjustment costs results in nonlinear hiring and investment decisions due to a real-options effect: Only when capital and labor productivity exceed certain thresholds do returns to investment and hiring exceed the option value of waiting. Moreover, the option value of waiting increases with uncertainty as firms raise their required return to hiring and investment. Thus when firms choose to wait, the result is a temporary decline in real investment and hiring activity.

Bloom's empirical proxy for uncertainty shocks is calibrated from stock market implied volatility. A recent set of findings, however, present challenges to this proxy. For example, Jurado et al. (2015) extract the unforecastable component of a large panel of real activity variables and find that their measure of economic uncertainty has little in common with stock market implied volatility. Along similar lines, Berger et al. (2017) identify uncertainty shocks as second-moment news shocks and find that shocks to uncertainty have no significant effect on the real economy, whereas shocks to stock market volatility do. Furthermore, they find that investors do not appear willing to pay a premium to hedge shocks to uncertainty, in contrast to the large variance risk premium they pay to hedge shocks to stock market volatility. Both papers suggest that stock market volatility and uncertainty about business conditions are not fully interchangeable, which leaves open the question as to how stock market volatility spills over into real activity.

The empirical evidence presented so far in our paper sheds some light on this question. In particular, our results support the notion that stock market implied volatility is related to the pricing of risk in stock and bond markets, which several models in the literature have linked to real economic activity. For example, the work of He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), and Adrian and Boyarchenko (2012) formalize the notion that fund flows and bank balance sheet constraints will translate into distortions of consumption and savings decisions, which would provide a link between stock market volatility and real activity. Under this view, one would therefore expect the VIX to forecast macroeconomic aggregates. Table 8 reports the output of SRR regressions on the five business cycle indicators that receive the largest weight in the Chicago Fed National Activity Index (CFNAI): industrial production (IP), IP manufacturing (IPMFG), manufacturing capacity utilization (CUMFG), change in goods-producing employment (LAGOODA), and total private nonfarm payroll series (LAPRIVA), with the market return included as the reference series.[22] We can see that the VIX strongly forecasts macro activity, with all of the business cycle indicators receiving the same sign. Furthermore, allowing for nonlinearity helps in terms of explanatory power.

Figure 13 shows the relationship between macroeconomic activity, as measured by the CFNAI's 3-month moving average, and $\phi_h(vix_t)$, estimated solely from asset return data as in Section 3.3 using our benchmark 6-month forecast horizon. The figure illustrates that the reversal of $\phi_h(vix_t)$ as VIX rises above its 99th percentile in the fall of 2008 (vertical line) presages the subsequent collapse in real activity. We therefore conjecture that very extreme observations of the VIX lead to a drastic update of expected cash flows, leading to the downward sloping $\phi_h(v_t)$ seen in prior figures. For Treasury returns, expectations about accommodative monetary policy might lead to the sharp rise in the relation between returns and the VIX, preserving the mirror image property even at extreme events. Because the entire cross-section of asset returns considered is pricing this reversal in expected returns, we emphasize that the shape of $\phi_h(v_t)$ is still precisely estimated for these extreme values of the VIX.

[22] We use the same stationarity-inducing transformations employed in the actual CFNAI calculation. For more details, see https://www.chicagofed.org/~/media/publications/cfnai/background/cfnai-technical-report-pdf.pdf.

5 Conclusion

We propose sieve reduced rank (SRR) regressions to extract a common nonlinear function of the VIX that jointly forecasts stock and bond returns at horizons up to two years. We conclude, based on new nonparametric tests, that the forecasting function $\phi_h(v)$ is the same across diverse sets of stock, bond, and credit returns, up to affine transformations. Intriguingly, the loadings of stock and bond returns on the common forecasting variable switch signs when comparing stocks and bonds. This is evidence of investor flight-to-safety: when the VIX rises above its median value, investors tend to reallocate from stocks to bonds, leading to an increase in expected returns for stocks and a compression of expected returns for bonds. We show that the shape of the functional form is robust across asset classes and across time. We can extract virtually indistinguishable shapes from only stocks or only bonds, or use subsamples of the data that exclude the 2008 crisis, or vice versa.

When we relate $\phi_h(v)$ to common return predictors, we find only very weak relationships, suggesting that the nonlinear function of equity market volatility captures economic forces that are complementary to previously documented forecasting variables. Cross-sectional regressions show that the loadings of future returns on $\phi_h(v)$ are cross-sectionally related risk factor loadings, suggesting that $\phi_h(v)$ is a price of risk variable in a dynamic asset pricing model. Our findings support the nonlinear pricing predictions of the asset management theory by Vayanos (2004) where flight-to-safety is associated with increases in equity market volatility, and the intermediary asset pricing theory of Adrian and Boyarchenko (2012) where value at risk constraints on banks give rise to a tight relationship between the pricing of risk and the level of aggregate volatility. We provide supportive evidence for both theories by analyzing global mutual fund flows and the value at risk constraints of major banking organizations.

References

Abel, A. B., 1990. Asset prices under habit formation and catching up with the Joneses. American Economic Review 80, 38–42.

Adrian, T., Boyarchenko, N., 2012. Intermediary leverage cycles and financial stability. Federal Reserve Bank of New York Staff Reports 567.

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

Adrian, T., Crump, R. K., Moench, E., 2015. Regression-based estimation of dynamic asset pricing models. Journal of Financial Economics 118, 211–244.

Andrews, D. W. K., 1987. Asymptotic results for generalized wald tests. Econometric Theory 3, 348–358.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Review of Financial Studies 20, 651–707.

Ang, A., Ulrich, M., 2012. Nominal bonds, real bonds, and equity. Netspar Discussion Paper 12/2011-103, 1–56.

Baele, L., Bekaert, G., Inghelbrecht, K., Wei, M., 2013. Flights to safety. Working Paper 19095, National Bureau of Economic Research.

Bali, T. G., Engle, R. F., 2010. The intertemporal capital asset pricing model with dynamic conditional correlations. Journal of Monetary Economics 57, 377–390.

Bansal, R., Kiku, D., Yaron, A., 2012. An empirical evaluation of the long-run risks model for asset prices. Critical Finance Review 1, 183–221.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Beber, A., Brandt, M. W., Kavajecz, K. A., 2009. Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. Review of Financial Studies 22, 925–957.

Bekaert, G., Engstrom, E., Grenadier, S., 2010. Stock and bond returns with moody investors. Journal of Empirical Finance 17, 867–894.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. Journal of Econometrics 183, 181–192.

Berenguer-Rico, V., Gonzalo, J., 2014. Summability of stochastic processes–a generalization of integration for non-linear processes. Journal of Econometrics 178, 331–341.

Berger, D., Dew-Becker, I., Giglio, S., 2017. Uncertainty shocks as second-moment news shocks. Working Paper 23796, National Bureau of Economic Research.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. Review of Economics and Statistics 72, 498–505.

Bollerslev, T., Osterrieder, D., Sizova, N., Tauchen, G., 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. Journal of Financial Economics 108, 409–424.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22, 4463–4492.

Bradley, R. C., 2007. Introduction to Strong Mixing Conditions, Volume 1. Kendrick Press.

Brunnermeier, M. K., Pedersen, L. H., 2009. Market liquidity and funding liquidity. Review of Financial studies 22, 2201–2238.

Brunnermeier, M. K., Sannikov, Y., 2014. A macroeconomic model with a financial sector. American Economic Review 104, 379–421.

Buffa, A. M., Vayanos, D., Woolley, P., 2014. Asset management contracts and equilibrium prices. Working paper 20480, National Bureau of Economic Research.

Bulinskii, A., Shashkin, A., 2004. Rates in the CLT for sums of dependent multiindexed random vectors. Journal of Mathematical Sciences 122, 3343–3358.

Caballero, R. J., Krishnamurthy, A., 2008. Collective risk management in a flight to quality episode. Journal of Finance 63, 2195–2230.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Campbell, J. Y., Cochrane, J. H., 2000. Explaining the poor performance of consumption-based asset pricing models. Journal of Finance 55, 2863–2878.

Campbell, J. Y., Giglio, S., Polk, C., 2013. Hard times. Review of Asset Pricing Studies 3, 95–132.

Campbell, J. Y., Shiller, R. J., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J. Y., Shiller, R. J., 1988b. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics 76, 5549–5632.

Chen, X., Christensen, T. M., 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. Journal of Econometrics 188, 447–465.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cline, R. E., Funderlic, R., 1979. The rank of a difference of matrices and associated generalized inverses. Linear Algebra and its Applications 24, 185–215.

Cochrane, J. H., 2011. Presidential address: Discount rates. Journal of Finance 66, 1047–1108.

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

Cochrane, J. H., Piazzesi, M., 2008. Decomposing the yield curve. AFA Atlanta Meetings Paper 2010, 1–54.

Constantinides, G. M., 1990. Habit formation: A resolution of the equity premium puzzle. Journal of Political Economy 98, 519–543.

Crump, R. K., Hotz, V. J., Imbens, G. W., Mitnik, O. A., 2008. Nonparametric tests for treatment effect heterogeneity. The Review of Economics and Statistics 90, 389–405.

De Boor, C., 1972. On calculating with b-splines. Journal of Approximation Theory 6, 50–62.

de Jong, R., Wang, C.-H., 2005. Further results on the asymptotics for nonlinear transformations of integrated time series. Econometric Theory 21, 413–430.

Doukhan, P., Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications 84, 313–342.

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

Friedman, J., Hastie, T., Tibshirani, R., 2009. The elements of statistical learning. Second Edition, Springer Series in Statistics.

Ghysels, E., Guérin, P., Marcellino, M., 2014. Regime switches in the risk–return trade-off. Journal of Empirical Finance 28, 118–138.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. Journal of Financial Economics 76, 509–548.

Granger, C. W. J., 1995. Modelling nonlinear relationships between extended-memory variables. Econometrica 63, 265–279.

Guo, H., Whitelaw, R. F., 2006. Uncovering the risk–return relation in the stock market. Journal of Finance 61, 1433–1463.

Gurkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.

Hansen, L., Hodrick, R., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. Journal of Political Economy 88, 829–853.

He, Z., Krishnamurthy, A., 2013. Intermediary Asset Pricing. American Economic Review 103, 732–770.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press.

Jurado, K., Ludvigson, S. C., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1216.

Koijen, R. S. J., Lustig, H. N., van Nieuwerburgh, S., 2013. The cross-section and time-series of stock and bond returns. Journal of Monetary Economics 88, 50–69.

Kyle, A. S., Obizhaeva, A. A., 2016. Market microstructure invariance: Empirical hypotheses. Econometrica 84, 1345–1404.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815–849.

Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Review of Financial Studies 21, 1607–1652.

Lettau, M., Wachter, J., 2011. The term structures of equity and interest rates. Journal of Financial Economics 101, 90–113.

Li, Q., Racine, J. S., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press.

Longstaff, F. A., 2004. The flight-to-liquidity premium in U.S. treasury bond prices. Journal of Business 77, 511–526.

Lundblad, C., 2007. The risk return tradeoff in the long run: 1836–2003. Journal of Financial Economics 85, 123–150.

Mamaysky, H., 2002. Market prices of risk and return predictability in a joint stock-bond pricing model. Yale ICF Working Paper 02-25, 1–62.

Marmer, V., 2008. Nonlinearity, nonstationarity, and spurious forecasts. Journal of Econometrics 142, 1–27.

Martin, I., 2017. What is the expected return on the market? The Quarterly Journal of Economics 132, 367–433.

Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Park, J. Y., Phillips, P. C. B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.

Park, J. Y., Phillips, P. C. B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Patton, A., Politis, D., White, H., 2009. Automatic block-length selection for the dependent bootstrap. Econometric Reviews 28, 372–375.

Pesaran, M. H., Pettenuzzo, D., Timmermann, A., 2006. Forecasting time series subject to multiple structural breaks. The Review of Economic Studies 73, 1057–1084.

Politis, D., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Reviews 23, 53–70.

Pötscher, B. M., 2004. Nonlinear functions and convergence to brownian motion: Beyond the continuous mapping theorem. Econometric Theory 20, 1–22.

Reinsel, G. C., Velu, R. P., 1998. Multivariate Reduced-Rank Regression. Springer, New York.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Rossi, A., Timmermann, A., 2010. What is the shape of the risk-return relation? AFA Atlanta Meetings Paper 2010, 1–58.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

Sundaresan, S. M., 1989. Intertemporally dependent preferences and the volatility of consumption and wealth. Review of Financial Studies 2, 73–89.

Van Tassel, P., Vogt, E., 2016. Global variance term premia and intermediary risk appetite. Staff Report 789, Federal Reserve Bank of New York.

Vayanos, D., 2004. Flight to quality, flight to liquidity, and the pricing of risk. Working Paper 10327, National Bureau of Economic Research.

Wei, M., Wright, J. H., 2013. Reverse regressions and long-horizon forecasting. Journal of Applied Econometrics 28, 353–371.

Weill, P.-O., 2007. Leaning against the wind. Review of Economic Studies 74, 1329–1354.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Figure 2: **P-values by Forecast Horizon: 1990 to 2014**

This figure plots p-values by forecast horizon for linear and polynomial VIX predictive regressions. The regressions $Rx^i_{t+h} = a^i_0 + a^i_1(vix_t) + \varepsilon^i_{t+h}$ and $Rx^i_{t+h} = b^i_0 + b^i_1(vix_t) + b^i_2(vix_t)^2 + b^i_3(vix_t)^3 + \varepsilon^i_{t+h}$ are each estimated by OLS for $h = 1, \ldots 24$, where i ranges over 1-year, 2-year, 3-year, 5-year, and 10-year Treasury excess returns and stock market excess returns. p-values for Wald tests of joint significance of slope coefficients $H_0 : b^i_1 = b^i_2 = b^i_3 = 0$ using Hodrick (1992) standard errors are reported. The sample period is 1990:1-2014:9.

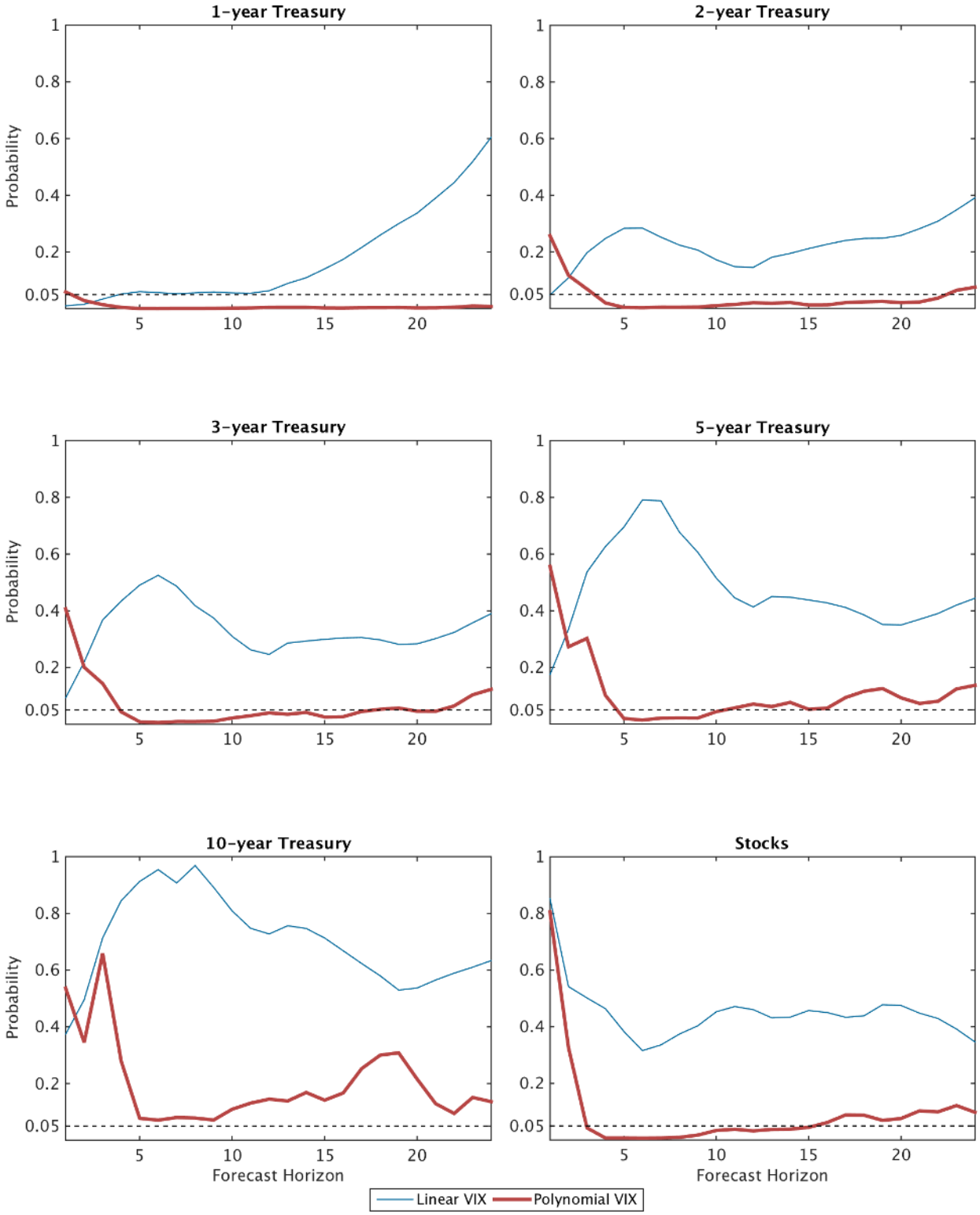


Figure 3: **P-values by Forecast Horizon: 1990 to 2007**

This figure plots p-values by forecast horizon for linear and polynomial VIX predictive regressions. The regressions $Rx^i_{t+h} = a^i_0 + a^i_1(vix_t) + \varepsilon^i_{t+h}$ and $Rx^i_{t+h} = b^i_0 + b^i_1(vix_t) + b^i_2(vix_t)^2 + b^i_3(vix_t)^3 + \varepsilon^i_{t+h}$ are each estimated by OLS for $h = 1, \ldots 24$, where i ranges over 1-year, 2-year, 3-year, 5-year, and 10-year Treasury excess returns and stock market excess returns. p-values for Wald tests of joint significance of slope coefficients $H_0 : b^i_1 = b^i_2 = b^i_3 = 0$ using Hodrick (1992) standard errors are reported. The sample period is 1990:1-2007:7.

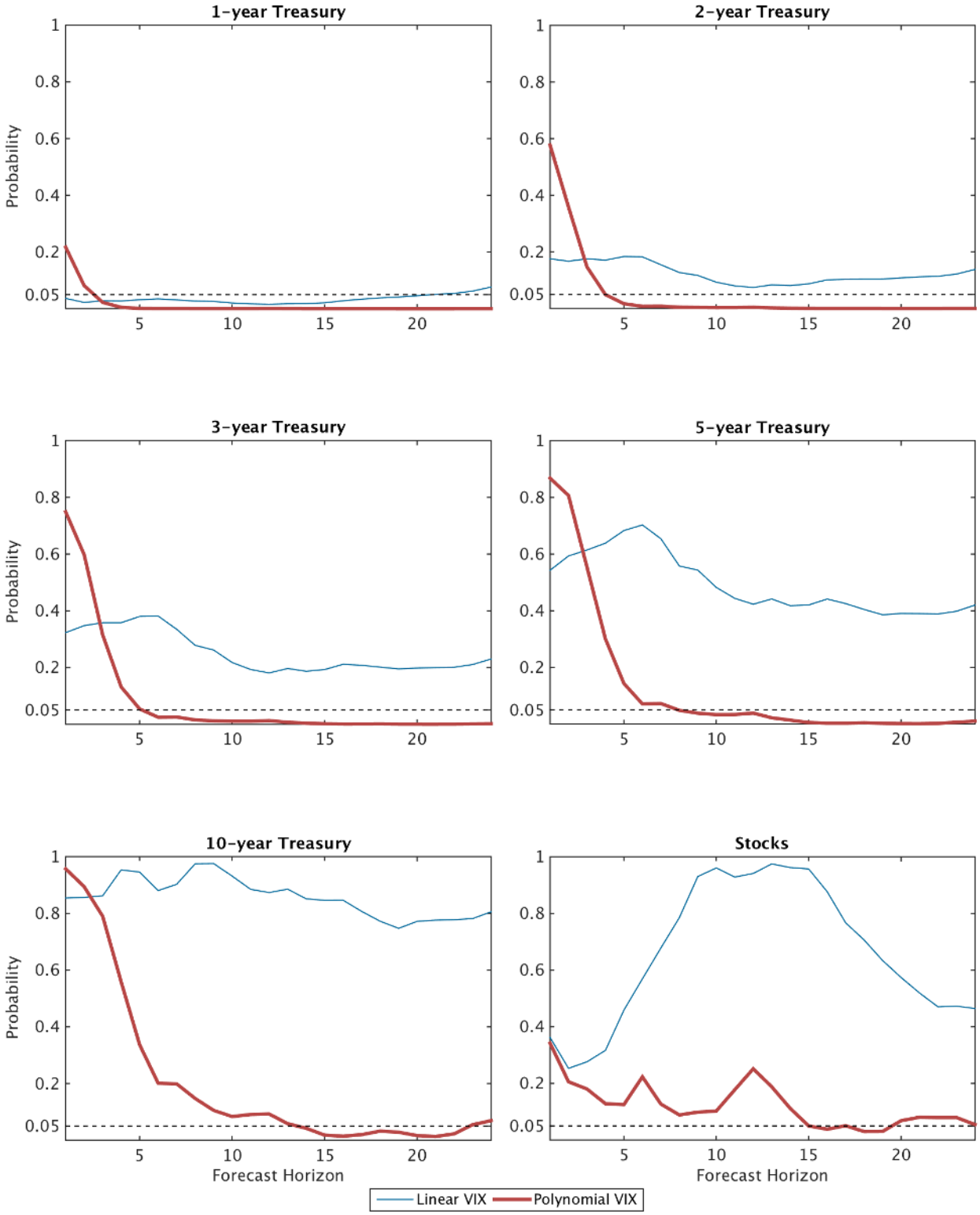


Figure 4: **Univariate Nonparametric and Polynomial Estimates of $\phi_h^i(v)$**

This figure shows sieve, polynomial, and kernel estimates of the nonlinear volatility function $\phi_h^i(v)$ from univariate predictive regressions $Rx_{t+h}^i = \phi_h^i(v_t) + \varepsilon_{t+h}$, where $E_t[Rx_{t+h}^i] = \phi_h^i(v_t)$. The superscript i indexes separate regressions in which the left hand side variable is either equity market excess returns or 1-year Treasury excess returns. In the top panel, the sample consists of monthly observations on $v_t = VIX_t$ from 1990:1 to 2014:9, whereas in the bottom panel, the sample consists of monthly observations from 1990:1 to 2007:7. In both panels, the forecast horizon plotted is $h = 6$ months. Within each panel, the left plot shows the nonparametric sieve estimate of $\phi_h^i(v_t)$, where the number of B-spline basis functions used in the estimation is chosen by out-of-sample cross validation. The middle plot shows a parametric cubic polynomial regression where $\phi_h^i(v_t) = a_0^i + a_1^i v_t + a_2^i v_t^2 + a_3^i v_t^3$, and the right plot shows $\phi_h^i(v_t)$ estimated by a Nadaraya-Watson kernel regression, where the bandwidth was chosen by Silverman's rule of thumb. The y-axis was rescaled by the unconditional standard deviation of Rx_t^i to display risk-adjusted returns.

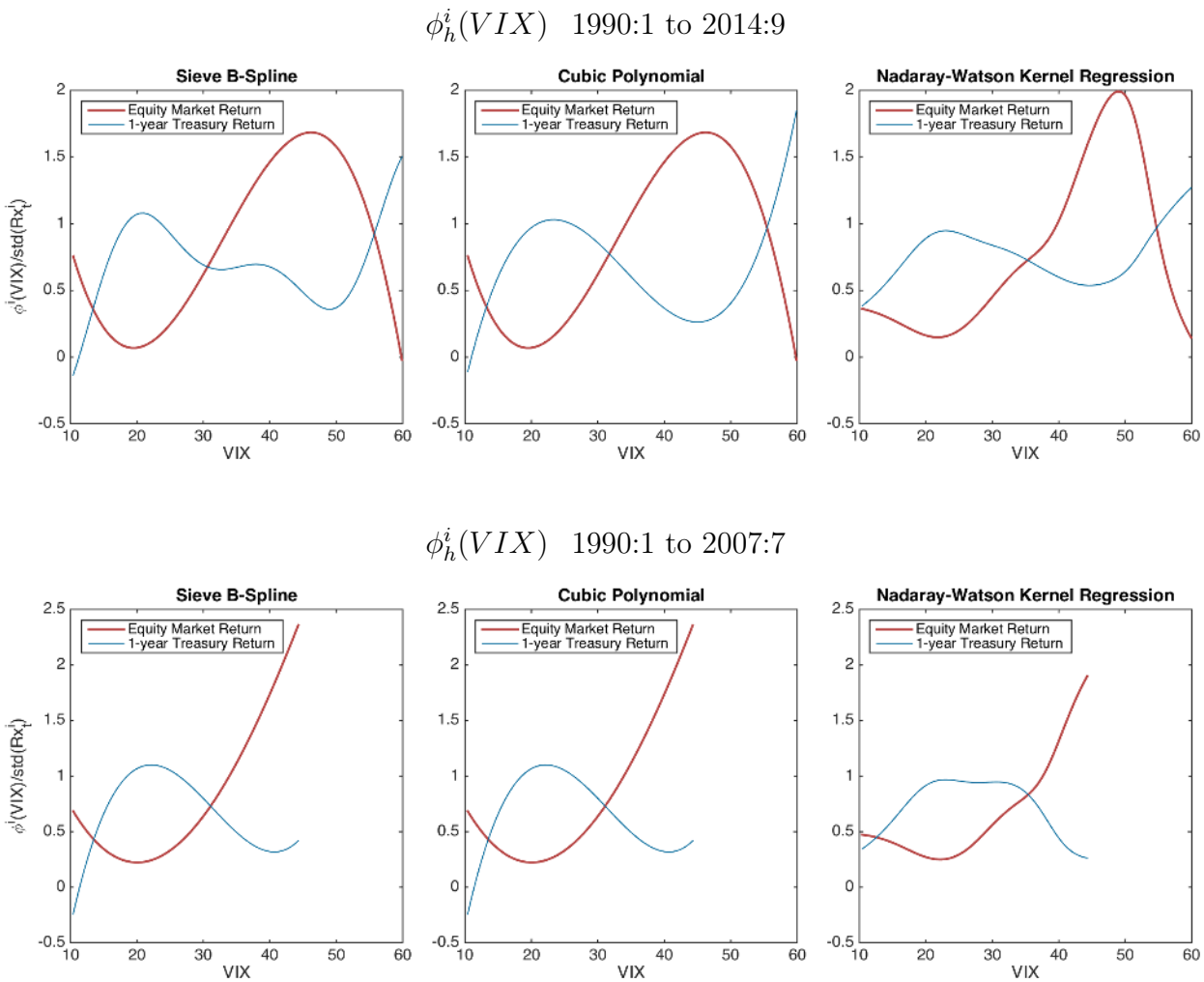


Figure 5: **SRR Regression Loadings:** $\hat{b}_h^i$

This figure plots SRR regression estimated portfolio loadings $\hat{b}_h^i$, where i ranges over the market return (MKT), 11 industry portfolio returns, constant-maturity Treasury returns with 1, 2, 5, 7, 10, 20, and 30 year maturities, and Barclays corporate bond portfolios. The shades of bars denote outcomes of the hypothesis test $\mathbb{H}_{2,0} : b_h^i \phi_h = 0$ of whether asset i loads signficantly on $\phi_h(v)$ (see Theorem 1). The samples consist of monthly observations from 1990:1 to 2014:9.

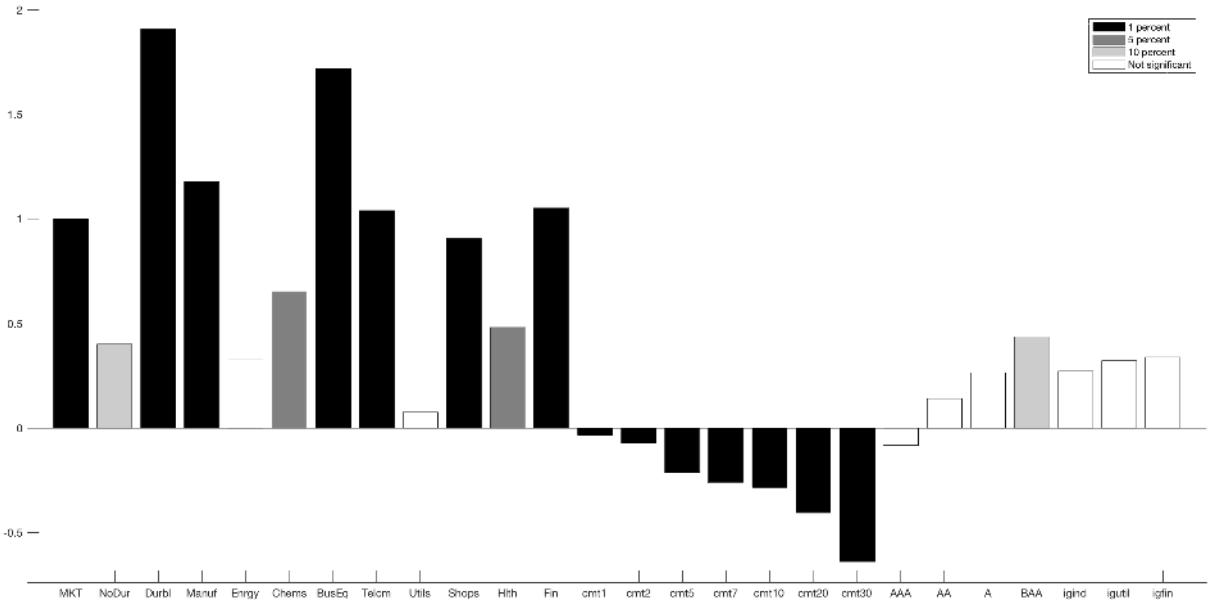


Figure 6: **SRR Regression: Expected Excess Returns for i^{th} Portfolio**

This figure plots normalized SRR regression estimated excess returns on portfolio i, $\hat{E}_t[Rx^i_{t+h}]/\hat{\sigma}(Rx^i_{t+h})$, where $\hat{E}_t[Rx^i_{t+h}] = \hat{\alpha}^i_h + \hat{b}^i_h\hat{\phi}_h(v_t)$, $\hat{\sigma}(Rx^i_{t+h})$ scales by unconditional excess return standard deviation, and where i ranges over the market return (MKT), 11 industry portfolio returns, constant-maturity Treasury returns with 1, 2, 5, 7, 10, 20, and 30 year maturities, and Barclays corporate bond portfolios. Red lines denote estimated excess returns with positive $\hat{b}^i_h$ loadings, and blue dashed lines denote estimated excess returns with negative $\hat{b}^i_h$ loadings. The forecast horizon is $h = 6$, and the sample consists of monthly observations from 1990:1 to 2014:9.

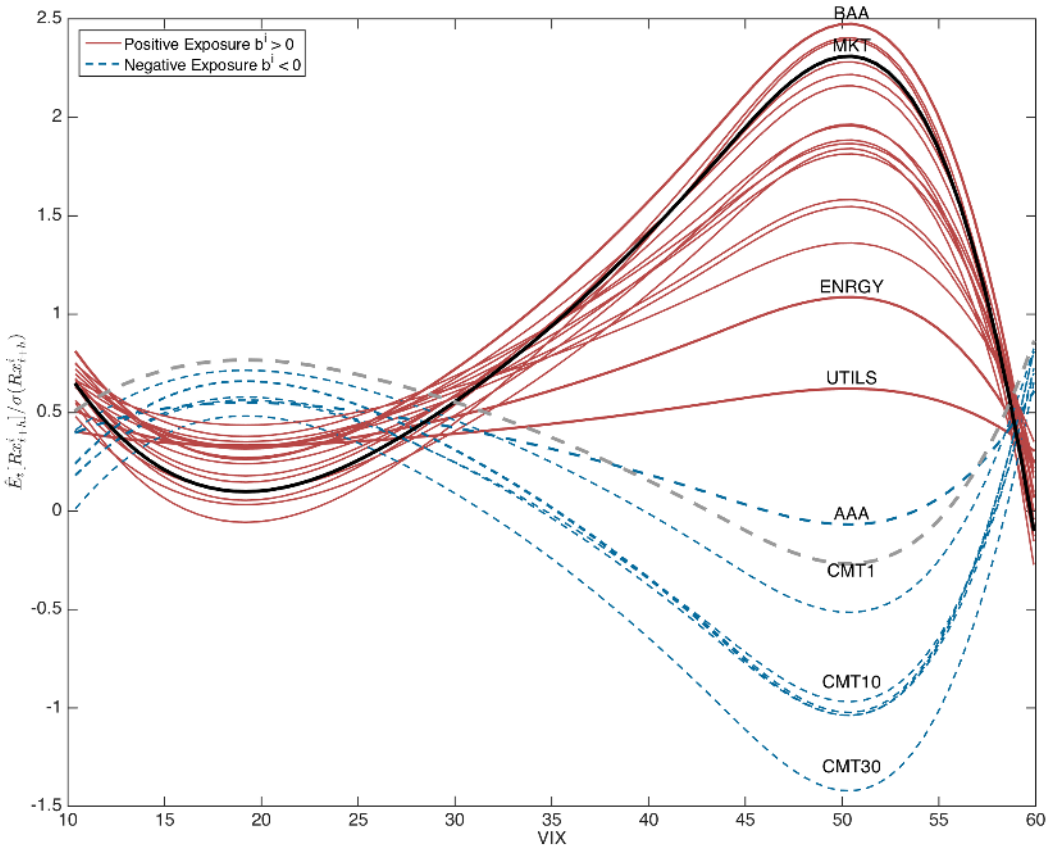


Figure 7: **SRR Regression $\phi_h(v)$: Separately Estimated from Equities and Treasuries**

This figure plots two versions of $\hat{\phi}_h(v)$ based on SRR regressions. The first estimates $\hat{\phi}_h(v)$ from the sieve reduced rank regression $Rx^i_{t+h} = a^i_h + b^i_h \phi_h(v) + \varepsilon^i_{t+h}$ where i ranges over equity industry and market portfolios only (red dashed). The second estimate of $\phi_h(v)$ comes from a sieve reduced rank regression, but where i ranges over Treasury return portfolios only (blue). The figure examines whether the two resulting nonlinear volatility functions $\hat{\phi}^{\text{Treas}}_h(v)$ and $\hat{\phi}^{\text{Equity}}_h(v)$ differ only by location and scale. This is tested by regressing $\hat{\phi}^{\text{Equity}}_h(v_t)$ on $\hat{\phi}^{\text{Treas}}_h(v_t)$ and a constant, and then plotting $\hat{\phi}^{\text{Equity}}_h(v)$ and $\hat{c}_1 + \hat{c}_2\hat{\phi}^{\text{Treas}}_h(v)$ alongside each other, where $\hat{c}_1$ and $\hat{c}_2$ are the regression coefficients. Dotted lines represent 95-percent confidence intervals from test $\mathbb{H}_{3,0}$ in Theorem 1 in the text. The forecast horizon is $h = 6$, and the sample consists of monthly observations from 1990:1 to 2014:9.

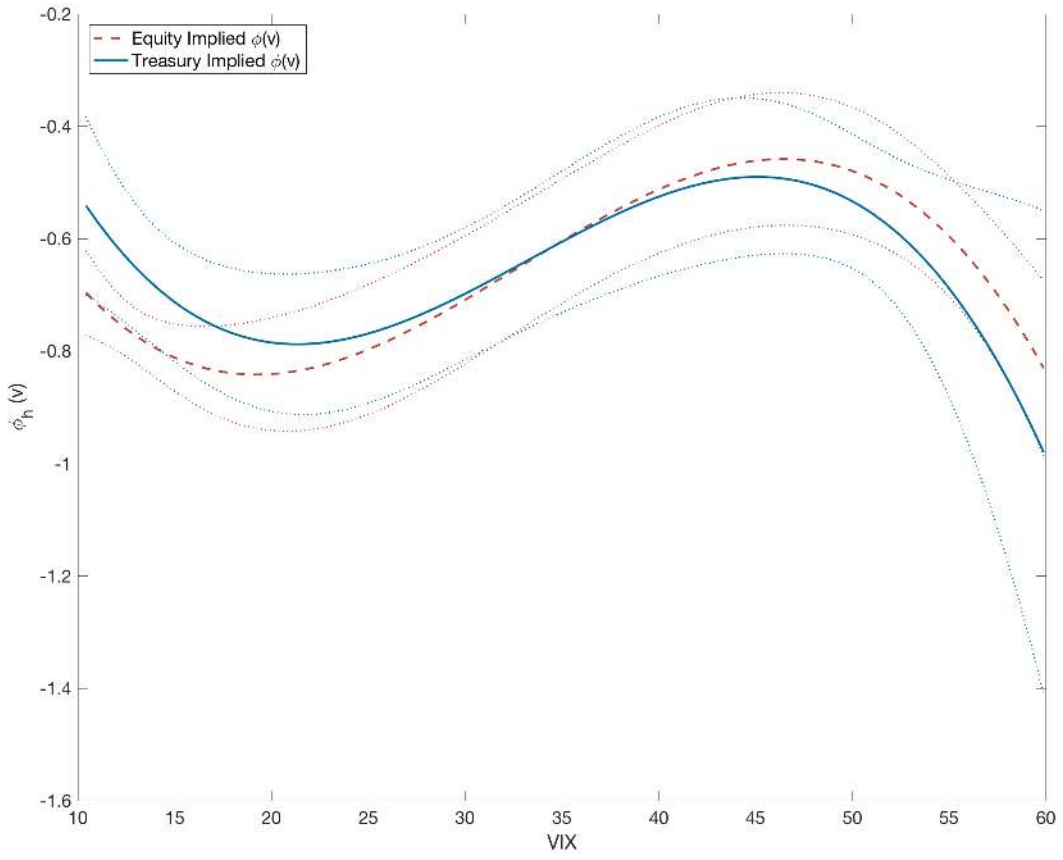


Figure 8: **Cross-Sectional Pricing**

This figure plots the results of the unrestricted joint forecasting regressions $Rx^i_{t+1} = a^i + b^i\phi(v_t) + \varepsilon^i_{t+1}$ against the restricted joint forecasting regressions $Rx^i_{t+1} = (\alpha^i + \beta^i\lambda_0) + \beta^i\lambda_1\phi(v_t) + \beta^i u_{t+1} + \varepsilon^i_{t+1}$, obtained from a dynamic asset pricing model with affine prices of risk. The innovations $u_{t+1} = Y_{t+1} - E_t[Y_{t+1}]$ correspond to the cross-sectional pricing factors $Y_t = (MKT_t, TSY1_t, \phi(v_t))$, where MKT is the return to the CRSP value-weighted equity market return, TSY1 is the return to the one year Treasury, and $\phi(v_t)$ is the nonlinear pricing factor for $v_t = vix_t$. Excess returns over i refer to 11 equity portfolios sorted by industry from Ken French's website, seven maturity-sorted Treasury portfolios, the six Barclay's industry and ratings sorted corporate bond portfolios, and the CRSP market return. To obtain estimates, the unrestricted regression is estimated by sieve reduced rank regression, yielding parametric estimates of a^i and b^i and a nonparametric estimate of $\phi(v_t)$. Then the restricted joint forecasting regression is estimated by taking $\phi(v_t)$ as given, making a^i and b^i from the unrestricted forecasts directly comparable to $(\alpha^i + \beta^i\lambda_0)$ and $\beta^i\lambda_1$ in the restricted regressions. The comparisons are scattered in the plots. The sample consists of monthly observations from 1990:1 to 2014:9.

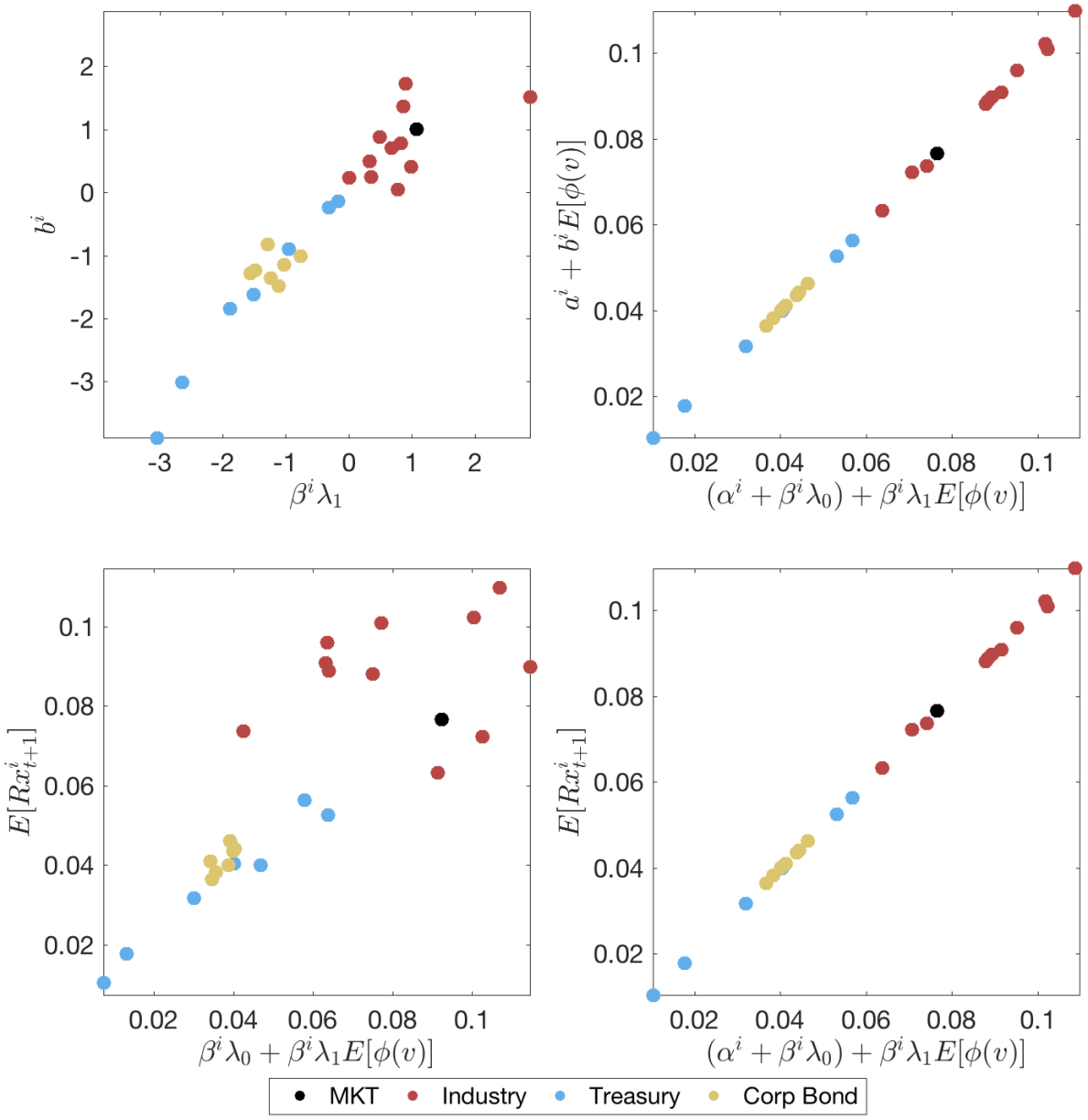


Figure 9: **Comparison to other Variables**

This figure plots a nonlinear function of the VIX, $\phi_6(v_t)$, estimated by sieve reduced rank regression against known price of risk, liquidity, and uncertainty factors: the 10-year Treasury yield (TSY10), the term spread between the 10-year and 3-month Treasury yield (TERM), the Cochrane-Piazzesi Factor (CP), the spread between Moody's Baa-rated corporate bonds and the 10-year Treasury yield (DEF), the log dividend yield (DY), the Lettau-Ludvigson CAY factor (CAY), the Kyle-Obizhaeva illiquidity factor (ILL), and the Jurado-Ludvigson-Ng macro uncertainty variable. To facilitate visual comparisons, all variables are demeaned and scaled by their unconditional standard deviations. The sample period is 1990:1-2014:9.

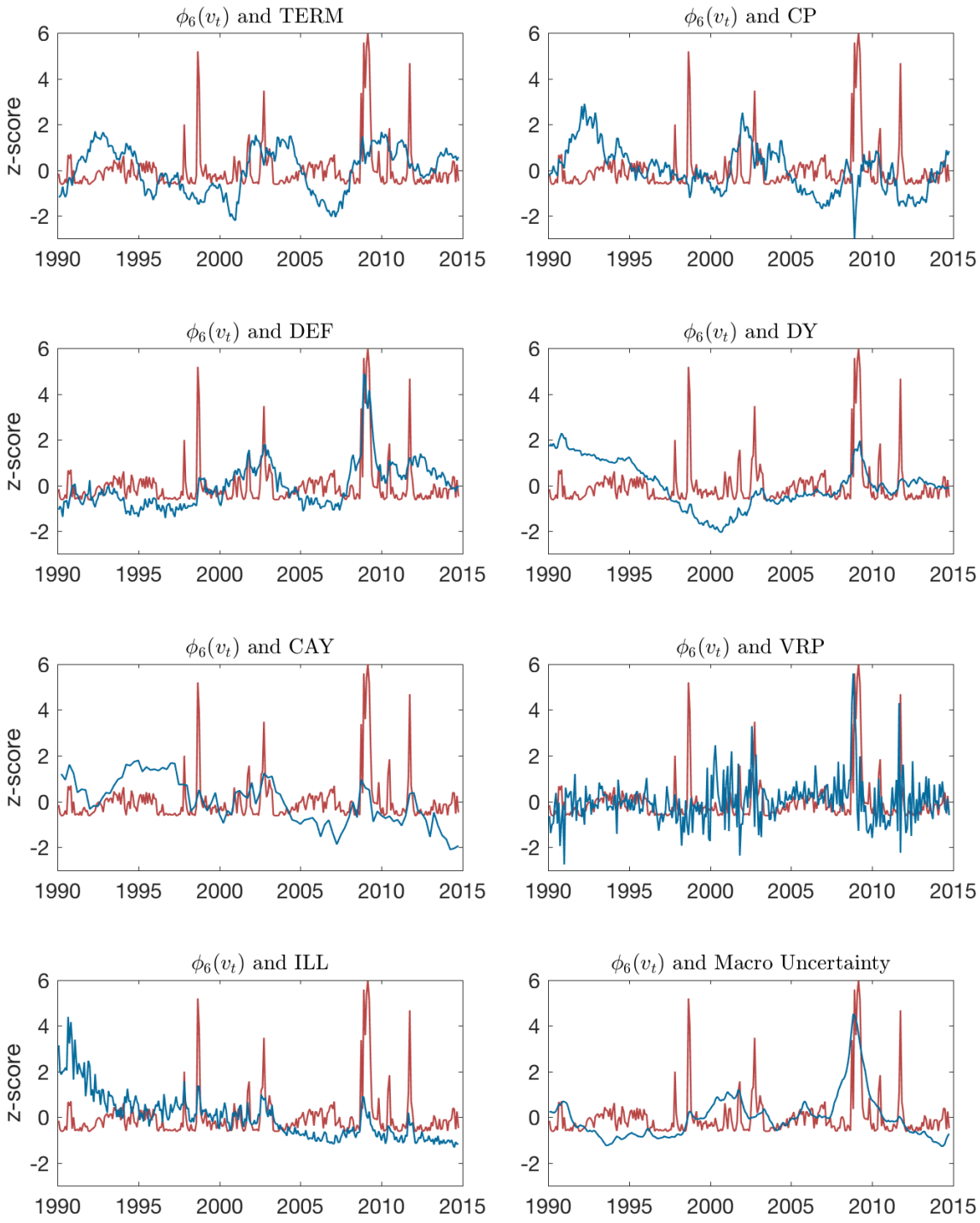


Figure 10: **Fund Flows, $\phi_h(vix_t)$, and Flight-to-Safety**

Using a panel of mutual fund flows, we estimate sieve reduced rank regressions $Flows^i_t = a^i + b^i\phi^{FF}(vix_t)+\varepsilon^i_t$. The top left panel shows the predicted fund flows $\hat{Flows}^i_t = \hat{a}^i+\hat{b}^i\hat{\phi}^{FF}(vix_t)$ for the subset of fund types that were found to load significantly on $\phi^{FF}(vix_t)$ (Table 7). The dashed vertical line indicates when government bond funds experience net inflows. The top right panel shows $\hat{\phi}^{FF}(v)$ alongside $\hat{\phi}_h(v)$ and its 95% pointwise confidence intervals estimated separately from sieve reduced rank regressions $Rx^i_{t+h} = a^i_h + b^i_h\phi_h(v_t) + \varepsilon^i_{t+h}$, where $h = 6$ months and $i = 1,\ldots,n$ ranges over the CRSP value-weighted market excess return and the seven CRSP constant maturity Treasury excess returns corresponding to 1, 2, 5, 7, 10, 20, and 30 years to maturity. Given the different units in each regression, $\phi^{FF}(v)$ was affine-translated to the scale of $\phi_h(v)$. The sample period is 1990:1 to 2014:9. The bottom panel plots values of the VIX above its sample median (left axis, solid blue line) next to combined stock fund outflows and government bond fund inflows (right axis, dashed red line). Stock fund outflows are the sum of US equity, non-US equity, and hybrid equity mutual fund outflows. The sample consists of monthly observations from 2000:1 to 2014:12. Source: ICI Trends in Mutual Fund Activity.

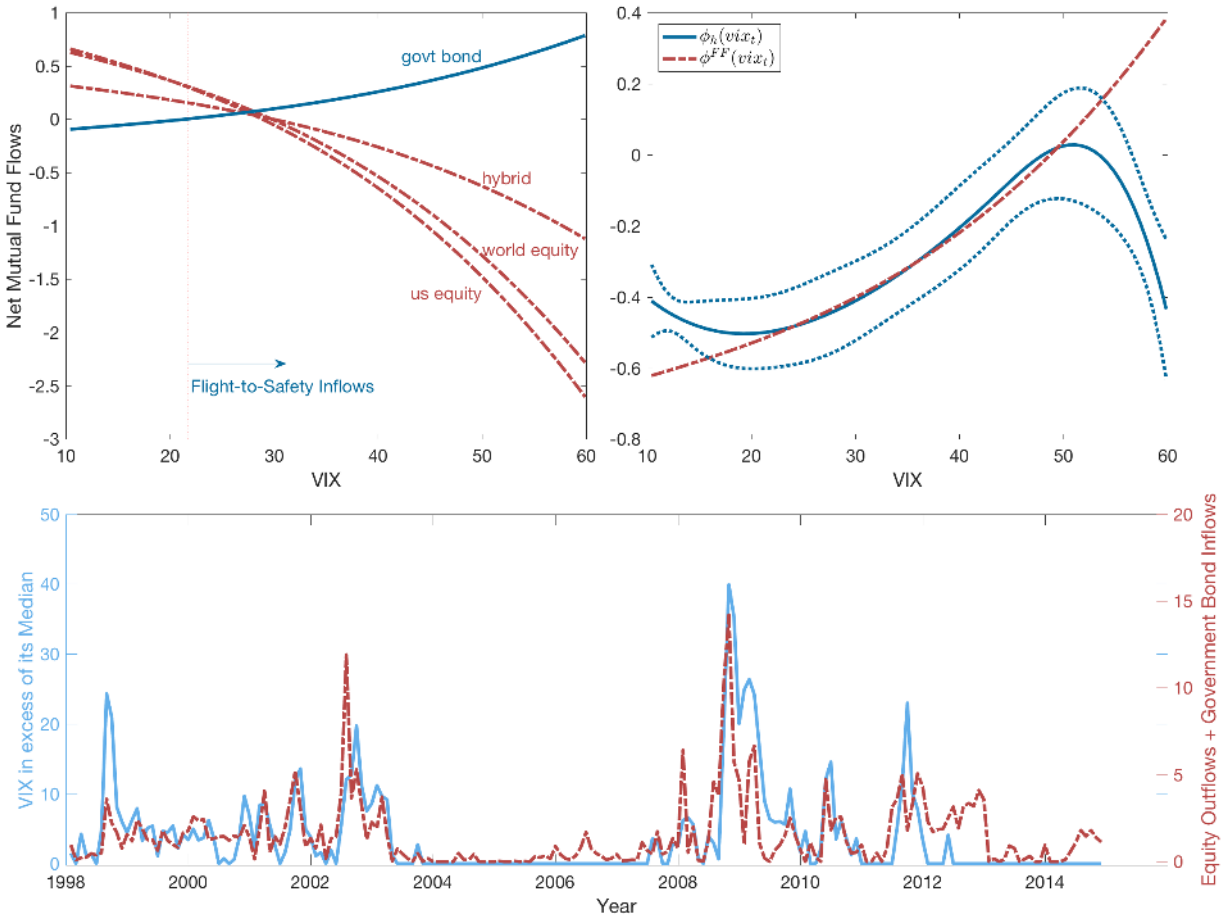
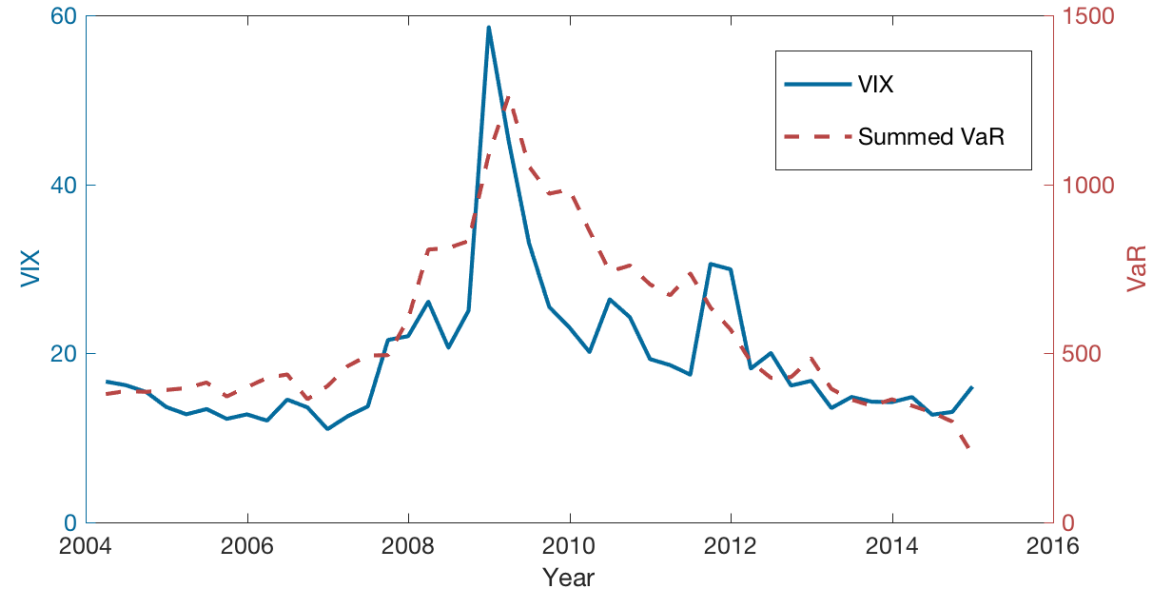


Figure 11: **SRR Regression: Value-at-Risk Cross-Section and the VIX**

The top panel of this figure plots the VIX (left axis, solid line) next to major dealer banks' summed Value-at-Risks (VaR) (right axis, dashed line) over time. The bottom panel shows the estimated $\phi^{VaR}(vix)$ from the contemporaneous sieve reduced rank regression of dealers' disaggregated VaRs on the VIX, $VaR_t^i = a^i + b^i\phi^{VaR}(vix_t) + \varepsilon_t^i$, where $i = 1, \ldots, 5$ indexes individual VaRs of the dealer banks that comprise the aggregate measure: Bank of America, Citigroup, Goldman Sachs, JP Morgan, and Morgan Stanley. Dotted lines represent 95-percent confidence intervals from test $\mathbb{H}_{3,0}$ in Theorem 1 in the text. The sample consists of quarterly observations from 2004:1 to 2014:4.



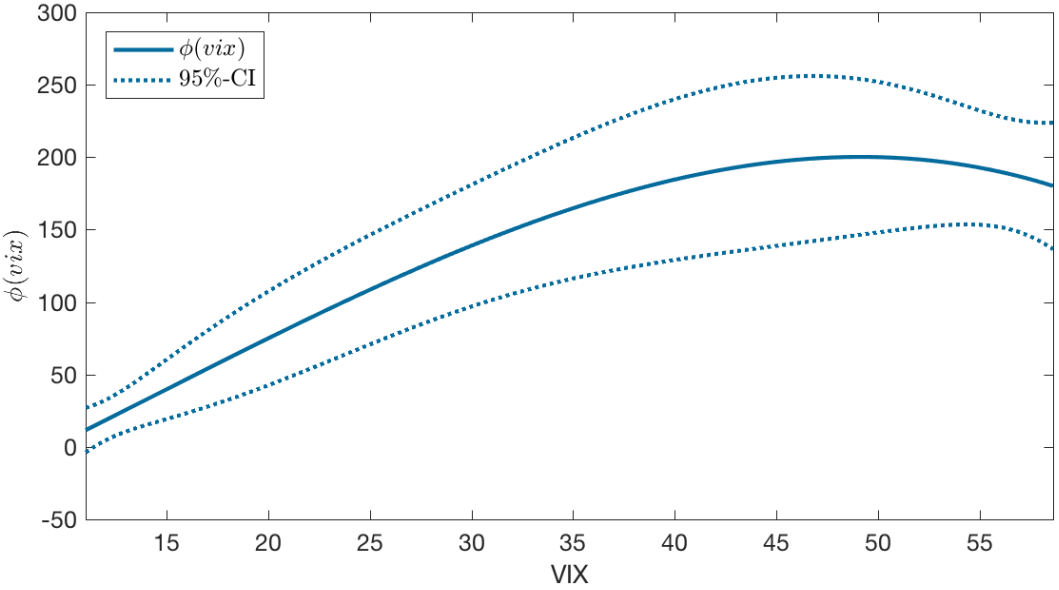


Figure 12: **Relation to Consumption-Based Asset Pricing**

The plot shows the SRR regression estimate of the nonlinear price of risk $\phi_h(vix_t)$, for $h = 6$, together with the maximum Sharpe ratio of the habit formation model by Campbell and Cochrane (1999, 2000) (top panel) and the long-run risk model of Bansal and Yaron (2004) (bottom panel). For the habit model, we use NIPA consumption data on nondurable goods and services and generate a time series s_t of the log surplus consumption ratio, yielding the maximal implied time-varying Sharpe Ratio as a function of s_t. For the long-run risk model, the maximal Sharpe ratio is a function of conditional consumption volatility, which is estimated as a 4-quarter moving average of squared consumption AR(1) innovations that have been projected on lagged innovations and the dividend yield. $\phi_h(vix_t)$ is plotted in solid red. The sample consists of quarterly observations from 1990:1 to 2014:4.

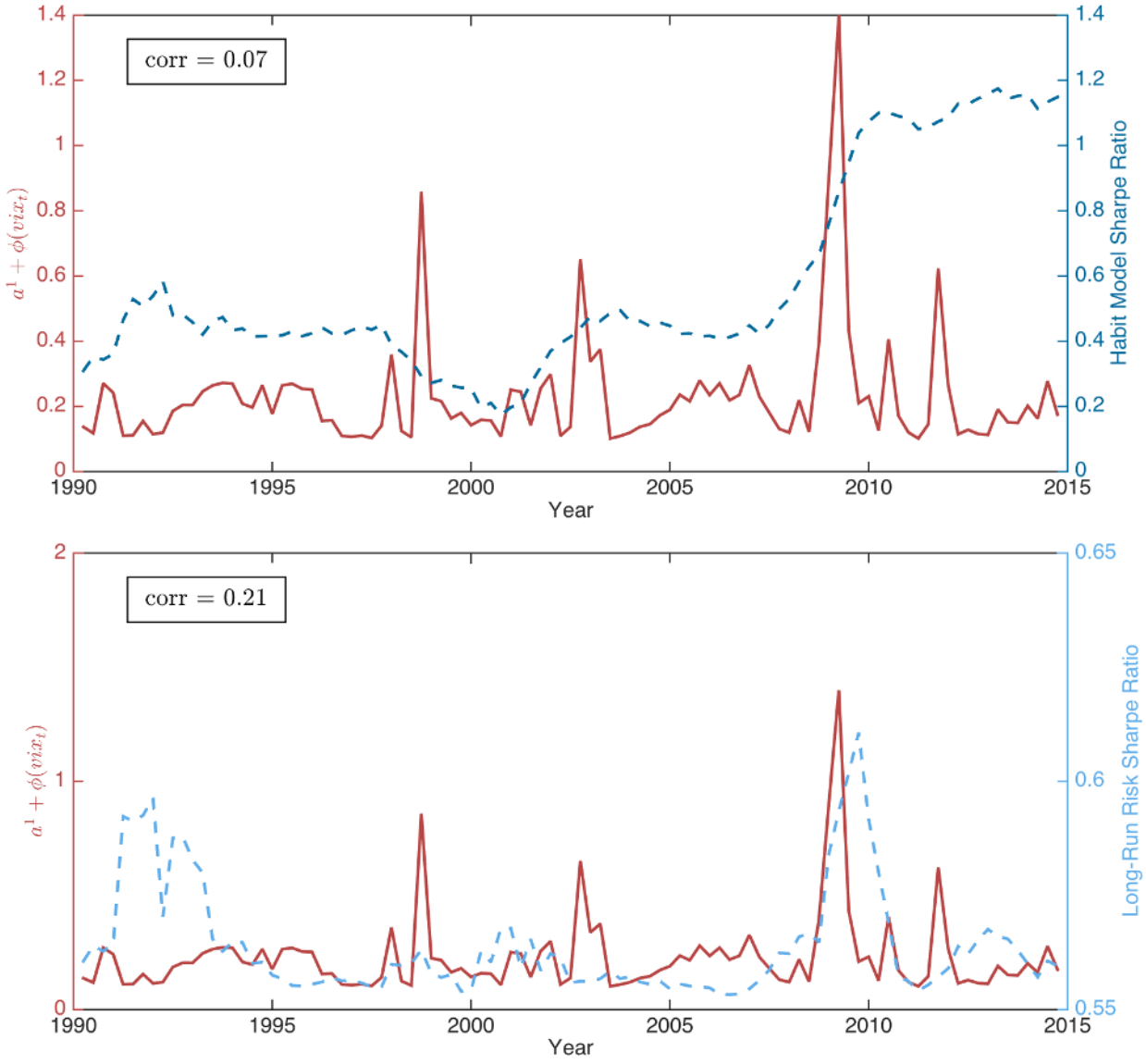


Figure 13: **Real Activity and** $\phi_h(vix_t)$

This plot shows the Federal Reserve Bank of Chicago's measure of real activity, the CFNAI-MA3, alongside $\phi(v_t)$ estimated from stocks and bonds alone. The sample consists of monthly observations from 1990:1 to 2014:9.

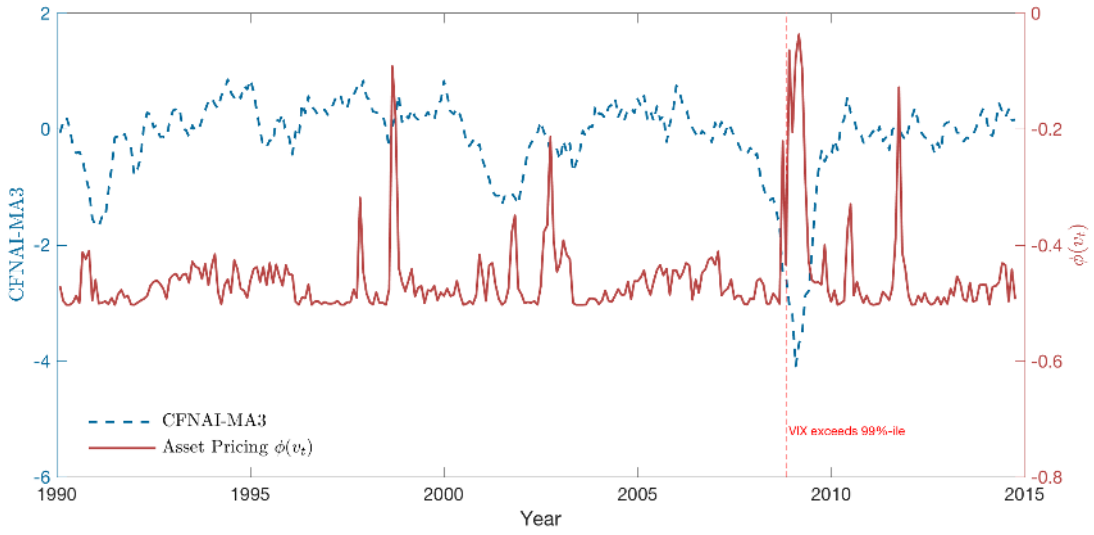


Table 1: **Excess Return Predictability: VIX and MOVE Polynomials: 1990 to 2014**

This table reports Hodrick (1992) t-statistics for coefficients from the regressions $Rx^i_{t+h} = a^i_h + \mathbf{b}^i_h(vix_t, vix^2_t, vix^3_t)' + \mathbf{c}^i_h(move_t, move^2_t, move^3_t)' + \mathbf{f}^i_h\mathbf{z_t} + \varepsilon^i_{t+h}$, where vix_t is the VIX equity implied volatility index at time t, and $move_t$ is the MOVE Treasury implied volatility index. The superscript i indexes separate regressions in which the left hand side variable is either the indicated Treasury excess return (top panel) or the equity market excess return (bottom panel). $\mathbf{z_t}$ consists of control variables representing the default spread between the 10-year Treasury yield and Moody's BAA corporate bond yield, the variance risk premium, the term spread between the 10-year and 3-month Treasury yield, and the log dividend yield. p-values report the outcome of the joint hypothesis test under the null of no predictability. The line labeled Linearity reports the p-values for the test of linearity in VIX, which corresponds to the null hypothesis that the coefficients on vix^2_t and vix^3_t are zero. The sample consists of monthly observations from 1990:1 to 2014:9.

1-year Treasury Excess Returns

	$h = 6$ months				$h = 12$ months				$h = 18$ months			
VIX^1	1.91	4.13		5.01	1.86	3.60		5.02	1.13	3.21		4.73
VIX^2		-4.08		-4.77		-3.61		-4.86		-3.37		-4.71
VIX^3		3.89		4.66		3.51		4.76		3.38		4.64
$MOVE^1$			0.26	-0.57			-0.58	-1.94			-0.23	-2.21
$MOVE^2$			0.17	0.81			1.00	2.14			0.54	2.43
$MOVE^3$			-0.35	-0.97			-1.16	-2.27			-0.69	-2.57
DEF				-0.99				-0.99				-1.23
VRP				2.49				2.84				3.87
TERM				-1.45				-2.76				-3.47
DY				3.34				3.58				3.24
const	1.40	-3.42	-0.14	0.58	1.65	-2.75	1.05	1.57	2.17	-2.19	1.19	1.63
p-value	0.057	0.001	0.089	0.000	0.064	0.005	0.303	0.000	0.260	0.004	0.845	0.000
Linearity		0.000		0.000		0.002		0.000		0.004		0.000

Stock Excess Returns

	$h = 6$ months				$h = 12$ months				$h = 18$ months			
VIX^1	1.00	-3.18		-2.68	0.74	-2.47		-1.98	0.78	-1.87		-1.35
VIX^2		3.36		2.90		2.61		2.22		2.01		1.54
VIX^3		-3.24		-2.68		-2.45		-2.15		-1.87		-1.48
$MOVE^1$			0.15	0.27			1.14	1.26			0.13	0.33
$MOVE^2$			-0.06	-0.23			-1.17	-1.35			-0.07	-0.41
$MOVE^3$			-0.01	-0.01			1.20	1.23			0.16	0.40
DEF				-0.58				-0.62				-0.34
VRP				-2.34				-2.03				-2.33
TERM				0.47				0.94				1.03
DY				1.17				1.51				1.63
const	-0.15	3.28	0.13	2.43	0.50	2.68	-0.65	2.06	0.58	2.10	0.15	2.01
p-value	0.316	0.007	0.991	0.018	0.460	0.032	0.674	0.075	0.439	0.088	0.810	0.112
Linearity		0.004		0.013		0.031		0.085		0.119		0.296

Table 2: **Nonlinear VIX Predictability using the Cross-Section: 1990 - 2014**

This table reports results from three predictive sieve reduced rank (SRR) regressions for each of $h = 6$, 12, and 18 month ahead forecasting horizons: (1) estimates of a_h^i and b_h^i from the regression $Rx_{t+h}^i = a_h^i + b_h^i v_t + \varepsilon_{t+h}^i$ of portfolio i's excess returns on linear $v_t = vix_t$; (2) estimates of a_h^i and b_h^i from the SRR regression $Rx_{t+h}^i = a_h^i + b_h^i \phi_h(v_t) + \varepsilon_{t+h}^i$ of portfolio i's excess return on the common nonparametric function $\phi_h(\cdot)$ of $v_t = vix_t$; (3) the same regression augmented with controls $\mathbf{f^i} \cdot (DEF_t, VRP_t, TERM_t, DY_t)'$ representing the default spread (DEF, 10-year Treasury yield minus Moody's BAA corporate bond yield), the variance risk premium (VRP, realized volatility minus VIX), the term spread (TERM, 10-year minus 3-month Treasury yields), and the S&P 500's (log) dividend yield. The index $i = 1, \ldots, n$ ranges over the CRSP value-weighted market excess return and the seven CRSP constant maturity Treasury excess returns corresponding to 1, 2, 5, 7, 10, 20, and 30 years to maturity. The sieve reduced rank regressions are introduced in Section 3 in the text. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a_i and $\mathbf{f^i}$ and for the χ^2-statistic on $b^i \phi_h(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $\mathbf{A}_h = 0$.

Horizon h = 6

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	−0.01	1.00	1.00*	1.00***	0.31	1.00***	0.05**	−1.42***	−0.01	0.17
cmt1	0.00	0.07*	−0.05*	−0.07***	−0.09**	−0.20***	0.00	0.03*	0.00*	0.02***
cmt2	0.01	0.09	−0.11*	−0.14***	−0.15*	−0.32***	0.00	0.08**	0.00	0.02**
cmt5	0.03	0.04	−0.26	−0.31***	−0.25	−0.60***	−0.02*	0.23**	0.01**	0.01
cmt7	0.04	0.04	−0.31	−0.38***	−0.27	−0.70***	−0.03**	0.32**	0.02**	0.00
cmt10	0.05	−0.08	−0.30	−0.37***	−0.25	−0.66***	−0.03**	0.39**	0.03***	0.01
cmt20	0.08	−0.22	−0.39	−0.49**	−0.23	−0.74**	−0.05***	0.51*	0.05***	0.03
cmt30	0.10	−0.52	−0.58	−0.68**	−0.29	−0.98**	−0.07***	0.70*	0.06***	0.06
Joint p-val		0.465		0.000		0.001				
Bootstrap p-val		0.494		0.018		0.001				

Horizon h = 12

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.64	1.00**	0.09	1.00***	0.03**	−0.70***	0.00*	0.18
cmt1	0.00	0.11*	−0.05	−0.10***	−0.08	−0.36***	0.00	0.03**	0.00**	0.02***
cmt2	0.01	0.22	−0.08	−0.18***	−0.12	−0.57***	0.00	0.05**	0.00	0.03***
cmt5	0.02	0.33	−0.18	−0.39***	−0.21	−1.03***	−0.01	0.08*	0.01	0.03
cmt7	0.02	0.35	−0.23	−0.48***	−0.24	−1.23***	−0.02*	0.12*	0.01**	0.02
cmt10	0.03	0.15	−0.22	−0.47***	−0.25	−1.25***	−0.02**	0.13*	0.02**	0.03
cmt20	0.05	0.12	−0.31	−0.65**	−0.27	−1.52**	−0.04**	0.16	0.03***	0.00
cmt30	0.06	−0.17	−0.44	−0.88**	−0.33	−1.92**	−0.05**	0.21	0.04***	0.01
Joint p-val		0.456		0.002		0.002				
Bootstrap p-val		0.472		0.038		0.000				

Horizon h = 18

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.44	1.00**	−0.03	1.00**	0.02**	−0.59***	0.01	0.18
cmt1	0.01	0.07	−0.04	−0.13***	−0.06	−0.60***	0.00	0.04***	0.00**	0.02***
cmt2	0.01	0.19	−0.07	−0.24***	−0.10	−0.95***	0.00	0.07***	−0.01	0.03***
cmt5	0.02	0.36	−0.14	−0.48**	−0.18	−1.65***	−0.01	0.11**	0.00	0.03**
cmt7	0.02	0.38	−0.16	−0.56**	−0.19	−1.87***	−0.01*	0.14**	0.00	0.03*
cmt10	0.03	0.23	−0.15	−0.52**	−0.20	−1.90***	−0.01*	0.15**	0.01*	0.04
cmt20	0.04	0.29	−0.19	−0.69*	−0.20	−2.16*	−0.02*	0.15	0.02**	0.01
cmt30	0.05	0.04	−0.28	−0.91**	−0.24	−2.63**	−0.03**	0.18	0.03**	0.00
Joint p-val		0.665		0.052		0.027				
Bootstrap p-val		0.662		0.071		0.001				

Table 3: **Nonlinear VIX Predictability using the Cross-Section: 1990 - 2007**

This table reports results from three predictive sieve reduced rank (SRR) regressions for each of $h = 6$, 12, and 18 month ahead forecasting horizons: (1) estimates of a_h^i and b_h^i from the regression $Rx_{t+h}^i = a_h^i + b_h^i v_t + \varepsilon_{t+h}^i$ of portfolio i's excess returns on linear $v_t = vix_t$; (2) estimates of a_h^i and b_h^i from the SRR regression $Rx_{t+h}^i = a_h^i + b_h^i \phi_h(v_t) + \varepsilon_{t+h}^i$ of portfolio i's excess return on the common nonparametric function $\phi_h(\cdot)$ of $v_t = vix_t$; (3) the same regression augmented with controls $\mathbf{f^i} \cdot (DEF_t, VRP_t, TERM_t, DY_t)'$ representing the default spread (DEF, 10-year Treasury yield minus Moody's BAA corporate bond yield), the variance risk premium (VRP, realized volatility minus VIX), the term spread (TERM, 10-year minus 3-month Treasury yields), and the S&P 500's (log) dividend yield. The index $i = 1, \ldots, n$ ranges over the CRSP value-weighted market excess return and the seven CRSP constant maturity Treasury excess returns corresponding to 1, 2, 5, 7, 10, 20, and 30 years to maturity. The sieve reduced rank regressions are introduced in Section 3 in the text. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a_i and $\mathbf{f^i}$ and for the χ^2-statistic on $b^i \phi_h(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $\phi_h = 0$.

Horizon h = 6

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.72	1.00**	0.25	1.00	−0.03	−0.84*	0.00	0.12
cmt1	0.00	0.23**	−0.09	−0.16***	−0.15	−0.59***	0.01	0.03	0.00*	0.03***
cmt2	0.00	0.29	−0.16	−0.28***	−0.22	−0.92***	0.01	0.10*	0.00	0.03***
cmt5	0.01	0.30	−0.31	−0.53***	−0.34	−1.57***	0.00	0.25*	0.01	0.03*
cmt7	0.03	0.18	−0.36	−0.62***	−0.35	−1.75***	−0.01	0.31*	0.02	0.02
cmt10	0.04	−0.23	−0.37	−0.62**	−0.32	−1.74***	−0.03	0.36*	0.02*	0.02
cmt20	0.06	−0.16	−0.42	−0.72**	−0.31	−1.95**	−0.05	0.46**	0.03**	−0.01
cmt30	0.05	−0.27	−0.51	−0.85*	−0.35	−2.25**	−0.06	0.58**	0.04**	−0.03
Joint p-val		0.045		0.000		0.003				
Bootstrap p-val		0.086		0.005		0.002				

Horizon h = 12

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.09	1.00	0.46	1.00*	−0.01	1.00	−0.03	−0.52*	0.00	0.16
cmt1	0.00	−2.09**	−0.08	−0.23***	−0.12	3.25***	0.00	0.03	0.00	0.03***
cmt2	0.00	−3.21*	−0.13	−0.40***	−0.18	5.27***	0.00	0.07*	0.00	0.03***
cmt5	0.00	−4.69	−0.24	−0.72***	−0.31	9.32***	0.00	0.12	0.01	0.04
cmt7	0.01	−3.98	−0.28	−0.84***	−0.33	10.77***	−0.01	0.13	0.01*	0.03
cmt10	0.03	−0.71	−0.27	−0.80***	−0.33	11.27***	−0.03*	0.14	0.02**	0.04
cmt20	0.04	−1.06	−0.33	−1.00***	−0.34	13.10***	−0.04**	0.14	0.03**	0.01
cmt30	0.04	−1.15	−0.40	−1.17**	−0.39	15.27***	−0.06**	0.16	0.04***	0.00
Joint p-val		0.010		0.000		0.002				
Bootstrap p-val		0.036		0.024		0.033				

Horizon h = 18

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.11	1.00	0.57	1.00***	0.13	1.00	−0.03	−0.54**	0.00	0.13
cmt1	0.00	−0.36**	−0.07	−0.17***	−0.11	−0.92***	0.00	0.05**	0.00	0.03***
cmt2	0.00	−0.59	−0.13	−0.30***	−0.19	−1.55***	0.00	0.11**	0.00	0.03***
cmt5	0.00	−0.93	−0.24	−0.55***	−0.35	−2.83***	0.01*	0.19**	0.00	0.05*
cmt7	0.01	−0.86	−0.27	−0.63***	−0.39	−3.28***	0.00**	0.22**	0.01*	0.05
cmt10	0.03	−0.25	−0.25	−0.59***	−0.40	−3.44***	−0.01**	0.24**	0.01**	0.06
cmt20	0.04	−0.45	−0.29	−0.70***	−0.41	−3.81***	−0.02**	0.20*	0.02**	0.03
cmt30	0.03	−0.43	−0.37	−0.85***	−0.50	−4.55***	−0.03**	0.25**	0.03***	0.03
Joint p-val		0.016		0.000		0.000				
Bootstrap p-val		0.057		0.011		0.003				

Table 4: **An LM Test of Flight-to-Safety**

This table reports the test statistic, $\hat{\mathcal{T}}_4$, and associated p-values of the test of whether expected stock and bond returns load on the same nonlinear function $\phi_h(v_t)$. The columns of the table correspond to three predictive sieve reduced rank regressions (SRRR) for each of $h = 6$, 12, and 18 month ahead forecasting horizons. The sieve reduced rank regressions are introduced in Section 3 in the text and the limiting distribution of $\hat{\mathcal{T}}_4$ is characterized in Theorem 2. For reference, the 90%, 95% and 99% quantile of a chi-square random variable with 14 degrees of freedom are 21.0641, 23.6848, and 29.1412. The full sample period is 1990:1-2014:9 and the pre-crisis sample period is 1990:1-2007:7.

	SRRR LM Test Statistic		
	Horizon $h = 6$	Horizon $h = 12$	Horizon $h = 18$
Full sample	19.914	20.094	20.204
	(0.133)	(0.127)	(0.124)
Pre-crisis	18.409	19.566	20.973
	(0.189)	(0.144)	(0.102)

Table 5: **Forecast Error Comparison: Mean-Only Model**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a mean-only model. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the running mean forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h \in \{6, 12, 18\}$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. We report the log of the ratio of unadjusted root MSFEs and mean absolute forecast errors (MAFEs), where the mean-only model is in the denominator. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Forecast Error Comparison						
	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
Log Ratio of RMSFEs, MAFEs (h=6)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	−0.01	0.00	0.00	0.01	0.00	0.00
cmt1	0.18	0.04	0.21	0.06	0.26	0.06
cmt2	0.16	0.06	0.21	0.08	0.23	0.05
cmt5	0.07	0.02	0.10	0.02	0.09	−0.01
cmt7	0.03	0.00	0.05	0.01	0.03	−0.03
cmt10	0.00	−0.02	0.02	−0.01	0.00	−0.04
cmt20	−0.02	−0.03	−0.01	−0.02	−0.02	−0.04
cmt30	−0.02	−0.03	−0.01	−0.03	−0.02	−0.04
Log Ratio of RMSFEs, MAFEs (h=12)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	−0.06	−0.03	−0.06	−0.01	−0.07	−0.01
cmt1	0.00	−0.01	0.01	0.00	−0.05	−0.05
cmt2	0.00	0.00	0.03	0.03	−0.05	−0.04
cmt5	−0.04	−0.06	−0.03	−0.05	−0.11	−0.16
cmt7	−0.05	−0.08	−0.04	−0.07	−0.11	−0.17
cmt10	−0.05	−0.07	−0.05	−0.09	−0.11	−0.18
cmt20	−0.06	−0.08	−0.07	−0.10	−0.09	−0.15
cmt30	−0.08	−0.08	−0.08	−0.09	−0.10	−0.13
Log Ratio of RMSFEs, MAFEs (h=18)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	−0.06	−0.03	−0.03	0.02	−0.03	0.01
cmt1	0.00	0.01	0.01	0.01	−0.03	−0.01
cmt2	0.00	0.01	0.04	0.07	0.00	0.04
cmt5	−0.02	−0.03	0.02	0.03	−0.03	−0.03
cmt7	−0.02	−0.04	0.01	0.01	−0.03	−0.06
cmt10	−0.02	−0.01	−0.01	0.00	−0.05	−0.07
cmt20	−0.05	−0.06	−0.04	−0.06	−0.06	−0.08
cmt30	−0.06	−0.06	−0.06	−0.07	−0.06	−0.06

Table 6: **Dynamic Asset Pricing: Factor Risk Exposures and Prices of Risk**

This table provides estimates of factor risk exposures and prices of risk from the dynamic asset pricing model $Rx^i_{t+1} = (\alpha^i + \beta^i\lambda_0) + \beta^i\lambda_1\phi(v_t) + \beta^i u_{t+1} + \varepsilon^i_{t+1}$, where i ranges over the test assets in the left column: MKT denotes the CRSP value-weighted market excess return, (NoDur ... Fin) are industry-sorted portfolio excess returns from Ken French's website, (cmt1, ..., cmt30) are constant maturity Treasury portfolio excess returns, and (AAA, ..., igfin) are Barclay's ratings and industry sorted corporate bond excess returns. In a first stage, $\phi(v_t)$ is estimated from a sieve reduced rank regression $Rx^i_{t+1} = a^i + b^i\phi(v_t) + \varepsilon^i_{t+1}$ jointly across i. The factor innovations $u_{t+1} = Y_{t+1} - E_t[Y_{t+1}]$ are then estimated from a VAR on the market (MKT), Treasury (TSY1), and nonlinear volatility factor ($\phi(v_t)$), for $v_t = vix_t$. In a second stage, coefficients are estimated jointly across all $i = 1, \ldots, n$ via a reduced rank regression. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

Exposures	β^i_{MKT}	β^i_{RF}	$\beta^i_{\phi(v)}$	$\beta^i\lambda_1$	$(\alpha^i + \beta^i\lambda_0)$
MKT	1.00***	−0.24***	0.02	1.08***	0.33***
NoDur	0.61***	0.45	0.01	0.49**	0.21***
Durbl	1.19***	−2.24**	1.90**	0.68	0.23
Manuf	1.09***	−0.86	0.85***	0.83***	0.30***
Enrgy	0.71***	−1.11	1.11	0.36	0.18
Chems	0.75***	−0.80	0.20	0.87***	0.30***
BusEq	1.44***	−1.76**	−1.47***	2.88***	0.80***
Telcm	0.94***	0.06	0.27	0.77*	0.25**
Utils	0.39***	0.10	0.59	0.01	0.08
Shops	0.86***	−0.64	0.10	0.99***	0.32***
Hlth	0.69***	1.04	0.12	0.33	0.18**
Fin	1.08***	1.26	−0.24	0.90**	0.30***
cmt1	0.00	0.73***	−0.07	−0.16**	−0.03*
cmt2	−0.01	1.40***	−0.14	−0.32**	−0.06*
cmt5	−0.03*	2.90***	0.16	−0.95***	−0.19***
cmt7	−0.04*	3.55***	0.77*	−1.52***	−0.32***
cmt10	−0.04	3.90***	1.19***	−1.88***	−0.41***
cmt20	−0.08*	4.76***	1.96***	−2.64***	−0.57***
cmt30	−0.12**	5.45***	2.23*	−3.03***	−0.67***
AAA	0.02	2.54***	0.68	−1.11***	−0.23***
AA	0.06***	2.28***	0.71	−1.02***	−0.21***
A	0.09***	2.00***	1.24***	−1.24***	−0.26***
BAA	0.12***	1.55***	1.58***	−1.29***	−0.26***
igind	0.08***	1.90***	1.67***	−1.48***	−0.31***
igutil	0.07**	1.99***	1.73***	−1.56***	−0.33***
igfin	0.11***	1.83***	0.57	−0.76***	−0.14**

Prices of Risk	*MKT*	*RF*	$\phi(v_t)$
λ_1	1.02***	−0.28**	−0.62**

Table 7: **Fund Flows and Nonlinear VIX**

This table reports results from the contemporaneous panel regressions of mutual fund flows into the funds of indicated type i (left column) on the common nonparametric function $\phi^{FF}(vix_t)$, $Flows_t^i = a^i + b^i\phi^{FF}(vix_t) + \varepsilon_t^i$. The panel regressions were estimated via the sieve reduced rank regressions introduced in Section 3 in the text. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a^i and for the χ^2-statistic on $b^i\phi^{FF}(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $(b^1, \ldots, b^n) \cdot \phi^{FF}(\cdot) = 0$.

	Sample: 1990 - 2014		Sample: 1990 - 2007	
	a^i	b^i	a^i	b^i
us equity	9361.10	−1.00***	4590.51	−1.00
world equity	8729.01	−0.89***	−2154.22	−1.86***
hybrid	4332.69	−0.44***	−994.43	−0.79***
corporate bond	1596.10*	−0.06	677.27***	0.01
HY bond	280.97	0.01	302.78	0.02
world bond	1976.57	−0.14*	−28.00	−0.06
govt bond	−1677.29	0.27***	2448.24	0.91***
strategic income	2591.58**	−0.02	2377.08	0.25
muni bond	575.98	0.01	384.06	−0.03
govt mmmf	−10 568.01	2.06	3836.77	0.96
nongovt mmmf	812.84	0.09	11 653.76	2.67
national mmmf	192.97	0.00	1410.59	0.21
state mmmf	639.50	−0.09***	560.80	0.06
Joint p-value		0.000		0.000

Table 8: **Business Cycle Panel Predictability**

This table reports results from two cross-sectional predictability regressions of various macro business cycle indicators on $h = 6$ month lagged functions of $v_t = vix_t$. The left panel is estimated on our full sample of monthly observations from 1990 to 2014, whereas the right panel is estimated on the pre-crisis subsample. The business cycle indicators represent the industrial production index (IP), the IP manufacturing index (IPMFG), the manufacturing capacity utilization index (CUMFG), the change in goods-producing employment (LAGOODA), and the total private nonfarm payroll series (LAPRIVA), which receive the largest weight within the Chicago Fed National Activity index. The leading reference asset is the market excess return. The left panel shows estimates of a_h^i and b_h^i from the sieve reduced rank regression $y_{t+h}^i = a_h^i + b_h^i \phi^{\text{macro}}(v_t) + \varepsilon_{t+h}^i$ of series i on the common nonparametric function $\phi_h^{\text{macro}}(\cdot)$. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a_i and for the χ^2-statistic on $b^i \phi_h^{\text{macro}}(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $(b^1, \ldots, b^n) \cdot \phi_h(\cdot) = 0$.

	Sample: 1990 - 2014		Sample: 1990 - 2007	
	a^i	b^i	a^i	b^i
MKT	−0.02	1.00	0.10	1.00
IP	0.01	−0.06***	0.00	−0.22***
IPMFG	0.01	−0.07***	0.00	−0.25***
CUMFG	0.00	−0.05***	−0.01	−0.26***
LAPRIVA	0.00	−0.04***	0.00	−0.14***
LAGOODA	0.01	−0.07***	−0.01	−0.26***
Joint p-value		0.000		0.000

Supplementary Appendix for "Nonlinearity and Flight-to-Safety in the Risk-Return Tradeoff for Stocks and Bonds"

Tobias Adrian, Richard K. Crump, & Erik Vogt

Supplementary Appendix

The Supplementary Appendix has five main sections. The first section, A.1, beginning on this page, provides details on the empirical implementation used throughout the paper including data sources. The section also provides additional results on forecast error comparisons to complement Section 3.4 and additional details on our theoretical setup. The second section, A.2, beginning on page 15, discusses the design and results of our Monte Carlo simulations. The third section, A.3, beginning on page 28, demonstrates the robustness of our results to using log(VIX) rather than VIX in our empirical applications. The fourth section, A.4, beginning on page 31, discusses the design and results of the block bootstrap resampling exercise we conducted. Finally, the fifth section, A.5, beginning on page 35, provides the proofs of the theoretical results introduced in the paper.

A.1 Details of Empirical Implementation

A.1.1 Data and Estimation

Our main results are based on excess returns for equities, Treasury bonds, and corporate bonds. For equities we use the value-weighted market return and the 12 industry-sorted portfolio returns, both from Ken French's website.[23] For Treasury bond returns we primarily use the constant maturity Treasury portfolios with maturities one, two, five, seven, ten, 20, and 30 years from the Center for Research in Securities Prices (CRSP). We also use returns (in Table 1 and Figures 2–4) and yields (Table 6 and Figure 8) from the Gurkaynak et al. (2007) data set available from the Board of Governors of the Federal Reserve's research data page.[24] For corporate bond returns, industry and rating sorted investment grade credit returns are obtained from Barclays. As our measure of the risk-free rate we use the CRSP 1-month risk-free rate. Our measure of equity volatility is the 1-month CBOE Volatility Index (VIX) (*FRED mnemonic*: `VIXCLS`) and for implied Treasury bond volatility is the 1-month Merrill Lynch Option Volatility Expectations (MOVE) index (*Haver mnemonic*: `SPMLV1@DAILY`).

The default spread is measured as the Moody's BAA corporate bond yield (*FRED mnemonic*: `DBAA`) less the 10-year Treasury yield (*FRED mnemonic*: `DGS10`) with each series obtained from the H15 statistical release from the Federal Reserve Board. The term spread is measured as the 10-year Treasury yield less the 3-month Treasury bill rate (*FRED mnemonic*: `DGS3MO`) each also from the H15 statistical release. The variance risk premium is calculated as in Bollerslev et al. (2009). We obtain the dividend yield of the S&P 500 from Haver and, in regressions, apply a log transformation to the series (*Haver mnemonic*: `SDY5COMM@USECON`). The CP factor is calculated as in Cochrane and Piazzesi (2005). The CAY factor is obtained from Martin Lettau's website.[25] The Kyle and Obizhaeva (2016) ILL factor is computed as a scaled ratio of S&P 500 intra-month realized volatility and corresponding total dollar volume. The realized volatility in this calculation is the square root of the sum of squared daily log returns within a calendar month.

[23] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[24] https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html
[25] http://faculty.haas.berkeley.edu/lettau/data_cay.html

All our main results are based on monthly data. To obtain monthly data we transform daily data to monthly data by using the last (non-missing) observation for the month. All data on returns are reported as annualized rates of return. Our full sample is January 1990–September 2014 while our pre-crisis sample covers January 1990–July 2007.

The B-spline basis in the paper is constructed via de Boor's algorithm (De Boor (1972)). Knots are placed with equal spacing on the empirical support of the VIX. With a choice of m we can estimate the parameters in the model by following the steps detailed in Section 3.1. In order to choose the specification for the nonparametric estimator of $\phi(\cdot)$, we use a mean-square forecast error (MSFE) cross validation procedure. The procedure is implemented as follows. We begin by splitting our full sample of T monthly observations into an initial test sample of $t^* = \lceil 0.6T \rceil$ observations, which are used to estimate the model. Beginning at $t = t^*$, we therefore obtain an estimate $\hat{E}_t^m[Rx_{t+h}^i]$, where $\hat{E}_t^m[\cdot]$ refers to a forecast conditional on m sieve basis functions and information up to time-t. The estimate is then compared to the realized Rx_{t+h}^i. We then keep incrementing t by one month and re-estimating the model, each time obtaining a forecast $\hat{E}_t^m[Rx_{t+h}^i]$ that we compare against the pseudo out-of-sample realization Rx_{t+h}^i. The number of sieve basis functions selected solves

$$\arg\min_m \sum_{t=t^*}^{T-h} \left(\mathrm{Rx}_{t+h} - \hat{E}_t^m[\mathrm{Rx}_{t+h}]\right)^2, \tag{A.1}$$

where we have stacked the test asset excess returns, $\mathrm{Rx}_{t+h} = (Rx_{t+h}^1, \ldots, Rx_{t+h}^n)'$.

In the paper, the number of B-spline basis functions m equals $s + k - 1$, where s is the assumed order of the spline and k is the number of interior knots. Following the terminology in (Li and Racine 2007, Chapter 15) or (Friedman et al. 2009, Chapter 5), the order of the spline is one plus the order of the piecewise polynomial between knots. For example, $s = 2$ corresponds to a piecewise linear spline, and $s = 4$ corresponds to a piecewise cubic spline. Under de Boor's algorithm, a cubic spline with 2 interior knots generates $s + k = 4 + 2 = 6$ basis functions. Because our regressions include a constant term, we drop one of the basis functions (the one on the left boundary of the support) to avoid multicollinearity. m in our setup therefore refers to the usual family of B-spline basis functions $s + k$ minus the one dropped to accommodate the constant. In the sieve literature, asymptotics for $m \to \infty$ typically assume that the spline order s is known and fixed, so that $m \to \infty$ is understood to mean that the number of knots $k \to \infty$ as the sample size grows (see also Chen (2007)). However, the sieve literature provides little guidance on the selection of s in practice. To remain agnostic, we therefore allow the search for the optimal m to consider both cubic and fourth order splines. It turns out that fixing s to be 4 or 5 ex ante or allowing cross validation to choose either one has no material impact on our results for both the full sample SRR regression and the pre-crisis sample. To illustrate this form of robustness, Table 9 below shows our main SRR regression table with s fixed at 4 and m ranging from $\{0, 1, \ldots, 5\}$, for a total of 6 different basis function configurations. The table confirms that under fixed s, the nonlinear specification is still significantly rejected across assets and forecast horizons, and that Treasuries and equities have differently signed loadings on $\phi(v)$. For completeness, we also report the results on the pre-crisis sample in Table 10. For our simulations and out-of-sample results, we perform cross-validations several thousand times. To ease the computational burden while allowing s to be 4 or 5, we therefore cap the number of

interior knots and allow k to range over $\{0, 1, 2\}$, for a total of $2 \times 3 = 6$ different basis function configurations. Our Monte Carlo results below verify that performing cross-validations over this range of m has good finite sample coverage properties for a variety of nonlinear functions while at the same time being computationally manageable.

Once we obtain the estimators as described in Section 3.2, we can then conduct inference based on Theorem 1 constructing the test statistics as in equations (A.6) and (A.7). Following footnote 9, note that the variance estimator is computed under the null hypothesis, i.e., that we impose the zero restrictions when calculating the estimator. As we discussed in Section 3.1 (see Footnote 6) we use critical values from the appropriate chi-square distribution rather than the standard normal distribution implied by the results in Theorems 1 and 2. This is a small-sample correction to improve the finite-sample properties of the inference procedure which we also employ in our simulations described in the next section. Finally note that in Table 1 and Figures 2–3 we make use of Hodrick (1992)'s variance estimator which is characterized in equations (1)–(8) of Hodrick (1992).

A.1.2 Forecast Error Comparison

As mentioned in Section 3.4, we utilize forecast error comparisons as a robustness check to ensure that our full-sample results are not an artifact of the flexibility of our nonparametric approach. In this section, we implement the tests of equal predictive accuracy of Clark and West (2007) as a further confirmation that the SRR regression estimates do not demonstrate signs of overfitting. The Clark and West (2007) test statistic is appropriate for cases where the models being compared are nested as they are in our setting. The test statistic compares the mean-squared errors between the encompassing model (SRR regression) and the nested model (regression with only a constant) but adds an adjustment term to appropriately center the test statistic. The intuition for this adjustment term is that under the null hypothesis that the nested model is the true model, the encompassing model will tend to have larger mean-square prediction errors because of the sampling uncertainty induced by estimating parameters whose true value is zero. In our setting, this is further complicated by the cross-sectional restrictions that we impose across assets. The nested model has equivalent forecast errors whether estimated equation by equation or jointly. Thus, when the null hypothesis holds, we are estimating parameters whose true value is zero but also imposing joint restrictions on these parameters. Therefore there is an additional source of deviation from a properly centered test statistic.

Tables 11, 12, and 13 display the outcomes from the Clark and West (2007) test applied to horizons h of 6, 12, and 18 months, respectively. The null hypothesis is that the true model is a mean-only model. The rows of each table correspond to our test assets whereas the columns correspond to the Clark and West (2007) test statistic calculated using two different choices of variance estimators across the three sample splits. In the first, third and fifth columns, the variance is estimated assuming a parametric form, a moving average model with $h-1$ lags (denoted by MA(h-1)) with all MA coefficients equal to unity.[26] This can be thought of as the analog of the Hodrick

[26]For an MA(h-1) process based on white noise innovations (with variance η^2), the long-run variance can be shown to be $\eta^2 \cdot h^2$. We then utilize a plug-in estimate of η^2 to calculate the variance estimator.

(1992) 1B standard errors in the out-of-sample testing setting of Clark and West (2007). Intuitively, this estimator is motivated by the observation that the dominant source of variation under the null hypothesis arises from the overlapping observations with known (and identical) coefficients. To our knowledge, this variance estimator has not been applied in the context of the Clark and West (2007) test statistic and may be of independent interest. In the second, fourth and sixth columns are the results using the variance estimator of Hansen and Hodrick (1980) (denoted by HH). The Clark and West (2007) procedure is based on a one-sided test and so we reject the null hypothesis for large positive values of the test statistic. Below each test statistic in the table is the associated p-value in parentheses.

For every sample split and each asset across all three horizons the test statistic is strictly positive. The p-values are lower using Hansen and Hodrick (1980) standard errors than using the MA(h-1)-based variance estimator. However, simulation results in Table 24 of the Appendix suggest that while the MA(h-1) variance estimator has better size properties than Hansen and Hodrick (1980), it often produces conservative inference (i.e., under-rejecting the null when it is true and so local power is likely to be below nominal size).[27] For $h = 6$ (Table 11) we find that p-values using HH are generally between 1% and 5% whereas those using the MA(h-1) estimator are generally between 10% and 15%. A similar pattern emerges for $h = 12$ (Table 12) where we can formally reject the mean-only model using HH and less so with MA(h-1). However, for a number of assets and sample splits we are able to reject the mean-only model using either variance estimator. For $h = 18$ (Table 13) the results are weaker than in the other two horizons, although the test statistics remain positive throughout. Along with the evidence presented in Table 5 these results show that the SRR model produces reasonable out-of-sample forecasts in real time and so our full-sample results do not appear to be an artifact of overfitting from the flexibility of our nonparametric approach.

Finally, we also report results comparing forecast errors of the SRR regression model to a model which is linear in the VIX. Table 14 shows the RMSFEs and MAFEs whereas Tables 15–17 show the results of the Clark and West (2007) test. Overall, these results are broadly in line with the comparison to the mean-only model.

A.1.3 Theoretical Setup

As discussed in the main text we utilize the "reverse regression" approach of Hodrick (1992) in our multivariate setting to ameliorate concerns about spurious inference induced by overlapping returns. For a similar derivation for the univariate case see Wei and Wright (2013). Consider the multivariate h-period long-horizon return regression of $y_{t+h}^{(h)} := y_{t+1} + y_{t+2} + \cdots + y_{t+h}$ on the predictor x_t. The best linear predictor is $\mathbf{C}_h = \mathbb{C}\left(y_{t+h}^{(h)}, x_t\right) \mathbb{V}\left(x_t\right)^{-1}$. Similarly, for the multivariate h-period reverse

[27] Table 24 in Appendix A.2 reports simulation results showing the finite-sample properties of this variance estimator against popular alternatives available in the literature and shows that the size control is better. These simulation results are for $h = 6$ only as the computational burden of simulating out-of-sample estimation with cross-validation at each period in the testing sample is substantial. In unreported results we have confirmed using the data-generating process in Clark and West (2007, DGP 1) that the MA(h-1) variance estimator controls size well although becomes more conservative in general as h grows, whereas popular alternatives are considerably oversized.

regression, the best linear predictor is $\mathbf{C} = \mathbb{C}\left(y_{t+1}, x_t^{(h)}\right)\mathbb{V}\left(x_t^{(h)}\right)^{-1}$ where $x_t^{(h)} := x_t + x_{t-1} + \cdots + x_{t-h}$. Then, if $(y_t, x_t')'$ is covariance stationary we have that $\mathbb{C}\left(y_{t+h}^{(h)}, x_t\right) = \mathbb{C}\left(y_{t+1}, x_t^{(h)}\right)$ and so $\mathbf{C}_h = \mathbf{C}\mathbb{V}\left(x_t^{(h)}\right)\mathbb{V}(x_t)^{-1}$. Thus, the ith row of $\mathbf{C}_h$ has all elements equal to zero if and only if the ith row of $\mathbf{C}$ has all elements equal to zero. Consequently, in our setting, tests of the null hypothesis of no predictability of asset i in the forward regression can be conducted using the same test in the reverse regression which formally justifies our empirical approach.

Now consider equation (3.4). For simplicity, the results in the Appendix are derived for the case $\mathbf{F}_h = 0$. The extension to the general case follows by similar derivations. The best linear predictor (pseudo-true value) in the forward regression has constant term $\mathbf{a}_h^*$ with linear coefficient $\mathbf{A}_h^* = \mathbf{b}_h^*\gamma_h^{*\prime}$. The corresponding pseudo-true value in the reverse regression is $\mathbf{a}^*$ and $\mathbf{A}^* = \mathbf{A}_h^*\mathbb{V}(X_{m,t})\mathbb{V}\left(X_{m,t}^{(h)}\right)^{-1} =: \mathbf{b}^*\gamma^{*\prime}$. Under our assumptions, we have that $\mathbf{b}^* = (1, \mathbf{b}_0^{*\prime})'$ and the pseudo-true value in both the forward and reverse regressions are of the reduced-rank form. Note that the parameters $\mathbf{a}_h^*$, $\mathbf{A}_h^*$, etc. are a function of m; however, we suppress this dependence for notational simplicity. Finally, define the pseudo-true value regression error from the reverse regression as $e_{t+1}^* := \mathrm{R}x_{t+1} - \mathbf{a}^* - \mathbf{b}^*\gamma^{*\prime}X_{m,t}^{(h)}$.

Table 9: **Robustness: SRR Regression with Fixed B-spline Order: 1990 - 2014**

This table verifies that the main results of the paper hold when the order of the B-spline is held fixed at $s = 4$ and the number of interior knots ranges from 0 to 5.

Horizon h = 6

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	−0.01	1.00	1.70	1.00***	0.59	1.00***	0.05**	−1.43***	−0.01	0.18
cmt1	0.00	0.07*	−0.11	−0.08***	−0.16	−0.21***	0.00	0.04*	0.00**	0.02***
cmt2	0.01	0.09	−0.23	−0.15***	−0.25	−0.34***	0.00	0.09*	0.00	0.02**
cmt5	0.03	0.04	−0.56	−0.36***	−0.45	−0.64***	−0.02*	0.24**	0.01**	0.01
cmt7	0.04	0.04	−0.68	−0.45***	−0.51	−0.74***	−0.03**	0.33**	0.02**	0.00
cmt10	0.05	−0.08	−0.71	−0.46***	−0.48	−0.72***	−0.03**	0.40**	0.03***	0.01
cmt20	0.08	−0.22	−0.91	−0.59***	−0.48	−0.80**	−0.05**	0.53*	0.05***	−0.04
cmt30	0.10	−0.52	−1.27	−0.81***	−0.62	−1.05***	−0.07***	0.72*	0.06***	−0.06
Joint p-val		0.465		0.000		0.024				
Bootstrap p-val		0.494		0.000		0.000				

Horizon h = 12

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.64	1.00**	0.09	1.00***	0.03**	−0.70***	0.00*	0.18
cmt1	0.00	0.11*	−0.05	−0.10***	−0.08	−0.36***	0.00	0.03**	0.00**	0.02***
cmt2	0.01	0.22	−0.08	−0.18***	−0.12	−0.57***	0.00	0.05**	0.00	0.03***
cmt5	0.02	0.33	−0.18	−0.39***	−0.21	−1.03***	−0.01	0.08*	0.01	0.03
cmt7	0.02	0.35	−0.23	−0.48***	−0.24	−1.23***	−0.02*	0.12*	0.01**	0.02
cmt10	0.03	0.15	−0.22	−0.47***	−0.25	−1.25***	−0.02**	0.13*	0.02**	0.03
cmt20	0.05	0.12	−0.31	−0.65**	−0.27	−1.52**	−0.04**	0.16	0.03***	0.00
cmt30	0.06	−0.17	−0.44	−0.88**	−0.33	−1.92**	−0.05**	0.21	0.04***	−0.01
Joint p-val		0.456		0.002		0.002				
Bootstrap p-val		0.472		0.038		0.000				

Horizon h = 18

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.44	1.00**	−0.03	1.00**	0.02**	−0.59***	0.01	0.18
cmt1	0.01	0.07	−0.04	−0.13***	−0.06	−0.60***	0.00	0.04***	0.00**	0.02***
cmt2	0.01	0.19	−0.07	−0.24***	−0.10	−0.95***	0.00	0.07***	−0.01	0.03***
cmt5	0.02	0.36	−0.14	−0.48**	−0.18	−1.65***	−0.01	0.11**	0.00	0.03**
cmt7	0.02	0.38	−0.16	−0.56**	−0.19	−1.87***	−0.01*	0.14**	0.00	0.03*
cmt10	0.03	0.23	−0.15	−0.52**	−0.20	−1.90***	−0.01*	0.15**	0.01*	0.04
cmt20	0.04	0.29	−0.19	−0.69*	−0.20	−2.16*	−0.02*	0.15	0.02**	0.01
cmt30	0.05	0.04	−0.28	−0.91**	−0.24	−2.63**	−0.03**	0.18	0.03**	0.00
Joint p-val		0.665		0.052		0.027				
Bootstrap p-val		0.662		0.071		0.001				

Table 10: **Robustness: SRR Regression with Fixed B-spline Order: 1990 - 2007**

This table verifies that the main results of the paper hold when the order of the B-spline is held fixed at $s = 4$ and the number of interior knots ranges from 0 to 5.

Horizon h = 6

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.03	1.00	0.72	1.00**	0.25	1.00	−0.03	−0.84*	0.00	0.12
cmt1	0.00	0.23**	−0.09	−0.16***	−0.15	−0.59***	0.01	0.03	0.00*	0.03***
cmt2	0.00	0.29	−0.16	−0.28***	−0.22	−0.92***	0.01	0.10*	0.00	0.03***
cmt5	0.01	0.30	−0.31	−0.53***	−0.34	−1.57***	0.00	0.25*	0.01	0.03*
cmt7	0.03	0.18	−0.36	−0.62***	−0.35	−1.75***	−0.01	0.31*	0.02	0.02
cmt10	0.04	−0.23	−0.37	−0.62**	−0.32	−1.74***	−0.03	0.36*	0.02*	0.02
cmt20	0.06	−0.16	−0.42	−0.72**	−0.31	−1.95**	−0.05	0.46**	0.03**	−0.01
cmt30	0.05	−0.27	−0.51	−0.85*	−0.35	−2.25**	−0.06	0.58**	0.04**	−0.03
Joint p-val		0.045		0.000		0.003				
Bootstrap p-val		0.086		0.005		0.002				

Horizon h = 12

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.09	1.00	0.46	1.00*	−0.01	1.00	−0.03	−0.52*	0.00	0.16
cmt1	0.00	−2.09**	−0.08	−0.23***	−0.12	3.25***	0.00	0.03	0.00	0.03***
cmt2	0.00	−3.21*	−0.13	−0.40***	−0.18	5.27***	0.00	0.07*	0.00	0.03***
cmt5	0.00	−4.69	−0.24	−0.72***	−0.31	9.32***	0.00	0.12	0.01	0.04
cmt7	0.01	−3.98	−0.28	−0.84***	−0.33	10.77***	−0.01	0.13	0.01*	0.03
cmt10	0.03	−0.71	−0.27	−0.80***	−0.33	11.27***	−0.03*	0.14	0.02**	0.04
cmt20	0.04	−1.06	−0.33	−1.00***	−0.34	13.10***	−0.04**	0.14	0.03**	0.01
cmt30	0.04	−1.15	−0.40	−1.17**	−0.39	15.27***	−0.06**	0.16	0.04***	0.00
Joint p-val		0.010		0.000		0.002				
Bootstrap p-val		0.036		0.024		0.033				

Horizon h = 18

	(1) Linear VIX		(2) Nonlinear VIX		(3) Nonlinear VIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.11	1.00	0.57	1.00***	0.13	1.00	−0.03	−0.54**	0.00	0.13
cmt1	0.00	−0.36**	−0.07	−0.17***	−0.11	−0.92***	0.00	0.05**	0.00	0.03***
cmt2	0.00	−0.59	−0.13	−0.30***	−0.19	−1.55***	0.00	0.11**	0.00	0.03***
cmt5	0.00	−0.93	−0.24	−0.55***	−0.35	−2.83***	0.01*	0.19**	0.00	0.05*
cmt7	0.01	−0.86	−0.27	−0.63***	−0.39	−3.28***	0.00**	0.22**	0.01*	0.05
cmt10	0.03	−0.25	−0.25	−0.59***	−0.40	−3.44***	−0.01**	0.24**	0.01**	0.06
cmt20	0.04	−0.45	−0.29	−0.70***	−0.41	−3.81***	−0.02**	0.20*	0.02**	0.03
cmt30	0.03	−0.43	−0.37	−0.85***	−0.50	−4.55***	−0.03**	0.25**	0.03***	0.03
Joint p-val		0.016		0.000		0.000				
Bootstrap p-val		0.057		0.011		0.003				

Table 11: **Clark-West Test: Mean-Only Model ($h = 6$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a mean-only model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the running mean forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 6$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (running mean) and encompassing model 2 (SRRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	1.18	1.70**	1.06	1.49*	1.09	1.59*
	(0.12)	(0.04)	(0.15)	(0.07)	(0.14)	(0.06)
cmt1	1.14	1.91**	1.09	1.84**	1.17	2.11**
	(0.13)	(0.03)	(0.14)	(0.03)	(0.12)	(0.02)
cmt2	1.26	2.23**	1.13	1.94**	1.32*	2.54***
	(0.10)	(0.01)	(0.13)	(0.03)	(0.09)	(0.01)
cmt5	1.19	2.18**	1.07	1.93**	1.18	2.31**
	(0.12)	(0.01)	(0.14)	(0.03)	(0.12)	(0.01)
cmt7	0.95	1.92**	0.86	1.75**	0.92	1.98**
	(0.17)	(0.03)	(0.19)	(0.04)	(0.18)	(0.02)
cmt10	0.97	2.10**	0.85	1.85**	0.91	2.09**
	(0.17)	(0.02)	(0.20)	(0.03)	(0.18)	(0.02)
cmt20	1.01	1.94**	0.94	1.80**	0.99	2.02**
	(0.16)	(0.03)	(0.17)	(0.04)	(0.16)	(0.02)
cmt30	1.27	1.94**	1.21	1.86**	1.25	2.01**
	(0.10)	(0.03)	(0.11)	(0.03)	(0.11)	(0.02)

Table 12: **Clark-West Test: Mean-Only Model ($h = 12$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a mean-only model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the running mean forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 12$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (running mean) and encompassing model 2 (SRRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	1.38*	1.79**	0.96	1.25	0.97	1.30*
	(0.08)	(0.04)	(0.17)	(0.11)	(0.17)	(0.10)
cmt1	0.87	1.07	0.71	0.83	1.10	1.31*
	(0.19)	(0.14)	(0.24)	(0.20)	(0.14)	(0.10)
cmt2	0.91	1.28	0.63	0.81	1.09	1.49*
	(0.18)	(0.10)	(0.26)	(0.21)	(0.14)	(0.07)
cmt5	1.29*	2.02**	1.09	1.51*	1.50*	2.27**
	(0.10)	(0.02)	(0.14)	(0.07)	(0.07)	(0.01)
cmt7	1.38*	2.31**	1.23	1.85**	1.57*	2.61***
	(0.08)	(0.01)	(0.11)	(0.03)	(0.06)	(0.00)
cmt10	1.37*	2.46***	1.27	2.07**	1.60*	2.90***
	(0.09)	(0.01)	(0.10)	(0.02)	(0.05)	(0.00)
cmt20	1.48*	2.61***	1.43*	2.45***	1.55*	2.96***
	(0.07)	(0.00)	(0.08)	(0.01)	(0.06)	(0.00)
cmt30	1.50*	2.57***	1.47*	2.55***	1.49*	2.87***
	(0.07)	(0.01)	(0.07)	(0.01)	(0.07)	(0.00)

Table 13: **Clark-West Test: Mean-Only Model ($h = 18$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a mean-only model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the running mean forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 18$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (running mean) and encompassing model 2 (SRRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	1.20	1.64*	0.54	0.91	0.58	1.02
	(0.11)	(0.05)	(0.29)	(0.18)	(0.28)	(0.15)
cmt1	0.71	1.24	0.60	1.00	0.78	1.21
	(0.24)	(0.11)	(0.27)	(0.16)	(0.22)	(0.11)
cmt2	0.73	1.63*	0.45	0.94	0.71	1.41*
	(0.23)	(0.05)	(0.33)	(0.17)	(0.24)	(0.08)
cmt5	0.92	2.62***	0.63	1.57*	0.87	1.96**
	(0.18)	(0.00)	(0.26)	(0.06)	(0.19)	(0.03)
cmt7	0.94	3.88***	0.70	2.39***	0.88	2.72***
	(0.17)	(0.00)	(0.24)	(0.01)	(0.19)	(0.00)
cmt10	0.98	3.12***	0.84	2.37***	0.97	2.46***
	(0.16)	(0.00)	(0.20)	(0.01)	(0.17)	(0.01)
cmt20	1.15	2.70***	1.05	2.49***	1.05	2.34***
	(0.13)	(0.00)	(0.15)	(0.01)	(0.15)	(0.01)
cmt30	1.30*	4.09***	1.22	4.56***	1.15	4.38***
	(0.10)	(0.00)	(0.11)	(0.00)	(0.13)	(0.00)

Table 14: **Forecast Error Comparison: Linear Model**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a linear VIX model. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the linear VIX forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h \in \{6, 12, 18\}$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. We report the log of the ratio of unadjusted root MSFEs and mean absolute forecast errors (MAFEs), where the linear VIX model is in the denominator. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Forecast Error Comparison						
	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
Log Ratio of RMSFEs, MAFEs (h=6)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	0.00	0.00	0.03	0.03	0.02	0.01
cmt1	0.21	0.06	0.25	0.08	0.29	0.08
cmt2	0.18	0.07	0.24	0.09	0.26	0.07
cmt5	0.07	0.01	0.10	0.02	0.09	−0.01
cmt7	0.03	−0.01	0.04	0.00	0.03	−0.03
cmt10	0.00	−0.03	0.01	−0.02	−0.01	−0.05
cmt20	−0.02	−0.03	−0.01	−0.03	−0.02	−0.04
cmt30	−0.01	−0.02	−0.01	−0.02	−0.01	−0.03
Log Ratio of RMSFEs, MAFEs (h=12)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	−0.03	0.01	−0.04	0.02	−0.04	0.03
cmt1	0.06	0.04	0.06	0.02	0.00	−0.02
cmt2	0.07	0.07	0.09	0.09	0.02	0.03
cmt5	0.00	−0.01	0.03	0.01	−0.06	−0.10
cmt7	−0.01	−0.03	0.00	−0.02	−0.06	−0.10
cmt10	−0.04	−0.07	−0.03	−0.06	−0.09	−0.15
cmt20	−0.05	−0.07	−0.05	−0.08	−0.08	−0.13
cmt30	−0.07	−0.06	−0.07	−0.08	−0.09	−0.11
Log Ratio of RMSFEs, MAFEs (h=18)						
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
MKT	0.00	0.03	−0.01	0.04	0.01	0.06
cmt1	0.06	0.03	0.06	0.03	0.02	0.00
cmt2	0.08	0.08	0.11	0.11	0.06	0.09
cmt5	0.06	0.04	0.09	0.09	0.04	0.03
cmt7	0.04	0.04	0.07	0.08	0.03	0.02
cmt10	0.00	0.01	0.02	0.04	−0.01	−0.01
cmt20	−0.03	−0.03	−0.02	−0.02	−0.03	−0.03
cmt30	−0.05	−0.05	−0.04	−0.05	−0.05	−0.05

Table 15: **Clark-West Test: Linear Model ($h = 6$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a linear VIX model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the linear VIX forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 6$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (linear VIX) and encompassing model 2 (SRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004). Note that ÒUndÓ signifies a case where the HH estimated variance is negative.

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	1.22	1.87**	0.90	1.36*	1.00	1.60*
	(0.11)	(0.03)	(0.18)	(0.09)	(0.16)	(0.05)
cmt1	0.34	1.32*	0.34	1.54*	0.39	3.64***
	(0.37)	(0.09)	(0.37)	(0.06)	(0.35)	(0.00)
cmt2	0.64	3.19***	0.52	2.86***	0.70	Und.
	(0.26)	(0.00)	(0.30)	(0.00)	(0.24)	(Und.)
cmt5	1.55*	2.39***	1.40*	2.07**	1.58*	2.57***
	(0.06)	(0.01)	(0.08)	(0.02)	(0.06)	(0.01)
cmt7	1.17	1.97**	1.07	1.80**	1.15	2.05**
	(0.12)	(0.02)	(0.14)	(0.04)	(0.12)	(0.02)
cmt10	1.01	2.10**	0.93	1.94**	0.97	2.18**
	(0.16)	(0.02)	(0.18)	(0.03)	(0.17)	(0.01)
cmt20	1.05	1.93**	1.00	1.85**	1.06	2.09**
	(0.15)	(0.03)	(0.16)	(0.03)	(0.15)	(0.02)
cmt30	1.26	1.97**	1.21	1.91**	1.27	2.15**
	(0.10)	(0.02)	(0.11)	(0.03)	(0.10)	(0.02)

Table 16: **Clark-West Test: Linear Model ($h = 12$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a linear VIX model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the linear VIX forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 12$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (linear VIX) and encompassing model 2 (SRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	0.85	1.24	0.84	1.21	0.75	1.08
	(0.20)	(0.11)	(0.20)	(0.11)	(0.23)	(0.14)
cmt1	0.37	0.54	0.43	0.61	0.80	1.35*
	(0.36)	(0.30)	(0.34)	(0.27)	(0.21)	(0.09)
cmt2	0.42	0.65	0.36	0.52	0.89	1.63*
	(0.34)	(0.26)	(0.36)	(0.30)	(0.19)	(0.05)
cmt5	0.86	1.37*	0.71	1.05	1.07	1.79**
	(0.20)	(0.08)	(0.24)	(0.15)	(0.14)	(0.04)
cmt7	1.01	1.69**	0.87	1.36*	1.14	1.92**
	(0.16)	(0.05)	(0.19)	(0.09)	(0.13)	(0.03)
cmt10	1.27	2.20**	1.11	1.71**	1.41*	2.33***
	(0.10)	(0.01)	(0.13)	(0.04)	(0.08)	(0.01)
cmt20	1.30*	2.17**	1.21	1.95**	1.32*	2.28**
	(0.10)	(0.02)	(0.11)	(0.03)	(0.09)	(0.01)
cmt30	1.32*	2.28**	1.27	2.20**	1.31*	2.44***
	(0.09)	(0.01)	(0.10)	(0.01)	(0.10)	(0.01)

Table 17: **Clark-West Test: Linear Model ($h = 18$)**

This table compares out-of-sample forecast errors of the sieve reduced rank (SRR) regression against a linear VIX model. The forecast error comparison follows the procedure in Clark and West (2007), which examines the difference in adjusted mean squared forecast errors (MSFEs) between nested models. We split our full monthly sample from 1990:1 to 2014:9 into an in-sample period $t = 1, \ldots, (t^* - 1)$ and an out-of-sample period $t = t^*, \ldots, T$. We then evaluate the SRR regression model and the linear VIX forecast $\hat{E}_{t^*}[Rx^i_{t^*+h}]$ and compare it against the $h = 18$ realized compound return $Rx^i_{t^*+h}$ for i indexing the market and seven constant maturity Treasury portfolios. The forecast errors of the nested model 1 (linear VIX) and encompassing model 2 (SRR regression model) are each squared, differenced, and adjusted

$$f_{t^*,t^*+h} \equiv \left(\hat{\varepsilon}^i_{1,t^*,t^*+h}\right)^2 - \left(\hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2 + \left(\hat{\varepsilon}^i_{1,t^*,t^*+h} - \hat{\varepsilon}^i_{2,t^*,t^*+h}\right)^2.$$

The table reports the t-statistic from regressions of $f_{t,t+h}$ on a constant for the out-of-sample period $t = t^*, \ldots, T$, where the model is re-estimated for each t on an expanding window. A positive number indicates that adjusted, squared forecast errors of the nested model are larger than those of the encompassing SRR regression model. MA(h-1) and HH denote test statistics formed using a parametric MA model or Hansen and Hodrick (1980), respectively. The difference is statistically significant if the t-statistic on this intercept exceeds one-sided standard critical values for 1% (***), 5% (**), and 10% (*) significance levels. Note that for each new $t = t^*, \ldots, T$, sieve expansion terms are chosen by a new cross-validation on the in-sample portion of the data. Each panel of the table corresponds to a different in-sample starting period $t^*/T = 0.4$ (Nov-1999), $t^*/T = 0.5$ (May-2002), and $t^*/T = 0.6$ (Nov-2004).

Clark-West Forecast Error Comparison

	In-sample Split $(t^*/T) = 0.4$		In-sample Split $(t^*/T) = 0.5$		In-sample Split $(t^*/T) = 0.6$	
	MA-h	HH	MA-h	HH	MA-h	HH
MKT	0.30	0.71	0.40	0.81	0.21	0.40
	(0.38)	(0.24)	(0.35)	(0.21)	(0.42)	(0.34)
cmt1	0.22	0.44	0.25	0.49	0.55	1.13
	(0.41)	(0.33)	(0.40)	(0.31)	(0.29)	(0.13)
cmt2	0.23	0.53	0.18	0.36	0.62	1.45*
	(0.41)	(0.30)	(0.43)	(0.36)	(0.27)	(0.07)
cmt5	0.51	1.45*	0.35	0.86	0.71	2.05**
	(0.30)	(0.07)	(0.36)	(0.19)	(0.24)	(0.02)
cmt7	0.62	2.45***	0.43	1.49*	0.67	2.31**
	(0.27)	(0.01)	(0.33)	(0.07)	(0.25)	(0.01)
cmt10	0.89	2.80***	0.64	1.77**	0.78	1.94**
	(0.19)	(0.00)	(0.26)	(0.04)	(0.22)	(0.03)
cmt20	1.01	2.68***	0.88	2.28**	0.86	2.13**
	(0.16)	(0.00)	(0.19)	(0.01)	(0.20)	(0.02)
cmt30	1.16	3.36***	1.06	3.24***	0.98	2.93***
	(0.12)	(0.00)	(0.15)	(0.00)	(0.16)	(0.00)

A.2 Simulation Results

A.2.1 Main Setup

We conduct Monte Carlo simulations to examine the finite-sample performance of the inference procedures in Theorem 1. Our simulations focus on the sieve-reduced rank system

$$Rx_{t+h}^i = a_h^i + b_h^i \cdot \phi_h(v_t) + \varepsilon_{t+h}^i, \tag{A.2}$$

proposed in equation (3.1) in the main text. Within the simulated environment, $i = 1$ indexes the market excess return MKT, and $i = 2, \ldots, 8$ indexes the seven constant-maturity Treasury portfolios $cmt1, cmt2, cmt5, cmt7, cmt10, cmt20, cmt30$ we consider in the paper, for a total of 8 test assets.

The DGPs for these return series depend on the null hypothesis in question. Since $\mathbb{H}_{1,0}$ tests the joint restriction that the VIX does not predict any of the test assets, we first simulate excess returns under the DGP

$$Rx_{t+h}^i = a_h^i + \varepsilon_{t+h}^i, \qquad i = 1, \ldots, 8 \tag{A.3}$$

1,000 times and assume that the researcher instead runs regressions of the form (A.2). For example, if the test of no predictability $\mathbb{H}_{1,0}$ is chosen to haven a nominal size of 5% and controls Type I error, then the null of no predictability should be rejected among 5% of simulated samples.

To estimate regressions of the form in equation (A.2) inside the simulation environment, we require a realistic data generating process (DGP) for $v_t = VIX_t$. To that end, we use results from Van Tassel and Vogt (2016), who provide evidence that the term structure of variance swap rates is affine in three factors: the level of the variance swap curve (PC1), its slope (PC2), and the level of realized variance (RV). Because the VIX is the square-root of the annualized 1-month variance swap rate multiplied by 100, simulating the three primitive variance factors produces a VIX. To be precise, we follow Van Tassel and Vogt (2016) and assume that the standardized $X_t = (RV_t, PC1_t, PC2_t)$ follows the VAR

$$X_{t+1} = \mu + \Phi X_t + v_{t+1}. \tag{A.4}$$

Using their empirical estimates of μ, Φ, and the variance matrix of the error terms, the affine pricing kernel and no-arbitrage arguments yield a $VIX_t = 100\sqrt{A_1 + B_1' X_t}$, where A_1 and B_1 are functions of these estimated parameters and the prices of risk. Figure 14 shows that this multi-component approach can produce an empirically realistic VIX given an appropriate set of shocks.

We then proceed by simulating Gaussian shocks to the systems in equations (A.3) and (A.4) jointly, matching the mean and covariance found in the data. This ensures that our simulated data mimic the properties observed in the sample such as leverage effects. Figure 15 shows the residual autocorrelations of this joint system as estimated from the data and shows them to be statistically indistinguishable from white noise. We follow this feature in the simulations by not imposing an autoregressive structure on the shocks. Note that for PC1, PC2, and RV, we are ignoring the first stage estimation error on $\hat{\mu}$ and $\hat{\Phi}$ for simplicity in the calculation of the autocorrelation confidence bands in Figure 15.

The simulation design for $\mathbb{H}_{2,0}$ requires adjustments. In particular, $\mathbb{H}_{2,0}$ tests the null hypothesis

Figure 14: **Diagnostics: Multi-Factor Model for VIX Dynamics**

This figure plots the true VIX against the Van Tassel and Vogt (2016) multi-factor model for VIX dynamics. In that model, VIX^2 is an affine function of three factors, given by realized variance, the level of variance swap term structure, and the slope of the variance swap term structure, which evolve according to a vector autoregression. The figure shows that the three factor model is an effective DGP for the VIX timeseries.

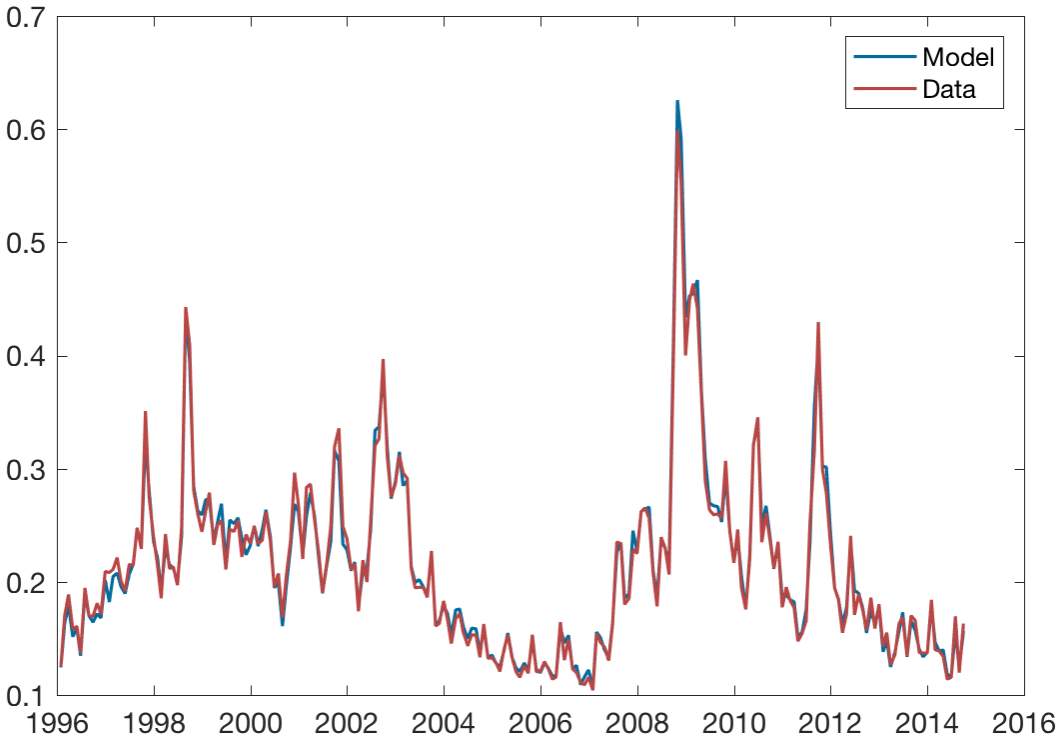
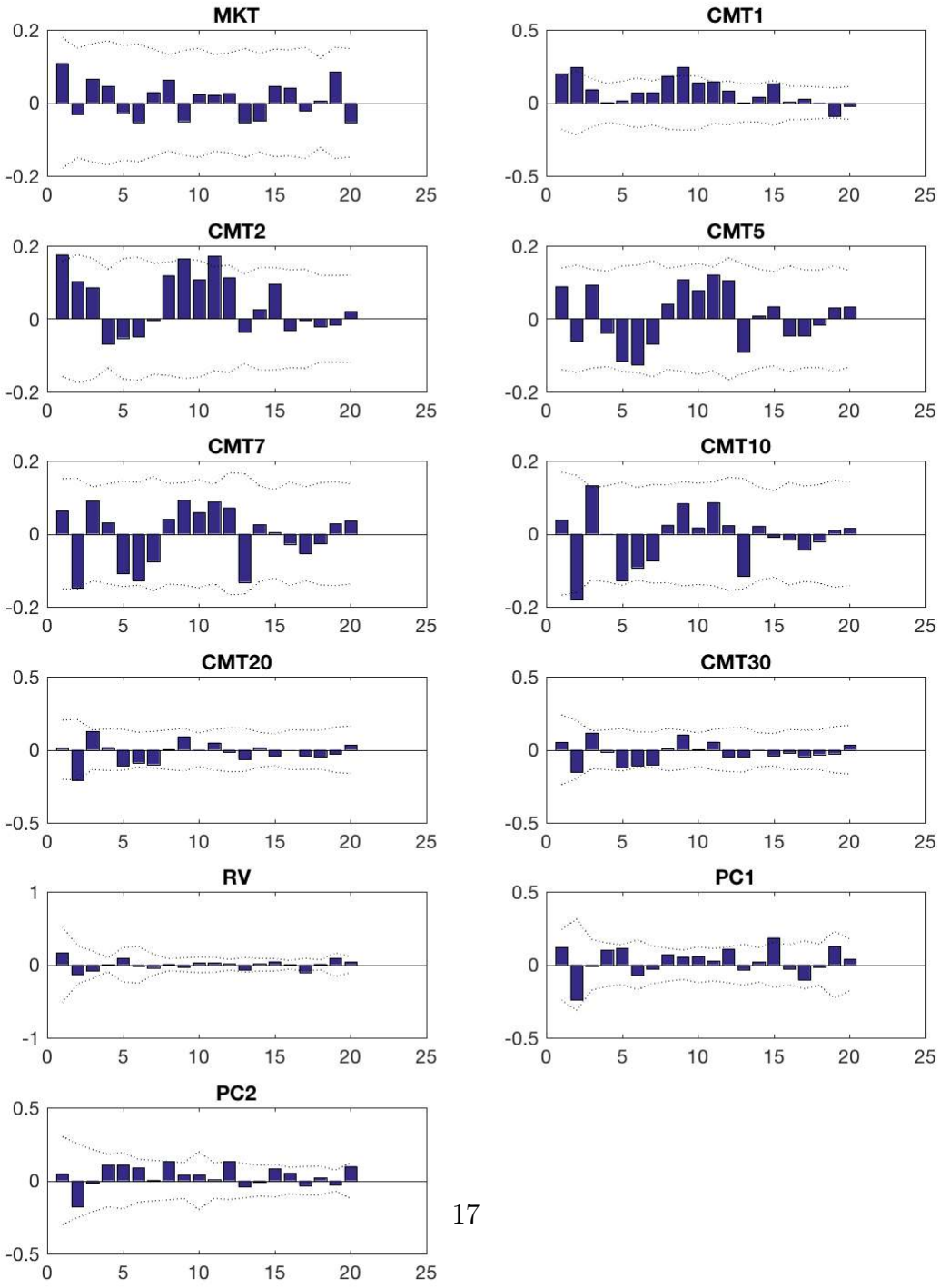


Figure 15: **Diagnostics: Multi-Factor Model Residuals**

This figure plots autocorrelation functions of excess returns along with factor residuals from the Van Tassel and Vogt (2016) three-factor VAR model for VIX dynamics. Dashed lines denote 95% confidence bands.



that asset i does not predict returns, but there is at least one asset $j \neq i$ that does.[28] Thus, the simulation design for $\mathbb{H}_{2,0}$ requires us to specify a function $\phi_h(v_t)$ on which b_h^j loads for some $j \neq i$. To that end, we consider six different specifications for $\phi_h(v_t)$, which we plot in Figure 16. The top left panel of the figure shows a cubic polynomial sieve estimate obtained directly from an SRR regression of our 8 test assets on lagged VIX. Beyond this data-determined ϕ_h, we also examine 5 other nonlinear functions to test various forms of non-monotonicities, concavities/convexities, and multiple inflection points.

In simulating DGPs under the null $\mathbb{H}_{2,0}$, we make use of the relationship between SRR regression loadings at the h-month horizon versus the 1-month horizon described in Section A.1.3 which allows us to simulate one-step ahead returns calibrated from h-month ahead regressions.

A.2.2 Simulation Design under Heteroskedasticity

We also repeat each simulation exercise under heteroskedasticity. Thus, when simulating the shocks to the joint system of excess returns (A.3) and volatility factors (A.4), we allow for a multivariate GARCH-CCC structure in the return equations (see Bollerslev (1990)). The GARCH-CCC parameters are estimated from actual returns and variance factors, thereby mimicking the correlation structure found in the data and, importantly, also preserving leverage effects between returns and volatility. We chose not to estimate time-varying correlations for parsimony, given that our variance factor observations start in only 1996. Following a referee's advice, we specified the time-varying volatilities in the GARCH process to be affine in the lagged VIX_{t-1}. We found this parsimonious specification to generate empirically realistic (in terms of moments and autocorrelations), and yet numerically stable sample paths in the simulation.

A.2.3 Cross-Validation Procedures

For each simulation exercise, we examine two different cross-validation procedures to select the number of sieve basis functions used to estimate ϕ. We refer to the first procedure as MSFE cross-validation (described in Section A.1) and to the second procedure as leave-k-out. The MSFE procedure selects the number of sieve basis functions that minimizes the out-of-sample mean-squared error, where the observation left out is always at the end of the sample. We do this to preserve any long-run dependencies in the data.

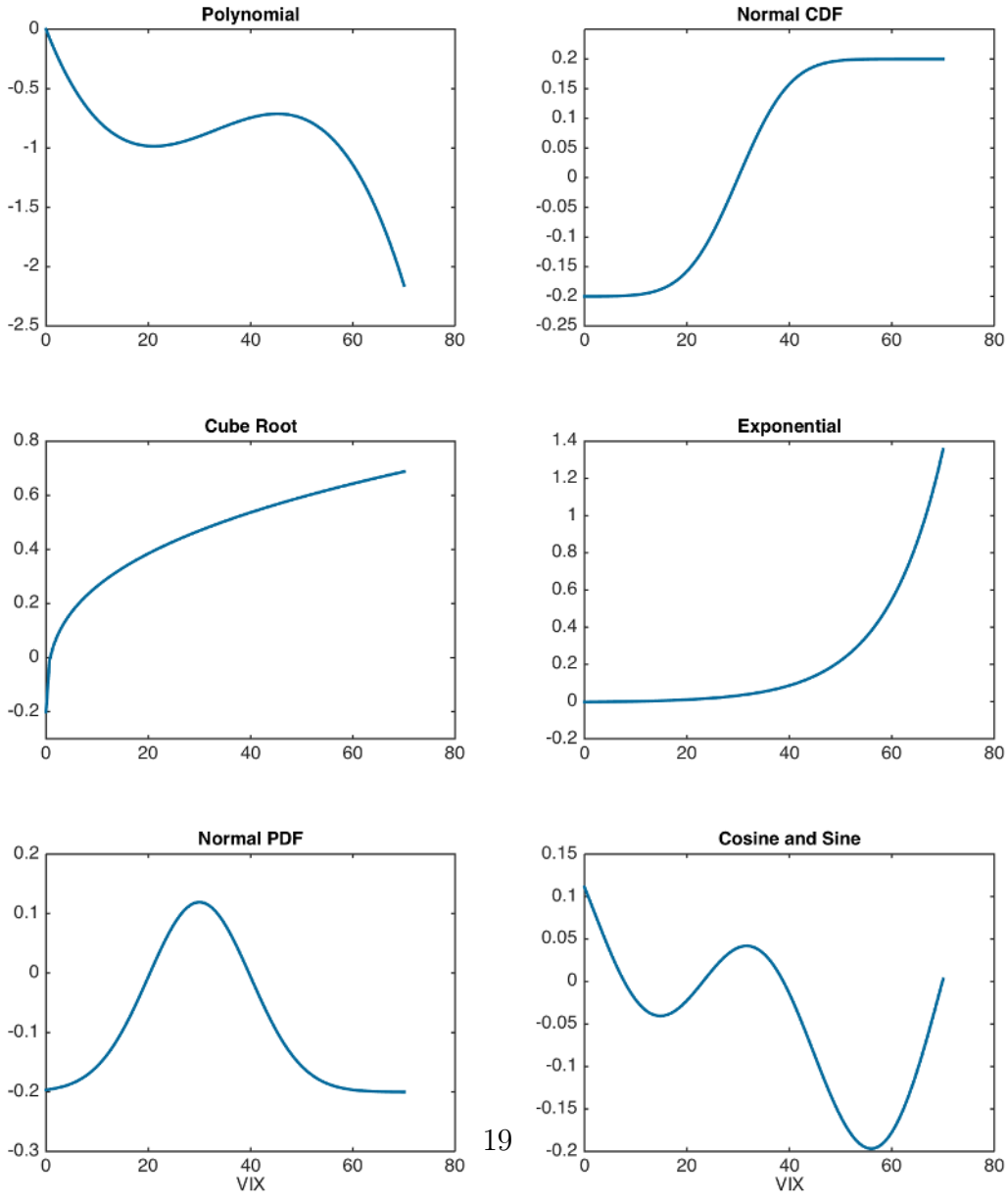
The leave-k-out cross-validation procedure leaves out k contiguous observations $[i-k+1,\ldots,i]$ for $i = k,\ldots,T$ and sets those aside for out-of-sample evaluation.[29] The remaining observations are used to form the estimate $\hat{E}_t^m[Rx_{t+h}^i]$, which are then compared against the realizations Rx_{t+h}^i for $t \in [i-k+1,\ldots,i]$. For our simulations, we set $k = 10$ months to account for time series dependencies in the DGP. Thus, the number of sieve basis functions selected via leave-k-out cross-

[28] Recall that the case where none of the assets predict returns is subsumed in test $\mathbb{H}_{1,0}$.

[29] Note that unlike the MSFE cross validation procedure, the sample left out of estimation in the leave-k-out procedure is not necessarily at the end of the sample.

Figure 16: **Monte Carlo DGPs for** $\phi(vix)$

This figure plots the candidate DGPs used in the Monte Carlo simulations. The polynomial function is $\phi(v) = a \cdot v + b \cdot v^2 + c \cdot v^3$, where the coefficients come from a sieve reduced rank regression of the market and seven maturity-sorted Treasury portfolios on the VIX in percent, using polynomial basis functions. The remaining functions are chosen to display concavity, convexity, multiple inflection points, and multiple local extrema, with each function translated and scaled to fit in the VIX domain and to take bounded values over this domain. The normal CDF is $a \cdot N_{cdf}(\mu, \sigma) - b$; the cube root is $x^{1/3} - a$; the exponential is $(\exp(ax) - b)/c$; the normal PDF is $a \cdot N_{pdf}(\mu, \sigma) - b$; and the trigonometric function is $(cos(a\pi x + \frac{\pi}{2}) + sin(b\pi x + \frac{\pi}{2}))/c$.



validation solves

$$\arg\min_{m} \sum_{i=k}^{T} \frac{1}{k} \sum_{t=i-k+1}^{i} \left(\mathrm{Rx}_{t+h} - \hat{E}_t^m[\mathrm{Rx}_{t+h}]\right)^2. \tag{A.5}$$

A.2.4 Results

Our simulation results are presented in Tables 18–23. Table 18 shows rejection rates for the joint null hypothesis $\mathbb{H}_{1,0}$ of no predictability. Table 19 shows the empirical size for the test of the null hypothesis $\mathbb{H}_{2,0}$ under MSFE cross-validation procedure, and Table 20 shows empirical size under leave-k-out cross-validation for $k = 10$. Tables 21 through 23 repeat the exercise with the heteroskedastic error structure described previously. All reported results are based on three common choices for the nominal size of the test, namely, 1%, 5% and 10%. The tables below show that the empirical size of our tests are very close to the empirical size across all of the different modeling specifications.

Table 18: **Monte Carlo: Reverse Regression under the Joint Null**

This table examines the finite sample performance of test $\mathbb{H}_{1,0}$ in Theorem 1. We simulate 1000 samples of excess returns under the joint null hypothesis $\mathbb{H}_{1,0}$ of no predictability. Each sample represents a multivariate time series of eight different assets with sample mean and covariance calibrated to excess returns to the equity market and seven constant maturity Treasury portfolios $cmt1, cmt2, cmt5, cmt7, cmt10, cmt20, cmt30$. The variance estimator $\hat{\mathrm{V}}_2$ used to form the test statistic, $\hat{\mathcal{T}}_2$, is constructed under the null hypothesis of no predictability. For each Monte Carlo simulation the order of sieve B-splines is chosen by MSFE cross-validation, where the last observations is left out from a series of expanding-window pseudo out-of-sample forecasts (first row). The second row reports results for leave-k-out cross-validation, where blocks of size $k = 10$ are removed from estimation and are set aside for out-of-sample evaluation.

Nominal Size:	1%	5%	10%
MSFE	0.010	0.052	0.107
Leave-k-out	0.010	0.053	0.113

Table 19: **Monte Carlo Rejection Frequencies: MSFE Cross-Validation**

This table examines the finite sample performance of test $\mathbb{H}_{2,0}$ in Theorem 1. We simulate 1000 samples under the null hypothesis $\mathbb{H}_{2,0}$ that asset i does not predict returns ($b_h^i\phi = 0$), but that the other assets do predict returns nonlinearly ($b_h^j\phi \neq 0$ for $j \neq i$). The type of nonlinearity ϕ is given by the DGPs in Figure 16. For each Monte Carlo simulation the order of sieve B-splines is chosen by MSFE cross-validation, where the last observations is left out from a series of expanding-window pseudo out-of-sample forecasts.

Nominal Size:	1%	5%	10%	1%	5%	10%
	Polynomial			Normal CDF		
MKT	0.006	0.061	0.111	0.008	0.057	0.109
cmt1	0.015	0.062	0.115	0.017	0.065	0.129
cmt2	0.015	0.064	0.117	0.015	0.062	0.121
cmt5	0.009	0.051	0.106	0.011	0.055	0.112
cmt7	0.006	0.040	0.097	0.012	0.055	0.122
cmt10	0.017	0.065	0.111	0.010	0.049	0.103
cmt20	0.018	0.066	0.126	0.018	0.061	0.119
cmt30	0.010	0.044	0.087	0.014	0.071	0.123
	Cube Root			Exponential		
MKT	0.017	0.056	0.097	0.014	0.059	0.127
cmt1	0.012	0.055	0.094	0.020	0.078	0.136
cmt2	0.010	0.056	0.106	0.018	0.067	0.122
cmt5	0.007	0.052	0.100	0.019	0.063	0.129
cmt7	0.009	0.044	0.096	0.017	0.062	0.113
cmt10	0.013	0.049	0.104	0.017	0.063	0.126
cmt20	0.011	0.057	0.110	0.021	0.067	0.128
cmt30	0.008	0.050	0.115	0.023	0.064	0.125
	Normal PDF			Cosine and Sine		
MKT	0.011	0.050	0.107	0.013	0.070	0.113
cmt1	0.013	0.055	0.105	0.010	0.042	0.097
cmt2	0.013	0.052	0.089	0.010	0.051	0.100
cmt5	0.009	0.055	0.089	0.018	0.060	0.119
cmt7	0.010	0.049	0.111	0.014	0.063	0.108
cmt10	0.004	0.046	0.100	0.012	0.063	0.104
cmt20	0.009	0.053	0.114	0.016	0.066	0.116
cmt30	0.015	0.060	0.112	0.011	0.050	0.102

Table 20: **Monte Carlo Rejection Frequencies: Leave-k-Out Cross-Validation**

This table examines the finite sample performance of test $\mathbb{H}_{2,0}$ in Theorem 1. We simulate 1000 samples under the null hypothesis $\mathbb{H}_{2,0}$ that asset i does not predict returns ($b_h^i\phi = 0$), but that the other assets do predict returns nonlinearly ($b_h^j\phi \neq 0$ for $j \neq i$). The type of nonlinearity ϕ is given by the DGPs in Figure 16. For each Monte Carlo simulation the order of sieve B-splines is chosen by leave-k-out cross-validation for $k = 10$.

Nominal Size:	1%	5%	10%	1%	5%	10%
	Polynomial			Normal CDF		
MKT	0.015	0.052	0.104	0.016	0.064	0.128
cmt1	0.016	0.052	0.099	0.013	0.050	0.110
cmt2	0.007	0.064	0.117	0.019	0.063	0.106
cmt5	0.012	0.056	0.117	0.016	0.063	0.118
cmt7	0.015	0.071	0.121	0.011	0.052	0.103
cmt10	0.013	0.051	0.101	0.018	0.061	0.122
cmt20	0.013	0.052	0.113	0.014	0.063	0.125
cmt30	0.012	0.064	0.103	0.009	0.048	0.101
	Cube Root			Exponential		
MKT	0.020	0.065	0.117	0.021	0.072	0.121
cmt1	0.016	0.051	0.101	0.024	0.077	0.130
cmt2	0.016	0.062	0.124	0.020	0.077	0.133
cmt5	0.013	0.054	0.111	0.016	0.068	0.128
cmt7	0.017	0.053	0.091	0.017	0.076	0.131
cmt10	0.013	0.047	0.103	0.025	0.078	0.139
cmt20	0.013	0.070	0.121	0.018	0.072	0.133
cmt30	0.012	0.057	0.114	0.011	0.060	0.122
	Normal PDF			Cosine and Sine		
MKT	0.009	0.054	0.104	0.018	0.074	0.136
cmt1	0.010	0.045	0.105	0.016	0.056	0.100
cmt2	0.010	0.049	0.106	0.006	0.042	0.094
cmt5	0.016	0.054	0.102	0.016	0.050	0.112
cmt7	0.017	0.063	0.121	0.012	0.052	0.102
cmt10	0.014	0.049	0.110	0.010	0.051	0.097
cmt20	0.013	0.055	0.107	0.011	0.062	0.116
cmt30	0.013	0.062	0.111	0.012	0.056	0.101

Table 21: **Monte Carlo: Reverse Regression under the Joint Null and Heteroskedasticity**

This table examines the finite sample performance of test $\mathbb{H}_{1,0}$ in Theorem 1. We simulate 1000 samples of excess returns under the joint null hypothesis $\mathbb{H}_{1,0}$ of no predictability. Each sample represents a multivariate time series of eight different assets with sample mean and covariance calibrated to excess returns to the equity market and seven constant maturity Treasury portfolios $cmt1, cmt2, cmt5, cmt7, cmt10, cmt20, cmt30$. The variance estimator $\hat{\mathrm{V}}_2$ used to form the test statistic, $\widehat{\mathcal{T}}_2$, is constructed under the null hypothesis of no predictability. For each Monte Carlo simulation the order of sieve B-splines is chosen by MSFE cross-validation, where the last observations is left out from a series of expanding-window pseudo out-of-sample forecasts (first row). The second row reports results for leave-k-out cross-validation, where blocks of size $k = 10$ are removed from estimation and are set aside for out-of-sample evaluation.

Nominal Size:	1%	5%	10%
MSFE	0.003	0.046	0.097
Leave-k-out	0.007	0.058	0.109

Table 22: **Heteroskedasticity Monte Carlo: MSFE Cross-Validation**

This table examines the finite sample performance of test $\mathbb{H}_{2,0}$ in Theorem 1. We simulate 1000 samples under the null hypothesis $\mathbb{H}_{2,0}$ that asset i does not predict returns ($b_h^i\phi = 0$), but that the other assets do predict returns nonlinearly ($b_h^j\phi \neq 0$ for $j \neq i$). The type of nonlinearity ϕ is given by the DGPs in Figure 16. For each Monte Carlo simulation the order of sieve B-splines is chosen by MSFE cross-validation, where the last observations is left out from a series of expanding-window pseudo out-of-sample forecasts.

Nominal Size:	1%	5%	10%	1%	5%	10%
	Polynomial			Normal CDF		
MKT	0.017	0.063	0.108	0.011	0.053	0.117
cmt1	0.013	0.060	0.110	0.010	0.058	0.106
cmt2	0.010	0.061	0.110	0.016	0.071	0.136
cmt5	0.016	0.054	0.107	0.021	0.067	0.127
cmt7	0.013	0.061	0.121	0.014	0.068	0.121
cmt10	0.018	0.069	0.129	0.021	0.079	0.139
cmt20	0.019	0.072	0.124	0.015	0.072	0.145
cmt30	0.011	0.061	0.126	0.018	0.063	0.117
	Cube Root			Exponential		
MKT	0.022	0.072	0.132	0.024	0.085	0.142
cmt1	0.017	0.058	0.130	0.016	0.080	0.141
cmt2	0.016	0.072	0.127	0.018	0.067	0.137
cmt5	0.024	0.078	0.151	0.026	0.079	0.147
cmt7	0.018	0.077	0.133	0.016	0.076	0.152
cmt10	0.024	0.087	0.151	0.030	0.081	0.158
cmt20	0.017	0.077	0.139	0.018	0.081	0.144
cmt30	0.017	0.070	0.121	0.032	0.086	0.147
	Normal PDF			Cosine and Sine		
MKT	0.013	0.072	0.130	0.019	0.061	0.115
cmt1	0.017	0.064	0.113	0.015	0.069	0.123
cmt2	0.010	0.061	0.116	0.013	0.067	0.125
cmt5	0.015	0.071	0.126	0.020	0.083	0.140
cmt7	0.014	0.060	0.108	0.023	0.067	0.121
cmt10	0.014	0.065	0.120	0.026	0.083	0.130
cmt20	0.024	0.080	0.142	0.023	0.079	0.144
cmt30	0.017	0.067	0.123	0.023	0.075	0.128

Table 23: **Heteroskedasticity Monte Carlo: Leave-k-Out Cross-Validation**

This table examines the finite sample performance of test $\mathbb{H}_{2,0}$ in Theorem 1. We simulate 1000 samples under the null hypothesis $\mathbb{H}_{2,0}$ that asset i does not predict returns ($b_h^i\phi = 0$), but that the other assets do predict returns nonlinearly ($b_h^j\phi \neq 0$ for $j \neq i$). The type of nonlinearity ϕ is given by the DGPs in Figure 16. For each Monte Carlo simulation the order of sieve B-splines is chosen by leave-k-out cross-validation for $k = 10$.

Nominal Size:	1%	5%	10%	1%	5%	10%
	Polynomial			Normal CDF		
MKT	0.015	0.052	0.095	0.009	0.061	0.125
cmt1	0.013	0.058	0.121	0.014	0.071	0.126
cmt2	0.015	0.055	0.115	0.025	0.072	0.122
cmt5	0.019	0.066	0.130	0.015	0.070	0.130
cmt7	0.015	0.060	0.116	0.014	0.070	0.136
cmt10	0.012	0.062	0.127	0.020	0.064	0.117
cmt20	0.014	0.055	0.108	0.019	0.076	0.119
cmt30	0.018	0.063	0.113	0.014	0.060	0.121
	Cube Root			Exponential		
MKT	0.012	0.065	0.111	0.026	0.085	0.126
cmt1	0.019	0.074	0.147	0.026	0.077	0.142
cmt2	0.015	0.072	0.117	0.023	0.076	0.133
cmt5	0.017	0.068	0.129	0.019	0.077	0.133
cmt7	0.016	0.067	0.127	0.030	0.097	0.166
cmt10	0.018	0.065	0.120	0.027	0.096	0.147
cmt20	0.016	0.072	0.127	0.025	0.100	0.154
cmt30	0.019	0.078	0.128	0.032	0.094	0.169
	Normal PDF			Cosine and Sine		
MKT	0.013	0.068	0.129	0.020	0.072	0.134
cmt1	0.008	0.048	0.103	0.033	0.096	0.155
cmt2	0.019	0.076	0.132	0.020	0.068	0.120
cmt5	0.023	0.081	0.136	0.020	0.082	0.146
cmt7	0.015	0.067	0.130	0.019	0.072	0.143
cmt10	0.017	0.070	0.134	0.016	0.064	0.126
cmt20	0.018	0.075	0.134	0.019	0.073	0.132
cmt30	0.016	0.065	0.124	0.018	0.064	0.117

Table 24: **Monte Carlo: Test of Equal Predictive Accuracy**

This table examines the finite sample performance of the Clark and West (2007) test statistic for $h = 6$. We simulate 1000 samples of excess returns under the joint null hypothesis $\mathbb{H}_{1,0}$ of no predictability. Each sample represents a multivariate time series of eight different assets with sample mean and covariance calibrated to excess returns to the equity market and seven constant maturity Treasury portfolios cmt1, cmt2, cmt5, cmt7, cmt10, cmt20, cmt30. The Clark and West (2007) test is implemented with three sample splits—40%, 50%, and 60%—and three variance estimators—the MA(h-1) approach discussed in Appendix A.1, Newey and West (1987) with h lags, and Hansen and Hodrick (1980). For each Monte Carlo simulation the order of sieve B-splines is chosen by MSFE cross-validation, where the last observations is left out from a series of expanding-window pseudo out-of-sample forecasts.

Nominal Size:	1%	5%	10%	1%	5%	10%	1%	5%	10%
MA(h-1)									
	40% in-sample			50% in-sample			60% in-sample		
MKT	0.007	0.069	0.189	0.006	0.066	0.172	0.003	0.052	0.154
cmt1	0.001	0.046	0.132	0.000	0.037	0.125	0.003	0.034	0.100
cmt2	0.000	0.034	0.123	0.002	0.023	0.102	0.001	0.028	0.073
cmt5	0.001	0.025	0.113	0.000	0.015	0.078	0.001	0.014	0.055
cmt7	0.001	0.024	0.108	0.000	0.017	0.068	0.001	0.019	0.057
cmt10	0.000	0.035	0.108	0.000	0.021	0.073	0.000	0.021	0.068
cmt20	0.001	0.032	0.122	0.001	0.026	0.094	0.000	0.030	0.082
cmt30	0.001	0.032	0.120	0.001	0.024	0.091	0.001	0.026	0.085
Newey-West									
	40% in-sample			50% in-sample			60% in-sample		
MKT	0.173	0.376	0.525	0.165	0.343	0.451	0.156	0.325	0.439
cmt1	0.120	0.300	0.444	0.119	0.262	0.412	0.104	0.259	0.380
cmt2	0.115	0.292	0.443	0.102	0.249	0.378	0.097	0.230	0.361
cmt5	0.098	0.284	0.429	0.094	0.223	0.356	0.087	0.213	0.336
cmt7	0.100	0.281	0.435	0.093	0.234	0.376	0.086	0.207	0.329
cmt10	0.109	0.279	0.426	0.090	0.237	0.380	0.086	0.223	0.337
cmt20	0.116	0.294	0.439	0.099	0.248	0.384	0.107	0.230	0.361
cmt30	0.114	0.285	0.451	0.099	0.267	0.373	0.102	0.239	0.372
Hansen-Hodrick									
	40% in-sample			50% in-sample			60% in-sample		
MKT	0.146	0.357	0.517	0.148	0.318	0.439	0.141	0.300	0.439
cmt1	0.105	0.295	0.439	0.102	0.250	0.392	0.098	0.243	0.372
cmt2	0.108	0.269	0.427	0.081	0.234	0.368	0.078	0.212	0.340
cmt5	0.075	0.263	0.432	0.070	0.197	0.339	0.062	0.181	0.310
cmt7	0.076	0.266	0.418	0.074	0.217	0.350	0.060	0.176	0.307
cmt10	0.086	0.267	0.406	0.074	0.208	0.352	0.060	0.199	0.309
cmt20	0.097	0.273	0.434	0.067	0.223	0.352	0.070	0.205	0.343
cmt30	0.099	0.267	0.427	0.075	0.240	0.349	0.077	0.220	0.352

A.3 Robustness to Using Log of VIX

Our preferred specification, presented in the main text, allows returns to be driven by a common, unknown function of the level of the VIX. Here we include an extension of our main empirical results, in which we use the log of the VIX rather than the VIX level itself, as a robustness check. In particular, Tables 25 and 26 report sieve reduced rank regressions of stock and bond returns on lagged values of the log of VIX. The tables show that the log of VIX does not linearly predict excess stock and bond returns, while a nonlinear transformation of the log of VIX does. The main conclusions of the paper are therefore upheld in the log of VIX case. For reference, we also plot the time series of $\phi_h(v_t)$ for $h = 6$ under the log specification in Figure 17.

Figure 17: $\phi_h(v_t)$ **for** $v_t = log(vix_t)$

This plot shows the common nonlinear function $\phi_h(v_t)$ estimated from stocks and bonds for $v_t = log(vix_t)$. The forecast horizon is $h = 6$. The sample consists of monthly observations from 1990:1 to 2014:9.

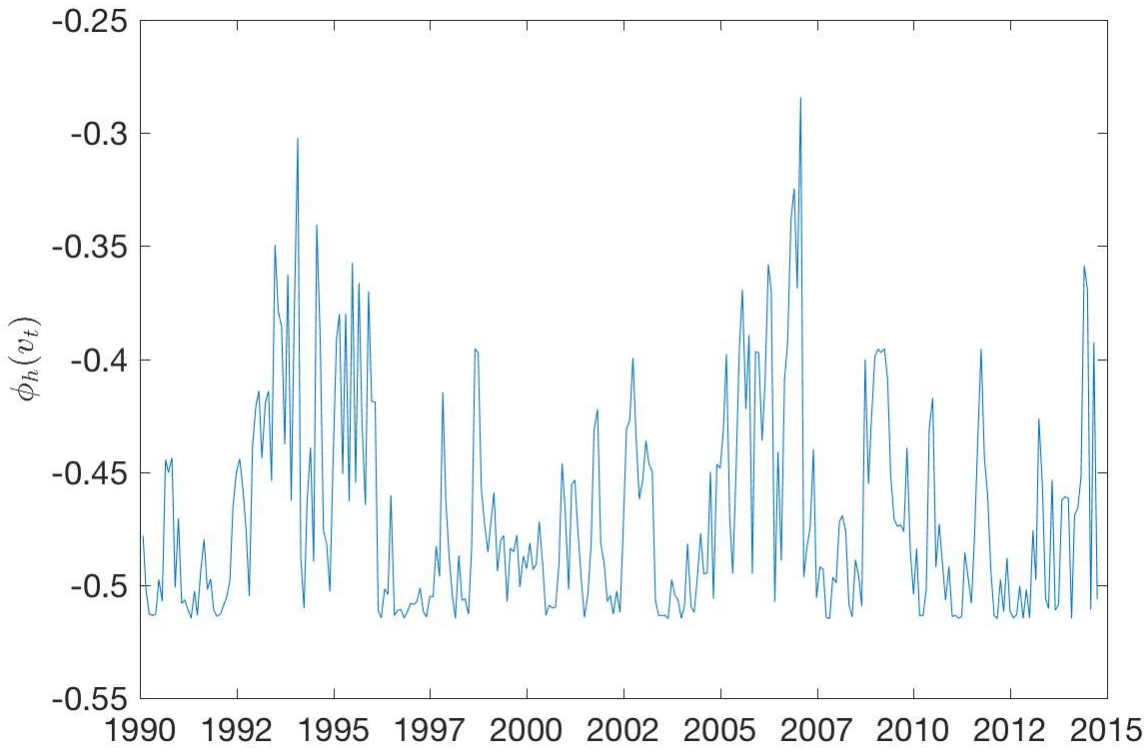


Table 25: **Nonlinear** $\log(VIX)$ **Predictability using the Cross-Section: 1990 - 2014**

This table reports results from three predictive sieve reduced rank regressions (SRRR) for each of $h = 6$, 12, and 18 month ahead forecasting horizons: (1) estimates of a_h^i and b_h^i from the SRRR $Rx_{t+h}^i = a_h^i + b_h^i v_t + \varepsilon_{t+h}^i$ of portfolio i's excess returns on linear $v_t = \log(vix_t)$; (2) estimates of a_h^i and b_h^i from the SRRR $Rx_{t+h}^i = a_h^i + b_h^i \phi_h(v_t) + \varepsilon_{t+h}^i$ of portfolio i's excess return on the common nonparametric function $\phi_h(\cdot)$ of v_t; (3) the same regression augmented with controls $\mathbf{f^i} \cdot (DEF_t, VRP_t, TERM_t, DY_t)'$ representing the default spread (DEF, 10-year Treasury yield minus Moody's BAA corporate bond yield), the variance risk premium (VRP, realized volatility minus VIX), the term spread (TERM, 10-year minus 3-month Treasury yields), and the S&P 500's (log) dividend yield. The index $i = 1, \ldots, n$ ranges over the CRSP value-weighted market excess return and the seven CRSP constant maturity Treasury excess returns corresponding to 1, 2, 5, 7, 10, 20, and 30 years to maturity. The sieve reduced rank regressions are introduced in Section 2 in the text. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a_i and $\mathbf{f^i}$ and for the χ^2-statistic on $b^i \phi_h(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $\mathbf{A}_h = 0$.

Horizon h = 6

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	−0.12	1.00	0.55	1.00**	0.03	1.00***	0.05**	−1.42***	−0.01	0.19
cmt1	−0.02	0.14**	−0.04	−0.10***	−0.05	−0.27***	0.00	0.04**	0.00*	0.02***
cmt2	−0.02	0.21	−0.08	−0.20***	−0.08	−0.43***	0.00	0.10**	0.00	0.02**
cmt5	−0.01	0.23	−0.18	−0.45***	−0.12	−0.79***	−0.02**	0.25**	0.01**	0.01
cmt7	−0.01	0.24	−0.21	−0.54***	−0.13	−0.92***	−0.03**	0.34**	0.02**	0.00
cmt10	0.03	0.04	−0.23	−0.57***	−0.12	−0.90***	−0.04**	0.42**	0.03***	0.01
cmt20	0.08	−0.11	−0.28	−0.72***	−0.08	−1.00**	−0.05***	0.54*	0.05***	−0.04
cmt30	0.14	−0.42	−0.40	−0.97***	−0.09	−1.30***	−0.08***	0.74*	0.06***	−0.06
Joint p-val		0.228		0.005		0.005				
Bootstrap p-val		0.274		0.009		0.000				

Horizon h = 12

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.00	1.00	0.30	1.00*	−0.07	1.00**	0.02**	−0.68***	0.00*	0.19
cmt1	−0.01	0.32**	−0.02	−0.14***	−0.03	−0.55***	0.00	0.04***	0.00**	0.02***
cmt2	−0.03	0.62*	−0.04	−0.26***	−0.05	−0.86***	0.00	0.07***	0.00	0.03***
cmt5	−0.05	1.05	−0.09	−0.55***	−0.08	−1.54***	−0.01	0.10**	0.01	0.02
cmt7	−0.05	1.14	−0.11	−0.67**	−0.08	−1.83***	−0.02*	0.14*	0.01**	0.02
cmt10	−0.01	0.72	−0.11	−0.68**	−0.09	−1.86**	−0.02**	0.16*	0.02**	0.03
cmt20	−0.01	0.83	−0.15	−0.92**	−0.07	−2.25**	−0.04**	0.19	0.03***	0.00
cmt30	0.01	0.50	−0.22	−1.22**	−0.08	−2.83**	−0.05**	0.25	0.04***	−0.02
Joint p-val		0.313		0.061		0.007				
Bootstrap p-val		0.351		0.042		0.002				

Horizon h = 18

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	−0.01	1.00	0.22	1.00	−0.11	1.00**	0.02**	−0.58***	0.01	0.18
cmt1	−0.01	0.20	−0.02	−0.19***	−0.03	−0.86***	0.00	0.04***	0.00**	0.02***
cmt2	−0.02	0.44	−0.03	−0.37**	−0.04	−1.37***	0.00	0.08***	−0.01	0.02***
cmt5	−0.04	0.87	−0.08	−0.78**	−0.08	−2.40***	−0.01	0.12***	0.00	0.03**
cmt7	−0.04	0.95	−0.09	−0.92**	−0.08	−2.74***	−0.01*	0.16**	0.00	0.03*
cmt10	−0.02	0.66	−0.09	−0.93**	−0.09	−2.83***	−0.02*	0.17**	0.01*	0.04*
cmt20	−0.02	0.86	−0.12	−1.26**	−0.08	−3.30*	−0.02*	0.17	0.02**	0.01
cmt30	0.00	0.60	−0.18	−1.64**	−0.09	−4.05**	−0.03**	0.21	0.03**	0.00
Joint p-val		0.573		0.630		0.064				
Bootstrap p-val		0.578		0.105		0.002				

Table 26: **Nonlinear** $\log(VIX)$ **Predictability using the Cross-Section: 1990 - 2007**

This table reports results from three predictive sieve reduced rank regressions (SRRR) for each of $h = 6$, 12, and 18 month ahead forecasting horizons: (1) estimates of a_h^i and b_h^i from the SRRR $Rx_{t+h}^i = a_h^i + b_h^i v_t + \varepsilon_{t+h}^i$ of portfolio i's excess returns on linear $v_t = \log(vix_t)$; (2) estimates of a_h^i and b_h^i from the SRRR $Rx_{t+h}^i = a_h^i + b_h^i \phi_h(v_t) + \varepsilon_{t+h}^i$ of portfolio i's excess return on the common nonparametric function $\phi_h(\cdot)$ of v_t; (3) the same regression augmented with controls $\mathbf{f^i} \cdot (DEF_t, VRP_t, TERM_t, DY_t)'$ representing the default spread (DEF, 10-year Treasury yield minus Moody's BAA corporate bond yield), the variance risk premium (VRP, realized volatility minus VIX), the term spread (TERM, 10-year minus 3-month Treasury yields), and the S&P 500's (log) dividend yield. The index $i = 1, \ldots, n$ ranges over the CRSP value-weighted market excess return and the seven CRSP constant maturity Treasury excess returns corresponding to 1, 2, 5, 7, 10, 20, and 30 years to maturity. The sieve reduced rank regressions are introduced in Section 2 in the text. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level for t-statistics on a_i and $\mathbf{f^i}$ and for the χ^2-statistic on $b^i \phi_h(\cdot)$ derived in Theorem 1. The joint test p-value reports the likelihood that the sample was generated from the model where $\mathbf{A}_h = 0$.

Horizon h = 6

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.01	1.00	0.34	1.00*	0.09	1.00	−0.03	−0.82*	0.00	0.12
cmt1	−0.03*	0.57***	−0.04	−0.18***	−0.06	−0.68***	0.01	0.02	0.00*	0.03***
cmt2	−0.04	0.79*	−0.07	−0.33***	−0.08	−1.07***	0.01	0.10*	0.00	0.03***
cmt5	−0.05	1.00	−0.13	−0.63***	−0.11	−1.83***	0.00	0.24*	0.01	0.03*
cmt7	−0.03	0.86	−0.16	−0.75***	−0.09	−2.05***	−0.02	0.30*	0.02	0.02
cmt10	0.03	0.08	−0.17	−0.78**	−0.07	−2.05***	−0.03	0.35*	0.02*	0.02
cmt20	0.03	0.31	−0.19	−0.92**	−0.03	−2.37**	−0.05	0.46**	0.03**	−0.01
cmt30	0.03	0.20	−0.23	−1.07**	−0.02	−2.72**	−0.07*	0.57**	0.04**	−0.03
Joint p-val		0.016		0.007		0.003				
Bootstrap p-val		0.046		0.010		0.006				

Horizon h = 12

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.14	1.00	0.23	1.00	0.00	1.00	−0.03	−0.51*	0.00	0.16
cmt1	−0.03**	−0.73***	−0.03	−0.30***	−0.05	2.31***	0.00	0.03	0.00	0.03***
cmt2	−0.05	−1.16**	−0.06	−0.54***	−0.08	3.76***	0.01	0.07*	0.00	0.03***
cmt5	−0.08	−1.79	−0.12	−1.00***	−0.13	6.67***	0.00	0.11*	0.01	0.04
cmt7	−0.07	−1.69	−0.14	−1.19***	−0.13	7.72***	−0.01	0.13*	0.01*	0.03
cmt10	−0.01	−0.79	−0.14	−1.20***	−0.12	8.13***	−0.03*	0.13*	0.02**	0.04
cmt20	−0.01	−1.04	−0.17	−1.48***	−0.09	9.44***	−0.04**	0.13*	0.03**	0.01
cmt30	−0.03	−1.19	−0.22	−1.75**	−0.11	11.04***	−0.06**	0.16*	0.04**	0.00
Joint p-val		0.005		0.001		0.007				
Bootstrap p-val		0.026		0.031		0.052				

Horizon h = 18

	(1) Linear lVIX		(2) Nonlinear lVIX		(3) Nonlinear lVIX and Controls					
	a^i	b^i	a^i	b^i	a^i	b^i	f^i_{DEF}	f^i_{VRP}	f^i_{TERM}	f^i_{DY}
MKT	0.22	1.00	0.27	1.00***	0.05	1.00	−0.03	−0.54**	0.00	0.14
cmt1	−0.03*	−0.28**	−0.03	−0.19***	−0.04	−1.01***	0.00	0.05**	0.00	0.03***
cmt2	−0.05	−0.47*	−0.05	−0.35***	−0.07	−1.72***	0.00	0.10**	0.00	0.03***
cmt5	−0.08	−0.77	−0.10	−0.65***	−0.13	−3.15***	0.01*	0.18**	0.00	0.05*
cmt7	−0.07	−0.77	−0.11	−0.76***	−0.14	−3.66***	0.00*	0.21**	0.01*	0.04
cmt10	−0.02	−0.40	−0.11	−0.73***	−0.14	−3.86***	−0.01**	0.23**	0.01**	0.06
cmt20	−0.03	−0.56	−0.12	−0.88***	−0.12	−4.28***	−0.02**	0.19**	0.02**	0.03
cmt30	−0.04	−0.62	−0.17	−1.08***	−0.15	−5.13***	−0.03**	0.24**	0.03***	0.03
Joint p-val		0.009		0.000		0.000				
Bootstrap p-val		0.044		0.019		0.002				

A.4 Block Bootstrap Results

In this section we provide results of a block bootstrap procedure.[30] In this resampling exercise we jointly block bootstrap only the equity and bond returns (i.e., we resample under the null of no predictability). Under our technical assumptions, we can allow for higher-order dependence in returns even under the null of no predictability. Thus, we want to choose an appropriate block size to best capture the time series and cross-sectional dependencies in the raw returns. To this end, we jointly bootstrap the returns and choose the optimal time series block size in a data-driven way utilizing results of Politis and White (2004) and Patton et al. (2009). We choose the maximum block length across assets suggested by the optimal choice derived in Politis and White (2004). The optimal choice is seven for our sample. Each bootstrapped sample can then be interpreted as representing a pseudo data set of raw returns. From here, our estimation procedure follows that of the paper repeated on each bootstrapped sample. That is, for a given bootstrapped sample, we first construct holding period returns for the particular forecast horizon h of interest. Then, for the same bootstrapped sample, we estimate the model on the holding period returns and construct the test statistic $\hat{\mathcal{T}}_1$ using the reverse regression estimates based on the observed VIX and control variables. It is important to emphasize that the observed VIX and control variables are not resampled. By resampling under the null hypothesis we can then compare the bootstrap-based distribution to that of the distribution we use to conduct inference in our main results. This exercise is a natural counterpart to the simulation study from above, but has the additional advantage of generating pseudo-data sets directly from the observed returns. Therefore it does not require explicit assumptions on the underlying data generating process. Figures 18, 19, and 20, plot the comparison of the bootstrapped test statistic distribution and the asymptotic distribution derived in Theorem 1 for the forecast horizons $h = 6, 12$, and 18 considered in the paper. The figures show a close alignment between the bootstrapped distributions and the asymptotic distribution for each forecast horizon.

[30] We thank our referees for suggesting this exercise.

Figure 18: **Block Bootstrapped Test Statistics $\hat{\mathcal{T}}_1$ for $h = 6$**

This figure shows block boostrapped test statistics of the null hypothesis $\mathbb{H}_{1,0}$ as a histogram in blue. The bootstrap test statistics are obtained by jointly block resampling the equity and Treasury returns. Next, $\hat{\mathcal{T}}_1$ is formed using the observed VIX and control variables. In red is the corresponding χ^2 distribution. The bootstrap block size is seven and all results are based on 10,000 replications. The sample consists of monthly observations from 1990:1 to 2014:9.

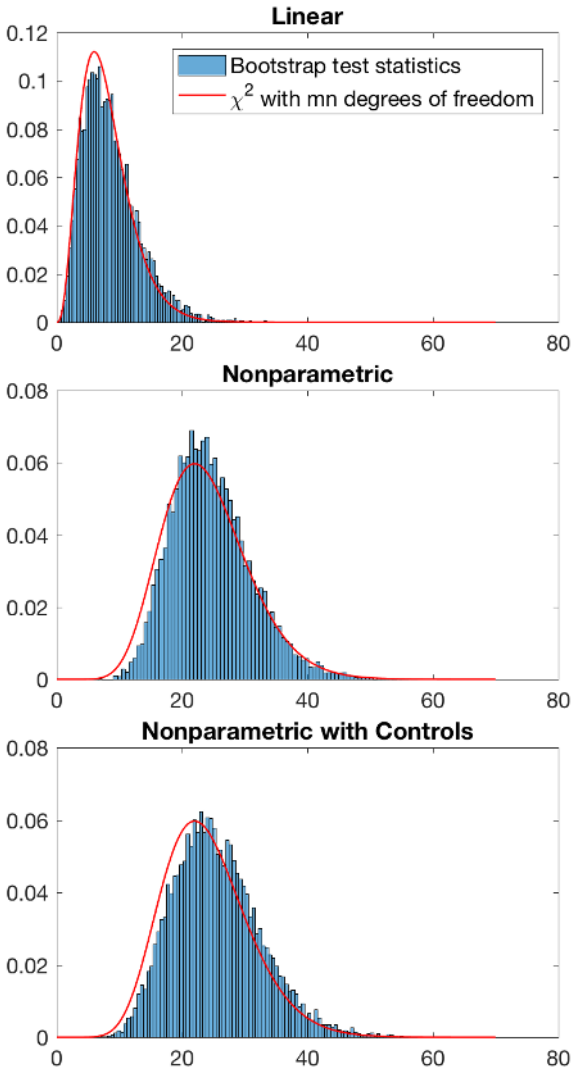


Figure 19: **Block Bootstrapped Test Statistics $\hat{\mathcal{T}}_1$ for $h = 12$**

This figure shows block boostrapped test statistics of the null hypothesis $\mathbb{H}_{1,0}$ as a histogram in blue. The bootstrap test statistics are obtained by jointly block resampling the equity and Treasury returns. Next, $\hat{\mathcal{T}}_1$ is formed using the observed VIX and control variables. In red is the corresponding χ^2 distribution. The bootstrap block size is seven and all results are based on 10,000 replications. The sample consists of monthly observations from 1990:1 to 2014:9.

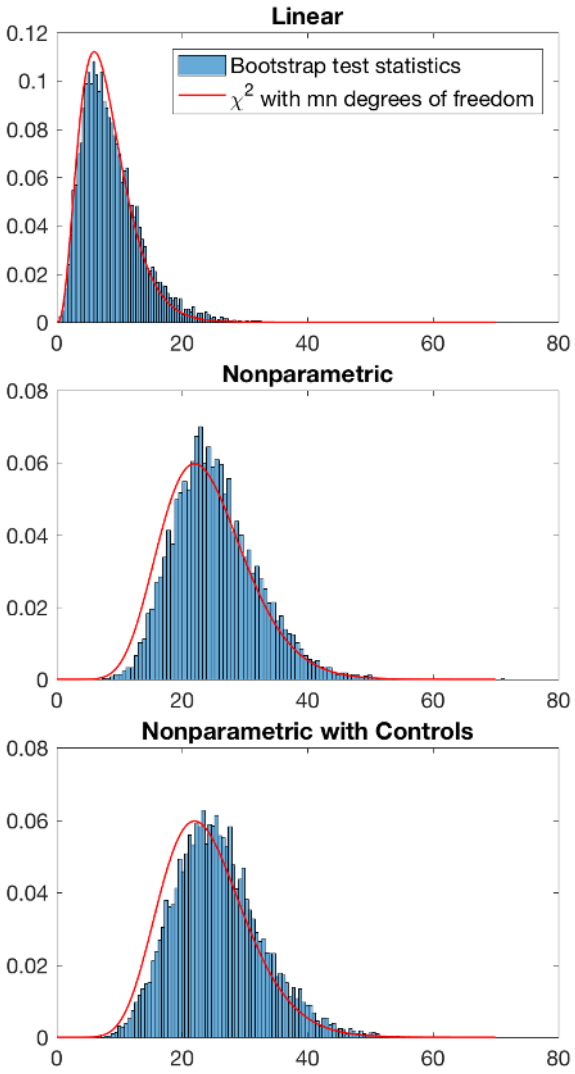
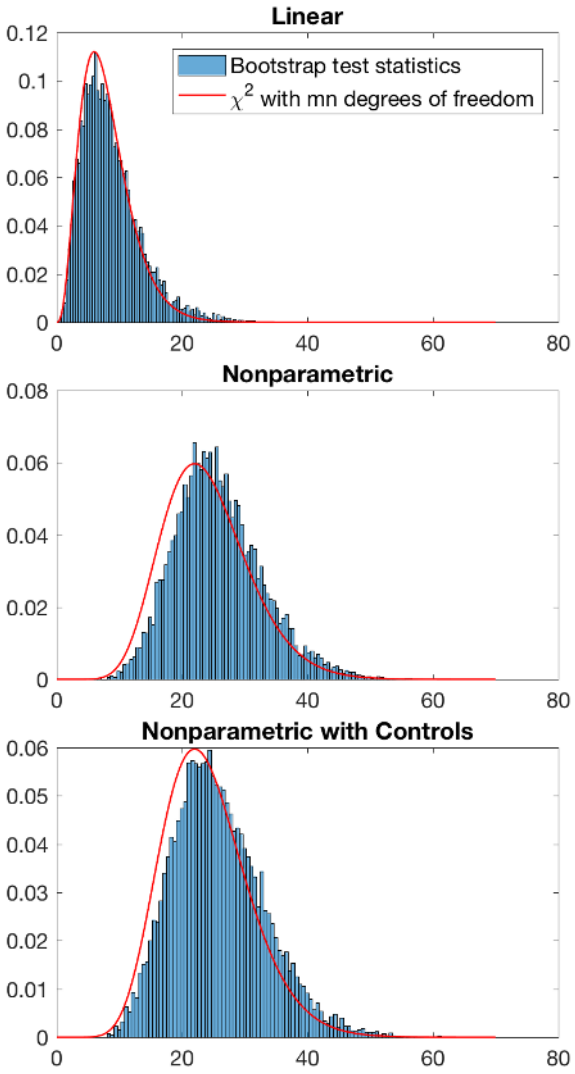


Figure 20: **Block Bootstrapped Test Statistics $\hat{\mathcal{T}}_1$ for $h = 18$**

This figure shows block boostrapped test statistics of the null hypothesis $\mathbb{H}_{1,0}$ as a histogram in blue. The bootstrap test statistics are obtained by jointly block resampling the equity and Treasury returns. Next, $\hat{\mathcal{T}}_1$ is formed using the observed VIX and control variables. In red is the corresponding χ^2 distribution. The bootstrap block size is seven and all results are based on 10,000 replications. The sample consists of monthly observations from 1990:1 to 2014:9.



A.5 Proofs

Assumptions and Preliminary Lemmas

To prove the theorem we need a number of assumptions. To impose conditions on time-series dependence we utilize the conditions of Doukhan and Louhichi (1999) along with certain mixing conditions (see, for example, Bradley (2007)). Before providing the assumptions let $\|\cdot\|$ denote the Euclidean matrix norm, i.e., for a matrix A, $\|A\| = \sqrt{\text{tr}\left(A'A\right)}$ and $\|\cdot\|_1$ denote the L_1 matrix norm, i.e., for a $n \times m$ matrix A with (i,j) element A_{ij}, $\|A\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|$.

Assumption A.1. *(i)* $\left(\text{Rx}_t', v_t\right)_{t\in\mathbb{N}}$ *are strictly stationary and satisfy equation (3.4) with* $\mathbb{E}\left\|\left(\text{Rx}_t', v_t\right)\right\|^{2+\delta}$ *bounded for some* $\delta > 0$ *and satisfy* $\left\|T^{-1}\sum_{t=1}^{T}\text{Rx}_{t+1}\text{Rx}_{t+1}' - \mathbb{E}\left[\text{Rx}_{t+1}\text{Rx}_{t+1}'\right]\right\| = O_p\left(T^{-\zeta}\right)$ *for some* $\zeta > 0$.

(ii) v_t *has a density that is bounded and bounded away from zero on a compact support* $\mathcal{V}$

(iii) $\{v_t\}_{t\in\mathbb{N}}$ *is algebraically beta-mixing at rate* φ

(iv) The sequence $m = m\left(T\right)$ *satisfies* $m \to \infty$ *and there exists a* $\nu > 0$ *and constants* C_1 *and* C_2 *such that* $C_1 T^{1/\nu} \leq m \leq C_2 T^{1/\nu}$ *and* $\nu < \frac{\varphi(1+\epsilon)}{1+\varphi}$ *for some* $\epsilon > 0$

(v) Let $\xi_{m,t} := \text{vec}\left(e_t^* \tilde{X}_{m,t-1}^{(h)\prime}\right)$ *and* $\mathcal{F}_L$ *the class of bounded Lipschitz functions. Then* $\{\xi_{m,t}\}_{t\in\mathbb{N}}$ *is* $(\vartheta, \mathcal{F}_L, \psi)$*-weak dependent in the sense that there exists a sequence* $\{\vartheta_u\}_{u\in\mathbb{N}}$ *such that* $\vartheta_u \to 0$ *when* $u \to 0$ *and for any* ℓ_1*-tuple* $(t_1, \ldots, t_{\ell_1})$ *and any* ℓ_2*-tuple* $\left(t_1', \ldots, t_{\ell_2}'\right)$ *with* $t_1 \leq \cdots \leq t_{\ell_1} < t_{\ell_1} + u = t_1' \leq \cdots \leq t_{\ell_2}'$

$$\left|\mathbb{C}\left(f\left(\xi_{m,t_1}, \ldots, \xi_{m,t_{\ell_1}}\right), g\left(\xi_{m,t_1'}, \ldots, \xi_{m,t_{\ell_2}'}\right)\right)\right| \leq \text{Lip}\left(f\right)\text{Lip}\left(g\right)\omega\left(\ell_1, \ell_2\right)\vartheta_u,$$

where $\omega : \mathbb{N}^2 \mapsto [0, \infty)$, $f \in \mathcal{F}_L : \mathbb{R}^{\ell_1} \mapsto \mathbb{R}$, $g \in \mathcal{F}_L : \mathbb{R}^{\ell_2} \mapsto \mathbb{R}$, *and*

$$\text{Lip}\left(h\right) = \sup_{y_1 \neq y_2} \frac{\left|h\left(y_1\right) - h\left(y_2\right)\right|}{\left\|y_1 - y_2\right\|_1}.$$

$\xi_{m,t}$ *also satisfies* $\mathbb{E}\left[\xi_{m,t}'\xi_{m,s}\right] = 0$ *for* $s \neq t$ *and*

$$Tm^{9/2} \sup_{t\in\{1,\ldots,T\}, f,g\in\mathcal{F}_L} \left|\mathbb{C}\left(f\left(\xi_{m,t}\right), g\left(\xi_{m,1}, \ldots, \xi_{m,t-1}, \xi_{m,t+1}, \ldots, \xi_{m,t_{\ell_2}'}\right)\right)\right| \to 0.$$

(vi) Let $\varsigma_{m,t} := \text{vech}\left(\xi_{m,t}\xi_{m,t}'\right)$. *Then* $\varsigma_{m,t}$ *is uniform mixing with mixing coefficients* $\{\varkappa_j\}$ *which satisfy* $\lim_{J\to\infty} J^{6\nu-1}U_J = C < \infty$ *where* $U_J = \sum_{j=1}^{J}\varkappa_j^{1/2}$.

(vii) Let $\Sigma\left(v_t\right) = \mathbb{E}\left[e_{t+1}^* e_{t+1}^{*\prime}\middle| v_t\right]$. *There exists a nonrandom, positive definite matrix* $\underline{\Sigma}$ *such that* $\Sigma\left(v_t\right) - \underline{\Sigma}$ *is positive semi-definite and* $\lambda_{\min}\left(\underline{\Sigma}\right) \geq c > 0$ *for some* c.

(viii) $\phi_h(\cdot)$ *belongs to a Holder space of smoothness* $p \geq 1$ *on the domain* $\mathcal{V}$

Define $\Omega_m \equiv \mathbb{E}\left[\tilde{X}_{m,t}^{(h)}\tilde{X}_{m,t}^{(h)\prime}\right]$ where $\tilde{X}_{m,t}^{(h)} = \left(1, X_{m,t}^{(h)\prime}\right)'$. Also define $\hat{\Omega}_{m,T} = T^{-1}\sum_{t=1}^{T}\tilde{X}_{m,t}^{(h)}X_{m,t}^{(h)\prime}$ and let $\kappa_{m,n}$ be the $mn \times mn$ commutation matrix which satisfies $\text{vec}(L') = \kappa_{m,n}\text{vec}(L)$ for an $m \times n$ matrix L. It will also be convenient to define $S_X = \begin{bmatrix}0_{m\times 1} & \mathrm{I}_m\end{bmatrix}$ so that $X_{m,t}^{(h)} = S_X\tilde{X}_{m,t}^{(h)}$ and let $s_{d,g}$ be the $d \times 1$ unit vector with a one for the gth element. In addition define

$$\Gamma_m := \mathbb{E}\left[\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)'\right] = \mathbb{E}\left[\left(\tilde{X}_{m,t}^{(h)}\tilde{X}_{m,t}^{(h)\prime}\otimes\mathbb{E}\left[e_{t+1}^{*}e_{t+1}^{*\prime}\middle|\, v_t\right]\right)\right].$$

Also note that all limits are taken as $T \to \infty$ and C is a generic positive constant. To simplify notation, note that all derivations are on the sequence of events where the minimum eigenvalues of $\hat{\Omega}_{m,T}$, $\hat{\Gamma}_{m,T}$ and $\hat{W}$ are bounded away from zero. Based on our results these conditions hold with probability approaching one.

Before proceeding, we will list a number of useful results that we will use throughout. The proofs rely on two technical results obtained in Chen and Christensen (2015). Recall that the B-spline basis is $\tilde{X}_{m,t}^{(h)} = \tilde{X}_{m}^{(h)}(v_t)$.

$$\sup_{v_t\in\mathcal{V}}\left\|\tilde{X}_{m,t}\right\| = O\left(m^{1/2}\right), \tag{CC1}$$

$$\left\|\Omega_{m,t}^{-1/2}\hat{\Omega}_{m,t}\Omega_{m,t}^{-1/2} - \mathrm{I}_{m+1}\right\| = O_p\left(\sqrt{T^{-\varphi/(1+\varphi)}m\log(m)}\right). \tag{CC2}$$

Then we have the following lemma.

Lemma 3. *Let Assumption A.1 hold. Then,*

(i) $\left\|\hat{\Omega}_{m,T} - \Omega_m\right\| = O_p\left(\sqrt{T^{-\varphi/(1+\varphi)}m\log(m)}\right)$;

(ii) $\mathbb{V}(\mathrm{Rx}_{t+1}) - \mathbb{E}\left[e_{t+1}^{*}e_{t+1}^{*\prime}\right] \geq 0$ *in a positive semi-definite sense;*

(iii) $\left\|\hat{D}_{\text{ols}} - D^{*}\right\| = O_p\left(m^{1/2}T^{-1/2}\right)$ *where* $D = \begin{bmatrix}\mathbf{a} & \mathbf{b}\gamma'\end{bmatrix}$;

(iv) $\sup_{v\in\mathcal{V}}\left\|\left(\hat{D}_{\text{ols}} - D^{*}\right)\tilde{X}_{m}^{(h)}(v)\right\| = O_p\left(mT^{-1/2}\right)$;

(v) $\mathbb{E}\left[\|e_t^{*}\|^{2+\varrho}\right] \leq Cm^{2+\varrho}$ *for* $\varrho \leq \delta$;

(vi) $\mathbb{E}\left|\gamma^{*\prime}\tilde{X}_{m,t}^{(h)}\right| < C$

(vii) $\mathbb{E}\left\|\mathbf{b}^{*}\gamma^{*\prime}\tilde{X}_{m,t}^{(h)}\right\| < C$

(viii) $\left\|T^{-1}\sum_{t=1}^{T}\text{vec}\left(\xi_{m,t+1}\xi_{m,t+1}' - \Gamma_m\right)\right\| = O_p\left(\sqrt{T^{6\nu-1}U_T}\right)$ *for* $\delta \geq 2$;

(ix) $\left\|\hat{\Gamma}_{m,T} - \Gamma_m\right\| = O_p\left(\sqrt{T^{6\nu-1}U_T}\right) + O_p\left(m^2 T^{-1/2}\right)$ *for* $\delta \geq 2$.

Proof of Lemma 3 (i) follows by CC2. (ii) follows by properties of the best linear predictor since $\mathbf{b}^*\gamma^{*\prime}\mathbb{V}\left(X_{m,t}^{(h)}\right)\gamma^*\mathbf{b}^{*\prime} \geq 0$ in a positive semi-definite sense. Next, consider (iii). Let $\mathbf{1}_{\Omega,T}$ be an indicator of the event that $\lambda_{\min}\left(\Omega_{m,t}^{-1/2}\hat{\Omega}_{m,T}\Omega_{m,t}^{-1/2}\right) \leq \frac{1}{2}$. By (i). Note that $\mathbf{1}_{\Omega,T} = 0$ with probability approaching one. Thus, $\left\|\hat{D}_{\text{ols}} - D^*\right\| = (1 - \mathbf{1}_{\Omega,T})\left\|\hat{D}_{\text{ols}} - D^*\right\| + o_p(1)$ and

$$(1 - \mathbf{1}_{\Omega,T})\left\|\hat{D}_{\text{ols}} - D^*\right\| \leq (1 - \mathbf{1}_{\Omega,T})\,\lambda_{\max}\left(\hat{\Omega}_{m,T}^{-1}\right)\left\|T^{-1}\sum\nolimits_{t=1}^{T} e_t^*\tilde{X}_{m,t-1}^{(h)\prime}\hat{\Omega}_{m,T}^{-1}\right\|.$$

Then, $\left\|T^{-1}\sum_{t=1}^{T} e_t^*\tilde{X}_{m,t-1}^{(h)\prime}\right\|^2 = O_p\left(mT^{-1}\right)$ by Markov's inequality, (ii), and since

$$\mathbb{E}\left\|T^{-1}\sum\nolimits_{t=1}^{T} e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right\|^2 = T^{-2}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[\left\|\tilde{X}_{m,t}^{(h)\prime}\right\|^2\left\|e_{t+1}^*\right\|^2\right] \leq CmT^{-1}.$$

Next, (iv) follows since

$$\sup_{v\in\mathcal{V}}\left\|\left(\hat{D}_{\text{ols}} - D^*\right)\tilde{X}_m^{(h)}(v_t)\right\| \leq C\sup_{v\in\mathcal{V}}\left\|\tilde{X}_m^{(h)}(v)\right\|\left\|\hat{D}_{\text{ols}} - D^*\right\| \leq CmT^{-1/2},$$

by CC1 and (iii). (v) follows by Minkowski's inequality since

$$\left(\mathbb{E}\left[\left\|e_{t+1}^*\right\|^{2+\varrho}\right]\right)^{\frac{1}{2+\varrho}} \leq \left(\mathbb{E}\left[\left\|\mathrm{R}x_{t+1}\right\|^{2+\varrho}\right]\right)^{\frac{1}{2+\varrho}} + \left(\mathbb{E}\left[\left\|D^*\tilde{X}_{m,t}^{(h)}\right\|^{2+\varrho}\right]\right)^{\frac{1}{2+\varrho}},$$

and $\left\|D^*\tilde{X}_{m,t}^{(h)}\right\|^2 \leq Cm^2$. For (vi),

$$\gamma^{*\prime}\tilde{X}_{m,t}^{(h)} = s_{n,1}'\mathrm{R}x_{t+1} - s_{n,1}'\mathbf{a}^* - s_{n,1}'e_{t+1}^*,$$

so that

$$\mathbb{E}\left|\gamma^{*\prime}\tilde{X}_{m,t}^{(h)}\right|^2 \leq C\cdot\mathbb{E}\left|s_{n,1}'\mathrm{R}x_{t+1}\right|^2 + C\cdot\left|s_{n,1}'\mathbf{a}^*\right|^2 + C\cdot\mathbb{E}\left|s_{n,1}'e_{t+1}^*\right|^2 \leq C\cdot\mathbb{E}\left|s_{n,1}'Rx_{t+1}\right|^2 + C\cdot\left|s_{n,1}'\mathbf{a}^*\right|^2.$$

(vii) follows by similar steps. Next consider $(viii)$. We have that

$$\begin{aligned}
&\mathbb{E}\left\|T^{-1}\sum\nolimits_{t=1}^{T}\mathrm{vec}\left(\xi_{m,t+1}\xi_{m,t+1}' - \Gamma_m\right)\right\|^2 \\
= \;& T^{-2}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[\mathrm{tr}\left(\mathrm{vec}\left(\xi_{m,t+1}\xi_{m,t+1}' - \Gamma_m\right)\mathrm{vec}\left(\xi_{m,t}\xi_{m,t}' - \Gamma_m\right)'\right)\right] \\
&+T^{-2}\mathbb{E}\left[\sum\nolimits_{1\leq|j|\leq T}\sum\nolimits_{\{t:t,t+j\in\{1,\ldots,T\}\}}\mathrm{tr}\left(\mathrm{vec}\left(\xi_{m,t+1}\xi_{m,t+1}' - \Gamma_m\right)\mathrm{vec}\left(\xi_{m,t+1+j}\xi_{m,t+1+j}' - \Gamma_m\right)'\right)\right].
\end{aligned}$$

The first term is

$$\begin{aligned}
& T^{-2}\sum_{t=1}^{T}\mathbb{E}\left[\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)'\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)\right] \\
\leq\ & T^{-1}\mathbb{E}\left[\operatorname{tr}\left(\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'\right)\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'\right)'\right)\right] \\
=\ & T^{-1}\mathbb{E}\left[\left\|\tilde{X}_{m,t}^{(h)}\right\|^4\left\|e_{t+1}^*\right\|^4\right] \\
\leq\ & CT^{-1}m^6,
\end{aligned}$$

by CC1 and (v). This term is $o(1)$ under our assumptions. The second term is

$$\begin{aligned}
& T^{-2}\mathbb{E}\left[\sum\nolimits_{1\leq|j|\leq T}\sum\nolimits_{\{t:t,t+j\in\{1,\ldots,T\}\}}\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)'\operatorname{vec}\left(\xi_{m,t+1+j}\xi_{m,t+1+j}'-\Gamma_m\right)\right] \\
\leq\ & 2T^{-1}\sum\nolimits_{1\leq|j|\leq T}\varkappa_j^{1/2}\mathbb{E}\sum\nolimits_{\ell=1}^{n(m+1)}\left|s_{n(m+1),\ell}'\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)\right|^2 \\
=\ & 2T^{-1}\sum\nolimits_{1\leq|j|\leq T}\varkappa_j^{1/2}\mathbb{E}\left[\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)'\operatorname{vec}\left(\xi_{m,t+1}\xi_{m,t+1}'-\Gamma_m\right)\right] \\
\leq\ & CT^{-1}\sum\nolimits_{1\leq|j|\leq T}\varkappa_j^{1/2}m^6 \\
\leq\ & CT^{6\nu-1}\sum\nolimits_{1\leq j\leq T}\varkappa_j^{1/2},
\end{aligned}$$

by properties of uniform mixing variables. Finally, for (ix) we have that $\hat{\Gamma}_{m,T}=\mathfrak{A}_1+\mathfrak{A}_2+\mathfrak{A}_2'+\mathfrak{A}_3$,where

$$\begin{aligned}
\mathfrak{A}_1 &= T^{-1}\sum\nolimits_{t=1}^{T}\operatorname{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\operatorname{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)'-\Gamma_m \\
\mathfrak{A}_2 &= T^{-1}\sum\nolimits_{t=1}^{T}\operatorname{vec}\left(\left(\hat{e}_{t+1}-e_{t+1}^*\right)\tilde{X}_{m,t}^{(h)\prime}\right)\operatorname{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)' \\
\mathfrak{A}_3 &= T^{-1}\sum\nolimits_{t=1}^{T}\operatorname{vec}\left(\left(\hat{e}_{t+1}-e_{t+1}^*\right)\tilde{X}_{m,t}^{(h)\prime}\right)\operatorname{vec}\left(\left(\hat{e}_{t+1}-e_{t+1}^*\right)\tilde{X}_{m,t}^{(h)\prime}\right)'.
\end{aligned}$$

By (v) we have that

$$\|\mathfrak{A}_1\|=O_p\left(\sqrt{T^{6\nu-1}\sum\nolimits_{1\leq j\leq T}\varkappa_j^{1/2}}\right).$$

Next consider $\mathfrak{A}_2$,

$$\|\mathfrak{A}_2\|\leq CT^{-1}\sum\nolimits_{t=1}^{T}\left(\left\|\tilde{X}_{m,t}^{(h)}\tilde{X}_{m,t}^{(h)\prime}\right\|\left\|\left(\hat{e}_{t+1}-e_{t+1}^*\right)e_{t+1}^{*\prime}\right\|\right),$$

which is $O_p\left(m^2T^{-1/2}\right)$ by CC1 and (iv). By similar steps, $\mathfrak{A}_3$ is of lower order than $\mathfrak{A}_2$.

■

Sieve Extremum Estimators

Let $\theta = (\mathbf{a}', \mathbf{b}', \gamma')'$ and note that our estimator solves,

$$\hat{\theta} = \arg\min_{\theta} \hat{\mathcal{Q}}_T\left(\theta; \hat{W}\right), \qquad \hat{\mathcal{Q}}_T\left(\theta; \hat{W}\right) = T^{-1}\sum\nolimits_{t=1}^{T}\left(\mathrm{Rx}_{t+1} - D\tilde{X}_{m,t}^{(h)}\right)' \hat{W}^{-1}\left(\mathrm{Rx}_{t+1} - D\tilde{X}_{m,t}^{(h)}\right).$$

Recall that $D = \begin{bmatrix}\mathbf{a} & \mathbf{b}\gamma'\end{bmatrix}$. Then,

$$\hat{\mathcal{Q}}_T\left(\theta; \hat{W}\right) = \hat{\mathcal{Q}}_{1T}\left(\theta; \hat{W}\right) + \hat{\mathcal{Q}}_{2T}\left(\hat{W}\right),$$

where

$$\begin{aligned}
\hat{\mathcal{Q}}_{1T}(\theta; W) &= T^{-1}\sum\nolimits_{t=1}^{T}\left(\left(\hat{D}_{\mathrm{ols}} - D\right)\tilde{X}_{m,t}^{(h)}\right)' W^{-1}\left(\hat{D}_{\mathrm{ols}} - D\right)\tilde{X}_{m,t}^{(h)} \\
\hat{\mathcal{Q}}_{2T}(W) &= T^{-1}\sum\nolimits_{t=1}^{T}\left(\mathrm{Rx}_{t+1} - \hat{D}_{\mathrm{ols}}\tilde{X}_{m,t}^{(h)}\right)' W^{-1}\left(\mathrm{Rx}_{t+1} - \hat{D}_{\mathrm{ols}}\tilde{X}_{m,t}^{(h)}\right).
\end{aligned}$$

Also define

$$\begin{aligned}
\mathcal{Q}_{1T}(\theta; W) &= \mathbb{E}\left[\left(\left(\hat{D}_{\mathrm{ols}} - D\right)\tilde{X}_{m,t}^{(h)}\right)' W^{-1}\left(\hat{D}_{\mathrm{ols}} - D\right)\tilde{X}_{m,t}^{(h)}\right] \\
\mathcal{Q}_{2T}(W) &= \mathbb{E}\left[\left(\mathrm{Rx}_{t+1} - D^*\tilde{X}_{m,t}^{(h)}\right)' W^{-1}\left(\mathrm{Rx}_{t+1} - D^*\tilde{X}_{m,t}^{(h)}\right)\right].
\end{aligned}$$

We have the following lemma:

Lemma 4. *Let Assumption A.1 hold. Also, assume that* $\left\|\hat{W} - W\right\| = O_p\left(T^{-\zeta}\right)$ *where* W *has eigenvalues bounded and bounded away from zero. Then,*

$$\sup_{\theta}\left|\hat{\mathcal{Q}}_T^{\mathrm{alt}}\left(\theta; \hat{W}\right) - \mathcal{Q}^{\mathrm{alt}}(\theta; W)\right| = O_p\left(mT^{-1}\right) + O_p\left(mT^{-1/2}\right) + O_p\left(mT^{-\zeta}\right) + O_p\left(T^{-\varphi/2(1+\varphi)}m^{3/2}\log(m)\right)$$

Proof of Lemma 4 We have

$$\begin{aligned}
\sup_{\theta}\left|\hat{\mathcal{Q}}_{1T}\left(\theta; \hat{W}\right) - \mathcal{Q}_{1T}\left(\theta; \hat{W}\right)\right| \leq\ & \sup_{\theta}\left|T^{-1}\sum\nolimits_{t=1}^{T}\left(\left(\hat{D}_{\mathrm{ols}} - D^*\right)\tilde{X}_{m,t}^{(h)}\right)' \hat{W}^{-1}\left(\hat{D}_{\mathrm{ols}} - D^*\right)\tilde{X}_{m,t}^{(h)}\right| \\
&+ \sup_{\theta}\left|2T^{-1}\sum\nolimits_{t=1}^{T}\left(\left(\hat{D}_{\mathrm{ols}} - D^*\right)\tilde{X}_{m,t}^{(h)}\right)' \hat{W}^{-1}(D^* - D)\tilde{X}_{m,t}^{(h)}\right| \\
&+ \sup_{\theta}\left|T^{-1}\sum\nolimits_{t=1}^{T}\left((D^* - D)\tilde{X}_{m,t}^{(h)}\right)' \hat{W}^{-1}(D^* - D)\tilde{X}_{m,t}^{(h)} - \mathcal{Q}_{1T}\left(\theta; \hat{W}\right)\right|.
\end{aligned}$$

The first term is

$$\sup_{\theta}\left|T^{-1}\sum\nolimits_{t=1}^{T}\left(\left(\hat{D}_{\mathrm{ols}} - D^*\right)\tilde{X}_{m,t}^{(h)}\right)' \hat{W}^{-1}\left(\hat{D}_{\mathrm{ols}} - D^*\right)\tilde{X}_{m,t}^{(h)}\right| \leq \lambda_{\max}\left(\hat{\Omega}_{m,T} \otimes \hat{W}^{-1}\right)\left\|\hat{D}_{\mathrm{ols}} - D^*\right\|^2,$$

and so is $O_p\left(mT^{-1}\right)$. The second term is

$$\begin{aligned}
&\sup_{\theta}\left|2T^{-1}\sum\nolimits_{t=1}^{T}\left(\left(\hat{D}_{\text{ols}}-D^{*}\right)\tilde{X}_{m,t}^{(h)}\right)'\hat{W}^{-1}\left(D^{*}-D\right)\tilde{X}_{m,t}^{(h)}\right| \\
\leq\ & 2\left|\text{vec}\left(\hat{D}_{\text{ols}}-D^{*}\right)'\left(\hat{\Omega}_{m,T}\otimes\hat{W}^{-1}\right)\text{vec}\left(\hat{D}_{\text{ols}}-D^{*}\right)\right|^{1/2}\times \\
& 2\sup_{\theta}\left|\text{vec}\left(D^{*}-D\right)'\left(\hat{\Omega}_{m,T}\otimes\hat{W}^{-1}\right)\text{vec}\left(D^{*}-D\right)\right|^{1/2}.
\end{aligned}$$

The first factor is $O_p\left(m^{1/2}T^{-1/2}\right)$ by the steps above and the second factor is $O_p\left(m^{1/2}\right)$. Thus, the second term is $O_p\left(mT^{-1/2}\right)$. Finally, by similar steps, the third term can be shown to be $O_p\left(mT^{-\zeta}\right)+O_p\left(T^{-\varphi/2(1+\varphi)}m^{3/2}\log\left(m\right)\right)$ and $\left|\hat{\mathcal{Q}}_{2T}\left(W\right)-\mathcal{Q}_{2T}\left(W\right)\right|=o_p\left(1\right)$ under our assumptions.

Define $\theta^{*}=\left(\mathbf{a}^{*},\mathbf{b}^{*},\gamma^{*\prime}\right)'$ with corresponding estimator $\hat{\theta}$. Also define $\theta_{h,0}=\left(\mathbf{a}_h',\mathbf{b}_h',\phi_h\left(\cdot\right)\right)$ with corresponding estimator $\hat{\theta}_h$. Based on these results we now have:

Lemma 5. *Let Assumptions A.1 hold along with the assumptions of Lemma 4. In addition assume that*

$$mT^{-1/2}+mT^{-\zeta}+T^{-\varphi/2(1+\varphi)}m^{3/2}\log\left(m\right)\to 0.$$

Then $\left\|\hat{\theta}-\theta^{*}\right\|=o_p\left(1\right)$ *and* $\left\|\hat{\theta}_h-\theta_{h,0}\right\|=o_p\left(1\right)$.

The proof of Lemma 5 follows by Lemma 4 along with Theorem 3.1 from Chen (2007). The other conditions can be verified straightforwardly.

Main Results

The hypotheses of interest are

$$\begin{array}{ll}
\mathbb{H}_{1,0}:\mathbf{A}_h=0, & \mathbb{H}_{1,A}:\mathbf{A}_h\neq 0 \\
\mathbb{H}_{2,0}:b_h^{j}\phi_h\left(\cdot\right)=0, & \mathbb{H}_{2,A}:b_h^{j}\phi_h\left(\cdot\right)\neq 0 \\
\mathbb{H}_{3,0}:\phi_h\left(\bar{v}\right)=0, & \mathbb{H}_{3,A}:\phi_h\left(\bar{v}\right)\neq 0
\end{array}$$

for some fixed $\bar{v}\in\mathcal{V}$, where $\mathcal{V}$ is the support of the random variable v_t, along with

$$\begin{aligned}
&\mathbb{H}_{4,0}:\exists\, b_h,\ \phi_h\left(\cdot\right)\ \text{s.t.}\ \mathbb{E}\left[\text{Rx}_{t+h}\middle|\mathcal{F}_t\right]=b_h\phi_h\left(v_t\right) \\
&\mathbb{H}_{4,A}:\nexists\, b_h,\ \phi_h\left(\cdot\right)\ \text{s.t.}\ \mathbb{E}\left[\text{Rx}_{t+h}\middle|\mathcal{F}_t\right]=b_h\phi_h\left(v_t\right).
\end{aligned}$$

We would like to prove the following two theorems:

Theorem 1: Let Assumption A.1 hold along with the assumptions of Lemma 5. Then,

(i) if $\nu<\left(\frac{\delta}{3\delta+21}\wedge\frac{\varphi(1+\varepsilon)}{3(1+\varphi)}\wedge\frac{1}{6}\right)$ and $\delta\geq 2$, $\left(\text{vec}\left(\hat{\mathbf{A}}\right)'\hat{\text{V}}_1^{-1}\text{vec}\left(\hat{\mathbf{A}}\right)-mn\right)\Big/\sqrt{2mn}\to_{d,\mathbb{H}_{1,0}}\mathcal{N}\left(0,1\right)$;

(*ii*) if $\nu < \left(\frac{\delta}{4\delta+23} \wedge \frac{1}{6}\right)$ and $\delta \geq 2$, $\left(\left(\hat{b}^{j}\right)^{2} \hat{\gamma}' \hat{\mathrm{V}}_2^{+} \hat{\gamma} - 1\right) \Big/ \sqrt{2} \rightarrow_{d,\mathbb{H}_{2,0}} \mathcal{N}(0,1)$, where $\hat{V}_2^{+}$ is the Moore-Penrose inverse of $\hat{V}_2$;

(*iii*) $\left(\hat{\phi}_{h,m}(\bar{v}) - \phi_h(\bar{v})\right) \Big/ \hat{\mathrm{V}}_3 \rightarrow_{d,\mathbb{H}_{3,0}} \mathcal{N}(0,1)$.

Theorem 2: Let Assumption A.1 hold along with the assumptions of Lemma 5. Then, if $\nu < \left(\frac{\delta}{4\delta+23} \wedge \frac{1}{6}\right)$ and $\delta \geq 2$,

$$\left(\left(\hat{\mathbf{A}} - \hat{b}\hat{\gamma}'\right)' \hat{V}_4^{+} \left(\hat{\mathbf{A}} - \hat{b}\hat{\gamma}'\right) - s\right) \Big/ \sqrt{s} \rightarrow_{d,\mathbb{H}_{4,0}} \mathcal{N}(0,1),$$

where $s = (m-1)(n-1)$ and $\hat{V}_4^{+}$ is the Moore-Penrose inverse of $\hat{V}_4$.

Here $\hat{\mathbf{A}}$ is the OLS estimator from the reverse regression and $\hat{b}^{j}, \hat{\gamma}$ are the reduced-rank estimators from the reverse regression. $\hat{\phi}_{h,m}(\bar{v})$ is formed via the forward regression.

Remark 1 To gain intuition for the test statistic in Theorem 2 consider the parametric reduced-rank regression,

$$y_t = Ax_t + e_t, \qquad t = 1, \ldots, T,$$

y_t is $(n \times 1)$, x_t is $((m+1) \times 1)$ and A is of reduced rank. For simplicity, suppose that $e_t \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$. Then an LM test is based on $\sum_t \mathrm{vec}(\hat{e}_t x_t')$ where $\hat{e}_t := y_t - \hat{A}_{\mathrm{rr}} x_t$ and $\hat{A}_{\mathrm{rr}}$ is the MLE with the reduced-rank restrictions imposed. Then by standard properties of the OLS estimator, $\hat{A}_{\mathrm{ols}}$,

$$\sum_t \mathrm{vec}\left(\hat{e}_t x_t'\right) = \sum_t \mathrm{vec}\left(\left(\hat{A}_{\mathrm{ols}} - \hat{A}_{\mathrm{rr}}\right) x_t x_t'\right) = \left(\sum_t x_t x_t' \otimes \mathrm{I}_n\right) \mathrm{vec}\left(\hat{A}_{\mathrm{ols}} - \hat{A}_{\mathrm{rr}}\right)$$

Under regularity conditions, $\mathrm{plim}_{T\to\infty} \sum_t x_t x_t'$ is nonsingular and so inference using the LM test statistic is equivalent to conducting inference based on the difference, $\left(\hat{A}_{\mathrm{ols}} - \hat{A}_{\mathrm{rr}}\right)$. Note that we can also interpret the test statistic as a minimum distance criterion function (see Adrian et al. (2015)).

Proof of Theorem 1(*i*)**:** The OLS estimator in the reverse regression is,

$$\hat{D}_{\mathrm{ols}} = \left[\hat{\mathbf{a}} \;\; \hat{\mathbf{A}}\right] = R\tilde{X}'\left(\tilde{X}\tilde{X}'\right)^{-1}, \qquad \tilde{X} = \left[\tilde{X}_{m,0}^{(h)} \;\; \cdots \;\; \tilde{X}_{m,T-1}^{(h)}\right],$$

where R is $n \times T$ and $\tilde{X}$ is $(m+1) \times T$. Then, if we define the selection matrix $S_A = \left[0_{mn \times n} \quad \mathrm{I}_{nm}\right]$,

$$\begin{aligned}
\sqrt{T}\mathrm{vec}\left(\hat{\mathbf{A}} - \mathbf{A}^*\right) &= \sqrt{T} S_A \mathrm{vec}\left(\hat{D}_{\mathrm{ols}} - D^*\right) \\
&= S_A \left(\Omega_m^{-1} \otimes I_n\right) T^{-1/2} \sum\nolimits_{t=1}^{T} \mathrm{vec}\left(e_{t+1}^* \tilde{X}_{m,t}^{(h)\prime}\right) \\
&\quad + S_A \left(\left(\hat{\Omega}_{m,T}^{-1} - \Omega_m^{-1}\right) \otimes I_n\right) T^{-1/2} \sum\nolimits_{t=1}^{T} \mathrm{vec}\left(e_{t+1}^* \tilde{X}_{m,t}^{(h)\prime}\right) \\
&= \mathfrak{T}_{11} + \mathfrak{T}_{12}.
\end{aligned}$$

First note that $\mathfrak{T}_{12}$ is of lower order than $\mathfrak{T}_{11}$ by Lemma 3(i). Now consider the re-scaled version of $\mathfrak{T}_{11}$:

$$\mathfrak{R}_{11} = \left[S_A \left(\Omega_m^{-1} \otimes I_n\right) \Gamma_m \left(\Omega_m^{-1} \otimes I_n\right) S_A'\right]^{-1/2} \left(\Omega_m^{-1} \otimes I_n\right) T^{-1/2} \sum\nolimits_{t=1}^{T} \mathrm{vec}\left(e_{t+1}^* \tilde{X}_{m,t}^{(h)\prime}\right).$$

Note that this is a sample average of mean zero random variables with the sum of variances equal to I_{nm}. Let $\mathcal{D}_{\mathcal{C}}$ be the class of convex sets on R^{nm} and $\mathcal{Z}$ a $nm \times 1$ Gaussian random vector with variance matrix I_{nm}. Then by Bulinskii and Shashkin (2004, Corollary 2) and Assumption A.1(v):

$$\begin{aligned}
&\sup_{D \in \mathcal{D}_{\mathcal{C}}} \left|\mathbb{P}\left(\mathfrak{R}_{11} \in D\right) - \mathbb{P}\left(\mathcal{Z} \in D\right)\right| \\
\leq\ & \left(n\left(m+1\right)\right)^{5/2} \left(\sum\nolimits_{t=1}^{T} \mathbb{E}\left\| \left[S_A \left(\Omega_m^{-1} \otimes I_n\right) \Gamma_m \left(\Omega_m^{-1} \otimes I_n\right) S_A'\right]^{-1/2} \left(\Omega_m^{-1} \otimes I_n\right) T^{-1/2} \xi_{m,t+1} \right\|^{2+\delta}\right)^{1/3} \\
=\ & Cm^{5/2} \left(\sum\nolimits_{t=1}^{T} \mathbb{E}\left\| \left[S_A \left(\Omega_m^{-1} \otimes \Sigma\right) S_A'\right]^{-1/2} \left(\Omega_m^{-1} \otimes I_n\right) T^{-1/2} \xi_{m,t+1} \right\|^{2+\delta}\right)^{1/3} \\
\leq\ & Cm^{5/2} \left(\sum\nolimits_{t=1}^{T} \mathbb{E}\left\| T^{-1/2} \xi_{m,t+1} \right\|^{2+\delta}\right)^{1/3} \\
\leq\ & CT^{-\delta/6} m^{5/2} \left(\left(\sqrt{m}\right)^{2+\delta} \mathbb{E}\left[\left\| e_{t+1}^* \right\|^{2+\delta}\right]\right)^{1/3} \\
\leq\ & CT^{-\delta/6} m^{5/2} \left(\left(\sqrt{m}\right)^{2+\delta} m^{2+\delta}\right)^{1/3}
\end{aligned}$$

where the last inequality follows by CC1 and Lemma 3(v). Then we have that

$$\sup_{D \in \mathcal{D}_{\mathcal{C}}} \left|\mathbb{P}\left(\mathfrak{R}_{11} \in D\right) - \mathbb{P}\left(\mathcal{Z} \in D\right)\right| \leq C\left(T^{-1} m^{(3+21/\delta)}\right)^{\delta/6},$$

which is $o\left(1\right)$ under the assumptions of Theorem 1. Thus if we define

$$\mathcal{T}_1^0 = T\mathrm{vec}\left(\hat{\mathbf{A}} - \mathbf{A}^*\right)' \left[\left(\Omega_m^{-1} \otimes I_n\right) \Gamma_m \left(\Omega_m^{-1} \otimes I_n\right)\right]^{-1} \mathrm{vec}\left(\hat{\mathbf{A}} - \mathbf{A}^*\right),$$

then we can show that $\left(2mn\right)^{-1/2}\left(\mathcal{T}_1^0 - mn\right) \to_d \mathcal{N}\left(0,1\right)$. To see this note first that we may define

the set $U_c = \{u \in \mathbb{R}^{mn} : u'u \leq c\}$. U_c is a convex set in $\mathbb{R}^{mn}$. Thus if $\mathcal{Z}_{mn} \sim \mathcal{N}(0, I_{mn})$ then

$$\begin{aligned} & \sup_c \left|\mathbb{P}\left(\mathcal{T}_1^0 \leq c\right) - \mathbb{P}\left(\mathcal{Z}'\mathcal{Z} \leq c\right)\right| \\ = & \sup_c \left|\mathbb{P}\left(\mathfrak{R}_{11} \in U_c\right) - \mathbb{P}\left(\mathcal{Z} \in U_c\right)\right| \\ \leq & \sup_{D \in \mathcal{D}_\mathcal{C}} \left|\mathbb{P}\left(\mathfrak{R}_{11} \in D\right) - \mathbb{P}\left(\mathcal{Z} \in D\right)\right|, \end{aligned}$$

which is $o(1)$ based on the previous result under our assumptions. Finally, let $\Phi(\cdot)$ be the CDF of the standard normal distribution. Then,

$$\begin{aligned} & \sup_c \left|\mathbb{P}\left((2mn)^{-1/2}\left(\mathcal{T}_1^0 - mn\right) \leq c\right) - \Phi(c)\right| \\ \leq & \sup_c \left|\mathbb{P}\left(\mathcal{T}_1^0 \leq mn + (2mn)^{1/2} c\right) - \mathbb{P}\left(\mathcal{Z}'\mathcal{Z} \leq mn + (2mn)^{1/2} c\right)\right| \\ & + \sup_c \left|\mathbb{P}\left(\mathcal{Z}'\mathcal{Z} \leq mn + (2mn)^{1/2} c\right) - \Phi(c)\right|. \end{aligned}$$

The first term is $o(1)$ by the previous result and the second term is

$$\sup_c \left|\mathbb{P}\left((2mn)^{-1/2}\left(\mathcal{Z}'\mathcal{Z} - mn\right) \leq c\right) - \Phi(c)\right|.$$

Let $\mathcal{Z}_i$ be the ith element of $\mathcal{Z}$ so that

$$(2mn)^{-1/2}\left(\mathcal{Z}'\mathcal{Z} - mn\right) = (2mn)^{-1/2} \sum\nolimits_{i=1}^{mn} \left(\mathcal{Z}_i^2 - 1\right).$$

Thus, by the Berry-Esseen inequality

$$\sup_c \left|\mathbb{P}\left((2mn)^{-1/2}\left(\mathcal{Z}'\mathcal{Z} - mn\right) \leq c\right) - \Phi(c)\right| \leq C (mn)^{-1/2},$$

and the result follows. It only remains to show that for

$$\hat{\mathcal{T}}_1^0 = T\text{vec}\left(\hat{\mathbf{A}} - \mathbf{A}^*\right)' \left[\left(\hat{\Omega}_{m,T}^{-1} \otimes I_n\right) \hat{\Gamma}_{m,T} \left(\hat{\Omega}_{m,T}^{-1} \otimes I_n\right)\right]^{-1} \text{vec}\left(\hat{\mathbf{A}} - \mathbf{A}^*\right), \tag{A.6}$$

we have that $m^{-1/2}\left|\hat{\mathcal{T}}_1^0 - \mathcal{T}_1^0\right| = o_p(1)$, where

$$\hat{\Gamma}_{m,T} = T^{-1} \sum\nolimits_{t=1}^{T} \text{vec}\left(\hat{e}_{t+1} \tilde{X}_{m,t}^{(h)\prime}\right) \text{vec}\left(\hat{e}_{t+1} \tilde{X}_{m,t}^{(h)\prime}\right)', \qquad \hat{e}_{t+1} = \text{Rx}_{t+1} - \hat{D}_{\text{ols}} \tilde{X}_{m,t}^{(h)}.$$

We have,

$$\begin{aligned}
& \sqrt{m}\left\|\left(\hat{\Omega}_{m,T}^{-1}\otimes \mathrm{I}_n\right)\hat{\Gamma}_{m,T}\left(\hat{\Omega}_{m,T}^{-1}\otimes \mathrm{I}_n\right)-\left(\Omega_m\otimes \mathrm{I}_n\right)\Gamma_m^{-1}\left(\Omega_m\otimes \mathrm{I}_n\right)\right\| \\
\leq\ & \sqrt{m}\left\|\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\Gamma_m\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\Gamma_m\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\left(\hat{\Gamma}_{m,T}-\Gamma_m\right)\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\left(\hat{\Gamma}_{m,T}-\Gamma_m\right)\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\Gamma_m\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\left(\hat{\Gamma}_{m,T}-\Gamma_m\right)\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\right\| \\
& +\sqrt{m}\left\|\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\left(\hat{\Gamma}_{m,T}-\Gamma_m\right)\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\right\| \\
=\ & \sum_{\ell=1}^{7}\left\|\mathfrak{N}_\ell\right\|.
\end{aligned}$$

So we need only show that these seven terms are $o_p(1)$. First note that,

$$\begin{aligned}
\left\|\mathfrak{N}_1\right\| &= \sqrt{m}\left\|\left(\left(\hat{\Omega}_{m,T}^{-1}-\Omega_m^{-1}\right)\otimes \mathrm{I}_n\right)\Gamma_m\left(\Omega_m^{-1}\otimes \mathrm{I}_n\right)\right\| \\
&= \sqrt{m}\lambda_{\max}\left(\Omega_m^{-1}\right)\lambda_{\max}\left(\Gamma_m\right)\left\|\Omega_m^{-1}\left(\hat{\Omega}_{m,T}-\Omega_m\right)\hat{\Omega}_{m,T}^{-1}\right\|\left\|\mathrm{I}_n\right\| \\
&\leq C\sqrt{m}\lambda_{\max}\left(\Omega_m^{-1}\right)^2\lambda_{\max}\left(\hat{\Omega}_{m,T}^{-1}\right)\lambda_{\max}\left(\Gamma_m\right)\left\|\hat{\Omega}_{m,T}-\Omega_m\right\| \\
&= O\left(1\right)O_p\left(1\right)O\left(1\right)o_p\left(1\right),
\end{aligned}$$

under our assumptions. $\mathfrak{N}_2-\mathfrak{N}_7$ follow by similar bounding arguments using Lemma 3.

Proof of Theorem 1(ii)**:** Note first that implicit in this hypothesis test is that $\operatorname{rank}\left(\mathbf{b}\phi_h\left(\cdot\right)\right)=1$ since the previous hypothesis has tested whether $\operatorname{rank}\left(\mathbf{b}\phi_h\left(\cdot\right)\right)=0$. Let $\theta=\left(\alpha,\mathbf{b}_0',\gamma'\right)'\in\Theta$ which is $d_\theta\times 1$ where $d_\theta=2n-1+m$. We solve for

$$\hat{\theta}=\arg\min_\theta\frac{1}{2}T^{-1}\sum_{t=1}^{T}e_{t+1}^{\mathrm{o}\prime}\hat{W}^{-1}e_{t+1}^{\mathrm{o}},\qquad e_t^{\mathrm{o}}\left(\theta\right)=\mathrm{R}\mathrm{x}_t-\alpha-b\gamma' X_{m,t-1}^{(h)}.$$

Define $\partial b/\partial b_0'=:\mathrm{J}_{n,n-1}$,

$$\mathfrak{s}_t\left(\theta\right)=\partial e_t^{\mathrm{o}}/\partial\theta'=-\left[\mathrm{I}_n\quad \gamma' X_{m,t-1}^{(h)}\mathrm{J}_{n,n-1}\quad bX_{m,t-1}^{(h)\prime}\right],$$

and

$$q_T\left(\vartheta\right)':=\left.\partial\mathcal{Q}_T\left(\theta;\hat{W}\right)\middle/\partial\theta'\right|_{\theta=\vartheta}=T^{-1}\sum_{t=1}^{T}e_{t+1}^{\mathrm{o}\prime}\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\vartheta\right),$$

for any $\vartheta \in \Theta$. Then,

$$q_T\left(\hat{\theta}\right) = 0 = q_T\left(\theta^*\right) + \left\{\dot{q}_T\left(\bar{\theta}\right)\right\}\left(\hat{\theta} - \theta^*\right),$$

where $\bar{\theta} \in \left[\hat{\theta}, \theta^*\right]$, $\theta^* = \left(\mathbf{a}^{*\prime}, \mathbf{b}_0^{*\prime}, \gamma^{*\prime}\right)'$ are the pseudo-true values and $\dot{q}_T\left(\theta\right) = \partial \mathcal{Q}_T\left(\theta; \hat{W}\right)/\partial\theta\partial\theta'$. To complete the result we need the following lemma:

Lemma 6. *Under our assumptions,*

$$\dot{q}_T\left(\theta\right) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathfrak{s}_{t+1}\left(\theta\right)'\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\theta\right) + \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\mathrm{I}_{d_\theta} \otimes e_{t+1}^{\mathrm{o}}\left(\theta\right)'\hat{W}^{-1}\right)\tilde{\mathfrak{s}}_{t+1}$$

where

$$\tilde{\mathfrak{s}}_t = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathrm{vec}\left(J_{n,n-1}\right) X_{m,t-1}^{(h)\prime} \\ 0 & \left(X_{m,t-1}^{(h)} \otimes \mathrm{I}_n\right) & 0 \end{bmatrix}.$$

Also,

$$\left\|\dot{q}_T\left(\bar{\theta}\right) - \dot{q}_T\left(\theta^*\right)\right\| = o_p\left(1\right),$$

and

$$\left\|\dot{q}_T\left(\theta^*\right) - H_T\left(\theta^*\right)\right\| = o_p\left(1\right), \qquad H_T\left(\theta\right) = T^{-1}\sum\nolimits_{t=1}^{T} \mathfrak{s}_{t+1}\left(\theta\right)' W^{-1}\mathfrak{s}_{t+1}\left(\theta\right).$$

By our mean-value expansion we have that,

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta} - \theta^*\right) &= -\left(\left[\dot{q}_T\left(\bar{\theta}\right)\right]^{-1} - \left[\dot{q}_T\left(\theta^*\right)\right]^{-1}\right)\sqrt{T}q_T\left(\theta^*\right) - \left(\left[\dot{q}_T\left(\theta^*\right)\right]^{-1} - H_T\left(\theta^*\right)^{-1}\right)\sqrt{T}q_T\left(\theta^*\right) \\
&\quad -H_T\left(\theta^*\right)^{-1}\sqrt{T}q_T\left(\theta^*\right) \\
&= \mathfrak{T}_{21} + \mathfrak{T}_{22} + \mathfrak{T}_{23}
\end{aligned}$$

The first and second terms are of lower order by Lemma 6. Thus we have that,

$$\sqrt{T}\left(\hat{\theta} - \theta^*\right) = H_T\left(\theta^*\right)^{-1} \times T^{-1/2}\sum\nolimits_{t=1}^{T} G^* \mathrm{vec}\left(e_{t+1}^* \tilde{X}_{m,t}^{(h)\prime}\right) + o_p\left(1\right), \qquad G^* = \begin{bmatrix} \left(s'_{(m+1),1} \otimes W^{-1}\right) \\ \left(\gamma^{*\prime} S_X \otimes J'_{n,n-1} W^{-1}\right) \\ \left(\mathbf{b}^{*\prime} W^{-1} \otimes S_X\right)\kappa_{n,m+1} \end{bmatrix}.$$

The leading term is a sample average of mean-zero random variables. Each summand has variance

$$\Psi_m := \mathbb{E}\left[G^* \mathrm{vec}\left(e_t^* \tilde{X}_{m,t-1}^{(h)\prime}\right) \mathrm{vec}\left(e_t^* \tilde{X}_{m,t-1}^{(h)\prime}\right)' G^{*\prime}\right] = G^* \Gamma_m G^{*\prime}.$$

Consider the rescaled statistic,

$$\mathfrak{R}_{21} = \mathsf{V}_\theta^{-1/2}\left\{\dot{q}_T\left(\theta^*\right)\right\}^{-1} T^{-1/2}\sum\nolimits_{t=1}^{T} G^* \mathrm{vec}\left(e_{t+1}^* \tilde{X}_{m,t}^{(h)\prime}\right),$$

where

$$\mathsf{V}_\theta = \left\{H_T\left(\theta^*\right)\right\}^{-1} \Psi_m \left\{H_T\left(\theta^*\right)\right\}^{-1}$$

This is now a sample average of mean-zero random variables with the sum of the variances equal to I_{d_θ}. Let $\mathcal{D}_\mathcal{C}$ be the class of convex sets in $\mathbb{R}^{d_\theta}$ and $\mathcal{Z}$ a Gaussian random vector with variance matrix I_{d_θ}. Then by Bulinskii and Shashkin (2004, Corollary 2) and Assumption A.1(v):

$$\begin{aligned}
&\sup_{D\in\mathcal{D}_\mathcal{C}} \left|\mathbb{P}\left(\mathfrak{R}_{21}\in D\right) - \mathbb{P}\left(\mathcal{Z}\in D\right)\right| \\
\leq\ & Cd_\theta^{5/2}\left(\sum\nolimits_{t=1}^{T}\mathbb{E}\left\|T^{-1/2}\mathsf{V}_\theta^{-1/2}\left[H_T\left(\theta^*\right)\right]^{-1}G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right\|^{2+\delta}\right)^{1/3} \\
\leq\ & Cm^{5/2}\left(\sum\nolimits_{t=1}^{T}\lambda_{\min}\left(\mathsf{V}_\theta^{-1/2}\right)\lambda_{\min}\left(H_T\left(\theta^*\right)\right)\mathbb{E}\left\|T^{-1/2}G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right\|^{2+\delta}\right)^{1/3} \\
\leq\ & Cm^{5/2}\left(T^{-\delta/2}T^{-1}\sum\nolimits_{t=1}^{T}\mathbb{E}\left\|G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right\|^{2+\delta}\right)^{1/3} \\
=\ & Cm^{5/2}\left(T^{-\delta/2}\mathbb{E}\left\|G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right\|^{2+\delta}\right)^{1/3}
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\left\|G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right\|^{2+\delta} &= \mathbb{E}\left[\left(\left\|W^{-1}e_{t+1}^*\right\|^2 + \left\|W^{-1}e_{t+1}^*X_{m,t}^{(h)\prime}\gamma^*\right\|^2 + \left\|X_{m,t}^{(h)}e_{t+1}^{*\prime}W^{-1}b^*\right\|^2\right)^{1+\delta/2}\right] \\
&\leq \mathbb{E}\left[\left(C_1\left\|e_{t+1}^*\right\|^2 + C_2\left(1+m\right)\left\|e_{t+1}^*X_{m,t}^{(h)\prime}\right\|^2\right)^{1+\delta/2}\right] \\
&\leq C\left(1+m+m^2\right)^{1+\delta/2}\mathbb{E}\left\|e_{t+1}^*\right\|^{2+\delta} \\
&\leq C\left(1+m+m^2\right)^{1+\delta/2}m^{2+\delta} \\
&\leq Cm^{2(2+\delta)}
\end{aligned}$$

by CC1 and Lemma 3(v). Thus, we have that

$$\sup_{D\in\mathcal{D}_\mathcal{C}}\left|\mathbb{P}\left(\mathfrak{R}_{21}\in D\right) - \mathbb{P}\left(\mathcal{Z}\in D\right)\right| \leq C\left(T^{-1}m^{(4+23/\delta)}\right)^{\delta/6},$$

which is $o\left(1\right)$ under the assumptions of the theorem. Next

$$s_{n,j}'\left(\hat{\mathbf{b}}\hat{\gamma}' - \mathbf{b}^*\gamma^*\right) = \left(\gamma^*\otimes s_{n,j}'\right)\left(\hat{\mathbf{b}}-\mathbf{b}^*\right) + \left(\mathrm{I}_m\otimes s_{n,j}'\mathbf{b}^*\right)\left(\hat{\gamma}-\gamma^*\right) + o_p\left(1\right) = S_{b\gamma,j}\left(\hat{\theta}-\theta^*\right),$$

where

$$S_{b\gamma,j} = \begin{bmatrix} 0 & \left(\gamma^*\otimes S_j'\right) & s_{n,j}'\mathbf{b}^*\cdot\mathrm{I}_m \end{bmatrix}.$$

Note next that the expression $S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}$ is rank one. To see this observe first that in the special case of homoskedasticity that,

$$\begin{aligned} S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j} &= \left(\left(S_X\hat{\Omega}_{m,T}S'_X\right)^{-1}\otimes\Sigma_{jj}\right) \\ &\quad -\left(\left[\left(S_X\hat{\Omega}_{m,T}S'_X\right)^{-1}-\gamma^*\left(\gamma^{*\prime}S_X\hat{\Omega}_{m,T}S'_X\gamma^*\right)^{-1}\gamma^{*\prime}\right]\otimes \right. \\ &\qquad \left.\left[\Sigma_{jj}-\left(b_j^*\right)^2\left(\mathbf{b}^{*\prime}W^{-1}\mathbf{b}^*\right)^{-1}\mathbf{b}^{*\prime}W^{-1}\Sigma W^{-1}\mathbf{b}^*\left(\mathbf{b}^{*\prime}W^{-1}\mathbf{b}^*\right)^{-1}\right]\right) \end{aligned}$$

Then, by Cline and Funderlic (1979, Corollary 3.3), under the null hypothesis, it follows that $S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}$ is rank one. For the heteroskedastic case, we can appeal to the sandwich form of the variance matrix and so it is also rank one in general under the null hypothesis. By the same steps we can also show that the plug-in variance estimator, $S_{\hat{b}\hat{\gamma},j}\hat{\mathsf{V}}_\theta S'_{\hat{b}\hat{\gamma},j}$, is rank one. Next note that,

$$\begin{aligned} \mathcal{T}_2^0 &= \left(\mathsf{V}_\theta^{-1/2}H_T\left(\theta^*\right)^{-1}T^{-1/2}\sum\nolimits_{t=1}^{T}G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right)'\mathsf{V}_\theta^{1/2} \\ &\quad \times S'_{b\gamma,j}\left(S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}\right)^+S_{b\gamma,j}\mathsf{V}_\theta^{1/2}\left(\mathsf{V}_\theta^{-1/2}H_T\left(\theta^*\right)^{-1}T^{-1/2}\sum\nolimits_{t=1}^{T}G^*\mathrm{vec}\left(e_{t+1}^*\tilde{X}_{m,t}^{(h)\prime}\right)\right). \end{aligned}$$

Define the set $U_c=\left\{u\in\mathbb{R}^{d_\theta}:u'\mathsf{V}_\theta^{1/2}S'_{b\gamma,j}\left(S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}\right)^+S_{b\gamma,j}\mathsf{V}_\theta^{1/2}u\leq c\right\}$, where V^+ is the Moore-Penrose inverse of a matrix V. Note that U_c is a convex set in $\mathbb{R}^{d_\theta}$. Then,

$$\begin{aligned} &\sup_c\left|\mathbb{P}\left(\mathcal{T}_2^0\leq c\right)-\mathbb{P}\left(\mathcal{Z}'\mathsf{V}_\theta^{1/2}S'_{b\gamma,j}\left(S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}\right)^+S_{b\gamma,j}\mathsf{V}_\theta^{1/2}\mathcal{Z}\leq c\right)\right| \\ =\ &\sup_c\left|\mathbb{P}\left(\mathfrak{R}_{21}\in U_c\right)-\mathbb{P}\left(\mathcal{Z}\in U_c\right)\right| \\ \leq\ &\sup_{D\in\mathcal{D}_c}\left|\mathbb{P}\left(\mathfrak{R}_{21}\in D\right)-\mathbb{P}\left(\mathcal{Z}\in D\right)\right|, \end{aligned}$$

which is $o\left(1\right)$ under our assumptions. Finally,

$$\begin{aligned} &\sup_c\left|\mathbb{P}\left(\sqrt{2}\left(\mathcal{T}_2^0-1\right)\leq c\right)-\Phi\left(c\right)\right| \\ =\ &\sup_c\left|\mathbb{P}\left(\mathcal{T}_2^0\leq 1+c\sqrt{2}\right)-\Phi\left(c\right)\right| \\ \leq\ &\sup_c\left|\mathbb{P}\left(\mathcal{T}_2^0\leq 1+c\sqrt{2}\right)-\mathbb{P}\left(\mathcal{Z}'\mathsf{V}_\theta^{1/2}S'_{b\gamma,j}\left(S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}\right)^+S_{b\gamma,j}\mathsf{V}_\theta^{1/2}\mathcal{Z}\leq 1+c\sqrt{2}\right)\right| \\ &+\sup_c\left|\mathbb{P}\left(\mathcal{Z}'\mathsf{V}_\theta^{1/2}S'_{b\gamma,j}\left(S_{b\gamma,j}\mathsf{V}_\theta S'_{b\gamma,j}\right)^+S_{b\gamma,j}\mathsf{V}_\theta^{1/2}\mathcal{Z}\leq 1+c\sqrt{2}\right)-\Phi\left(c\right)\right|. \end{aligned}$$

The first term goes to zero by the steps above and the second term satisfies, by the Berry-Esseen

inequality, and the results above,

$$\sup_c \left| \mathbb{P}\left(\mathcal{Z}' \mathsf{V}_\theta^{1/2} S'_{b\gamma,j} \left(S_{b\gamma,j} \mathsf{V}_\theta S'_{b\gamma,j} \right)^+ S_{b\gamma,j} \mathsf{V}_\theta^{1/2} \mathcal{Z} \leq 1 + c\sqrt{2} \right) - \Phi(c) \right| \leq C m^{-1/2}.$$

Thus, we have that $2^{-1/2}\left(\mathcal{T}_2^0 - 1\right) \to_d \mathcal{N}(0,1)$. Finally, we need only show that $\left|\hat{\mathcal{T}}_2^0 - \mathcal{T}_2^0\right| = o_p(1)$ where

$$\hat{\mathcal{T}}_2^0 = T\left(\hat{b}^j\hat{\gamma} - b^{*j}\gamma^*\right)' \left(S_{\hat{b}\hat{\gamma},j} \hat{\mathsf{V}}_\theta S'_{\hat{b}\hat{\gamma},j} \right)^+ \left(\hat{b}^j\hat{\gamma} - b^{*j}\gamma^*\right), \tag{A.7}$$

and

$$\hat{\mathsf{V}}_\theta = \left\{ H_T\left(\hat{\theta}\right) \right\}^{-1} \hat{\Psi}_{m,T} \left\{ H_T\left(\hat{\theta}\right) \right\}^{-1}.$$

By the previous results, Andrews (1987, Theorem 2), and since $S_{\hat{b}\hat{\gamma},j} \hat{\mathsf{V}}_\theta S'_{\hat{b}\hat{\gamma},j}$ is rank one under the null hypothesis then we need only show that $\left\| \hat{\mathsf{V}}_\theta - \mathsf{V}_\theta \right\| = o_p(1)$. In order to do so, note that we have already shown that $\left\| H_T\left(\hat{\theta}\right) - H_T\left(\theta^*\right) \right\| = o_p(1)$ in the proof of Lemma 6. Next to show $\hat{\Psi}_{m,T}$ is consistent recall that

$$\hat{\Psi}_m = \hat{G}\hat{\Gamma}_m\hat{G}', \qquad \hat{G} := \begin{bmatrix} \left(s'_{(m+1),1} \otimes \hat{W}^{-1} \right) \\ \left(\hat{\gamma}' S_X \otimes \hat{W}^{-1} \right) \\ \left(\hat{\mathbf{b}}' \hat{W}^{-1} \otimes \mathrm{I}_m \right) \kappa_{n,m+1} \end{bmatrix}.$$

Thus we need only show that $\left\| \hat{G} - G^* \right\| = o_p(1)$. Note that

$$\begin{aligned} \left\| \hat{G} - G^* \right\|^2 &= \left\| \left(s'_{(m+1),1} \otimes \left(\hat{W}^{-1} - W^{-1} \right) \right) \right\|^2 \\ &\quad + \left\| \left(\hat{\gamma}' S_X \otimes \hat{W}^{-1} \right) - \left(\gamma^{*\prime} S_X \otimes W^{-1} \right) \right\|^2 \\ &\quad + \left\| \left(\left(\hat{\mathbf{b}}' \hat{W}^{-1} - \mathbf{b}^{*\prime} W^{-1} \right) \otimes \mathrm{I}_m \right) \kappa_{n,m+1} \right\|^2 \\ &= \mathfrak{O}_1 + \mathfrak{O}_2 + \mathfrak{O}_3 \end{aligned}$$

$\mathfrak{O}_1$ is

$$\left\| \left(s'_{(m+1),1} \otimes \left(\hat{W}^{-1} - W^{-1} \right) \right) \right\|^2 \leq \lambda_{\max}\left(W^{-1}\right)^2 \lambda_{\max}\left(\hat{W}^{-1}\right)^2 \left\| \hat{W} - W \right\|^2,$$

and so is $o_p(1)$. $\mathfrak{O}_2$ is

$$\begin{aligned} & \left\| \left(\hat{\gamma}' S_X \otimes \hat{W}^{-1} \right) - \left(\gamma^{*\prime} S_X \otimes W^{-1} \right) \right\|^2 \\ \leq\ & C \left\| \left(\hat{\gamma} - \gamma^* \right)' S_X \right\|^2 \left\| \hat{D}_R^{-1} \right\|^2 + C \left\| \gamma^{*\prime} S_X \right\|^2 \left\| \hat{W}^{-1} - W^{-1} \right\|^2 + C \left\| \left(\hat{\gamma} - \gamma^* \right)' S_X \right\|^2 \left\| \hat{W}^{-1} - W^{-1} \right\|^2 \\ =\ & \mathfrak{O}_{21} + \mathfrak{O}_{22} + \mathfrak{O}_{23} \end{aligned}$$

Then $\mathfrak{O}_{21}$

$$\begin{aligned}\mathfrak{O}_{21} &= \left\|\left(\hat{\gamma}-\gamma^*\right)' S_X\right\|^2 \left\|\hat{W}^{-1}\right\|^2 \\ &\leq n\cdot\lambda_{\min}\left(\hat{W}\right)^2 \left\|\left(\hat{\gamma}-\gamma^*\right)' S_X\right\|^2 \\ &\leq n\cdot\lambda_{\min}\left(\hat{W}\right)^2 \left\|\hat{\gamma}-\gamma^*\right\|^2,\end{aligned}$$

and so is $o_p(1)$ and $\mathfrak{O}_{22}$ is

$$\begin{aligned}\mathfrak{O}_{22} &= \left\|\gamma^{*\prime} S_X\right\|^2 \left\|\hat{W}^{-1}-W^{-1}\right\|^2 \\ &\leq \lambda_{\min}\left(\hat{W}\right)\lambda_{\min}(W)\left\|\gamma^*\right\|^2 \left\|\hat{W}-W\right\|^2 \\ &\leq Cm\lambda_{\min}\left(\hat{W}\right)\lambda_{\min}(W)\left\|\hat{W}-W\right\|^2,\end{aligned}$$

and so is $O_p\left(m^{1/2}T^{-\zeta/2}\right)$ which is $o_p(1)$ under our assumptions. Next,

$$\begin{aligned}\mathfrak{O}_{23} &= \left\|\left(\hat{\gamma}-\gamma^*\right)' S_X\right\|^2 \left\|\hat{W}^{-1}-W^{-1}\right\|^2 \\ &\leq \left\|\left(\hat{\gamma}-\gamma^*\right)\right\|^2 \times \lambda_{\min}(W)\lambda_{\min}\left(\hat{W}\right)\left\|\hat{W}-W\right\|^2,\end{aligned}$$

and so is $o_p(1)$. Finally, by similar steps note that $\mathfrak{O}_3$

$$\begin{aligned}\mathfrak{O}_3^{1/2} &= \left\|\mathrm{I}_m\otimes\left(\hat{\mathbf{b}}'\hat{D}_R^{-1}-\mathbf{b}^{*\prime}D_R^{-1}\right)\right\| \\ &\leq \sqrt{m}\left\|\left(\hat{\mathbf{b}}-\mathbf{b}^*\right)'\hat{W}^{-1}\right\| + \sqrt{m}\left\|\mathbf{b}^{*\prime}\left(\hat{W}^{-1}-W^{-1}\right)\right\| + \sqrt{m}\left\|\left(\hat{\mathbf{b}}-\mathbf{b}^*\right)'\left(\hat{W}^{-1}-W^{-1}\right)\right\| \\ &= O_p\left(m^{3/2}T^{-1}\right)+O_p\left(m^{1/2}T^{-\zeta}\right),\end{aligned}$$

which is $o_p(1)$ under our assumptions. Thus $\left\|\hat{\Psi}_{m,T}-\Psi_m\right\| = o_p(1)$ and the result follows.

Proof of Lemma 6 The first result follows by standard properties of matrix differentiation. Next note that we have,

$$\begin{aligned}&\dot{q}_T\left(\bar{\theta}\right)-\dot{q}_T\left(\theta^*\right) \\ =\ & \frac{1}{T}\sum_{t=1}^T\left[\mathfrak{s}_{t+1}\left(\bar{\theta}\right)'\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\bar{\theta}\right)-\mathfrak{s}_{t+1}\left(\theta^*\right)'\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\theta^*\right)\right]+\frac{1}{T}\sum_{t=1}^T\left(\mathrm{I}_{d_\theta}\otimes\left(e_{t+1}^{\mathrm{o}}\left(\bar{\theta}\right)-e_{t+1}^*\right)'\hat{W}^{-1}\right)\tilde{\mathfrak{s}}_{t+1}.\end{aligned}$$

Consider the second term first. We have

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{I}_{d_\theta}\otimes\left(e_{t+1}^{\mathrm{o}}\left(\bar{\theta}\right)-e_{t+1}^{*}\right)'\hat{W}^{-1}\right)\tilde{\mathfrak{s}}_{t+1}\right\|^2\\
\leq\ & C\lambda_{\max}\left(\hat{W}^{-1}\right)^2\left\|\left(D^*-\bar{D}\right)\hat{\Omega}_{m,T}S_X'\right\|^2\\
\leq\ & C\lambda_{\max}\left(\hat{W}^{-1}\right)^2\lambda_{\max}\left(\hat{\Omega}_{m,T}\right)^2\left\|\hat{D}-D^*\right\|^2,
\end{aligned}$$

which is $o_p\left(1\right)$. Now consider the first term,

$$\mathfrak{D}:=\frac{1}{T}\sum_{t=1}^{T}\left[\mathfrak{s}_{t+1}\left(\bar{\theta}\right)'\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\bar{\theta}\right)-\mathfrak{s}_{t+1}\left(\theta^*\right)'\hat{W}^{-1}\mathfrak{s}_{t+1}\left(\theta^*\right)\right],$$

which may be written as a partitioned matrix with elements

$$\begin{aligned}
\mathfrak{D}_{11} &= 0\\
\mathfrak{D}_{12} &= \left(\bar{\gamma}-\gamma^*\right)'\left(T^{-1}\sum_{t=1}^{T}S_X\tilde{X}_{m,t}^{(h)}\right)\hat{W}^{-1}\mathrm{J}_{n,n-1}\\
\mathfrak{D}_{13} &= \hat{W}^{-1}\left(\bar{\mathbf{b}}-\mathbf{b}^*\right)\left(T^{-1}\sum_{t=1}^{T}\tilde{X}_{m,t}^{(h)\prime}S_X'\right)\\
\mathfrak{D}_{22} &= T^{-1}\sum_{t=1}^{T}\left(\bar{\gamma}'S_X\tilde{X}_{m,t}^{(h)}\right)^2\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}-T^{-1}\sum_{t=1}^{T}\left(\gamma^{*\prime}S_X\tilde{X}_{m,t}^{(h)}\right)^2\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}\\
\mathfrak{D}_{23} &= T^{-1}\sum_{t=1}^{T}\mathrm{J}_{n,n-1}'\hat{W}^{-1}\bar{\mathbf{b}}\tilde{X}_{m,t}^{(h)\prime}S_X'\left(\bar{\gamma}'S_X\tilde{X}_{m,t}^{(h)}\right)-T^{-1}\sum_{t=1}^{T}\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathbf{b}^*\tilde{X}_{m,t}^{(h)\prime}S_X'\left(\gamma^{*\prime}S_X\tilde{X}_{m,t}^{(h)}\right)\\
\mathfrak{D}_{33} &= T^{-1}\sum_{t=1}^{T}\left[\bar{\mathbf{b}}'\hat{W}^{-1}\bar{\mathbf{b}}-\mathbf{b}^{*\prime}\hat{W}^{-1}\mathbf{b}^*\right]S_X\tilde{X}_{m,t}^{(h)}X_{m,t}^{(h)\prime}S_X'.
\end{aligned}$$

Note that

$$\begin{aligned}
\left\|\mathfrak{D}_{12}\right\|^2 &= \left\|\left(\bar{\gamma}-\gamma^*\right)'\left(T^{-1}\sum_{t=1}^{T}S_X\tilde{X}_{m,t}^{(h)}\right)\hat{W}^{-1}\mathrm{J}_{n,n-1}\right\|^2\\
&\leq \left(n-1\right)\lambda_{\min}\left(\hat{W}\right)\mathrm{tr}\left(\left(\bar{\gamma}-\gamma^*\right)'S_X\hat{\Omega}_{m,T}s_{(m+1),1}s_{(m+1),1}'\hat{\Omega}_{m,T}S_X'\left(\bar{\gamma}-\gamma^*\right)'\right)\\
&= \left(n-1\right)\lambda_{\min}\left(\hat{W}\right)\lambda_{\max}\left(\hat{\Omega}_{m,T}\right)^2\left\|\bar{\gamma}-\gamma^*\right\|^2,
\end{aligned}$$

and so is $o_p\left(1\right)$. $\mathfrak{D}_{13}$ is $o_p\left(1\right)$ by similar steps. Next we have

$$\begin{aligned}
\left\|\mathfrak{D}_{22}\right\| &= \left\|T^{-1}\sum_{t=1}^{T}\left(\bar{\gamma}'S_X\tilde{X}_{m,t-1}^{(h)}\right)^2\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}-T^{-1}\sum_{t=1}^{T}\left(\gamma^{*\prime}S_X\tilde{X}_{m,t-1}^{(h)}\right)^2\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}\right\|\\
&\leq \left\|T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^*\right)'S_X\tilde{X}_{m,t-1}^{(h)}\right)^2\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}\right\|\\
&\quad+2\left\|T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^*\right)'S_X\tilde{X}_{m,t-1}^{(h)}\right)\gamma^{*\prime}S_X\tilde{X}_{m,t-1}^{(h)}\mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}\right\|.
\end{aligned}$$

The first term is

$$\left\| T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^{*}\right)'S_X\tilde{X}_{m,t-1}^{(h)}\right)^2 \mathrm{J}_{n,n-1}'\hat{W}^{-1}\mathrm{J}_{n,n-1}\right\| \leq \lambda_{\max}\left(\hat{\Omega}_{m,T}\right)\left\|\hat{\gamma}-\gamma^{*}\right\|^2\times(n-1)\,\lambda_{\min}\left(\hat{W}\right),$$

and so is $o_p(1)$. The second term is

$$\begin{aligned}&\left\| T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^{*}\right)'S_X\tilde{X}_{m,t-1}^{(h)}\right)\gamma^{*\prime}S_X\tilde{X}_{m,t-1}^{(h)}\hat{W}^{-1}\right\|^2\\ \leq\ & \lambda_{\max}\left(\hat{W}^{-1}\right)^2\left(T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^{*}\right)'S_X\tilde{X}_{m,t-1}^{(h)}\right)^2\right)\times T^{-1}\sum_{t=1}^{T}\left(\gamma^{*\prime}S_X\tilde{X}_{m,t-1}^{(h)}\right)^2.\end{aligned}$$

The second factor is

$$T^{-1}\sum_{t=1}^{T}\left(\left(\bar{\gamma}-\gamma^{*}\right)'S_X\tilde{X}_{m,t}^{(h)}\right)^2 \leq \lambda_{\max}\left(\hat{\Omega}_{m,T}\right)\left\|\bar{\gamma}-\gamma^{*}\right\|^2,$$

and the third factor has expectation

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\gamma^{*\prime}S_X\tilde{X}_{m,t}^{(h)}\right)^2=\mathbb{E}\left[\left(\gamma^{*\prime}X_{m,t}^{(h)}\right)^2\right]\leq C,$$

by Lemma 3(vi). Thus, $\|\mathfrak{D}_{22}\|=o_p(1)$. $\mathfrak{D}_{23}$ and $\mathfrak{D}_{33}$ follow by similar arguments.

Finally,

$$\begin{aligned}\dot{q}_T(\theta^{*}) \;=\;& \frac{1}{T}\sum_{t=1}^{T}\mathfrak{s}_{t+1}(\theta^{*})'W^{-1}\mathfrak{s}_{t+1}(\theta^{*})+\frac{1}{T}\sum_{t=1}^{T}\mathfrak{s}_{t+1}(\theta^{*})'\left(\hat{W}^{-1}-W^{-1}\right)\mathfrak{s}_{t+1}(\theta^{*})\\ &+\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{I}_{d_\theta}\otimes e_{t+1}^{*\prime}\hat{W}^{-1}\right)\tilde{\mathfrak{s}}_{t+1}.\end{aligned}$$

The second term is of lower order than the first term. The third term satisfies,

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{I}_{d_\theta}\otimes e_{t+1}^{*\prime}\hat{W}^{-1}\right)\tilde{\mathfrak{s}}_{t+1}\right\|^2 \leq C\left\|\hat{W}^{-1}T^{-1}\sum_{t=1}^{T}e_{t+1}^{*}X_{m,t-1}^{(h)\prime}\right\|^2 \leq C\lambda_{\min}\left(\hat{W}\right)^2\left\|T^{-1}\sum_{t=1}^{T}e_{t+1}^{*}X_{m,t}^{(h)\prime}\right\|^2.$$

By the steps in the proof of Lemma 3 (iii) we have that $\left\|T^{-1}\sum_{t=1}^{T}e_{t+1}^{*}X_{m,t}^{(h)\prime}\right\|^2=O_p\left(mT^{-1}\right)$ and $\lambda_{\min}\left(\hat{W}\right)=O_p(1)$ so this term is $o_p(1)$ by our assumptions and the result follows.

Proof of Theorem 1(iii)**:** This result follows directly from the results of Chen and Christensen (2015).

Proof of Theorem 2 Based on our results above we have that

$$\sqrt{T}\mathrm{vec}\left(\hat{\mathbf{A}}_{\mathrm{ols}}-\mathbf{A}^{*}\right)=S_A\left(\Omega_m^{-1}\otimes I_n\right)T^{-1/2}\sum_{t=1}^{T}\mathrm{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)+o_p(1),$$

$$\sqrt{T}\left(\hat{\theta}-\theta^{*}\right)=H_{T}\left(\theta^{*}\right)^{-1}\times T^{-1/2}\sum\nolimits_{t=1}^{T}G^{*}\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)+o_{p}\left(1\right),$$

and

$$\text{vec}\left(\hat{\mathbf{b}}\hat{\gamma}^{\prime}-\mathbf{b}^{*}\gamma^{*\prime}\right)=S_{b\gamma}\left(\hat{\theta}-\theta^{*}\right)+o_{p}\left(1\right),\qquad S_{b\gamma}=\left[0\quad\left(\gamma^{*}\otimes\mathrm{J}_{n,n-1}\right)\quad\left(\mathrm{I}_{m}\otimes\mathbf{b}^{*}\right)\right].$$

Then,

$$\begin{aligned}\text{vec}\left(\hat{\mathbf{A}}-\hat{\mathbf{b}}\hat{\gamma}^{\prime}\right) &= S_{A}\left(\Omega_{m}^{-1}\otimes\mathrm{I}_{n}\right)T^{-1}\sum\nolimits_{t=1}^{T}\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)-\left(\hat{\mathbf{b}}\hat{\gamma}^{\prime}-\mathbf{A}^{*}\right)+o_{p}\left(1\right)\\ &= S_{A}\left(\Omega_{m}^{-1}\otimes\mathrm{I}_{n}\right)T^{-1}\sum\nolimits_{t=1}^{T}\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)\\ &\quad -S_{b\gamma}H_{T}\left(\theta^{*}\right)^{-1}G^{*}T^{-1}\sum\nolimits_{t=1}^{T}\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)\\ &= \left(S_{A}\left(\Omega_{m}^{-1}\otimes\mathrm{I}_{n}\right)-S_{b\gamma}H_{T}\left(\theta^{*}\right)^{-1}G^{*}\right)T^{-1}\sum\nolimits_{t=1}^{T}\text{vec}\left(e_{t+1}^{*}\tilde{X}_{m,t}^{(h)\prime}\right)\\ &= \left(S_{A}\left(\Omega_{m}^{-1}\otimes\mathrm{I}_{n}\right)-S_{b\gamma}H_{T}\left(\theta^{*}\right)^{-1}G^{*}\right)\left(\Omega_{m}\otimes\mathrm{I}_{n}\right)\text{vec}\left(\hat{\mathbf{A}}-\mathbf{A}^{*}\right)+o_{p}\left(1\right)\end{aligned}$$

Let us, without loss of generality, change the normalization scheme so that the first element of γ^{*} is equal to one and no restrictions are placed on b^{*}. Thus, $\gamma^{*}=(1,\gamma_{0}^{*\prime})^{\prime}$. It can be shown that we can rewrite the above relationship as

$$\sqrt{T}\text{vec}\left(\hat{\mathbf{A}}-\hat{\mathbf{b}}\hat{\gamma}^{\prime}\right)=\left(\mathrm{I}_{mn}-\mathcal{H}_{m}\right)S_{A}\sqrt{T}\text{vec}\left(\hat{\mathbf{A}}-\mathbf{A}^{*}\right)+o_{p}\left(1\right),$$

$$\mathcal{H}_{m}:=\left(\Upsilon\otimes\mathrm{I}_{n}\right)+\left(\left(\mathrm{I}_{m}-\Upsilon\right)\mathrm{J}_{m,m-1}\tilde{\gamma}\otimes\mathbf{b}^{*}\left(\mathbf{b}^{*\prime}W^{-1}\mathbf{b}^{*}\right)^{-1}\mathbf{b}^{*\prime}W^{-1}\right),$$

where $\tilde{\gamma}:=\left[-\gamma_{0}^{*}\quad\mathrm{I}_{m-1}\right]$ and

$$\Upsilon:=\gamma^{*}\gamma^{*\prime}\left(S_{X}\Omega_{m}^{-1}S_{X}\right)^{-1}\left[\gamma^{*\prime}\left(S_{X}\Omega_{m}^{-1}S_{X}\right)^{-1}\gamma^{*}\right]^{-1}.$$

Note that the sample counterpart of $\left(S_{X}\Omega_{m}^{-1}S_{X}\right)^{-1}$ is $XM_{\iota}X$ where $M_{\iota}=\mathrm{I}_{T}-T^{-1}\cdot\iota_{T}\iota_{T}^{\prime}$. We can then form the test statistic as

$$\mathcal{T}_{4}^{0}=T\text{vec}\left(\hat{\mathbf{A}}-\hat{\mathbf{b}}\hat{\gamma}^{\prime}\right)^{\prime}\mathrm{V}_{4}^{+}\text{vec}\left(\hat{\mathbf{A}}-\hat{\mathbf{b}}\hat{\gamma}^{\prime}\right),$$

where

$$\mathrm{V}_{4}=\left(\mathrm{I}_{mn}-\mathcal{H}_{m}\right)S_{A}\left(\Omega_{m}^{-1}\otimes I_{n}\right)\Gamma_{m}\left(\Omega_{m}^{-1}\otimes I_{n}\right)^{\prime}S_{A}^{\prime}\left(\mathrm{I}_{mn}-\mathcal{H}_{m}\right)^{\prime}.$$

To show that V_{4} has rank s we can appeal to Cline and Funderlic (1979, Corollary 3.3) under the assumption of conditional homoskedasticity. Then the result follows immediately in the more general case because of the sandwich form of V_{4}. By similar steps as in the proof of Theorem 1 we

then have that $(2s)^{-1/2}\left(\mathcal{T}_4^0 - s\right) \to_d \mathcal{N}(0,1)$. A feasible test statistic is then formed using

$$\hat{\mathcal{T}}_4^0 = T\text{vec}\left(\hat{\mathbf{A}} - \hat{\mathbf{b}}\hat{\gamma}'\right)' \hat{\text{V}}_4^{+}\text{vec}\left(\hat{\mathbf{A}} - \hat{\mathbf{b}}\hat{\gamma}'\right),$$

where

$$\hat{\text{V}}_4 = \left(\text{I}_{mn} - \widehat{\mathcal{H}}_m\right) S_A \left(\hat{\Omega}_m^{-1} \otimes I_n\right) \hat{\Gamma}_m \left(\hat{\Omega}_m^{-1} \otimes I_n\right)' S_A' \left(\text{I}_{mn} - \widehat{\mathcal{H}}_m\right)'.$$

By our previous results we need only show that $m^{-1/2}\left|\hat{\mathcal{T}}_4^0 - \mathcal{T}_4^0\right| = o_p(1)$. This follows by the same steps as in the proof of Theorem 1.