The VIX, the Variance Premium, and Expected Returns

Osterrieder, Daniela; Ventosa-Santaulària, Daniel; Vera-Valdés, J Eduardo

Published in:
Journal of Financial Econometrics

DOI (link to publication from Publisher):
10.1093/jjfinec/nby008

Publication date:
2019

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
Osterrieder, D., Ventosa-Santaulària, D., & Vera-Valdés, J. E. (2019). The VIX, the Variance Premium, and Expected Returns. *Journal of Financial Econometrics*, *17*(4), 517-558. https://doi.org/10.1093/jjfinec/nby008

The VIX, the Variance Premium, and Expected Returns*

Daniela Osterrieder[‡§] Daniel Ventosa-Santaulària[¶] J. Eduardo Vera-Valdés[§]

Abstract: Existing studies find conflicting estimates of the risk-return relation. We show that the trade-off parameter is inconsistently estimated when VIX measures risk. The inconsistency arises from a misspecified, unbalanced, and endogenous return regression. These problems are eliminated if risk is captured by the variance premium instead. Yet, the variance premium is unobserved. Accordingly, we propose a GMM estimator that produces consistent estimates without observing the variance premium. Using this method, we find a positive risk-return trade-off and long-run return predictability. Our approach outperforms commonly used risk-return estimation methods, and reveals a significant link between the variance premium and economic uncertainty.

Keywords: Risk-return trade-off, variance premium, return prediction, persistent predictor, fractional integration, implied variance, integrated variance

JEL codes: G12, G13, C22, C26, C51

*The authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

‡ *Corresponding Author*: Rutgers Business School - Newark & New Brunswick, Department of Finance and Economics, Rutgers - The State University of New Jersey, 1 Washington Park, Newark, NJ 07102.
Email: dosterrieder@business.rutgers.edu. Phone: +1 973 353 2728. Fax: +1 973 353 1006.

§ CREATES and Aarhus University

¶ Centro de Investigación y Docencia Económicas (CIDE)

1 Introduction

The risk-return trade-off is a central concept in modern finance theory[1]. Yet, there is, at best, only weak empirical evidence on the positive risk premium implied by the risk-return trade off[2]. This lack of convincing evidence lead the popular press to question this central tenet of modern finance[3]. Our paper first shows a problem in the existing empirical studies that result in conflicting evidence, then provides an econometric solution to this problem, and finally presents empirical evidence of positive risk premium implied by the risk-return trade-off with the estimation problem resolved.

The Chicago Board Options Exchange (CBOE) volatility index, VIX, measures the implied market volatility of the U.S. stock market over the next month. Often implied volatility or variance is viewed as a measure for time-varying economic uncertainty or the aggregate risk level (see e.g. Bloom, 2009 and Bali and Peng, 2006). If risk or uncertainty were indeed captured by the implied variance, the mainstream risk-return trade-off theory would suggest that (at fixed levels of risk aversion) a higher level of implied variance corresponded to higher expected excess returns. Alternatively, implied market variance and in particular the VIX is commonly referred to as the "investor fear gauge" (Whaley, 2000), and may thus be viewed as a popular indicator of aggregate risk aversion (Bekaert et al., 2013). In this case the risk-return trade-off theory would again suggest a positive relation between implied variance and expected excess returns, at fixed levels of risk.

Recently, implied market variance and with it VIX have fallen on hard times, in the sense that empirical studies strongly challenge their role as variables that help gauge the risk-return trade-off. Given the lack of consensus, new theories have been brought forward suggesting that it is rather the variance premium (VP) that is positively related to the market risk premium. VP is the wedge between risk-neutral and objective expectations of future integrated market variance. Structural

[1]See e.g. Shim and Siegel (2008) for a textbook reference.

[2]For instance, Pastor, Sinha, and Swaminathan (JF 2008), Lundblad (JFE 2007), Ludvigson and Ng (JFE 2007), Guo and Whitelaw (JF 2006), Bali and Peng (JAE 2006), Ghysels, Santa-Clara, and Valkanov (JFE 2005), Goyal and Santa-Clara (JF 2003), Scruggs and Glabadanidis (JFQA 2003), Harrison and Zhang (REStat 1999), Scruggs (JF 1998), Chan, Karolyi, and Stulz (JFE 1992), and Harvey (JFE 1989) uncover a positive risk-return trade-off. In contrast, the latter trade-off is found to be negative in e.g. Brandt and Kang (JFE 2004), Whitelaw (JF 1994, RFS 2000), Glosten, Jagannathan, and Runkle (JF 1993), Nelson (ECTA 1991), Breen, Glosten, and Jagannathan (JF 1989), Turner, Startz, and Nelson (JFE 1989), and Campbell (JFE 1987). We discuss the conflicting existing evidence in more detail below.

[3]See e.g. the press article by John Authers in the Financial Times titled "Risk-return relationship has been upended" (2014), available at
`http://www.ft.com/cms/s/0/bc78d710-6371-11e4-9a79-00144feabdc0.html#axzz4DetL1gMt`.

models of Drechsler and Yaron (2011) and Bollerslev et al. (2012) show that VP is linked to economic uncertainty and that the latter commands a non-negligible equity risk premium. Conversely, the model of Bekaert and Engstrom (2010) shows that VP is an indicator of aggregate risk aversion and hence is positively related to the equity premium. Financial market data strongly support the suggested relation.

In this study, we contribute to the literature in three respects. Motivated by recent theories and empirical facts, we set up a stylized data generating process (DGP) in which the variance premium is the variable that truly drives conditionally expected future excess returns. Even though our empirical analysis reveals a stronger relation between VP and economic uncertainty compared to the link between VP and risk aversion, this is inconsequential for the development of the econometric results in the study. For the most part of the paper we will therefore refer to VP simply as *risk*. We show that empirical analyses where VIX is used to evaluate the risk-return trade-off result in a misspecified, unbalanced, and endogenous regression. The return regression is unbalanced since the right-hand side variable (VIX) exhibits autocorrelation patterns that are too strong to line up with the erratic almost noise-type behavior of returns on the left-hand side. The regression is endogenous since VIX is an imperfect predictor of the equity premium. In particular, VIX measures risk with a sizable and persistent error, which leads to the errors-in-variables problem. The new result that we provide in this paper demonstrates that even in this very "unfavorable" regression specification, the researcher can still use standard techniques of statistical inference to test for the significance of a risk-return trade-off. Intuitively speaking, this approach works since under the null hypothesis of no trade-off the problems of unbalancedness and endogeneity disappear. However, the ordinary least squares (OLS) estimator of the trade-off parameter is inconsistent, meaning that the VIX cannot be used to gauge the magnitude of the risk-return trade-off. The first part of our paper formalizes this intuition with the necessary mathematical proofs.

An obvious solution to avoid the errors-in-variables problem above would be to rely on VP as a predictor instead of VIX and to estimate the trade-off by OLS. Yet VP, that is the difference between the risk-neutral and the physical expectation of future integrated variance, is inherently latent. The first term, the implied market variance, is observable for the U.S. market by the squared VIX if the return variance is integrated over 30 days. The second term, the objective expectation of the integrated market variance, cannot be observed, however. Measuring it firstly requires an estimate for the integrated variance. Secondly, a probabilistic model for the dynamics of integrated variances of asset returns needs to be assumed and estimated. This is the crux of the literature on VP and risk-return modeling. To emphasize its importance Bekaert and Hoerova (2014) dedicate an entire research article to the issue, analyzing an abundance of "state-of-the-art" dynamic variance models. Thus, while many scholars agree on the importance of measuring VP to gauge risk, a consensus on modeling the objective expectation of the integrated market variance is largely absent from the literature[4]. The model uncertainty as well as the estimation error in the resulting estimate for the unobserved VP will directly affect the estimation of the risk-return trade-off parameter, likely biasing the results both in sample as well as out of sample. To avoid these consequences, our second contribution is to suggest a generalized method of moments (GMM) estimator that consistently estimates the relation between VP and the equity premium, without observing VP itself. To the best of our knowledge, we are the first to show that the risk-return trade-off parameter can be estimated without the necessity of observing, measuring, or estimating risk itself. The proposed GMM estimation approach allows for standard statistical inference on the parameters, and we further develop methods to establish the validity and the relevance of the instruments.

Our third contribution is empirical. Using data on the S&P 500 we demonstrate that there is ample empirical support for the DGP assumed here. Relying on the proposed GMM estimation technique, we then find evidence for a positive significant risk-return trade-off relation. To that end we identify two valid and relevant instruments that are closely related to the ex-post variance risk premium of Bollerslev et al. (2009) and the jump component of the stock price process. The uncovered risk-return trade-off is of sizable magnitude. We find that for a unit increase in risk

[4]Bali and Peng (2006) and Todorov (2010) rely on an ARMA specification as the econometric model for variance prediction; Bollerslev et al. (2013) rely on a co-fractional VAR model; Bollerslev et al. (2009), Du and Kapadia (2012), Bollerslev et al. (2014), Camponovo et al. (2012), Kelly and Jiang (2014), and Vilkov and Xiao (2013) use the realized variance over the past month as a proxy for the forward conditional expectation; Bollerslev et al. (2012) and Bollerslev et al. (2014) consider the HAR-RV model of Corsi (2009); Han and Zhou (2011), Bali and Zhou (2016), Bekaert et al. (2013), Drechsler and Yaron (2011), and Bekaert and Engstrom (2010) predict realized variances with the past month's realized variance and implied variance; Bekaert and Hoerova (2014) examine many different variance predictors, including lagged implied variances, lagged jumps, and lagged returns.

investors demand an increase in the equity premium of 2% annually. This number is seven times larger than the corresponding inconsistent OLS estimate. We continue to demonstrate that there is significantly positive excess return predictability in VP at different horizons, from one day to half a year. Even though VP remains latent throughout our study we confirm that return predictability in VP is maximized at a four-months investment horizon, which is in line with studies that estimate VP[5]. We show that our estimation technique leads to stronger predictability of excess returns over these horizons relative to models that estimate the objective expectation of the integrated market variance and hence VP, both in sample as well as out of sample. We argue that the main reason for this improvement is that the GMM approach avoids the estimation error in the estimate for VP that traditional OLS approaches produce. Lastly, we inspect the degree of correlation between the latent VP measure uncovered here and popular indicators for both, economic uncertainty and risk aversion. Our empirical results tend to favor the models that relate VP to economic uncertainty, in the sense that all empirical correlations are positive and of considerable magnitude. In contrast, the correlation results between VP and risk aversion do not unanimously point in one direction.

We view our paper in the context of empirical analyses of a risk-return trade-off on aggregate stock markets. There is a large stream of literature that relies on the VIX as a risk measure and then produces estimates of the trade-off parameter. These existing empirical studies of the trade-off have largely produced relationships of either signs and magnitudes, however. For instance, Bali and Peng (2006) relate the lagged VIX to the S&P 500 cash index and to CRSP value-weighted excess returns and find a positive trade-off parameter at a daily horizon. Similarly, Bollerslev and Zhou (2006) rely on the VIX to predict S&P 500 returns and find a positive risk-return trade-off parameter at a monthly horizon, Bollerslev et al. (2009) find the same result at monthly and quarterly horizons, Eraker and Wang (2015) have a positive estimate for horizons from half a year up to two years, and Bekaert and Hoerova (2014) discover the same at quarterly and annual horizons. In contrast, negative signs were estimated e.g. by Eraker and Wang (2015) and Bekaert and Hoerova (2014) at a monthly horizon, by Vilkov and Xiao (2013) at weekly, monthly, and annual horizons, and by Bollerslev et al. (2009) at an annual horizon. Our paper provides one possible explanation for this mixed evidence that the empirical literature has produced to date, by providing a formal argument

[5] See e.g. Bollerslev et al. (2009) and Bollerslev et al. (2014).

that the underlying return regressions may be problematic and can lead to inconsistent estimates.

Our paper also ties in with the literature on return predictability that is driven by VP. There are numerous empirical studies that find the positive relation between VP and the expected (excess) returns to hold. Among many others, the positive relation is found to hold in the data by Bollerslev et al. (2013), Han and Zhou (2011), Bollerslev et al. (2009), Drechsler and Yaron (2011), Du and Kapadia (2012), Eraker and Wang (2015), Bali and Zhou (2016), Camponovo et al. (2012), Kelly and Jiang (2014), Vilkov and Xiao (2013), Bollerslev et al. (2014), and Bekaert and Hoerova (2014). The common factor in all these studies is that the unobserved VP is replaced by an estimate or proxy. Our work differs from the standard methodology in the literature by keeping VP unobserved. We can still estimate the risk-return trade-off parameter by the GMM approach that we suggest. In the empirical analysis we then show that this method is preferable as it produces larger risk-return estimates and stronger predictability.

Finally, our paper is related to the literature on return predictions with persistent regressors, where the latter in our case is the VIX. Such predictive regressions are known to suffer from biased OLS slope estimates (see e.g. Stambaugh, 1986, 1999) and/or nonstandard statistical inference on the parameters (see e.g. Maynard and Phillips, 2001). To deal with the second problem, based on the work of for instance Campbell and Yogo (2006), Cavanagh et al. (1995), and Stock (1991) researchers have relied on confidence intervals computed using Bonferroni bounds. Predictability tests relying on this methodology are known to be conservative. Instead, in this paper we show that standard inference remains valid even in the presence of a persistent regressor with long-memory dynamics. To tackle the first problem of biased estimates, several econometric methods for bias correction have been proposed, for example by Kothari and Shanken (1997) and Lewellen (2004). If the regressor possesses long memory, another popular solution is to filter the series to remove the persistence, which has been advocated by e.g. Maynard et al. (2013) and Christensen and Nielsen (2007). The GMM approach that we suggest here is an innovative alternative to filtering that eliminates the persistence without requiring exact knowledge of the strength of serial dependence. Intuitively, our method works because the multiplication of the persistent regressor, VIX, with a less persistent instrument destroys the long memory in the series.

The plan for the rest of the article is as follows. Section 2 introduces the DGP and provides a description of the data underlying our empirical investigations, pointing out the support in the data for the assumed DGP. Section 3 discusses OLS risk-return trade-off regressions, where VIX is used as a predictor. Section 4 details the risk-return relation that can be uncovered by GMM estimation with a latent risk factor. In Section 5 we investigate the relevance and validity of our identified instruments. These results, in turn, motivate our analysis of the long-run relation between risk and return presented in Section 6. Section 6 also discusses the comparative advantages of our approach relative to the status quo of the literature. Section 7 empirically analyses the relation between latent VP and economic uncertainty and risk aversion. Section 8 concludes.

2 DGP and Initial Data Statistics

We propose a simple framework for the DGP of excess returns and risk that incorporates many well-known empirical properties of the data. We let the variance premium VP_t, $t = 1, 2, \ldots, T$, be an $I(0)$ process and assume that it is not observed by the researcher. The conditional expectation of the integrated market variance taken under the objective probability measure $\mathsf{E}_t^P(IV_{t,t+\tau})$, where the horizon τ equals 30 days, is also a latent process but strongly persistent, $I(d)$. Conversely the conditional expectation of the integrated market variance taken under the equivalent martingale measure $\mathsf{E}_t^Q(IV_{t,t+\tau}) = VIX_t^2$ is observable for the U.S. market. Since the latter is the sum of VP_t and $\mathsf{E}_t^P(IV_{t,t+\tau})$ this implies that the observed VIX_t^2 is $I(d)$. Excess returns on financial markets $r_t^{(e)}$ are generated as an $I(0)$ predictive function of VP_t with prediction coefficient β and level α, such that $\mathsf{E}\left(r_{t+1}^{(e)}|\mathcal{I}_t\right) = \alpha + \beta VP_t$. $\mathcal{I}_t$ is the information set of the informed investor at time t. Equations (1)-(4) detail the assumed DGP.

$$VP_t = \phi(L)\varepsilon_t \tag{1}$$

$$VIX_t^2 \equiv VP_t + \mathsf{E}_t^P(IV_{t,t+\tau}) \tag{2}$$

$$r_{t+1}^{(e)} = \alpha + \beta VP_t + \xi_{t+1} \tag{3}$$

$$\mathsf{E}_t^P(IV_{t,t+\tau}) = (1-L)^{-d}\eta_t, \tag{4}$$

where $0 \leq d < 1/2$. A vector consisting of noise processes ε_t, ξ_t, η_t, and additional shocks $\upsilon_{k,t}$, for $k = 1, 2, \ldots, K$, is vector independently and identically distributed (i.i.d.) with mean zero and a diagonal variance matrix with elements σ^2_ε, σ^2_ξ, σ^2_η, and $\sigma^2_{\upsilon_k}$. We assume that $\phi(L) = \sum_{i=0}^{\infty} \phi_i L^i$ with $\sum_{i=0}^{\infty} i|\phi_i| < \infty$ and $\phi(1) \neq 0$. The variance of VP_t is $\sigma^2_{VP} = \sigma^2_\varepsilon \sum_{i=0}^{\infty} \phi_i^2$. The variance of $\mathsf{E}_t^P(IV_{t,t+\tau})$ is $\sigma^2_P = \sigma^2_\eta \Gamma(1-2d)/(\Gamma(1-d))^2$.

The DGP (1)-(4) incorporates many stylized empirical facts as well as theoretical results. Expected excess returns are time-varying and are positively related to risk if $\beta > 0$, which is in line with empirical findings as well as new theoretical underpinnings (see references in Section 1). Consistent with empirical regularities of observed excess returns, $r_t^{(e)}$ is generated as a stationary $I(0)$ process that exhibits some short-memory dynamics but the impact of shocks decays quickly. In contrast, the conditional variance series VIX_t^2 and $\mathsf{E}_t^P(IV_{t,t+\tau})$ are strongly persistent, an often observed property of financial data. The volatility index and nonparametric realized variance measures empirically exhibit strong temporal dependence (see e.g. Bollerslev et al., 2012, and references therein), but also when estimating conditional variances with (G)ARCH-type models, the ARCH coefficient or the sum of the ARCH and the GARCH term are typically found to be close to one (for a summary, see e.g. Bollerslev et al., 1992). We describe the persistence in the variance series as stationary long memory, or $I(d)$ with $d \in [0, 1/2)$. It is well documented in the literature that fractionally integrated models with $d \in (0, 1/2)$ fit the dynamics of both, observed as well as model-implied conditional variances, very well (see, among others, Ding et al., 1993, Baillie et al., 1996, Andersen and Bollerslev, 1997, Comte and Renault, 1998, Bollerslev et al., 2013). Finally, the difference between our two variance series, VP_t, is $I(0)$. It possesses less memory than its two components, suggesting that the two variance series fractionally cointegrate as found by e.g. Christensen and Nielsen (2006) and Bandi and Perron (2006).

Our data support the proposed DGP. In particular, we focus on the S&P 500 stock market index. We consider daily data for the period from February 3, 2000 until June 30, 2014, resulting in a large number of $T = 3622$ observations that is particularly convenient for this study. We rely on the volatility index, $\text{VIX}_{\text{CBOE},t}$, to measure the expected integrated volatility over the next month under the equivalent martingale measure. We obtain the series $\text{VIX}_{\text{CBOE},t}$, which is quoted on the

CBOE, from the WRDS database. We transform the data series into maturity-scaled variance units by

$$VIX_t^2 = \frac{30}{365}\mathrm{VIX}^2_{\mathrm{CBOE},t}. \tag{5}$$

Now VIX_t^2 is in line with $\mathsf{E}_t^Q(IV_{t,t+\tau})$, where τ corresponds to 30 calendar days. We further obtain two variance measures that contain information about the integrated variance. Our first measure is the realized return variance, $\mathrm{RV}_{\mathrm{RL},t}$, computed on the basis of intradaily observations spaced into 5-minute intervals and subsampled at a 1-minute frequency. Under certain regularity conditions, $\mathrm{RV}_{\mathrm{RL},t}$ converges to the daily quadratic variation of returns, as shown by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002). If there are jumps in prices, the daily quadratic variation is the sum of $IV_{t,t+1}$ and daily jumps. Our second measure is the bipower variation, $\mathrm{BV}_{\mathrm{RL},t}$, of Barndorff-Nielsen and Shephard (2004), which converges to the integrated variance of returns $IV_{t,t+1}$. The series $\mathrm{RV}_{\mathrm{RL},t}$, and $\mathrm{BV}_{\mathrm{RL},t}$, as well as daily prices on the S&P 500, $P_t^{(\mathrm{open})}$ and $P_t^{(\mathrm{close})}$, are obtained from the Oxford-Man Institute's "Realised Library"[6].

Whereas VIX_t^2 is related to the return variation over the next month, the raw series $\mathrm{RV}_{\mathrm{RL},t}$ and $\mathrm{BV}_{\mathrm{RL},t}$ measure daily variation. To align the three measures, we modify the latter two as follows.

$$RV_t = \sum_{i=1}^{22}\left(\mathrm{RV}_{\mathrm{RL},t+i}\times 100^2 + \left\{\left[\ln\frac{P_{t+i+1}^{(\mathrm{open})}}{P_{t+i}^{(\mathrm{close})}}\right]\times 100\right\}^2\right) \tag{6}$$

$$BV_t = \sum_{i=1}^{22}\left(\mathrm{BV}_{\mathrm{RL},t+i}\times 100^2 + \left\{\left[\ln\frac{P_{t+i+1}^{(\mathrm{open})}}{P_{t+i}^{(\mathrm{close})}}\right]\times 100\right\}^2\right). \tag{7}$$

The two series thus contain information about the unobserved $IV_{t,t+\tau}$. Finally, we measure $r_{t+1}^{(e)}$ as daily annualized continuously compounded excess returns (measured in percentages)

$$r_{t+1}^{(e)} = 100\times\ln\left[\left(\frac{P_{t+1}^{(\mathrm{close})}}{P_t^{(\mathrm{close})}}\right)^{252}\right] - r_t^{(f)}. \tag{8}$$

We obtain the daily 3-month T-Bill rate from the FRED database[7] and convert it into annualized continuously compounded rates $r_t^{(f)}$.

[6]Available at `http://realized.oxford-man.ox.ac.uk/`.

[7]Available at `https://research.stlouisfed.org/fred2/series/DTB3`

Our DGP (1)-(4) implies that $r_t^{(e)}$ is a short-memory $I(0)$ process whereas VIX_t^2 is a long-memory $I(d)$ process. Table 1 shows summary statistics for our data. For annualized daily excess returns we find that the autocorrelation estimates are very close to zero, suggesting that there is very little persistence in the series. Conversely, for VIX_t^2 we find that the first three autocorrelation estimates are very close to 1. Even after 22 trading days the serial correlation is still very strong, such that roughly 75% of a shock's impact remains. If physical expectations are taken under the rational information set, it follows that the temporal dependence of the realized $IV_{t,t+\tau}$ proxies in (6)-(7) give us an indication of the unknown dynamics of $\mathsf{E}_t^P(IV_{t,t+\tau})$. Autocorrelation estimates for RV_t and BV_t in Table 1 are very similar to the estimates for VIX_t^2, suggesting that the variance series share similar persistent dynamics. For further evidence of the apparently distinct dynamics of the three variance series from stock returns see also Figure 1, where we plot the autocorrelations of the four processes. Whereas shocks to daily returns die out immediately, shocks to RV_t, BV_t, and VIX_t^2 are highly persistent. As opposed to the $I(0)$ excess return process, it takes many lags to revert the effect of a shock to the variance.

We estimate the respective fractional integration order, d_i, $i = \{RV, BV, VIX, r\}$, of the four series, RV_t, BV_t, VIX_t^2, and $r_t^{(e)}$, jointly for efficiency. It is common to rely on semiparametric techniques for the estimation of d_i. The exact local Whittle (EW) due to Shimotsu and Phillips (2005) is particularly attractive, since it is consistent and asymptotically normally distributed for any value of d_i. Nielsen and Shimotsu (2007) derive a multivariate version of the EW, which we apply for the joint estimation of d_{RV}, d_{BV}, d_{VIX}, and d_r[8].

Table 2 summarizes our results. The realized variance and the bipower variation are integrated of the order $I(0.32)$. At a 5% significance level, we reject that $d_i = 0$ and $d_i = 1$ for both series, yet we fail to reject that $d_i = 0.5$. The point estimate for the memory of the variance index, VIX_t^2, is somewhat higher, $\hat{d}_{VIX} = 0.40$. According to the t-test of Nielsen and Shimotsu (2007) for the equality of d_i, we cannot reject that the three variance series are integrated of the same order, however. Excess returns are integrated of the approximate order zero, and we fail to reject $d_i = 0$, but reject $d_i = 0.5$ and $d_i = 1$.

[8]The consistency and asymptotic properties of the EW estimator rely on the knowledge of the true mean of the data generating process. As this value is not known in practical applications, we modify the EW to account for this uncertainty, relying on the two-step feasible EW estimator of Shimotsu (2010).

One shortcoming of the approach above is that the EW is not explicitly robust to the presence of additive perturbations, which are present in three variance processes, RV_t, BV_t, and VIX_t^2, under the assumed DGP. That is

$$\begin{aligned}
VIX_t^2 &= \mathsf{E}_t^P(IV_{t,t+\tau}) &&+ VP_t \\
RV_t &\to \mathsf{E}_t^P(IV_{t,t+\tau}) &&+ \left(\text{jumps}_{t,t+\tau} + \text{expectations error}^{(R)}\right) \\
BV_t &\to \underbrace{\mathsf{E}_t^P(IV_{t,t+\tau})}_{I(d)\text{ Process}} &&+ \underbrace{\text{expectations error}^{(B)}}_{\text{Additive Perturbations}}.
\end{aligned}$$

In addition, with the EW estimation we did not restrict the integration orders of RV_t, BV_t, and VIX_t^2 to be the same, which they must be if the perturbations are integrated of an order $< d$. We adopt the trivariate version of the modified EW estimator of Sun and Phillips (2004) (TEW) for the vector $X_t \equiv [RV_t,\ BV_t,\ VIX_t^2]'$. Implementation and estimation details are in Appendix A. We find that $\hat{d}_P$ =0.39. The exact asymptotic properties of the TEW are unknown, yet Sun and Phillips (2004) conjecture that the distribution of $\hat{d}$ is normal and that standard errors are bound between [0.12,0.16]. The estimated fractional order of $\mathsf{E}_t^P(IV_{t,t+\tau})$ is different from zero and statistically indistinguishable from the non-robust estimates in Table 2. Our data thus lend support to the proposed DGP.

3 Estimating the Risk-Return Trade-Off by OLS

In the framework (1)-(4) the coefficient β has precisely the interpretation of a risk-return trade-off coefficient. Given the aforementioned discrepancies in both the sign and the magnitude of the corresponding estimate, it will be our main interest to consistently estimate β. The coefficient also captures short-run predictability of excess returns if it exists. A consistent estimator therefore will allow us to evaluate whether the variance premium indeed commands a positive equity premium and whether returns are predictable. Further, our aim is to be able to conduct valid statistical inference on the parameters of (3).

Evaluating the relation between risk and return, the correct specification to estimate would be to regress $r_{t+1}^{(e)}$ on VP_t. Yet, VP_t is not observed by the researcher, but VIX_t^2 is not latent. It is common to assume that the econometrician's information set $\mathcal{A}_t$ satisfies $\mathcal{A}_t \subseteq \mathcal{I}_t$ (see e.g. Nagel, 2013). Whereas the fully informed investor may know VP_t, we assume that the researcher only observes VIX_t^2. The latter will be inclined to evaluate the following regression

$$r_{t+1}^{(e)} = a + b\,VIX_t^2 + e_{t+1}, \tag{9}$$

which is unbalanced since the integration orders of the regressor and the regressand differ (Banerjee et al., 1993). The empirical work on the risk-return trade-off, as well as most of the existing theoretical contribution on the econometric properties of predictive regressions in general, impose the assumption that the true predictor VP_t and the observable predictor VIX_t^2 are the same or perfectly correlated. A very different idea is considered by Ferson et al. (2003) and Deng (2014). They demonstrate the risk of spurious inference in predictive regressions, where the expected (demeaned) return is assumed to be independent of the predictor. In our case that would make $\beta\,VP_t$ independent of VIX_t^2. Note that both setups can be viewed as extremes of our DGP, where the first scenario arises if $\sigma_\eta^2 = 0$, and the second scenario occurs if $\beta = 0$ and/or $\sigma_\varepsilon^2 = 0$. Instead of imposing these extreme setups, we consider the predictor in our model to be *imperfect*. Similarly to Pastor and Stambaugh (2009) and Binsbergen and Koijen (2010) we assume that the observed variable VIX_t^2 contains relevant information about the expected return, but it is imperfectly correlated with the latter, which strictly speaking leaves regression (9) misspecified. Besides being misspecified and unbalanced, the econometrician's model (9) is endogenous. The regression residuals of (9) are composed of two elements, that is $e_{t+1} = -\beta \mathsf{E}_t^P(IV_{t,t+\tau}) + \xi_{t+1}$. Thus, e_{t+1} will be naturally correlated with the observed variance measure VIX_t^2 with $\mathsf{Cov}(e_{t+1}, VIX_t^2) = -\beta\sigma_P^2$.

The results from the empirical literature on the risk-return trade-off using the observable risk measure VIX_t^2 are largely inconclusive to date (see reference in Section 1). Analyses in the field typically evaluate a predictive regression such as (9) by OLS. Theorem 1 below aids our understanding of the likely causes of finding a risk-return trade-off of either sign and magnitude. Define two matrices $\mathbf{X}$ and $\mathbf{y}$ of size $(T-1)\times 2$ and $(T-1)\times 1$ respectively by

$$\mathbf{X} \equiv \begin{pmatrix} 1 & 1 & \dots & 1 \\ VIX_1^2 & VIX_2^2 & \dots & VIX_{T-1}^2 \end{pmatrix}' \tag{10}$$

$$\mathbf{y} \equiv \begin{pmatrix} r_2^{(e)} & r_3^{(e)} & \dots & r_T^{(e)} \end{pmatrix}'. \tag{11}$$

Theorem 1 summarizes our results for both hypotheses, the presence and absence of return predictability from VP_t. A proof of Theorem 1 can be found in Appendix C. Small sample simulations supporting all of our results can be found in the online appendix to this paper.

Theorem 1. *Let VP_t, VIX_t^2, $r_t^{(e)}$, and $\mathsf{E}_t^P(IV_{t,t+\tau})$ be generated by (1)-(4). Estimate Regression (9) by OLS, resulting in*

$$\hat{\mathfrak{b}}_{OLS} \equiv \left(\hat{a},\ \hat{b}\right)' = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right). \tag{12}$$

Let $\xrightarrow{P}$ denote convergence in probability, and $\xrightarrow{D}$ convergence in distribution. As $T \to \infty$:

1. *If $\beta = 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad\qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_\xi^2}{\sigma_{VP}^2+\sigma_P^2}\right)$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad\qquad t_b \xrightarrow{D} \mathcal{N}(0,1).$$

   $t_a = \hat{a}/\sqrt{\mathsf{Var}(\hat{a})}$ and $t_b = \hat{b}/\sqrt{\mathsf{Var}(\hat{b})}$ denote the t-statistics associated with $\hat{a}$ and $\hat{b}$, respectively, and $\mathcal{N}(\cdot,\cdot)$ is the normal distribution. In addition, it holds that $s^2 \xrightarrow{P} \sigma_\xi^2$, where $s^2 = (T-3)^{-1}\sum_{t=2}^T \hat{e}_t^2$ is the variance of the OLS residuals.

2. *If $\beta \neq 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad\qquad \hat{b} \xrightarrow{P} \beta\frac{\sigma_{VP}^2}{\sigma_{VP}^2+\sigma_P^2}$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma_\xi^2 + \beta^2\frac{\sigma_{VP}^2\sigma_P^2}{\sigma_{VP}^2+\sigma_P^2}\right)^{1/2}} \qquad\qquad T^{-1/2}t_b \xrightarrow{P} \frac{\beta\sigma_{VP}^2}{\left(\beta^2\sigma_{VP}^2\sigma_P^2 + \sigma_\xi^2(\sigma_{VP}^2+\sigma_P^2)\right)^{1/2}},$$

   where $s^2 \xrightarrow{P} \sigma_\xi^2 + \beta^2\frac{\sigma_{VP}^2\sigma_P^2}{\sigma_{VP}^2+\sigma_P^2}$.

The first part of Theorem 1 summarizes the case in which there is no risk-return trade-off, i.e. $\beta = 0$. In this situation, the OLS slope estimate $\hat{b}$ correctly converges to zero and to a normal distribution at the usual rate $T^{-1/2}$. More importantly, under the premise that there is a risk premium in the market, the second part of Theorem 1 shows that OLS produces an inconsistent estimate for β. In finite sample simulations the estimate $\hat{b}$ is of either sign and value, which is in line with the findings in empirical studies that use VIX_t^2 as a predictor. Asymptotically, the OLS slope estimate $\hat{b}$ is biased towards zero, implying that in large samples the researcher would underestimate the implied predictive power from VP_t on $r_{t+1}^{(e)}$. We view this inconsistency of the OLS estimator for β as one possible explanation for the widely different estimates that existing finite-sample studies of (9) have produce to date.

Given the unbalancedness and endogeneity issues in (9) it may not be too surprising to the reader that the OLS estimator for β is inconsistent. The asymptotic bias of the estimator towards zero is in line with e.g. Maynard and Phillips (2001). What is truly new and largely different from the extant literature on return predictions with persistent regressors is the finding that standard statistical inference can be carried out. In particular, Theorem 1 shows that the t-statistic associated with $\hat{b}$ converges asymptotically to a standard normal limiting distribution that is free of nuisance parameters under the null hypothesis that $\beta = 0$. In small sample simulations we find that the size of a simple t-test on the parameter β is always very close to the nominal size of 5%. Under the alternative hypothesis, the t-statistic t_b diverges asymptotically at rate $T^{1/2}$. Simulations suggest that a t-test generally has very good power in finite samples. The implication of these results is that one can draw valid statistical inference on the significance of β. Thus, even in the unbalanced, misspecified, and endogenous regression framework considered here, the t-statistic can be considered a useful tool to draw inference on the significance of the predictability of $r_{t+1}^{(e)}$ from the latent VP_t.

The empirical evidence in our data indeed lends support to the unbalancedness of Regression (9). The t-tests for $H_0 : d_i = d_j$ in Table 2 indicate that we reject the hypothesis that variance series and returns are integrated of the same order. Nevertheless, it is common to predicting tomorrow's excess returns with today's VIX_t^2 by OLS. Table 3 outlines the results from estimating

(9) by OLS with our dataset[9]. The estimated risk-return trade-off parameter is small and positive, equal to 0.27. The estimate is statistically different from zero. Since we know from Theorem 1 that valid inference can be carried out, we conclude that the latent variance premium VP_t, which is one of the components of the observed VIX_t^2, significantly predicts returns. The estimated coefficient is rather small, however, and we deduce from Theorem 1 that the estimate is inconsistent and asymptotically biased towards zero. The researcher could be tempted to make the erroneous conclusion that the estimate $\hat{b} = 0.27$ implies that an increase in yesterday's perceived risk by one standard deviation leads to an increase of tomorrow's annualized excess return expectations of 12.68%. To put these numbers into perspective, an increase in the VIX_t^2 of one standard deviation equal to 47.36 corresponds to a very large increase, for instance more than doubling the average of what the econometrician may perceive as risk, or leaping from the median VIX_t^2 to the 88th quantile. The resulting (inconsistently) estimated effect on returns would then seem rather moderate, corresponding only to an increase from the 50th to the 53rd quantile of the empirical return distribution.

4 Estimating the Risk-Return Trade-Off by GMM

The risk-return trade-off parameter cannot be estimated by an OLS regression of (9). A possible solution could be to make VP_t observable, i.e. replacing it by an estimate $\hat{VP}_t$, as is commonly done in the literature[4]. Yet, the model uncertainty and estimation error in $\hat{VP}_t$ would directly impact the OLS estimator $\hat{b}$, implying that the estimation of the risk-return trade-off with this approach would be prone to error. Instead, we suggest to resolve the problems of the OLS regression by relying on a GMM approach. Assume that the researcher has access to a valid and relevant $I(0)$ instrument, i.e. a variable that is strongly correlated with VP_t but not with the variance $\mathsf{E}_t^P(IV_{t,t+\tau})$ and the innovation ξ_{t+1}[10]. Theorem 2 summarizes the asymptotic properties of the GMM estimates of (9).

Theorem 2. *Let VP_t, VIX_t^2, $r_t^{(e)}$, and $\mathsf{E}_t^P(IV_{t,t+\tau})$ be generated by* (1)-(4)*. Assume there exist*

[9]Our DGP assumes that VP_t, $\mathsf{E}_t^P(IV_{t,t+\tau})$, and hence also VIX_t^2 have a mean of zero. Henceforth we therefore consider all variables, except the excess return series, in deviation of their sample averages.

[10]An instrument that is neither correlated with $\mathsf{E}_t^P(IV_{t,t+\tau})$ nor with ξ_{t+1} will by definition also be unrelated to the error term of the unbalanced regression (9), e_{t+1}.

K instruments

$$q_{k,t} = \rho_k VP_t + \upsilon_{k,t}, \quad k = 1, 2, \ldots, K, \tag{13}$$

where $\rho_k \neq 0 \ \forall k$. Define

$$\mathbf{Q} \equiv \begin{pmatrix} 1 & 1 & \ldots & 1 \\ q_{1,1} & q_{1,2} & \cdots & q_{1,T-1} \\ \vdots & \ddots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,T-1} \end{pmatrix}'. \tag{14}$$

Estimate Regression (9) *by GMM using $q_{k,t}$ as instruments. The GMM estimate is given by*

$$\hat{\mathfrak{b}}_{GMM} \equiv \left(\hat{a},\ \hat{b}\right)' = \left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{y}\right). \tag{15}$$

Then, as $T \to \infty$:

1. *If $\beta = 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_\xi^2\left(\sigma_{VP}^2\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2} + 1\right)}{\sigma_{VP}^4\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}\right)$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad t_b \xrightarrow{D} \mathcal{N}(0,1),$$

where $s^2 \xrightarrow{P} \sigma_\xi^2$.

2. *If $\beta \neq 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad \hat{b} \xrightarrow{P} \beta$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma_\xi^2 + \beta^2\sigma_P^2\right)^{1/2}} \qquad T^{-1/2}t_b \xrightarrow{P} \beta\left(\frac{\sigma_{VP}^4\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{\left(\sigma_\xi^2 + \beta^2\sigma_P^2\right)\left(\sigma_{VP}^2\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2} + 1\right)}\right)^{1/2},$$

where $s^2 \xrightarrow{P} \sigma_\xi^2 + \beta^2\sigma_P^2$.

Theorem 2 shows that in the absence of a risk-return trade-off, the GMM estimate $\hat{b}$ converges to a normal distribution with zero mean at the standard rate $T^{-1/2}$. More importantly, Theorem

2 demonstrates that GMM estimation results in a consistent estimator for β. In finite sample simulations the average relative bias, $\hat{b}/\beta$, is very small, bound between 1 and 1.05 across the set of chosen parameter values. Intuitively, the GMM approach to estimation works since firstly the use of a relevant but exogenous instrument resolves the endogeneity issue. Secondly, in computing the GMM method we multiply the $I(d)$ regressor innovation with an $I(0)$ instrument. Lemma 1 in Appendix B shows that this multiplication destroys the long memory and the resulting process has standard short-memory dynamics. Hence, under the maintained assumption that the DGP follows (1)-(4), the predictive power of the latent variable VP_t on $r_t^{(e)}$ can be correctly estimated if the researcher finds a relevant and valid $I(0)$ instrument as in (13). The proof of Theorem 2 can be found in Appendix D.

Theorem 2 further implies that the statistical significance of β can be correctly inferred from a simple t-test. Under the null hypothesis that $H_0 : \beta = 0$ the t-statistic of the GMM estimate $\hat{b}$, t_b, converges to a standard normal distribution. Simulations under the null hypothesis that $H_0 : \beta = 0$ show that the size of the test is close to the nominal level of 5%, albeit marginally undersized for very small T. The statistic t_b diverges at rate $T^{1/2}$ under $H_1 : \beta \neq 0$. The finite sample power of the t-test is very close to 100% across the scenarios that we consider in the simulations. The researcher will thus be very likely to detect predictability and hence the risk-return trade-off if it is present.

Inspired by the results in Theorem 2, we identify a set of $I(0)$ instruments for GMM estimation in our data. To that end we partly rely on the realized measures RV_t and BV_t. To avoid problems that could arise in the successive estimation due to a look-ahead bias, we shift the two variance series backwards such that they capture the quadratic variance and the integrated variance over the past month, respectively. Denote the shifted series by $\tilde{RV}_t$ and $\tilde{BV}_t$.

The existing literature provides substantial evidence that there is a linear long-run relation between $\tilde{RV}_t$ and VIX_t^2 that is $I(0)$. For instance, Bandi and Perron (2006) and Christensen and Nielsen (2006) find evidence of fractional cointegration between the two series. Furthermore, if the cointegrating vector is equal to $[-1,\ 1]'$, then the resulting cointegrating series corresponds to the monthly ex-post realized variance risk premium, VRP_t, as defined by Bollerslev et al. (2009)[11]. The

[11]Note that VRP_t is different from the true latent variance premium VP_t in (1)-(4), unless $d = 1$ and $RV_t =$

latter argue that VRP_t may be viewed as bet on pure volatility; as such it is reasonable to expect that the measure is closely linked to risk VP_t. Bollerslev et al. (2009) and Bollerslev et al. (2013) also present evidence that VRP_t can predict aggregate market returns, which is further motivation for considering the measure to be a relevant instrument in our framework.

Besides the cointegrating relation between $\tilde{RV}_t$ and VIX_t^2, we expect there to be a long-run relation between $\tilde{RV}_t$ and $\tilde{BV}_t$, as both series capture the monthly integrated variance of stock returns over the past month. Following the arguments in Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007), and Huang and Tauchen (2005), the cointegrating relation between $\tilde{RV}_t$ and $\tilde{BV}_t$ represents the contribution of price jumps to the variance, if the cointegrating vector is equal to $[1,\ -1]'$. For instance, Andersen et al. (2007) find that the jump component exhibits a much lower degree of persistence than the two series $\tilde{RV}_t$ and $\tilde{BV}_t$, providing evidence for a fractional cointegration relation. Jumps are closely related to VP_t; for instance Bollerslev and Todorov (2011) demonstrate the the variance premium can be decomposed into a diffusive part and a discontinuous (jump) element. We thus anticipate jumps to be a relevant instrument for risk.

We investigate the potential cointegration relation by a restricted version of the co-fractional vector autoregressive model of Johansen (2008, 2009) and Johansen and Nielsen (2012), given by

$$\Delta^d \tilde{X}_t = \varphi \left[\theta' \left(1 - \Delta^d\right) \tilde{X}_t \right] + \sum_{i=1}^{n} \Gamma_i \Delta^d \left(1 - \Delta^d\right)^i \tilde{X}_t + u_t, \tag{16}$$

where $\tilde{X}_t \equiv [\tilde{RV}_t,\ \tilde{BV}_t,\ VIX_t^2]'$. We rely on model (16) because it allows us to identify a cointegration relation between the variables, while at the same time explicitly accounting for possible dynamics at higher frequencies, which may be present due to the overlapping nature of $\tilde{RV}_t$ and $\tilde{BV}_t$[12]. Given the identification problems of the model (see Carlini and Santucci de Magistris, 2013), we initially fix the cointegration rank $r = 2$ and estimate (16) by restricted maximum likelihood. Subsequently, we test for cointegration. For $\hat{d} = 0.38$ (SE($\hat{d}$)=0.03) and $n = 3$ we find the

---

$IV_{t,t+\tau}$.

[12]The Matlab code for the maximum-likelihood estimation of the parameters of model (16) has been provided by Nielsen and Morin (2012).

two instruments

$$q_t = \begin{pmatrix} q_{1,t} \\ q_{2,t} \end{pmatrix} = \hat{\theta}' \tilde{X}_t = \begin{pmatrix} 1 & -1.07 & 0 \\ -1.07 & 0 & 1 \end{pmatrix} \tilde{X}_t. \tag{17}$$

If we estimate a restricted version of our benchmark co-fractional model, where $\theta_{(2,1)} = -1$ and $\theta_{(1,2)} = -1$, we obtain a LR statistic of 19.99. This implies that we reject the restriction and the parameters θ are very precisely estimated. While statistically different, numerically $q_{2,t}$ is very close to the ex-post realized variance risk premium VRP_t of Bollerslev et al. (2009). Similarly, $q_{1,t}$ differs only very marginally from the pure jump contribution, i.e. the squared jump sizes over the past month. More precisely, $q_{1,t} \approx \sum_{i=1}^{22} \sum_{j=1}^{N_{t-i+1}} \psi_{t-i+1,j}^2$, where $\psi_{t,j}$ is the size of the jth jump on day t, and N_t denotes the total number of jumps in a day.

Table 3 lists the outcomes of the GMM estimation of Regression (9), using $q_{1,t}$ and $q_{2,t}$ from (17) as instruments. If we predict $r_{t+1}^{(e)}$ by VIX_t^2 using the two identified instruments and GMM estimation, we obtain a statistically significant slope estimate of $\hat{b} = 1.93$. This estimate is more than seven times larger than the corresponding inconsistent OLS estimate. Hence, we find strong evidence that there is an unobservable risk component, VP_t, contained in the VIX_t^2 series that positively predicts future daily stock returns, and that risk-return trade-off thus is positive.

We find that for a unit increase in risk investors demand an increase in the equity premium of approximately 2% annually. Putting this number into perspective, the value $\hat{b} = 1.93$ implies that a large increase in yesterday's VIX_t^2 of one standard deviation (=47.36), that is solely caused by an increase in VP_t by the same amount, implies a 91.37% increase in tomorrow's annualized predicted excess returns. That is, the equity premium almost doubles in reaction to such large changes in risk. The estimated effect on excess returns corresponds to a leap from the median to the 68th quantile of the return distribution, or it is equivalent to a return increase of 4.47 times the annualized standard deviation of excess returns. Our results lend strong support to the new theories of risk-return trade-off in that the variance premium captures risk and this risk is priced in aggregate markets resulting in sizable equity premiums.

5 Robustness

The results in the previous section depend on the adequacy of the assumptions made in the DGP. We review these assumptions here, and present several robustness checks.

For the findings to hold, it is necessary that the instruments $q_{k,t}$ are not irrelevant. To see this, let $q_{k,t} = \upsilon_{k,t}$ in Theorem 2 and estimate Regression (9) by GMM using $q_{k,t}$ as instruments. Then, as $T \to \infty$, $\hat{b} = O_p(1)$. From small sample simulations we infer that estimating (9) by GMM with an irrelevant instrument leads to an inconsistent and inefficient estimator of the risk-return trade-off parameter. To avoid such an outcome, we suggest a simple testing procedure. Assume that the researcher has identified a candidate instrument. Recall that the instrument follows the DGP given in (13), $q_{k,t} = \rho_k \, VP_t + \upsilon_{k,t}$. As VP_t is unobserved the researcher cannot simply regress the instrument on VP_t to conduct inference on the value of ρ_k and thus on the instrument relevance. Instead, $q_{k,t}$ can be regressed on the observed VIX_t^2 by OLS, however. By Theorem 1 it holds that the slope coefficient of this regression is an inconsistent estimate of ρ_k, yet valid statistical inference using a t-test can be carried out. Thus, relying on a simple OLS t-test the researcher can infer whether the instrument is statistically irrelevant.

Applying this approach to the two instruments identified in Section 4, we find no reason for concern. Regressing $q_{1,t}$ on VIX_t^2, the corresponding t-statistic, $t_{\hat{\rho}_1}$, is equal to 4.49. The *jump instrument* is thus a relevant instrument. Carrying out the same analysis for $q_{2,t}$, we find the respective value for $t_{\hat{\rho}_2}$ to be equal to 26.56, suggesting that also the *variance risk premium instrument* is strongly relevant.

Besides being relevant, the instruments $q_{k,t}$ further need to be valid. For an instrument to be valid, it may not be correlated with the residuals of the GMM regression of Equation (9), e_{t+1}. This implies that it may neither correlate with $\mathsf{E}_t^P(IV_{t,t+\tau})$ nor with ξ_{t+1}. In simulations we generate an instrument with innovations $\upsilon_{k,t} = \kappa_k \mathsf{E}_t^P(IV_{t,t+\tau}) + \mu_t$, where μ_t is an i.i.d. sequence. If $\kappa_k \neq 0$, this instrument is invalid as it violates the former assumption. We find that relying on such an invalid instrument leads to the same outcome as when estimating Regression (9) by simple OLS, i.e. standard inference is valid, but the risk-return trade-off parameter estimator is inconsistent. Alternatively, consider an instrument that violates the latter assumption, i.e. it is linearly related to unexpected returns ξ_{t+1}. In practice, using such an invalid instrument should be avoided at all costs. From our simulations we conclude that the size of a t-test on the significance of the risk-return trade-off coefficient is approximately 100%. The power of the test is also close to 100% in most instances, yet in extreme cases it may drop down to as low as 31.42%. The estimation of (9) by GMM further is strongly inconsistent.

A common approach to test for the validity of an instrument is to rely on Sargan's $\mathcal{J}$ test (Sargan, 1958). Corollary 1 summarizes the asymptotic behavior of the $\mathcal{J}$ test for our DGP (1)-(4).

Corollary 1. *Let VP_t, VIX_t^2, $r_t^{(e)}$, and $\mathsf{E}_t^P(IV_{t,t+\tau})$ be generated by* (1)-(4). *Assume there exist K instruments, generated by* (13). *Estimate the following second-stage regression by OLS*

$$\hat{\mathbf{e}} = \mathbf{Q}\varpi + \mathbf{v}, \tag{18}$$

where $\hat{\mathbf{e}}$ is the vector of regression residuals from estimating Equation (9) *by GMM. ϖ is a $(K+1)$ OLS coefficient vector and $\mathbf{v}$ is a vector of innovations. Compute the uncentered R^2 of Regression* (18) *as $R_u^2 = 1 - \frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}$. Define a test statistic for the validity of the instruments as*

$$\mathcal{J} \equiv TR_u^2. \tag{19}$$

Then, as $T \to \infty$:

$$\mathcal{J} \xrightarrow{D} \sum_{j=1}^{K-1} \lambda_j \chi_j^2(1),$$

where $\chi_j^2(1)$ are $K-1$ independent $\chi^2(1)$ distributed random variables. The weights λ_j are the eigenvalues of the $(K \times K)$ matrix $\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2\prime}$, which are defined in Appendix E in Equations (E6) *and* (E13)*, respectively.*

A proof of Corollary 1 can be found in Appendix E. The corollary shows that even though the true predictor VP_t is not observable, we can still test whether $q_{k,t}$ is a valid instrument. The statistical inference on the $\mathcal{J}$-statistic can be based on simulated p-values, following the approach suggested in Jagannathan and Wang (1996).

For the two instruments that we identified for our data set in Section 4, the implied $\mathcal{J}$-statistic in Table 3 is equal to 1.85. The corresponding simulated $p-$value is 0.44. We thus strongly fail to reject the null hypothesis and conclude that the *jump instrument* and the *variance risk premium instrument* are valid.

The adequacy of the proposed GMM approach further hinges on the assumption that the instruments are $I(0)$. In practice, it is fairly straightforward for the researcher to verify this condition. For example, the integration order of the instruments can be estimated by relying on the semiparametric approaches in Section 2 (see Shimotsu and Phillips, 2005 or Sun and Phillips, 2004); based on these estimates the null hypothesis that $d = 0$ can be evaluated. Alternatively, one can rely on hypothesis tests such as e.g. the KPSS test (Kwiatkowski et al., 1992). Since we identified our instruments in Section 4 by the co-fractional model, we can rely on a third alternative here. Johansen (2008) states that model (16) has a solution and $q_t = \theta' \tilde{X}_t \sim I(0)$ if the following conditions are satisfied. Firstly, the cointegration rank r needs to be smaller than 3. The value of the likelihood-ratio (LR) statistic of Johansen and Nielsen (2012) that provides a test for $H_0 : r \leq 2$ against $r \leq 3$ is equal to 2.76; thus we fail to reject the null hypothesis. Secondly, it must hold that $|\varphi'_\perp \left(I_{3\times 3} - \sum_{i=1}^n \Gamma_i\right) \theta_\perp| \neq 0$. In our estimation this value is equal to -1.57, i.e. different from zero. Thirdly, the roots c of the characteristic polynomial $|(1-c)I_{3\times 3} - \varphi\theta' c - (1-c)\sum_{i=1}^n \Gamma_i c^i| = 0$ must be either equal to one or lie outside a complex disk $\mathbb{C}_d$. Figure 2 shows that all roots fulfill this final condition. Hence, q_t are are integrated of order zero.

Our DGP further presumes that the variance premium is the only predictor of excess returns. This may be a rather stylized representation, since the extant literature suggest that other factors, such as e.g. the dividend-price ratio or the *cay* factor (see Lettau and Ludvigson, 2001) offer some return predictability. If there are such omitted factors, they are part of the error term ξ_{t+1} in (3). As a result, ξ_{t+1} may be serially correlated. Our derivations in Appendix F show that as long as ξ_{t+1} remains independent of VP_t and admits a linear representation with one-summable coefficients, our GMM estimation results continue to hold. The estimator for the risk-return trade-off is consistent, and standard inference can be carried out. Only the standard errors need to be adjusted to allow for serially correlation. Table 3 reports the robust HAC standard errors for our data. The risk-return trade-off remains significant.

Alternatively, it is perceivable that the potentially omitted variables in ξ_{t+1} are correlated with VP_t. In this situation, the reported GMM estimates may be biased. Thus, we need to confirm that ξ_{t+1} is not correlated with VP_t. Note that the $\mathcal{J}$-test from Corollary 1 may be viewed as a test of the joint null hypothesis that VP_t is orthogonal to $\mathsf{E}_t^P(IV_{t,t+\tau})$, that VP_t is orthogonal to unexpected returns ξ_{t+1}, that $\upsilon_{k,t}$ is orthogonal to $\mathsf{E}_t^P(IV_{t,t+\tau})$, and that $\upsilon_{k,t}$ is orthogonal to ξ_{t+1}, which we fail to reject for our data with a $p-$value of 0.44. Appendix F shows that the results of the $\mathcal{J}-$test continue to hold, even if ξ_{t+1} is not i.i.d. We may thus conclude that there seems to be no evidence in the data that potential omitted variables affect our results.

Finally, our DGP suggests that the observed predictor, VIX_t^2, is a fractionally integrated process $I(d)$. In Section 2 we present evidence for our data set that suggests that both, VIX_t^2 as well as $\mathsf{E}_t^P(IV_{t,t+\tau})$, are indeed $I(d)$ processes. Of course, these are empirical findings and thus possibly prone to a small statistical error. As a thought experiment, assume that instead of being $I(d)$, the variance process is autoregressive with roots very close to the value of one. Such a process may have a similar autocorrelation structure as the one presented in Figure 1(iii), but it leaves VIX_t^2 an $I(0)$ process. Yet, even if this were the case, all econometric results from Theorems 1 and 2 and Corollary 1 would be robust to this. Our findings are based on the assumption that $0 \leq d < 1/2$, thus including the $I(0)$ representation for VIX_t^2.

6 Long-Horizon Return Predictability

If the relation between excess returns and the lagged variance premium in Equation (3) holds for daily data, as our results so far suggest, we would expect it to hold also for longer horizon returns. That is, we can assume

$$r_{t+h}^{(e)} = \alpha_h + \beta_h VP_t + \xi_{t+h}. \tag{20}$$

We find consistent estimates for parameters of this long-run relation by estimating the regression $r_{t+h}^{(e)} = a_h + b_h VIX_t^2 + e_{t+h}$ by our proposed GMM approach, relying on the instruments $q_{1,t}$ and $q_{2,t}$ in (17). We measure cumulative returns $r_{t+h}^{(e)}$ by $\frac{1}{h}\sum_{i=1}^{h} r_{t+i}^{(e)}$, where $r_{t+i}^{(e)}$ are the log excess returns defined in (8). Given the overlapping nature of the cumulative returns, inference will be based on

Hansen and Hodrick standard errors as is commonly done in the literature (see e.g. Campbell et al., 1997).

Figure 3(i) plots the estimated prediction coefficient $\hat{b}_h$. The estimate shows a steady decline from the initial value of 1.93 as the horizon h increases. For all horizons of up to 126 days, i.e. six months, the coefficient remains statistically different from zero at a significance level of 5%. For horizons one month ($h = 21$), three months, and six months, respectively, we find $\hat{b}_h$=0.57, 0.42, and 0.31. These numbers are qualitatively very similar to Bollerslev et al. (2009), albeit somewhat larger for $h = 21$. Thus, for a unit increase in today's risk VP_t, the investors demand an immediate increase in tomorrow's equity premium of roughly 2%, but the effect of the same increase on the equity premium a month later is only 30% of this number; three months later it is merely 22%, and six months from now only 16% of the initial increase. We conclude that long-run excess return expectations are not strongly impacted by shocks to VP_t, but still significantly so.

We further empirically investigate relation (20) in relatively tranquil periods compared to turbulent times. To that end, we include a dummy variable in the GMM regression to capture the Financial Crisis from February 27, 2007 to March 2, 2009[13]. First we look at the estimated risk-return trade-off parameter in 'normal' periods, where most likely overall market risk is lower. Compared to the entire sample period, the estimated coefficient drops significantly initially, but it decays slower over horizons, as Figure 3(ii) shows. The estimated effect remains small and statistically significant for all horizons from one day to six months. A possible interpretation of these findings is that markets are generally less nervous during 'normal' times. The investors do not demand an immediate high compensation in tomorrow's returns for higher levels of risk, but such a shock does lead investors of all horizons up to half a year to require a modest increase in the equity premium. Conversely, as Figure 3(iii) shows, in crisis periods investors react in the opposite way. Short-term estimates $\hat{b}_h$ are very large and significant up to the horizon of roughly one month. That is, for the same increase in risk as during 'normal' periods, the immediately required equity premium in turbulent times is much larger, for instance equal to 3.54 for the next day. The effect

[13] 2007/02/27 is the start of the official crisis timeline of the Federal Reserve Bank of St. Louis FED (`https://www.stlouisfed.org/financial-crisis/full-timeline`) corresponding to the Freddie Mac Press Release. 2009/03/02 corresponds to the U.S. Treasury's and Federal Reserve Board's announcement to participate in the AIG restructuring plan; one day later the two launched the TALF program; seven days later the S&P 500 closed at the low point of 676.53.

tapers off relatively quickly, however, becoming insignificant for horizons longer than one month and shorter than approximately 3.5 months. In the long run for $h > 75$, the effect is very small but significant.

The implied predictability of excess returns at different horizons h is equal to

$$R_h^2 = \frac{\hat{b}_h^2 \hat{\sigma}_{VP}^2}{\hat{\sigma}_{r,h}^2}, \tag{21}$$

where $\hat{\sigma}_{r,h}^2$ is the sample variance of cumulative returns and $\hat{\sigma}_{VP}^2$ is the sample variance of the variance premium. As VP_t is latent, the sample variance cannot be computed. Nevertheless, we can gauge how predictability evolves over different horizons for hypothetical values of $\hat{\sigma}_{VP}^2$. Figure 4 summarizes the behavior. Independently of the true value of $\hat{\sigma}_{VP}^2$, we find that predictability increases (not entirely monotonically) from horizons of one day to $h = 82$ days and decreases thereafter, showing a hump-shaped pattern. This initial increase in predictive power is also found by for instance Drechsler and Yaron (2011) for the first three months. Interestingly, we find that for any value of the $\hat{\sigma}_{VP}^2$, the predictability is maximized at almost exactly four months. This is in line with the findings for the U.S. market in Bollerslev et al. (2009). The evidence is also in line with the international evidence provided by Bollerslev et al. (2014) who find that R^2 is maximized at a four-month horizon for many of the global markets. We conclude that the variance premium is a good predictor for the equity premium at short and intermediate horizons; for long-horizon returns its predictive power decays.

6.1 Comparative Predictability

Much empirical work has been dedicated to the analysis of the return predictability implied by VP_t (see references in Section 1). To the best of our knowledge the methodology of previous work differs substantially from the methods proposed in this paper. It is common in the literature to firstly find an estimate for the integrated variance $IV_{t,t+\tau}$ (or the quadratic variation); realized measures such as RV_t and BV_t are particularly popular. Secondly, a model for the dynamics of $IV_{t,t+\tau}$ is specified and successively estimated, producing estimates for the latent $\mathsf{E}_t^P(IV_{t,t+\tau})$. An estimate for the variance premium is then obtained by subtracting the estimate for $\mathsf{E}_t^P(IV_{t,t+\tau})$ from VIX_t^2.

Finally, this estimated variance premium, $\hat{VP}_t$, is used as a lagged predictor in a return regression, typically estimated by OLS. Naturally, it is to be expected that the estimation error and the model uncertainty inherent in $\hat{VP}_t$ will impact the OLS estimate for the risk-return trade-off. Relying on the proposed GMM approach instead, we can avoid this problem.

We compare our estimation approach to the ones common in the literature. One of the most popular models for variance dynamics is the martingale model, which has first been employed by Bollerslev et al. (2009). In this case the latent $\mathsf{E}_t^P(IV_{t,t+\tau})$ is simply replaced by $\tilde{RV}_t$. This model has been criticized since the dynamics of RV_t do not seem to resemble martingales. Instead, the HAR-RV model of Corsi (2009) has been found to fit the realized variance dynamics particularly well. It is an autoregressive model of the order 22 for daily realized variance measures with restrictions on the parameters. For our data, we estimate the HAR-RV for the one-day realized variance measures (including the overnight return), and form expectations for the variance over the next month from this model and the estimated parameters. Drechsler and Yaron (2011) among others follow a different approach, suggesting that realized monthly variances can be described as linear functions of the previous month's variance and VIX_t^2. We replicate this approach with our data, estimating the regression $\tilde{RV}_{t+22} = \gamma_0 + \gamma_1 VIX_t^2 + \gamma_2 \tilde{RV}_t + u_{t+22}$ and forming corresponding expectations. Given that long-memory models seem to fit the variance dynamics well, we lastly also estimate an ARFIMA model for the realized series. Again, we estimate the model for the one-day realized variance and bipower variation, and successively form expectations for the variance over the next month. The information criteria (BIC and AIC) both support a pure fractional noise specification, ARFIMA(0,0.39,0) for daily realized variances and ARFIMA(0,0.38,0) for the daily bipower variation.

Having generated an array of different estimates for $\mathsf{E}_t^P(IV_{t,t+\tau})$, we construct the estimated variance premium $\hat{VP}_t$ and estimate a predictive return regression by OLS. Assuming that the sample variance of the true latent VP_t is approximately equal to the sample variance of the estimate $\hat{VP}_t$, i.e. $\hat{\sigma}_{VP} \approx \hat{\sigma}_{\hat{VP}}$, we can compute a relative R^2 measure as

$$RP_h = \frac{R^2_{h,GMM}}{R^2_{h,OLS}} = \frac{\hat{b}^2_{h,GMM}\hat{\sigma}^2_{VP}/\hat{\sigma}^2_{r,h}}{\hat{b}^2_{h,OLS}\hat{\sigma}^2_{\hat{VP}}/\hat{\sigma}^2_{r,h}} \approx \frac{\hat{b}^2_{h,GMM}}{\hat{b}^2_{h,OLS}}. \tag{22}$$

The measure RP_h is plotted in Figure 5 for different horizons h. The proposed GMM estimation approach outperforms the competing models in the sense that it implies a stronger return predictability in sample. At almost all horizons the ARFIMA models imply the lowest comparative R^2, resulting in a maximum $RP_h = 39.19$ at the horizon of roughly three months. Somewhat more predictability is implied when variance expectations are derived from the Drechsler and Yaron (2011) regression, but still substantially less than the GMM method. The HAR-RV model for most horizons performs better than the two previous approaches, nevertheless still producing RP_h measures that vary from 1.48 to 12.06. Overall, relying on the BV_t measure relative to the RV_t measure for return variances results in a lower predictive power. Investigating the statistical significance of the slope estimate $\hat{b}_{h,OLS}$, we confirm that none of these models produce an estimate $\hat{VP}_t$ that significantly predicts returns at a 5% level, apart from the initial one to seven days.

The only OLS approach that has a predictive power that comes close to the GMM method, especially in the long run of approximately half a year, is when $\hat{VP}_t$ is the result of a martingale model for realized variances. Nevertheless, the GMM method still produces a 38% increase in the fit relative to the martingale model for RV_t at a short horizon of $h = 8$, and a 23% increase at the four month horizon ($h = 82$). The martingale models are also the only competing models that result in a statistically significant slope estimate for all horizons from one to 126 days. This is an interesting finding given the criticism that the model does not represent variance dynamics well. Yet, the martingale is the only model that produces a proxy for $\mathsf{E}_t^P(IV_{t,t+\tau})$ without estimation. We conclude therefore that model uncertainty does not impact the discovery of predictability much. On the other hand, the estimation error that is contained in all of the other competing models seems to impact predictability regressions rather severely.

6.2 Out-of-Sample Predictability

Our results so far suggest that our GMM approach produces more precise estimates for the risk-return trade-off parameter β, resulting in a better in-sample fit relative to the traditional approaches from the literature. We have shown that these estimates do not require observations on risk VP_t. However, to generate out-of-sample (OOS) forecasts for future excess returns, we need an observable measure for VP_t. We compute cumulative return forecasts as $\hat{r}^{(e)}_{T_{IS}+h} = \hat{a}_h + \hat{b}_h \hat{VP}_{T_{IS}}$, where T_{IS} is the number of in-sample observations. Estimates $\hat{a}_h$ and $\hat{b}_h$ are obtained by in-sample cumulative return GMM regression on VIX_t^2, relying only on data up to T_{IS}. $\hat{VP}_{T_{IS}}$ is the variance premium proxy resulting from one of the models described in Section 6.1. Obviously, our 'clean' estimate for the risk-return trade-off parameter then scales not only the true latent variance premium VP_t, but also the estimation and model error inherent in the proxy $\hat{VP}_t$.

For the first $h-$step ahead prediction, we consider the trough of the Financial Crisis on 2009/03/02 as the end of the in-sample period. The remaining OOS forecasts are produced with a rolling-window approach. We first evaluate the OOS predictions in terms of efficiency, that is we analyze the trade-off between the bias in the level of the forecast and the uncertainty in the forecast, which we measure by the root mean squared error (RMSE). For the majority of the 126 horizons, the lowest RMSE is achieved when $\hat{VP}_{T_{IS}} = VIX_{T_{IS}}^2 - \tilde{BV}_{T_{IS}}$. The RMSE ranges anywhere from 17.65 for $h = 1$ to 1.49 for $h = 126$, continuously decreasing as h increases. To put these numbers into perspective, we contrast these findings to the RMSE from a historical mean model as in Welch and Goyal (2008). We always achieve a higher efficiency; the reduction in RMSE relative to the historical mean model is between 1% and 8.5%, where the lowest gain is at $h = 1$ and the largest at 81 days, which again corresponds to a horizon of roughly four months.

We compare these forecasts, where $\hat{VP}_{T_{IS}} = VIX_{T_{IS}}^2 - \tilde{BV}_{T_{IS}}$, to return predictions from the traditional OLS approach. More precisely, we produce a competing set of OOS forecasts, where the estimates $\hat{a}_h$ and $\hat{b}_h$ are obtained by in-sample OLS estimation, replacing VP_t by a proxy $\hat{VP}_t$. Figure 6 shows that the suggested GMM approach leads to a forecasting efficiency gain at almost all horizons. The gain relative to all OLS models is again maximized at a horizon of four months. At $h = 81$ days our approach leads to a reduction in RMSE of 11% relative to the OLS model with $\hat{VP}_t$ resulting from the Drechsler and Yaron (2011) model for BV_t. Just like in the in-sample analysis, the only serious competitors from the OLS models are the martingale models. For few intermediate and the very long horizons, we find a very small improvement in RMSE from the latter two models. Investigating this further, we find that for these horizons the OLS models result in a lower average forecast error, but higher forecast uncertainty.

As a last step we analyze how much OOS predictability the models imply. That is, how much variation does the forecast produce relative to the variation of cumulative returns? Figure 7 plots the OOS R-squared, $R^2_{OOS} = \mathsf{Var}(\hat{r}^{(e)}_{T_{IS}+h})/\mathsf{Var}(r^{(e)}_{T_{IS}+h})$. We observe that the models where the risk-return trade-off parameter is estimated by in-sample GMM produce more volatility in the forecasts, which is necessary to match the variation in realized cumulative returns. For most horizons, the maximal forecast variation is implied when $\hat{VP}_t$ follows from the ARFIMA models and β is estimated by GMM; this is closely followed by the HAR-RV and GMM estimation. Thus, on the one hand models that presumably fit the dynamics of realized variances best, generate estimates for VP_t that best match the variation in returns. On the other hand, these ARFIMA models for VP_t combined with in-sample OLS estimation of β have the worst OOS fit. Hence, in the OLS framework the large predictor volatility, which is needed to produce sufficient variation in the forecast, at the same time harms the in-sample estimation of the risk-return trade-off parameter severely. This can also be seen by looking at the martingale models. With GMM in-sample estimation they produce rather little forecast variation, meaning that $\hat{VP}_t$ is not volatile enough. Yet, the small variation in $\hat{VP}_t$ leads to the relatively most accurate OLS estimates of β. As before, we conclude that the estimation error strongly biases the OLS estimation of the risk-return trade-off, and that the proposed GMM estimation approach can help alleviate these shortcomings.

7 Risk Aversion or Economic Uncertainty

Up to this point, we simply refereed to VP as *risk*. As mentioned in Section 1, there are disagreeing views in the literature on whether the variance premium captures economic uncertainty or risk aversion. The models of Drechsler and Yaron (2011) and Bollerslev et al. (2012) are cast in the long-run risk-type framework pioneered Bansal and Yaron (2004). They imply that VP_t is intrinsically related to economic uncertainty. Representative agents with recursive utility are assumed to have a strong preference for an early resolution of uncertainty and thus dislike increases in time-varying economic uncertainty. These two assumptions are necessary to produce a positive time-varying variance premium. In contrast, within an external-habit type framework established by Campbell and Cochrane (1999), Bekaert and Engstrom (2010) show that VP_t is linked to aggregate risk aversion. More precisely, they model consumption growth as being driven by good and bad shocks and an increase in the relative importance of the former (latter) shocks decreases (increases) the risk aversion. This time-varying importance of different shocks is also what generates the positive time-varying variance premium. In a similar spirit, Bekaert and Hoerova (2016) assume that the stock-return distribution has three different states: good, bad, and crash. In their model, an increase in risk aversion implies a higher weight on the crash state, which in turn leads to an increase in the variance premium. We contribute to the discussion by linking our latent measure for VP_t to popular indicators from both fields, economic uncertainty and risk aversion.

As VP_t is a latent variable in our study, computing a correlation between the variance premium and various measures of either risk aversion or economic uncertainty is not straightforward. Yet, based on our previously reported econometric results, we can define the following pseudo correlation measure. Let z_t be a time series that is stationary with an integration order $d < 1/2$ and zero mean $\mathsf{E}(z_t) = 0$[14]. If z_t is not correlated with the innovations of the instruments, $\upsilon_{k,t}$, we can compute a pseudo correlation measure as

$$\begin{aligned}\widehat{P\mathsf{Corr}}(VP_t, z_t) &= \begin{pmatrix} 0 & 1 \end{pmatrix} T^{1/2} \left(\mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1/2} \\ &\xrightarrow{P} \frac{1}{\sigma_{VP}}\mathsf{Corr}(VP_t, z_t), \qquad (23)\end{aligned}$$

where $\mathbf{z}$ is the $(T \times 1)$ vector of elements z_t. We can consistently estimate the scaled correlation between latent VP_t and z_t. The scaling factor is the inverse standard deviation of VP_t. The correlation is bounded between -1 and 1; hence we know that asymptotically the measure is bounded within $[-1/\sigma_{VP}, 1/\sigma_{VP}]$. Since the sign of σ_{VP} is always positive, the pseudo correlation has the same sign as the correlation itself in large samples, which helps us in interpreting the estimate. The measure does not depend on the measurement units of z_t in the limit; we can thus compare the pseudo correlation across different economic and financial series. Under the null hypothesis that VP_t and z_t are unrelated, it further holds that the pseudo correlation converges to a normal distribution with zero mean at the standard rate $T^{-1/2}$. The proof follows easily from the derivations in Appendix D and Lemma 1. Yet, the asymptotic variance depends on unknown nuisance parameters,

[14]In practice, we subtract the time-series average from all measures z_t that we consider.

which is why we conduct inference based on bootstrap confidence intervals.

As a point of reference, we first compute $\widehat{P\mathsf{Corr}}(VP_t, z_t)$, where z_t are all the commonly used estimates for VP_t that we discussed in the previous section. Table 4 reports the findings. All pseudo correlations are positive and strongly statistically significant. We find the highest correlation estimate of 5.13×10^{-2} for $\widehat{P\mathsf{Corr}}(VP_t,\, \hat{VP}_t)$ if $\hat{VP}_t$ results from the martingale model for the realized variance, that is when $\mathsf{E}_t^P(IV_{t,t+\tau})$ is merely replaced by $\tilde{RV}_t$. Given our previous findings and the fact that one of our instruments is closely related to this estimate for VP_t, this result is not surprising. In what follows, we refer to this value as the benchmark correlation. The lowest pseudo correlation of 2.86×10^{-2} is found if $\mathsf{E}_t^P(IV_{t,t+\tau})$ is replaced by the ARFIMA model for BV_t.

We now turn to popular indicator series for economic uncertainty. A measure that is designed to capture the uncertainty in overall economic activity is proposed by Bali and Zhou (2016). It is the conditional variance of the Chicago Fed National Activity Index, CV_{CFNAI}. Positive (negative) values of the index signify that the U.S. economy is growing at a faster (slower) rate relative to its historical trend. As in Bekaert and Hoerova (2016), we compute the conditional variance as a GARCH(1,1) prediction of the index. The pseudo correlation between CV_{CFNAI} and VP_t is 1.74×10^{-2}. While it positively correlates with the latent variance premium, we cannot reject that the correlation is statistically insignificant, however. Whereas CV_{CFNAI} is based on one underlying economic indicator, the macroeconomic uncertainty series $MUS^{(i)}$, where $i = \{1, 3, 12\}$ months, of Jurado et al. (2015) merge the information of 132 i-period conditional volatilities of mostly macroeconomic indicators. We find $\widehat{P\mathsf{Corr}}(VP_t, MUS^{(i)}) = 1.49 \times 10^{-2}$, 1.46×10^{-2}, and 1.33×10^{-2} for conditional variances over the next one month, three months, and one year respectively. The correlations are positive, of considerable magnitude relative to the benchmark correlation, and statistically significant. The fact that the correlation is strongest for the one-month series is to be expected, since our variance premium is the difference between risk neutral and objective expectations of integrated variance over the next 30 days. Bekaert et al. (2013) define a further uncertainty measure, UC, by isolating the objective conditional variance component from the VIX. Thus, this measure is specifically related to the uncertainty in financial markets. As before, we find a fairly high positive correlation between this measure of economic uncertainty and VP_t, with $\widehat{P\mathsf{Corr}}(VP_t, UC) = 2.82 \times 10^{-2}$. The estimate is statistically different from zero. Finally, economic uncertainty may also be closely related to uncertainty about economic policy. The US Economic Policy Uncertainty News-Based Index, EPU, computed by Baker, Bloom, and Davis is based on newspaper archives from Access World New's NewsBank service. The newspapers range from large national papers such as USA Today to small local newspapers across the U.S. We find a statistically significant positive pseudo correlation of 1.45×10^{-2}. In contrast to EPU, which focuses on overall economic policy uncertainty, the Equity Market-related Economic Uncertainty Index, $EMEUI$, computed by the same authors, is based on news pertaining to equity markets. The pseudo correlation between VP_t and $EMEUI$ is 1.37×10^{-2} and it is strongly statistically significant. We conclude that we find considerable evidence that the latent measure VP_t positively covaries with economic uncertainty. All pseudo correlations are positive and of sizable magnitude relative to the benchmark correlation, amounting to 26% to 55% of the benchmark correlation. All measures of economic uncertainty, with the one exception of CV_{CFNAI}, have a statistically significant correlation with VP_t.

Next we look at common indicators for aggregate risk aversion. From Bekaert et al. (2013) we rely on their risk aversion series, RA, which is the difference between VIX and their uncertainty component UC. The result is rather disappointing. The pseudo correlation is of substantial magnitude, equal to -2.46×10^{-2}, but it has the wrong sign and it is statistically insignificant. The RA series is monthly and ends in August 2010. Interestingly, if we compute the benchmark correlation for this first part of the sample at the same monthly frequency, we also find a negative pseudo correlation. This shows that our latent VP_t does not simply replicate the information contained in $\hat{VP}_t$ from the martingale model for RV_t. Froot and O'Connell (2003) compute the State Street Investor Confidence Index, $SSICCONF$[15]. It is a measure of investors' risk tolerance or sentiment and it is based on the theory that increased (decreased) holdings of risky assets in international markets signal a higher (lower) risk appetite. If VP_t represents risk aversion, we expect its pseudo correlation with $SSICCONF$ to be negative. The estimate in Table 4 is indeed negative, but its value of -3.70×10^{-3} is very small relative to the benchmark value and it is not statistically different from zero. The Credit Suisse Risk Appetite Index, $CSRAI$, the Standard Chartered Risk

[15]For an overview of some of these indices, see also Illing and Aaron (2005).

Appetite Index, $SCGRRAI$, and the Westpac US Risk Aversion Index, $WPFSI$, which employs the IMF methodology to identify risk aversion, are all examples of practitioners' indices for risk aversion computed by aggregating information from financial markets, such as e.g. bank-sector beta and the TED spread. The pseudo correlations with these three series are 7.11×10^{-4}, 1.95×10^{-3}, -1.70×10^{-3}, respectively. Thus, each correlation has the wrong sign, is very small relative to the benchmark, and is statistically insignificant. The Westpac Risk Aversion Index, $WPRAI$, is another indicator that takes a global perspective based inter alia on movements in major currency exchange rate markets and bond spreads in emerging economies. Here we find the expected positive significant pseudo correlation of 8.70×10^{-3}. Note however that the estimate is decidedly small, amounting to only 17% of the benchmark correlation. The Global Risk Aversion Indicator from the European Central Bank, $RAECB$, combines the information from five currently available risk aversion indicators by computing the first principal component. The pseudo correlation between this encompassing measure of risk aversion and VP_t is positive, strongly statistically significant, and of reasonable magnitude equal to 1.97×10^{-2}. Finally, financial market stress has been linked to the concept of risk aversion. In periods of stress, such as e.g. the recent Financial Crisis, we tend to observe an increased demand for safe securities. This can be interpreted as a sign that investors are less tolerant towards risk. Increased risk aversion has also been described as the transmission mechanism of financial stress (see e.g. Kumar and Persaud, 2002). When correlating VP_t with the Global Financial Stress Index from Bank of America Merrill Lynch, $GFSI$, we find the expected positive estimate that is significant and with a value of 1.66×10^{-2} of notable magnitude relative to the benchmark. Another indicator for financial stress with a more U.S.-based focus is the St. Louis Fed Financial Stress Index, $STLFSI$, which is the aggregate of seven interest rate series, six yield spreads, and five other indicators. We find a significantly positive, albeit rather small pseudo correlation of 9.45×10^{-3}. Interpreting the latter two pseudo correlations as evidence in favor of the hypothesis that VP_t captures risk aversion is not without controversy, however. While linked to risk aversion, financial stress can also be viewed as an indicator of economic uncertainty, or even Knightian uncertainty, as argued among others by Bekaert and Hoerova (2016). To summarize, we do not find strong convincing evidence that our latent VP_t captures risk aversion. Most correlation estimates are either very small, statistically insignificant, have the wrong sign, or z_t is not an irrefutable measure of risk aversion. The only exception to this rule is the correlation between VP_t and $RAECB$.

Our inference based on the pseudo correlation measure in (23) relies on the assumption that z_t is integrated of an order $d < 1/2$. Whereas we find no convincing evidence that the economic-uncertainty indices have an integration order greater than or equal to 0.5, the outcomes for some of the risk-aversion series are less clear. In particular, $CSRAI$, $WPFSI$, and $STLFSI$ seem to have a $d \geq 1/2$. For robustness, we also compute $\widehat{P\mathsf{Corr}}(VP_t, \Delta^{0.5} z_t)$ and $\widehat{P\mathsf{Corr}}(VP_t, \Delta z_t)$ for these measures[16]. The estimates hardly change and their values remain very small relative to the benchmark pseudo correlation. This robustness exercise therefore does not alter our general conclusion that VP_t seems to be related to economic uncertainty, yet not necessarily to risk aversion.

8 Concluding Remarks

This paper presents a novel stylized DGP that accounts for many theoretical and empirical features of the risk-return trade-off literature, such as for instance the persistence in the observed risk measure VIX and the stationary noise-type behavior of excess aggregate market returns. Assuming that the researcher estimates a misspecified, unbalanced, and endogenous predictive regression to gauge the risk-return trade-off empirically, where the regressor is an imperfect measure of the true risk measure VP, we show that OLS estimation results in an inconsistent estimator for the trade-off parameter. Nevertheless, standard statistical inference based on t-tests remains valid. To avoid the problem of obtaining an inconsistent estimate for the trade-off coefficient, we propose an GMM estimation method. If the econometrician has access to a valid and relevant $I(0)$ instrument, GMM estimation results in a consistent estimate for the parameter and standard statistical inference on predictability can be carried out.

We apply the methods outlined in this paper to the investigation of the risk-return trade-off and predictability of daily excess returns on the S&P 500 stock market. Relying on a fractional cointegration analysis, we provide one suggestion of how $I(0)$ instruments can be identified. We find

[16]For the sake of conciseness, we omit the robustness results here, but they are available from the authors upon request.

evidence of significant return predictability and a positive risk-return trade-off, using the suggested GMM approach. Our approach to isolate the return predictability contained in VP by GMM outperforms traditional methods, both in sample as well as out of sample. In particular, we show that GMM estimation is preferable to the traditional OLS methods that specify an estimate for the latent risk VP, as it is less prone to be impacted by model uncertainty and estimation error. Finally, we use the techniques developed in the paper to define a correlation estimator that measures the degree of dependence between latent VP and popular indicators for economic uncertainty and risk aversion. Our results lend support to the hypothesis that the variance premium is closely linked to economic uncertainty.

While we specifically focus on the estimation of the risk-return trade-off parameter, the theoretical developments in this article apply more generally to the prediction literature with persistent imperfect regressors. As such, we believe that we are the first to show that the persistent endogenous predictor problem, where the predictor has long-memory $I(d)$ dynamics, can be readily solved by identifying instruments that only possess short memory, $I(0)$. In particular, whenever the observed predictor may be viewed as the sum of a latent $I(0)$ signal and a latent $I(d)$ noise, we can rely on the GMM estimation method proposed.

Appendix

A TEW Estimation

Let $X_t \equiv [RV_t,\ BV_t,\ VIX_t^2]'$. Each element of X_t is (asymptotically) the sum of the $I(d)$ process $\mathsf{E}_t^P(IV_{t,t+\tau})$ and a perturbation. We adopt the trivariate version of the modified EW estimator of Sun and Phillips (2004) (TEW) to find the fractional integration order d_P of $\mathsf{E}_t^P(IV_{t,t+\tau})$. The underlying assumption of the TEW estimation approach in our setup is that the spectral density of X_t at frequency λ is given by

$$f_X(\lambda) \sim DGD' + H \quad \text{as } \lambda \to 0+, \tag{A1}$$

where $D = (\text{diag}[\lambda^{-d},\ \lambda^{-d},\ \lambda^{-d}])$, and G is a positive semidefinite matrix given by $g\Pi$, where Π is a (3×3) matrix of ones and scalar $g > 0$, such that $f_\eta(\lambda) \sim g$ as $\lambda \to 0+$. H is a (3×3) positive definite matrix that approximates the spectral density of the perturbations as we approach frequency zero. The perturbations may be correlated across the series.

We estimate the parameters of (A1) by maximum likelihood. From Sun and Phillips (2004) and the application of *Sylvester's Determinant Theorem* and the fundamental Lemma in Miller (1981), we obtain estimates by minimizing the negative log-likelihood given by

$$\frac{1}{m}\sum_{j=1}^{m}\{\ln|H| + \ln(1 + h_j' H^{-1} h_j) + tr(D_j H^{-1} D_j I_{\Delta^d(X)}(\lambda_j)) - \frac{1}{1+\ell_j} tr(D_j H^{-1} h_j h_j' H^{-1} D_j I_{\Delta^d(X)}(\lambda_j))\}. \tag{A2}$$

The (3×1) vector h_j is given by $h_j = [\sqrt{g}\lambda_j^{-d},\ \sqrt{g}\lambda_j^{-d},\ \sqrt{g}\lambda_j^{-d}]'$, m is the size of the spectral window, $\ell_j = tr(GD_jH^{-1}D_j)$, and $I_{\Delta^d(X)}(\lambda_j)$ is the periodogram of the filtered series X_t.

B Useful Lemma

Lemma 1 will prove useful for the derivations of the results in this paper.

Lemma 1. *Let a_t and b_t be two independent processes given by $a_t = \phi(L)\varepsilon_t$ and $b_t = (1-L)^{-d}\eta_t$ where $\phi(L) = \sum_{i=0}^{\infty}\phi_i L^i$ with $\sum_{i=0}^{\infty} i|\phi_i| < \infty$, $\phi(1) \neq 0$ and $(1-L)^d = \sum_{i=0}^{\infty}\gamma_i L^i$ with $\gamma_i = \Gamma(i+d)/(\Gamma(d)\Gamma(i+1))$, $0 \le d < \frac{1}{2}$ and $\epsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$, $\eta_t \sim i.i.d.(0, \sigma_\eta^2)$.*
Define $\mathcal{Z}_t = a_t b_t$; then, $T^{-1/2}\sum_{t=1}^{T}\mathcal{Z}_t/\bar{\sigma} \xrightarrow{D} \mathcal{N}(0,1)$ where $\bar{\sigma}_T^2 := var[T^{-1/2}\sum_{t=1}^{T}\mathcal{Z}_t] \to \bar{\sigma}^2$ as $T \to \infty$.

Proof: Let a_t, b_t and $\mathcal{Z}_t$ be as above and let $\mathcal{F}_t$ be the σ-algebra generated by $\{\varepsilon_t, \eta_t, \varepsilon_{t-1}, \eta_{t-1}, \cdots\}$. Note that, given independence, $\mathcal{Z}_t$ is a stationary ergodic process and that $\{\mathcal{Z}_t, \mathcal{F}_t\}$ is an adapted stochastic sequence with $\mathsf{E}[\mathcal{Z}_t^2] = \mathsf{E}[a_t^2 b_t^2] = \sigma_a^2\sigma_b^2 < \infty$ where $\sigma_a^2 = \mathsf{E}[a_t^2]$, $\sigma_b^2 = \mathsf{E}[b_t^2]$.

The lemma follows from Theorem 5.16 in White (2002), where we prove directly that

$$\sum_{m=1}^{\infty}\left(\mathsf{E}\left[\mathsf{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2\right]\right)^{1/2} < \infty.$$

First note that

$$\mathsf{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2 = \mathsf{E}\left[\left(\sum_{i=0}^{\infty}\phi_i\varepsilon_{-i}\right)\left(\sum_{i=0}^{\infty}\gamma_i\eta_{-i}\right)\Big|\mathcal{F}_{-m}\right]^2 = \left(\sum_{i=m}^{\infty}\phi_i\varepsilon_{-i}\right)^2\left(\sum_{i=m}^{\infty}\gamma_i\eta_{-i}\right)^2.$$

Thus,

$$\begin{aligned}\sum_{m=1}^{\infty}\left(\mathsf{E}\left[\mathsf{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2\right]\right)^{1/2} &= \sum_{m=1}^{\infty}\left(\sigma_\varepsilon^2\sigma_\eta^2\sum_{i=m}^{\infty}{\phi_i}^2\sum_{i=m}^{\infty}{\gamma_i}^2\right)^{1/2} \leq \sum_{m=1}^{\infty}\left(\sigma_\varepsilon^2\sigma_b^2\sum_{i=m}^{\infty}{\phi_i}^2\right)^{1/2} \\ &\leq \sigma_\varepsilon\sigma_b\sum_{m=1}^{\infty}\left(\sum_{i=m}^{\infty}|\phi_i|\right) = \sigma_\varepsilon\sigma_b\left(\sum_{i=0}^{\infty}i|\phi_i|\right) < \infty.\end{aligned}$$

Note in particular that Lemma 1 proves that multiplying the long-memory process by an $I(0)$ process reduces the order of convergence to the one of a short-memory process.

C Proof of Theorem 1

Throughout the appendices, we rely on the same notation as in the main text. Let $\hat{(\cdot)}$ denote an estimator and introduce the following additional notation for Appendix C, D, and E

$$\begin{aligned} x_t &= VIX_t^2 \\ z_t &= \mathsf{E}_t^P(IV_{t,t+\tau}) \\ \mathbf{e} & \quad \text{a } (T-1)\times 1 \text{ vector given by } [e_2,\ e_3\ \ldots,\ e_T]' \\ \boldsymbol{\iota}_{(i)} & \quad \text{the } i\text{th unit vector} \\ \hat{\mathsf{b}}_{OLS} &= [\hat{a}_{OLS},\ \hat{b}_{OLS}]' \text{ and } \mathsf{b} = [\alpha,\ \beta]' \\ \mathbf{n} & \quad \text{a normally distributed random vector or scalar} \\ \boldsymbol{\rho} &= \text{a } K\times 1 \text{ vector given by } [\rho_1,\ \rho_2\ \ldots,\ \rho_K]' \\ \boldsymbol{\Sigma}_\upsilon &= \text{a } K\times K \text{ diagonal matrix, where the diagonal elements are equal to } \sigma_{\upsilon_k}^2 \\ \boldsymbol{\Gamma}_j &= \text{The autocovariance matrix at lag } j. \end{aligned}$$

To derive the asymptotic behavior of the estimators $\hat{a}_{OLS}$ and $\hat{b}_{OLS}$, along with the associated t-statistics, it is necessary to obtain the limit expression of the sums that define them. These are summarized in Table C1, along with their respective convergence rates. All of the convergence rates (see the *underbraced* expressions) can be found in Tsay and Chung (2000) or Hayashi (2000) except for the normalization ratios of $\sum VP_tz_t$ and $\sum\xi_{t+1}z_t$, which follow from Lemma 1.

We start by showing the convergence results for some linear combinations. First we consider the asymptotic properties of $(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1}$.

$$\text{plim}(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1} = \left(\text{plim}\,\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1} = \left(\text{plim}\,\frac{1}{T}\begin{bmatrix} T-1 & \sum x_t \\ \sum x_t & \sum x_t^2\end{bmatrix}\right)^{-1}$$

$\sum x_t$	$=$	$\underbrace{\sum VP_t}_{O_p(T^{1/2})} + \underbrace{\sum z_t}_{O_p(T^{d+1/2})}$
$\sum x_t^2$	$=$	$\underbrace{\sum VP_t^2}_{O_p(T)} + \underbrace{\sum z_t^2}_{O_p(T)} + 2\underbrace{\sum VP_t z_t}_{O_p(T^{1/2})}$
$\sum e_{t+1}$	$=$	$-\beta \sum z_t + \underbrace{\sum \xi_{t+1}}_{O_p(T^{1/2})}$
$\sum e_{t+1}^2$	$=$	$\beta^2 \sum z_t^2 + \underbrace{\sum \xi_{t+1}^2}_{O_p(T)} - 2\beta \underbrace{\sum \xi_{t+1} z_t}_{O_p(T^{1/2})}$
$\sum x_t e_{t+1}$	$=$	$-\beta \sum z_t^2 - \beta \sum VP_t z_t + \sum \xi_{t+1} z_t + \underbrace{\sum \xi_{t+1} VP_t}_{O_p(T^{1/2})}$

Table C1: Expressions for sums in Theorem 1.

$$= \begin{bmatrix} 1 & 0 \\ 0 & \sigma_{VP}^2 + \sigma_P^2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_{VP}^2 + \sigma_P^2} \end{bmatrix} \tag{C1}$$

Next, we focus on the dynamics of $\frac{1}{T}\mathbf{e}'\mathbf{e}$. Note that

$$\text{plim} \frac{1}{T}\mathbf{e}'\mathbf{e} = \text{plim} \frac{1}{T} \sum e_{t+1}^2 = \text{plim} \frac{1}{T} \sum \left(\beta^2 z_t^2 + \xi_{t+1}^2 - 2\beta\xi_{t+1} z_t\right) = \beta^2 \sigma_P^2 + \sigma_\xi^2. \tag{C2}$$

Finally, we show the asymptotic behavior of $\frac{1}{T}\mathbf{X}'\mathbf{e}$.

$$\frac{1}{T}\mathbf{X}'\mathbf{e} = \frac{1}{T} \begin{bmatrix} \sum e_{t+1} \\ \sum x_t e_{t+1} \end{bmatrix} = \begin{bmatrix} -\beta \frac{1}{T} \sum z_t + \frac{1}{T} \sum \xi_{t+1} \\ -\beta \frac{1}{T} \sum z_t^2 - \beta \frac{1}{T} \sum VP_t z_t + \frac{1}{T} \sum \xi_{t+1} z_t + \frac{1}{T} \sum \xi_{t+1} VP_t \end{bmatrix}. \tag{C3}$$

If $\underline{\beta \neq 0}$, it follows that plim $\frac{1}{T}\mathbf{X}'\mathbf{e} = [0, \ -\beta\sigma_P^2]'$. Conversely, if $\underline{\beta = 0}$, we find that

$$\frac{1}{T}\mathbf{X}'\mathbf{e} = \frac{1}{T^{1/2}} \begin{bmatrix} \frac{1}{T^{1/2}} \sum \xi_{t+1} \\ \frac{1}{T^{1/2}} \sum x_t \xi_{t+1} \end{bmatrix} = \frac{1}{T^{1/2}} \begin{bmatrix} \frac{1}{T^{1/2}} \sum \xi_{t+1} \\ \frac{1}{T^{1/2}} \sum \xi_{t+1} z_t + \frac{1}{T^{1/2}} \sum \xi_{t+1} VP_t \end{bmatrix}. \tag{C4}$$

The expression (C4) involves the random variables ξ_{t+1} and $\xi_{t+1}(VP_t + z_t)$, both of which are strictly stationary and ergodic and fulfill the conditions outlined in Lemma 1. The term thus has a zero mean and a constant variance. Hence, by the Central Limit Theorem (CLT) in Lemma 1 the term converges in distribution to

$$\mathcal{N}\left(0, \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j = \begin{bmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\xi^2(\sigma_{VP}^2 + \sigma_P^2) \end{bmatrix}\right) \tag{C5}$$

at rate $T^{1/2}$.

C.1 Asymptotic Properties of the OLS Estimator

Note that $\hat{\mathtt{b}}_{OLS} - \mathtt{b} = \left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1} \frac{1}{T}\mathbf{X}'\mathbf{e}$. Using the results above we find that if $\underline{\beta \neq 0}$

$$\hat{\mathtt{b}}_{OLS} - \mathtt{b} \xrightarrow{P} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix} \begin{bmatrix} 0 \\ -\beta\sigma^2_P \end{bmatrix} = \begin{bmatrix} 0 \\ -\beta\frac{\sigma^2_P}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix}, \tag{C6}$$

and therefore $\hat{a}_{OLS} \xrightarrow{P} \alpha$ and $\hat{b}_{OLS} \xrightarrow{P} \beta\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P}$. Conversely, if $\underline{\beta = 0}$ we find

$$T^{1/2}(\hat{\mathtt{b}}_{OLS} - \mathtt{b}) = \left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1} \frac{1}{T^{1/2}}\mathbf{X}'\mathbf{e} \xrightarrow{D} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix} \mathbf{n} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2_\xi & 0 \\ 0 & \frac{\sigma^2_\xi}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix}\right). \tag{C7}$$

C.2 Asymptotic Properties of the Estimator of the Error Variance

Note that $s^2 = \frac{1}{T-3}\hat{\mathbf{e}}'\hat{\mathbf{e}} = \frac{1}{T-3}\mathbf{e}'\mathbf{e} - \frac{1}{T-3}\mathbf{e}'\mathbf{X}\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbf{e}$. Using the results above we find that if $\underline{\beta \neq 0}$

$$s^2 \xrightarrow{P} \beta^2\sigma^2_P + \sigma^2_\xi - [0,\ -\beta\sigma^2_P] \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix} \begin{bmatrix} 0 \\ -\beta\sigma^2_P \end{bmatrix} = \sigma^2_\xi + \beta^2\sigma^2_P\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P} \tag{C8}$$

Conversely, if $\underline{\beta = 0}$ we find that $s^2 \xrightarrow{P} \sigma^2_\xi$.

C.3 Asymptotic Properties of the t-statistics

Note that the t-statistic for a test of either hypothesis $H_0 : \alpha = 0$ or $H_0 : \beta = 0$ can be written as $t_{(i)} = \boldsymbol{\iota}'_{(i)}\hat{\mathtt{b}}_{OLS}(s^2\frac{1}{T}\boldsymbol{\iota}'_{(i)}(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\iota}_{(i)})^{-1/2}$. Using the results above we find that if $\underline{\beta \neq 0}$

$$\begin{aligned} T^{-1/2}\operatorname{plim} t_a &= \alpha\left(\sigma^2_\xi + \beta^2\sigma^2_P\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P}\right)^{-1/2}\left([1,\ 0]\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)^{-1/2} \\ &= \frac{\alpha}{\sqrt{\sigma^2_\xi + \beta^2\sigma^2_P\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P}}} \end{aligned} \tag{C9}$$

$$\begin{aligned} T^{-1/2}\operatorname{plim} t_b &= \beta\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P}\left(\sigma^2_\xi + \beta^2\sigma^2_P\frac{\sigma^2_{VP}}{\sigma^2_{VP}+\sigma^2_P}\right)^{-1/2}\left([0,\ 1]\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2_{VP}+\sigma^2_P} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)^{-1/2} \\ &= \frac{\beta\sigma^2_{VP}}{\sqrt{\sigma^2_\xi(\sigma^2_{VP}+\sigma^2_P) + \beta^2\sigma^2_P\sigma^2_{VP}}}. \end{aligned} \tag{C10}$$

Conversely, if $\underline{\beta = 0}$ we find that $t_a \xrightarrow{P} \alpha/\sqrt{\sigma^2_\xi}$, which readily follows from (C9) above. For t_b we find that

$$t_b = T^{1/2}\boldsymbol{\iota}'_{(2)}(\hat{\mathtt{b}}_{OLS} - \mathtt{b})(s^2\boldsymbol{\iota}'_{(2)}(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\iota}_{(2)})^{-1/2} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2_\xi}{\sigma^2_{VP}+\sigma^2_P}\frac{1}{\sigma^2_\xi}(\frac{1}{\sigma^2_{VP}+\sigma^2_P})^{-1}\right) = \mathcal{N}(0,1) \tag{C11}$$

D Proof of Theorem 2

As for the proof of Theorem 1 in Appendix C, it is necessary to obtain the limit expression of the sums that appear in the definitions of the GMM estimates and the associated t-ratios. Most of these expressions are summarized in Table C1. The remaining sums can be found in Table D2.

$\sum q_{k,t}$	$=$	$\rho_k \sum VP_t + \underbrace{\sum \upsilon_{k,t}}_{O_p(T^{1/2})}$
$\sum q_{k,t}^2$	$=$	$\rho_k^2 \sum VP_t^2 + \underbrace{\sum \upsilon_{k,t}^2}_{O_p(T)} + 2 \underbrace{\rho_k \sum VP_t \upsilon_{k,t}}_{O_p(T^{1/2})}$
$\sum q_{k,t} q_{j,t}$	$=$	$\rho_k \rho_j \sum VP_t^2 + \rho_k \sum VP_t \upsilon_{j,t} + \rho_j \sum VP_t \upsilon_{k,t} + \underbrace{\sum \upsilon_{k,t}\upsilon_{j,t}}_{O_p(T^{1/2})}$
$\sum e_{t+1} q_{k,t}$	$=$	$-\beta\rho_k \sum VP_t z_t + \rho_k \sum \xi_{t+1} VP_t - \beta \underbrace{\sum z_t \upsilon_{k,t}}_{O_p(T^{1/2})} + \underbrace{\sum \xi_{t+1}\upsilon_{k,t}}_{O_p(T^{1/2})}$
$\sum x_t q_{k,t}$	$=$	$\rho_k \sum VP_t^2 + \sum VP_t \upsilon_{k,t} + \rho_k \sum VP_t z_t + \sum z_t \upsilon_{k,t}$

Table D2: Expressions for sums in Theorem 2 with $j \neq k$; $k = 1, \cdots, K$.

We start by showing the convergence results for some linear combinations. First we consider the asymptotic properties of $(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}$.

$$\operatorname{plim}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1} = (\operatorname{plim}\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1} = \left(\operatorname{plim}\frac{1}{T} \begin{bmatrix} T-1 & \sum q_{1,t} & \sum q_{2,t} & \cdots & \sum q_{K,t} \\ \sum q_{1,t} & \sum q_{1,t}^2 & \sum q_{1,t} q_{2,t} & \cdots & \sum q_{1,t} q_{K,t} \\ \sum q_{2,t} & \sum q_{2,t} q_{1,t} & \sum q_{2,t}^2 & \cdots & \sum q_{2,t} q_{K,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum q_{K,t} & \sum q_{K,t} q_{1,t} & \sum q_{K,t} q_{2,t} & \cdots & \sum q_{K,t}^2 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\rho}\boldsymbol{\rho}'\sigma_{VP}^2 + \boldsymbol{\Sigma}_\upsilon \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_\upsilon^{-1} - \frac{\sigma_{VP}^2 \boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}}{1+\sigma_{VP}^2 \boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}, \tag{D1}$$

where the last step follows by the Sherman-Morrison Identity. Next, we focus on the dynamics of $\frac{1}{T}\mathbf{Q}'\mathbf{X}$. Note that

$$\operatorname{plim}\frac{1}{T}\mathbf{Q}'\mathbf{X} = \operatorname{plim}\frac{1}{T} \begin{bmatrix} T-1 & \sum x_t \\ \sum q_{1,t} & \sum q_{1,t} x_t \\ \sum q_{2,t} & \sum q_{2,t} x_t \\ \vdots & \vdots \\ \sum q_{K,t} & \sum q_{K,t} x_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \mathbf{0} & \sigma_{VP}^2 \boldsymbol{\rho} \end{bmatrix}. \tag{D2}$$

Finally, we show the asymptotic behavior of $\frac{1}{T}\mathbf{Q}'\mathbf{e}$.

$$\frac{1}{T}\mathbf{Q}'\mathbf{e} = \frac{1}{T}\begin{bmatrix} \sum e_{t+1} \\ \sum q_{1,t}e_{t+1} \\ \sum q_{2,t}e_{t+1} \\ \vdots \\ \sum q_{K,t}e_{t+1} \end{bmatrix} = \begin{bmatrix} -\beta\frac{1}{T}\sum z_t + \frac{1}{T}\sum \xi_{t+1} \\ -\beta\frac{1}{T}\sum q_{1,t}z_t + \frac{1}{T}\sum q_{1,t}\xi_{t+1} \\ -\beta\frac{1}{T}\sum q_{2,t}z_t + \frac{1}{T}\sum q_{2,t}\xi_{t+1} \\ \vdots \\ -\beta\frac{1}{T}\sum q_{K,t}z_t + \frac{1}{T}\sum q_{K,t}\xi_{t+1} \end{bmatrix}. \tag{D3}$$

If $\underline{\beta \neq 0}$, it follows that plim $\frac{1}{T}\mathbf{Q}'\mathbf{e} = \mathbf{0}'$. Conversely, if $\underline{\beta = 0}$, we find that

$$\frac{1}{T}\mathbf{Q}'\mathbf{e} = \frac{1}{T^{1/2}}\begin{bmatrix} \frac{1}{T^{1/2}}\sum \xi_{t+1} \\ \frac{1}{T^{1/2}}\sum q_{1,t}\xi_{t+1} \\ \frac{1}{T^{1/2}}\sum q_{2,t}\xi_{t+1} \\ \vdots \\ \frac{1}{T^{1/2}}\sum q_{K,t}\xi_{t+1} \end{bmatrix} = \frac{1}{T^{1/2}}\begin{bmatrix} \frac{1}{T^{1/2}}\sum \xi_{t+1} \\ \frac{1}{T^{1/2}}\sum \rho_1\, VP_t\xi_{t+1} + \frac{1}{T^{1/2}}\sum \upsilon_{1,t}\xi_{t+1} \\ \frac{1}{T^{1/2}}\sum \rho_2\, VP_t\xi_{t+1} + \frac{1}{T^{1/2}}\sum \upsilon_{2,t}\xi_{t+1} \\ \vdots \\ \frac{1}{T^{1/2}}\sum \rho_K\, VP_t\xi_{t+1} + \frac{1}{T^{1/2}}\sum \upsilon_{K,t}\xi_{t+1} \end{bmatrix} \tag{D4}$$

The expression (D4) involves the random variables ξ_{t+1} and $\xi_{t+1}(\rho_k\, VP_t + \upsilon_{k,t})$, both of which are strictly stationary and ergodic and fulfill the conditions outlined in Lemma 1. The term thus has a zero mean and a constant variance. Hence, by the CLT in Lemma 1 the term converges in distribution to

$$\mathcal{N}\left(0, \sum_{j=\infty}^{\infty}\mathbf{\Gamma}_j = \sigma_\xi^2\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\rho}\boldsymbol{\rho}'\sigma_{VP}^2 + \mathbf{\Sigma}_\upsilon \end{bmatrix}\right) \tag{D5}$$

at rate $T^{1/2}$.

D.1 Asymptotic Properties of the GMM Estimator

Note that $\hat{\mathtt{b}}_{GMM} - \mathtt{b} = \left(\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{e}$. Using the results above we find that if $\underline{\beta \neq 0}$

$$\hat{\mathtt{b}}_{GMM} - \mathtt{b} \xrightarrow{P} \left(\begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \boldsymbol{\rho}'\sigma_{VP}^2 \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Sigma}_\upsilon^{-1} - \frac{\sigma_{VP}^2\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}}{1+\sigma_{VP}^2\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \mathbf{0} & \boldsymbol{\rho}\sigma_{VP}^2 \end{bmatrix}\right)^{-1} \tag{D6}$$

$$\times \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \boldsymbol{\rho}'\sigma_{VP}^2 \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Sigma}_\upsilon^{-1} - \frac{\sigma_{VP}^2\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}}{1+\sigma_{VP}^2\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma_{VP}^4\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and therefore $\hat{a}_{GMM} \xrightarrow{P} \alpha$ and $\hat{b}_{GMM} \xrightarrow{P} \beta$. Conversely, if $\underline{\beta = 0}$ we find

$$T^{1/2}(\hat{\mathtt{b}}_{GMM} - \mathtt{b}) = \left(\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T^{1/2}}\mathbf{Q}'\mathbf{e}$$

$$\xrightarrow{P} \left(\begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \boldsymbol{\rho}'\sigma_{VP}^2 \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Sigma}_\upsilon^{-1} - \frac{\sigma_{VP}^2\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}}{1+\sigma_{VP}^2\boldsymbol{\rho}'\mathbf{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \mathbf{0} & \boldsymbol{\rho}\sigma_{VP}^2 \end{bmatrix}\right)^{-1}$$

$$\times \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \boldsymbol{\rho}'\sigma^2_{VP} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_\upsilon^{-1} - \frac{\sigma^2_{VP}\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}}{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix} \mathbf{n} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \frac{\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}}{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix} \mathbf{n}$$

$$= \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \frac{\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}}{\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix} \mathbf{n} \sim \mathcal{N}\left(0, \sigma^2_\xi \begin{bmatrix} 1 & 0 \\ 0 & \frac{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\right). \tag{D7}$$

D.2 Asymptotic Properties of the Estimator of the Error Variance

Note that $s^2 = \frac{1}{T-3}\hat{\mathbf{e}}'\hat{\mathbf{e}} = \frac{1}{T-3}\mathbf{e}'\mathbf{e} - \frac{1}{T-3}\mathbf{e}'\mathbf{X}(\hat{\mathtt{b}}_{GMM} - \mathtt{b}) - (\hat{\mathtt{b}}_{GMM} - \mathtt{b})'\frac{1}{T-3}\mathbf{X}'\mathbf{e} + (\hat{\mathtt{b}}_{GMM} - \mathtt{b})'\frac{1}{T-3}\mathbf{X}'\mathbf{X}(\hat{\mathtt{b}}_{GMM} - \mathtt{b})$. Using the results above and in Appendix C we find that if $\underline{\beta \neq 0}$

$$s^2 \xrightarrow{P} (\beta^2\sigma^2_P + \sigma^2_\xi) - \left([0,\ -\beta\sigma^2_P]\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) - \left([0,\ 0]\begin{bmatrix} 0 \\ -\beta\sigma^2_P \end{bmatrix}\right) + \left([0,\ 0]\begin{bmatrix} 1 & 0 \\ 0 & \sigma^2_{VP} + \sigma^2_P \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$
$$= \beta^2\sigma^2_P + \sigma^2_\xi \tag{D8}$$

Conversely, if $\underline{\beta = 0}$ we find that $s^2 \xrightarrow{P} \sigma^2_\xi$.

D.3 Asymptotic Properties of the t-statistics

Note that the t-statistic can be written as $t_{(i)} = \boldsymbol{\iota}'_{(i)}\hat{\mathtt{b}}_{GMM}(s^2\frac{1}{T}\boldsymbol{\iota}'_{(i)}(\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X})^{-1}\boldsymbol{\iota}_{(i)})^{-1/2}$ for a test of either hypothesis $H_0 : \alpha = 0$ or $H_0 : \beta = 0$. Using the results above we find that if $\underline{\beta \neq 0}$

$$T^{-1/2}t_a \xrightarrow{P} \alpha\left(\sigma^2_\xi + \beta^2\sigma^2_P\right)^{-1/2}\left([1,\ 0]\begin{bmatrix} 1 & 0 \\ 0 & \frac{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)^{-1/2} = \frac{\alpha}{\sqrt{\sigma^2_\xi + \beta^2\sigma^2_P}} \tag{D9}$$

$$\begin{aligned} T^{-1/2}t_b &\xrightarrow{P} \beta\left(\sigma^2_\xi + \beta^2\sigma^2_P\right)^{-1/2}\left([0,\ 1]\begin{bmatrix} 1 & 0 \\ 0 & \frac{1+\sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)^{-1/2} \\ &= \beta\left(\frac{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{(\sigma^2_\xi + \beta^2\sigma^2_P)(1 + \sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho})}\right)^{1/2}. \end{aligned} \tag{D10}$$

Conversely, if $\underline{\beta = 0}$ we find that $t_b \xrightarrow{P} \alpha/\sqrt{\sigma^2_\xi}$, which readily follows from (D9) above. For t_b we find that

$$t_b = T^{1/2}\boldsymbol{\iota}'_{(2)}(\hat{\mathtt{b}}_{GMM} - \mathtt{b})(s^2\boldsymbol{\iota}'_{(2)}(\frac{1}{T}\mathbf{X}'\mathbf{Q}(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X})^{-1}\boldsymbol{\iota}_{(2)})^{-1/2} \xrightarrow{D} \mathbf{n}\left(\frac{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^2_\xi(1 + \sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho})}\right)^{1/2}$$
$$\sim \mathcal{N}\left(0, \frac{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^2_\xi(1 + \sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho})}\sigma^2_\xi\frac{1 + \sigma^2_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}{\sigma^4_{VP}\boldsymbol{\rho}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\rho}}\right) = \mathcal{N}(0, 1) \tag{D11}$$

E Proof of Corollary 1

For convenience introduce the following additional notation for Appendix E

$$\mathbf{q} \quad = \quad \text{a } K \times 1 \text{ vector given by } \left[\sum q_{1,t},\ \sum q_{2,t},\ \ldots,\ \sum q_{K,t}\right]'$$

$$\begin{aligned}
\mathbf{B} \quad &= \quad \text{a } K \times K \text{ matrix given by } \begin{bmatrix} \sum q_{1,t}^2 & \sum q_{1,t}q_{2,t} & \cdots & \sum q_{1,t}q_{K,t} \\ \sum q_{2,t}q_{1,t} & \sum q_{2,t}^2 & \cdots & \sum q_{2,t}q_{K,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sum q_{K,t}q_{1,t} & \sum q_{K,t}q_{2,t} & \cdots & \sum q_{K,t}^2 \end{bmatrix} \\
\mathbf{S} \quad &= \quad \mathbf{B} - \mathbf{q}\mathbf{q}'/(T-1) \\
\sum \overrightarrow{qx} \quad &= \quad \text{a } K \times 1 \text{ vector given by } \left[\sum q_{1,t}x_t,\ \sum q_{2,t}x_t,\ \ldots,\ \sum q_{K,t}x_t\right]' \\
\sum \overrightarrow{qe} \quad &= \quad \text{a } K \times 1 \text{ vector given by } \left[\sum q_{1,t}e_{t+1},\ \sum q_{2,t}e_{t+1},\ \ldots,\ \sum q_{K,t}e_{t+1}\right]' \\
\gamma_j^{(a\times b)} \quad &= \quad \text{The autocovariance of a series } a_t \times b_t \text{ at lag} j.
\end{aligned}$$

Recall that the test statistic in Corollary 1 is given by $\mathcal{J} = T(1-\hat{\mathbf{v}}'\hat{\mathbf{v}}/\hat{\mathbf{e}}'\hat{\mathbf{e}})$, where $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathsf{b}}_{GMM}$ and $\hat{\mathbf{v}} = \hat{\mathbf{e}} - \mathbf{Q}\hat{\varpi}$. Note that we can re-write the $\mathcal{J}-$statistic as follows

$$\mathcal{J} = T\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}} - \hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}} = \frac{\hat{\mathbf{e}}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\hat{\mathbf{e}}}{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}} = \frac{\mathbf{e}'\mathbf{Q}\mathbf{L}}{\sqrt{\frac{T-3}{T}s^2}}\left[\mathbf{I} - \mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\right]\frac{\mathbf{L}'\mathbf{Q}'\mathbf{e}}{\sqrt{\frac{T-3}{T}s^2}}, \qquad \text{(E1)}$$

where $\mathbf{L}$ is a $(K+1)\times(K+1)$ matrix such that $\mathbf{L}\mathbf{L}' = (\mathbf{Q}'\mathbf{Q})^{-1}$. Since it holds that

$$(\mathbf{Q}'\mathbf{Q})^{-1} = \begin{bmatrix} T-1 & \mathbf{q}' \\ \mathbf{q} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{T-1} + \frac{1}{(T-1)^2}\mathbf{q}'\mathbf{S}^{-1}\mathbf{q} & -\frac{1}{T-1}\mathbf{q}'\mathbf{S}^{-1} \\ -\frac{1}{T-1}\mathbf{S}^{-1}\mathbf{q} & \mathbf{S}^{-1} \end{bmatrix},$$

we can write $\mathbf{L}$ as

$$\mathbf{L} = \begin{bmatrix} \frac{1}{(T-1)^{1/2}} & -\frac{1}{T-1}\mathbf{q}'\mathbf{S}^{-1/2} \\ \mathbf{0} & \mathbf{S}^{-1/2} \end{bmatrix}.$$

Hence, $\mathcal{J}$ in (E1) is the squared form of a linear combination of a $(K+1)\times 1$ vector and a $(K+1)\times(K+1)$ symmetric and idempotent matrix. We note that

$$\mathbf{L}'\mathbf{Q}'\mathbf{X} = \begin{bmatrix} \frac{1}{(T-1)^{1/2}} & \mathbf{0}' \\ -\frac{1}{T-1}\mathbf{S}^{-1/2}\mathbf{q} & \mathbf{S}^{-1/2} \end{bmatrix}\begin{bmatrix} T-1 & \sum x_t \\ \mathbf{q} & \sum\overrightarrow{qx} \end{bmatrix} = \begin{bmatrix} (T-1)^{1/2} & \frac{\sum x_t}{(T-1)^{1/2}} \\ \mathbf{0} & \mathbf{S}^{-1/2}(-\frac{\sum x_t}{T-1}\mathbf{q} + \sum\overrightarrow{qx}) \end{bmatrix}. \qquad \text{(E2)}$$

Using (E2), we can re-write the idempotent and symmetric matrix in the definition of $\mathcal{J}$ in (E1) as follows

$$\mathbf{I} - \mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{I} - \frac{\mathbf{S}^{-1/2}\left(-\frac{\sum x_t}{T-1}\mathbf{q} + \sum\overrightarrow{qx}\right)\left(-\frac{\sum x_t}{T-1}\mathbf{q}' + \sum\overrightarrow{qx}'\right)\mathbf{S}^{-1/2}}{\left(-\frac{\sum x_t}{T-1}\mathbf{q}' + \sum\overrightarrow{qx}'\right)\mathbf{S}^{-1}\left(-\frac{\sum x_t}{T-1}\mathbf{q} + \sum\overrightarrow{qx}\right)} \end{bmatrix}. \qquad \text{(E3)}$$

Next we need to find the probability limit of the matrix $\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}$. To that end, note that it holds that

$$\operatorname{plim}\frac{1}{T}\mathbf{S}=\operatorname{plim}\frac{1}{T}\mathbf{B}-\operatorname{plim}\frac{1}{T(T-1)}\mathbf{q}\mathbf{q}'=\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v. \tag{E4}$$

Thus, we find the following probability limit $\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}$

$$\begin{aligned}
&\xrightarrow{P}\mathbf{I}-\operatorname{plim}\frac{1}{(T-1)^{1/2}}\mathbf{L}'\mathbf{Q}'\mathbf{X}\operatorname{plim}\left(\frac{1}{T-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\operatorname{plim}\frac{1}{(T-1)^{1/2}}\mathbf{X}'\mathbf{Q}\mathbf{L}\\
&\xrightarrow{P}\mathbf{I}-\begin{bmatrix}1 & 0\\ \mathbf{0} & \left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\sigma_{VP}^2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}{\sigma_{VP}^4\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}\end{bmatrix}\begin{bmatrix}1 & \mathbf{0}'\\ 0 & \sigma_{VP}^2\boldsymbol{\rho}'\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\end{bmatrix}\\
&\xrightarrow{P}\begin{bmatrix}0 & \mathbf{0}'\\ \mathbf{0} & \mathbf{I}-\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}\boldsymbol{\rho}'\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\end{bmatrix},
\end{aligned}\tag{E5}$$

where we have used result (D6) above. Note that the lower-right submatrix

$$\mathbf{M}\equiv\mathbf{I}-\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}\boldsymbol{\rho}'\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2} \tag{E6}$$

in (E5) is a symmetric and idempotent matrix of size $K\times K$. It therefore holds that

$$\begin{aligned}
\operatorname{rank}(\mathbf{M})=\mathsf{tr}(\mathbf{M})&=\mathsf{tr}(\mathbf{I})-\mathsf{tr}\left(\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}\boldsymbol{\rho}'\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1/2}\right)\\
&=K-\mathsf{tr}\left(\frac{1+\sigma_{VP}^2\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}}\boldsymbol{\rho}'\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}'+\boldsymbol{\Sigma}_v\right)^{-1}\boldsymbol{\rho}\right)=K-1.
\end{aligned}\tag{E7}$$

The second part of the proof shows that $(\mathbf{L}'\mathbf{Q}'\mathbf{e})/\sqrt{(1-3/T)s^2}$ converges to a normal distribution. Note that

$$\mathbf{L}'\mathbf{Q}'\mathbf{e}=\begin{bmatrix}\frac{1}{(T-1)^{1/2}} & \mathbf{0}'\\ -\frac{1}{T-1}\mathbf{S}^{-1/2}\mathbf{q} & \mathbf{S}^{-1/2}\end{bmatrix}\begin{bmatrix}\sum e_{t+1}\\ \sum\overrightarrow{qe}\end{bmatrix}=\begin{bmatrix}\frac{1}{(T-1)^{1/2}}\sum e_{t+1}\\ -(\frac{1}{T}\mathbf{S})^{-1/2}\frac{1}{T^{1/2}}\mathbf{q}\frac{1}{T-1}\sum e_{t+1}+(\frac{1}{T}\mathbf{S})^{-1/2}\frac{1}{T^{1/2}}\sum\overrightarrow{qe}\end{bmatrix}. \tag{E8}$$

Only the lower-right $K\times K$ submatrix of $\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}$ is non-zero, as Eq. (E3) demonstrates. Thus the first scalar element of the vector in (E8), $\sum e_{t+1}/\sqrt{T-1}$, cancels out from $\mathcal{J}$ in (E1). We focus on the bottom K elements of the vector in (E8). The first part of the sum converges to zero; that is $-(\frac{1}{T}\mathbf{S})^{-1/2}\frac{1}{T^{1/2}}\mathbf{q}\frac{1}{T-1}\sum e_{t+1}\xrightarrow{P}0$. The second part has an asymptotic normal distribution by the CLT in Lemma 1. That is

$$\frac{1}{T^{1/2}}\sum\overrightarrow{qe}\xrightarrow{D}\mathcal{N}\left(\mathbf{0},\boldsymbol{\rho}\boldsymbol{\rho}'\beta^2\sum_{j=-\infty}^{\infty}\gamma_j^{(VP\times z)}+\boldsymbol{\rho}\boldsymbol{\rho}'\sigma_{VP}^2\sigma_\xi^2+\beta^2\sigma_P^2\boldsymbol{\Sigma}_v+\sigma_\xi^2\boldsymbol{\Sigma}_v\right), \tag{E9}$$

and hence it follows that

$$(\frac{1}{T}\mathbf{S})^{-1/2}\frac{1}{T^{1/2}}\sum \overrightarrow{q}\vec{e} \overset{D}{\to} \left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\mathbf{n} \tag{E10}$$

$$\sim\mathcal{N}\left(\mathbf{0}, \left(\sigma_\xi^2 + \beta^2\sigma_P^2\right)\mathbf{I} + \beta^2\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\boldsymbol{\rho}' 2\sum_{j=1}^{\infty}\gamma_j^{(VP\times z)}\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\right).$$

Finally, from Section D it follows that $\sqrt{(1-3/T)s^2} \overset{P}{\to} (\beta^2\sigma_P^2 + \sigma_\xi^2)^{1/2}$ for any value of β. Hence, the bottom K elements of the vector $(\mathbf{L}'\mathbf{Q}'\mathbf{e})/\sqrt{(1-3/T)s^2}$ converge to the normal distribution

$$\overset{D}{\to}\mathcal{N}\left(\mathbf{0}, \mathbf{I} + \frac{1}{\sigma_\xi^2 + \beta^2\sigma_P^2}\beta^2\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\boldsymbol{\rho}' 2\sum_{j=1}^{\infty}\gamma_j^{(VP\times z)}\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\right). \tag{E11}$$

$\mathcal{J}$ then is asymptotically distributed as follows

$$\mathcal{J} \overset{D}{\to} \mathbf{n}'\mathbf{M}\mathbf{n}, \tag{E12}$$

where $\mathbf{n}$ has a $K-$variate normal distribution with asymptotic variance matrix

$$\mathbf{A} \equiv \mathbf{I} + \frac{1}{\sigma_\xi^2 + \beta^2\sigma_P^2}\beta^2\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}\boldsymbol{\rho}\boldsymbol{\rho}' 2\sum_{j=1}^{\infty}\gamma_j^{(VP\times z)}\left(\sigma_{VP}^2\boldsymbol{\rho}\boldsymbol{\rho}' + \boldsymbol{\Sigma}_v\right)^{-1/2}. \tag{E13}$$

Let $\mathbf{m}$ be the K-dimensional random vector of standard normal distribution. Then $\mathbf{n} = (\mathbf{A}^{1/2})'\mathbf{m}$. It follows that

$$\mathcal{J} \overset{D}{\to} \mathbf{m}'\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2'}\mathbf{m}. \tag{E14}$$

Recall that $\mathbf{M}$ has rank $K-1$ as shown above. The matrix $\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2'}$ is symmetric and positive definite, and hence also has rank $K-1$. Following the arguments in Jagannathan and Wang (1996) we know that $\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2'}$ has $K-1$ positive eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_{K-1}$. There exists a diagonal $(K\times K)$ matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{K-1}, 0)$, and an orthogonal matrix $\mathbf{J}$, such that we can write

$$\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2'} = \mathbf{J}'\boldsymbol{\Lambda}\mathbf{J}. \tag{E15}$$

Finally, let $\mathbf{o} = \mathbf{J}\mathbf{m}$. Then $\mathbf{o}$ is standard normally distributed and hence it follows that

$$\mathcal{J} \overset{D}{\to} \mathbf{o}'\boldsymbol{\Lambda}\mathbf{o} = \sum_{j=1}^{K-1}\lambda_j\chi_j^2(1), \tag{E16}$$

where $\chi_j^2(1)$ are $K-1$ independent $\chi^2(1)$ distributed random variables.

The asymptotic distribution of $\mathcal{J}$ is unknown, but we can simulated p-values as suggested by Jagannathan and Wang (1996) once the eigenvalues λ_j are estimated. To that end, we require a consistent estimator of $\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{1/2'}$. Above, we show that the lower right $(K\times K)$ submatrix of $\mathbf{I} - \mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}$ is consistent for $\mathbf{M}$. An estimate for $A^{1/2}$ is the upper triangular matrix following from a Cholesky decomposition of $\frac{1}{s^2}(\frac{1}{T}S)^{-1/2}\hat{\boldsymbol{\Omega}}(\frac{1}{T}S)^{-1/2'}$, where $\hat{\boldsymbol{\Omega}}$ is a consistent estimator of size $(K\times K)$ for the asymptotic variance in (E9).

We can compute $\hat{\boldsymbol{\Omega}}$ as the lower right $(K \times K)$ submatrix of the usual HAC estimator, i.e. of the $(K+1) \times (K+1)$ estimator $\hat{\hat{\boldsymbol{\Omega}}} = \sum_{j=-T}^{T} \kappa\left(\frac{j}{n(T)}\right) \hat{\boldsymbol{\Gamma}}_j$, where κ is the kernel and $n(T)$ is the bandwidth. We define the autocovariance estimates $\hat{\boldsymbol{\Gamma}}_j$ as

$$\hat{\boldsymbol{\Gamma}}_j = \frac{1}{T} \begin{bmatrix} 1 \\ q_{1,t} \\ \vdots \\ q_{K,t} \end{bmatrix} \begin{bmatrix} 1 & q_{1,t-j} & \dots & q_{K,t-j} \end{bmatrix} \hat{e}_{t+1}\hat{e}_{t+1-j}. \tag{E17}$$

Following the previous derivations above, it is then straightforward to show that the long-run variance in (E9) can simply be consistently estimated from the kernel estimator $\hat{\hat{\boldsymbol{\Omega}}}$, assuming that $\mathsf{E}(q_{k,t}q_{k,t-j}x_t x_{t-j})$ exists.

F Allowing ξ_{t+1} to be serially correlated

In this section we relax one of the assumptions of the DGP, and let $\xi_t = \psi(L)\mu_t$, where μ_t is i.i.d. with mean zero and constant variance. Let the coefficients of the moving average filter, ψ_i, be one-summable.

It is clear that this modification does not affect the representation of plim$(\frac{1}{T}\mathbf{Q}'\mathbf{Q})^{-1}$ and plim $\frac{1}{T}\mathbf{Q}'\mathbf{X}$ in (D1) and (D2). If $\beta \neq 0$, it further continues to hold that $\frac{1}{T}\mathbf{Q}'\mathbf{e} \xrightarrow{P} \mathbf{0}'$, and hence $\hat{\mathsf{b}}_{GMM} - \mathsf{b} \xrightarrow{P} 0$. Yet, if $\beta = 0$, we find that

$$\frac{1}{T^{1/2}}\mathbf{Q}'\mathbf{e} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} \sum_{j=-\infty}^{\infty} \gamma_j^{(\xi)} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\rho}\boldsymbol{\rho}' \sum_{j=-\infty}^{\infty} \gamma_j^{(\xi \times VP)} + \sigma_\xi^2 \boldsymbol{\Sigma}_v \end{bmatrix}\right), \tag{F1}$$

which implies that

$$T^{1/2}\left(\hat{\mathsf{b}}_{GMM} - \mathsf{b}\right) \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} \sum_{j=-\infty}^{\infty} \gamma_j^{(\xi)} & \mathbf{0}' \\ \mathbf{0} & \frac{\boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}\sum_{j=-\infty}^{\infty}\gamma_j^{(\xi \times VP)} + \sigma_\xi^2}{\sigma_{VP}^4 \boldsymbol{\rho}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\rho}} \end{bmatrix}\right). \tag{F2}$$

A consistent estimator for the asymptotic variance in (F2) is given by $\hat{\mathbf{H}} \equiv$

$$\left(\frac{1}{T}\mathbf{X}'\mathbf{Q}\left(\frac{1}{T}\mathbf{Q}'\mathbf{Q}\right)^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbf{Q}\left(\frac{1}{T}\mathbf{Q}'\mathbf{Q}\right)^{-1}\hat{\hat{\boldsymbol{\Omega}}}\left(\frac{1}{T}\mathbf{Q}'\mathbf{Q}\right)^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X}\left(\frac{1}{T}\mathbf{X}'\mathbf{Q}\left(\frac{1}{T}\mathbf{Q}'\mathbf{Q}\right)^{-1}\frac{1}{T}\mathbf{Q}'\mathbf{X}\right)^{-1}, \tag{F3}$$

where $\hat{\hat{\boldsymbol{\Omega}}}$ is the consistent HAC estimator in Appendix E. Replace the t-statistic in Appendix D.3 for the slope by the robust t-statistic: $t_b = T^{1/2}\boldsymbol{\iota}_2'\hat{\mathsf{b}}_{GMM}(\boldsymbol{\iota}_{(2)}'\hat{\mathbf{H}}\boldsymbol{\iota}_{(2)})^{-1/2}$. Then, under the null hypothesis that $\beta = 0$, this robust statistic converges to a standard normal distribution. Note that if further continues to holds that $s^2 \xrightarrow{P} \beta^2\sigma_P^2 + \sigma_\xi^2$.

The serial correlation in ξ_t affects only the asymptotic variance of $\frac{1}{T^{1/2}}\sum \overrightarrow{qe}$ in the derivation of the large-sample behavior of the $\mathcal{J}-$statistic in Appendix E. In particular, (E9) becomes

$$\frac{1}{T^{1/2}}\sum \overrightarrow{q\hat{e}} \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\rho}\boldsymbol{\rho}'\beta^2 \sum_{j=-\infty}^{\infty} \gamma_j^{(VP\times z)} + \boldsymbol{\rho}\boldsymbol{\rho}' \sum_{j=-\infty}^{\infty} \gamma_j^{(VP\times\xi)} + \beta^2\sigma_P^2\boldsymbol{\Sigma}_v + \sigma_\xi^2\boldsymbol{\Sigma}_v\right), \tag{F4}$$

for which $\hat{\boldsymbol{\Omega}}$ remains a consistent estimator. The asymptotic distribution of $\mathcal{J}$ continues to be the sum of $K-1$ weighted $\chi^2(1)$ variables.

References

Andersen, T. B., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics* **6**(1): 43–76.

Andersen, T. G. and Bollerslev, T. (1997). Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *Journal of Finance* **52**(3): 975–1005.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility, *Review of Economics and Statistics* **89**(4): 701–720.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **74**(1): 3–30.

Bali, T. G. and Peng, L. (2006). Is there a Risk-Return Trade-Off? Evidence from High-Frequency Data, *Journal of Applied Econometrics* **21**(8): 1169–1198.

Bali, T. G. and Zhou, H. (2016). Risk, Uncertainty, and Expected Returns, *Journal of Financial and Quantitative Analysis* **51**(3): 707–735.

Bandi, F. M. and Perron, B. (2006). Long Memory and the Relation Between Implied and Realized Volatility, *Journal of Financial Econometrics* **4**(4): 636–670.

Banerjee, A., Dolado, J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data (Advanced Texts in Econometrics*, Oxford University Press.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating Quadratic Variation Using Realised Variance, *Journal of Applied Econometrics* **17**: 457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics* **2**(1): 1–37.

Bekaert, G. and Engstrom, E. (2010). Asset Return Dynamics under Bad Environment-Good Environment Fundamentals. Forthcoming in: *Journal of Political Economy*.

Bekaert, G. and Hoerova, M. (2014). The VIX, the Variance Premium and Stock Market Volatility, *Journal of Econometrics* **183**(2): 181–192.

Bekaert, G. and Hoerova, M. (2016). What do asset prices have to say about risk appetite and uncertainty?, *Journal of Banking and Finance* **67**: 103–118.

Bekaert, G., Hoerova, M. and Lo Duca, M. (2013). Risk, Uncertainty and Monetary Policy, *Journal of Monetary Economics* **60**(7): 771–788.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, USA.

Binsbergen, J. H. and Koijen, R. S. J. (2010). Predictive Regressions: A Present-Value Approach, *Journal of Finance* **65**(4): 1439–1471.

Bloom, N. (2009). The Impact of Uncertainty Shocks, *Econometrica* **77**(3): 623–685.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH Modeling in Finance, *Journal of Econometrics* **52**: 5–59.

Bollerslev, T., Marrone, J., Xu, L. and Zhou, H. (2014). Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence, *Journal of Financial and Quantitative Analysis* **49**(3): 633–661.

Bollerslev, T., Osterrieder, D., Sizova, N. and 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. and Sizova, N. (2012). Volatility in Equilibrium: Asymmetries and Dynamic Dependencies, *Review of Finance* **16**(73): 31–80.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* **22**(11): 4463–4492.

Bollerslev, T. and Todorov, V. (2011). Tails, Fears, and Risk Premia, *Journal of Finance* **66**(6): 2165–2211.

Bollerslev, T. and Zhou, H. (2006). Volatility Puzzles: a Simple Framework for Gauging Return-Volatility Regressions, *Journal of Econometrics* **13**: 123–150.

Brandt, M. W. and Kang, Q. (2004). On the Relationship between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach, *Journal of Financial Economics* **72**(2): 217–257.

Breen, W., Glosten, L. R. and Jagannathan, R. (1989). Economic Significance of Predictable Variations in Stock Index Returns, *Journal of Finance* **44**: 1177–1189.

Campbell, J. Y. (1987). Stock returns andthe term structure, *Journal of Financial Economics* **18**(2): 373–399.

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

Campbell, J. Y., Lo, A. W. and MacKinlay, C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey, USA.

Campbell, J. Y. and Yogo, M. (2006). Efficient Tests of Stock Return Predictability, *Journal of Financial Economics* **81**: 27–60.

Camponovo, L., Scaillet, O. and Trojani, F. (2012). Predictive Regression and Robust Hypothesis Testing: Predictability Hidden by Anomalous Observations. SSRN Working Paper.

Carlini, F. and Santucci de Magistris, P. (2013). On the Identification of Fractionally Cointegrated VAR Models with the F(d) Condition. CREATES Research Paper **2013-44**.

Cavanagh, C., Elliott, G. and Stock, J. (1995). Inference in Models with Nearly Integrated Regressors, *Econometric Theory* **11**(5): 1131–1147.

Christensen, B. J. and Nielsen, M. O. (2006). Asymptotic Normality of Narrow-Band Least Squares in the Stationary Fractional Cointegration Model and Volatility Forecasting, *Journal of Econometrics* **133**: 343–371.

Christensen, B. J. and Nielsen, M. O. (2007). The Effect of Long Memory in Volatility on Stock Market Fluctuations, *Review of Economics and Statistics* **89**(4): 684–700.

Comte, F. and Renault, E. (1998). Long Memory in Continuous-Time Stochastic Volatility Models, *Mathematical Finance* **8**(4): 291–323.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics* **7**(2): 174–196.

Deng, A. (2014). Understanding Spurious Regression in Financial Economics, *Journal of Financial Econometrics* **12**(1): 122–150.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* **1**: 83–106.

Drechsler, I. and Yaron, A. (2011). Whats Vol Got to Do with It, *Review of Financial Studies* **24**(1): 1–45.

Du, J. and Kapadia, N. (2012). Tail and Volatility Indices from Option Prices. Working Paper, University of Massachusetts, Amhurst.

Eraker, B. and Wang, J. (2015). A Non-linear Dynamic Model of the Variance Risk Premium, *Journal of Econometrics* **187**(2): 547–556.

Ferson, W. E., Sarkissian, S. and Simin, T. T. (2003). Spurious Regressions in Financial Economics?, *Journal of Finance* **58**(4): 1393–1414.

Froot, K. A. and O'Connell, P. G. J. (2003). The Risk Tolerance of International Investors, *NBER Working Paper Series* **10157**.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return trade-off after all, *Journal of Financial Economics* **76**(3): 509–548.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**(5): 1779–1801.

Goyal, A. and Santa-Clara, P. (2003). Idiosyncratic Risk Matters!, *Journal of Finance* **58**(3): 975–1007.

Guo, H. and Whitelaw, R. F. (2006). Uncovering the RiskReturn Relation in the Stock Market, *Journal of Finance* **61**(3): 1433–1463.

Han, B. and Zhou, Y. (2011). Variance Risk Premium and Cross-Section of Stock Returns. Working Paper, Shanghai University of Finance and Economics.

Harrison, P. and Zhang, H. H. (1999). An investigation of the risk and return relation at long horizons, *Review of Economics and Statistics* **81**(3): 399–408.

Harvey, C. R. (1989). Time-varying Conditional Covariances in Tests of Asset Pricing Models, *Journal of Financial Economics* **24**(2): 289–317.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Huang, X. and Tauchen, G. (2005). The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics* **3**: 456–499.

Illing, M. and Aaron, M. (2005). A Brief Survey of Risk-Appetite Indexes, *Bank of Canada Financial System Review* pp. 37–43.

Jagannathan, R. and Wang, Z. (1996). The Conditional CAPM and the Cross-Section of Expected Returns, *Journal of Finance* **1**(51): 3–53.

Johansen, S. (2008). A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes, *Econometric Theory* **24**: 651676.

Johansen, S. (2009). Representation of Cointegrated Autoregressive Processes with Application to Fractional Processes, *Econometric Reviews* **28**: 121–145.

Johansen, S. and Nielsen, M. O. (2012). Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model, *Econometrica* **80**(6): 2667–2732.

Jurado, K., Ludvigson, S. and Ng, S. (2015). Measuring Uncertainty, *American Economic Review* **105**(3): 1177–1216.

Kelly, B. and Jiang, H. (2014). Tail Risk and Asset Prices, *Review of Financial Studies* **27**(10): 2841–2871.

Kothari, S. P. and Shanken, J. (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time Series Analysis, *Journal of Financial Economics* **18**: 169–203.

Kumar, M. S. and Persaud, A. (2002). Pure Contagion and Investors' Shifting Risk Appetite: Analytical Issues and Empirical Evidence, *International Finance* **5**(3): 401–436.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics* **54**(1-3): 159–178.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* **56**(3): 815–849.

Lewellen, J. (2004). Predicting Returns with Financial Ratios, *Journal of Financial Economics* **74**: 209–235.

Ludvigson, S. and Ng, S. (2007). The Empirical Risk-Return Relation: A Factor Analysis Approach, *Journal of Financial Economics* **83**(1): 171–222.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836-2003, *Journal of Financial Economics* **85**(1): 123–150.

Maynard, A. and Phillips, P. C. B. (2001). Rethinking an Old Empirical Puzzle: Econometric Evidence on the Forward Discount Anomaly, *Journal of Applied Econometrics* **16**: 671–708.

Maynard, A., Smallwood, A. and Wohar, M. E. (2013). Long Memory Regressors and Predictive Testing: A Two-Stage Rebalancing Approach, *Econometric Reviews* **32**(3): 318–360.

Meddahi, N. (2002). A Theoretical Comparison Between Integrated and Realized Volatility, *Journal of Applied Econometrics* **17**: 479–508.

Miller, K. S. (1981). On the Inverse of the Sum of Matrices, *Mathematics Magazine* **54**(2): 67–72.

Nagel, S. (2013). Empirical Cross-Sectional Asset Pricing, *Annual Review of Financial Economics* **5**: 167–199.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* **59**: 347–370.

Nielsen, M. O. and Morin, L. (2012). FCVARmodel.m: A Matlab Software Package for Estimation and Testing in the Fractionally Cointegrated VAR Model. QED Working Paper **1273**.

Nielsen, M. O. and Shimotsu, K. (2007). Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach, *Journal of Econometrics* **141**(2): 574–59.

Pastor, L., Sinha, M. and Swaminathan, B. (2008). Estimating the Intertemporal RiskReturn Trade-Off Using the Implied Cost of Capital, *Journal of Finance* **63**(6): 2859–2897.

Pastor, L. and Stambaugh, R. F. (2009). Predictive Systems: Living with Imperfect Predictors, *Journal of Finance* **64**(4): 1583–1628.

Sargan, D. (1958). The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica* **26**: 393–415.

Scruggs, J. T. (1998). Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach, *Journal of Finance* **53**: 575–602.

Scruggs, J. T. and Glabadanidis, P. (2003). Risk Premia and the Dynamic Covariance between Stock and Bond Returns, *Journal of Financial and Quantitative Analysis* **38**(2): 295–316.

Shim, J. K. and Siegel, J. G. (2008). *Financial Management*, 3rd edn, Barron's Business Library, chapter 7 "Understanding Return and Risk", pp. 133–134.

Shimotsu, K. (2010). Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend, *Econometric Theory* **26**(2): 501–540.

Shimotsu, K. and Phillips, P. C. B. (2005). Exact Local Whittle Estimation of Fractional Integration, *Annals of Statistics* **33**: 1890–1933.

Stambaugh, R. F. (1986). Bias in Regressions with Lagged Stochastic Regressors, *University of Chicago, Graduate School of Business, Working Paper Series* **156**.

Stambaugh, R. F. (1999). Predictive Regressions, *Journal of Financial Economics* **54**(3): 375–421.

Stock, J. (1991). Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series, *Journal of Monetary Economics* **28**(3): 435–459.

Sun, Y. and Phillips, P. C. B. (2004). Understanding the Fisher Equation, *Journal of Applied Econometrics* **19**(7): 869–886.

Todorov, V. (2010). Variance Risk Premium Dynamics: The Role of Jumps, *Review of Financial Studies* **23**(1): 345–383.

Tsay, W. J. and Chung, C. F. (2000). The Spurious Regression of Fractionally Integrated Processes, *Journal of Econometrics* **96**(1): 155–182.

Turner, C. M., Startz, R. and Nelson, C. R. (1989). A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market, *Journal of Financial Economics* **25**: 3–22.

Vilkov, G. and Xiao, Y. (2013). Option-Implied Information and Predictability of Extreme Returns. SAFE Working Paper **5**.

Welch, I. and Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* **21**(4): 1455–1508.

Whaley, R. E. (2000). The Investor Fear Gauge, *Journal of Portfolio Management* **26**(3): 12–17.

White, H. (2002). *Asymptotic Theory for Econometricians*, Academic Press.

Whitelaw, R. (2000). Stock Market Risk and Return: An Equilibrium Approach, *Review of Financial Studies* **13**(3): 521–547.

Whitelaw, R. F. (1994). Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns, *Journal of Finance* **49**(2): 515–541.

Tables

Table 1: Summary Statistics

The table reports summary statistics of the three variance series, excess returns on the S&P 500, and the instruments (from Febuary 3, 2000 to May 28, 2014). All variance series are in squared percentage form and scaled by maturity. The statistics for excess returns are annualized percentages. $q_{1,t}$ denotes the *jump instrument* and $q_{2,t}$ is the *variance risk premium instrument*.

`Summary Statistics`

			Autocorrelation			
	Average	Std. Dev.	1	2	3	22
$r_t^{(e)}$	0.2447	20.4558	-0.0770	-0.0547	0.0233	0.0328
RV_t	25.1492	40.4750	0.9972	0.9920	0.9849	0.6989
BV_t	22.7706	37.8044	0.9971	0.9918	0.9844	0.6917
VIX_t^2	43.7903	47.3600	0.9697	0.9481	0.9337	0.7473
$q_{1,t}$	0.7000	1.5949	0.9742	0.9505	0.9194	0.3404
$q_{2,t}$	16.6596	20.4266	0.8378	0.7153	0.6320	0.0888

Table 2: Long-Memory Estimates

The upper panel of the table reports estimates of d using the multivariate EW estimator of Nielsen and Shimotsu (2007) for $Y_t = [RV_t,\ BV_t,\ VIX_t^2,\ r_t^{(e)}]'$. The size of the spectral window is set to $m = T^{0.35}$; the choice is based on a graphical analysis of the slope of the log periodograms as suggested by Beran (1994). $t_{d=0}$, $t_{d=0.5}$, and $t_{d=1}$ denote the respective t-statistics of element i of Y_t given by $2\sqrt{m}(\hat{d}_i - d)$. The lower panel of the table summarizes the t-statistics corresponding to the null hypothesis $d_i = d_j$ for $i \neq j$. Nielsen and Shimotsu (2007) define the t-statistic as

$$t_{d_i=d_j} = \frac{\sqrt{m}\left(\hat{d}_i - \hat{d}_j\right)}{\sqrt{\frac{1}{2}\left(1 - \frac{\hat{\iota}_{i,j}^2}{\hat{\iota}_{i,i}\hat{\iota}_{j,j}}\right) + h(T)}},$$

where $\hat{\iota}_{i,j} = \frac{1}{m}\sum_{l=1}^{m} \text{real}\left\{I(\lambda_l)\right\}$ and $I(\lambda_l)$ is the periodogram of a (4×1) vector with elements $\Delta^{\hat{d}_i} Y_{t,i}$ at frequency λ_l. $h(T)$ is a tuning parameter, which we set equal to $(\ln(T))^{-1.3}$. The resulting statistic $t_{d_i=d_j}$ should be compared to critical values from a standard normal distribution.

	Estimates for d			
	RV_t	BV_t	VIX_t^2	$r_t^{(e)}$
$\hat{d}$	0.3234	0.3170	0.3961	0.1107
$t_{d=0}$	2.5872	2.5359	3.1684	0.8854
$t_{d=0.5}$	-1.4128	-1.4641	-0.8316	-3.1146
$t_{d=1}$	-5.4128	-5.4641	-4.8316	-7.1146

	$t_{d_i=d_j}$ statistics with $h(T) = 0.0721$			
	RV_t	BV_t	VIX_t^2	$r_t^{(e)}$
RV_t	–	0.3134	-0.9900	2.1295
BV_t		–	-1.0559	2.0591
VIX_t^2			–	2.3276
$r_t^{(e)}$				–

Table 3: OLS and GMM Estimation Results

The table summarizes the estimation results when the misspecified unbalanced and endogenous regression (9) is evaluated by OLS and GMM, respectively. SE denotes the usual standard error of the estimates that is not robust to heteroskedasticity or serial correlation. HAC-SE reports standard errors based on HAC covariance estimation using a Bartlett kernel that are robust to serial correlation and heteroskedastictiy. $\mathcal{J}$ is Sargan's statistic from Corollary 1. The corresponding p-value is obtained from 200,000 simulations of independent $\chi^2(1)$ variables multiplied by the eigenvalue estimate 5.513.

	`OLS Regression of` (9)	`GMM Regressions of` (9)
$\hat{a}$	0.3520	0.3544
SE($\hat{a}$)	5.3775	5.5338
HAC-SE($\hat{a}$)	6.6770	19.2892
$\hat{b}$	0.2678	1.9292
SE($\hat{b}$)	0.1137	0.2883
HAC-SE($\hat{b}$)	0.0960	0.8894
$\mathcal{J}$-statistic		1.8536
p-value($\mathcal{J}$)		0.4380

Table 4: Pseudo Correlation Measure

The table reports the pseudo correlation between z_t and latent VP_t as defined in Equation (23). For z_t, we first consider several estimates for the variance premium $\hat{VP}_t$, which are discussed in Section 6.1. Next, we let z_t denote a number of commonly used indicators for economic uncertainty and risk aversion, all of which are described in Section 7. The latter data is obtained from the following sources:

- $EMEUI$: from `https://fred.stlouisfed.org/series/WLEMUINDXD`
- EPU: "EPUCNUSD Index" from Bloomberg
- CV_{CFNAI}: GARCH(1,1) prediction on "CFNAI" from `https://fred.stlouisfed.org/series/CFNAI`
- UC: "uc"-series from `http://mariehoerova.net/`
- $MUS^{(i)}$: "Macro Uncertainty Series" ($h = \{1, 3, 12\}$) from `http://www.columbia.edu/~sn2294/pub.html`
- $STLFSI$: from `https://fred.stlouisfed.org/series/STLFSI`
- $GFSI$: "GFSI Index" (BofA Merrill Lynch GFSI) from Bloomberg
- $SSICCONF$: "SSICCONF Index "from Bloomberg
- $RAECB$: from `http://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=280.RDF.D.U2.Z0Z.4F.EC.U2_GRAI.HST`
- $CSRAI$: "RAIIHRVU Index" ("CS Risk Appetite HOLT Relative Value USD Index") from Bloomberg
- $SCGRRAI$: "SCGRRAI Index "from Bloomberg
- $WPRAI$: "WRAIRISK Index" from Bloomberg
- $WPFSI$: "WRAISTRS Index" from Bloomberg
- RA: "ra"-series from `http://mariehoerova.net/`

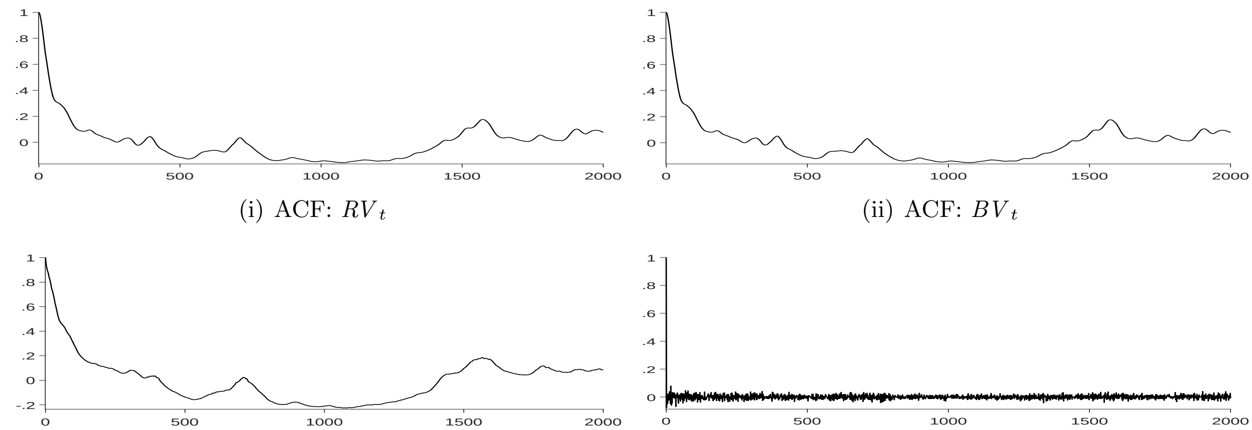
For all risk-aversion and economic-uncertainty series we merge the series with our daily data set by finding our date that is closest to the date stamp in the respective series. $MUS^{(i)}$ are the only series where an exact date stamp is missing; we match it with the observation in our daily data set that is closest to the 15th day of a month. We report 95% confidence intervals in brackets, obtained from 9999 block-bootstrap samples (the length of a block corresponds roughly to half a year for all series). To bootstrap the VIX-series, we first filter the series by $\hat{d} = 0.3961$, then create the bootstrap sample, and then apply the inverse filter to the new series. ***, **, * signify that the pseudo correlation is different from zero at a 1%, 5%, and 10% significance level, respectively.

z_t	$\widehat{P\mathsf{Corr}}(VP_t, z_t)$	T	Start	End
	`Different Estimates` $\hat{VP}_t$ `(see Section 6.1)`			
$\hat{VP}_t$: Martingale for RV	0.0513 ***	3622	2/3/2000	6/30/2014
	[0.0356, 0.1452]			
$\hat{VP}_t$: Martingale for BV	0.0496 ***	3622	2/3/2000	6/30/2014
	[0.0298, 0.1430]			
$\hat{VP}_t$: HAR-RV for RV	0.0343 ***	3621	2/3/2000	6/27/2014
	[0.0210, 0.1148]			
$\hat{VP}_t$: HAR-RV for BV	0.0325 ***	3621	2/3/2000	6/27/2014
	[0.0184, 0.1081]			
$\hat{VP}_t$: Drechsler & Yaron for RV	0.0345 ***	3622	2/3/2000	6/30/2014
	[0.0234, 0.1160]			
$\hat{VP}_t$: Drechsler & Yaron for BV	0.0328 ***	3622	2/3/2000	6/30/2014

	[0.0218, 0.1122]			
$\hat{VP}_t$: ARFIMA for RV	0.0296 ***	3621	2/3/2000	6/27/2014
	[0.0189, 0.1037]			
$\hat{VP}_t$: ARFIMA for BV	0.0286 ***	3621	2/3/2000	6/27/2014
	[0.0171, 0.1048]			
Different Popular Measures for Economic Uncertainty				
$EMEUI$	0.0137 ***	3622	2/3/2000	6/30/2014
	[0.0048, 0.0498]			
EPU	0.0145 **	3622	2/3/2000	6/30/2014
	[0.0004, 0.0464]			
CV_{CFNAI}	0.0174	174	2/3/2000	6/30/2014
	[-0.0318, 0.0638]			
UC	0.0282 *	128	2/3/2000	8/31/2010
	[-0.0039, 0.0866]			
$MUS^{(1)}$	0.0149 **	173	2/15/2000	6/16/2014
	[0.0020, 0.0610]			
$MUS^{(3)}$	0.0146 **	173	2/15/2000	6/16/2014
	[0.0018, 0.0564]			
$MUS^{(12)}$	0.0133 *	173	2/15/2000	6/16/2014
	[-0.0009, 0.0551]			
Different Popular Measures for Risk Aversion				
$STLFSI$	0.0095 **	752	2/4/2000	6/27/2014
	[0.0035, 0.0407]			
$GFSI$	0.0166 **	3622	2/3/2000	6/30/2014
	[0.0054, 0.0695]			
$SSICCONF$	-0.0037	174	2/3/2000	6/30/2014
	[-0.0401, 0.0281]			
$RAECB$	0.0197 ***	3622	2/3/2000	6/30/2014
	[0.0094, 0.0727]			
$CSRAI$	0.0007	3491	2/3/2000	6/30/2014
	[-0.0175, 0.0174]			
$SCGRRAI$	0.0020	2933	11/1/2002	6/30/2014
	[-0.0179, 0.0222]			
$WPRAI$	0.0087 ***	3502	2/3/2000	6/30/2014
	[0.0052, 0.0375]			
$WPFSI$	-0.0017	3502	2/3/2000	6/30/2014
	[-0.0174, 0.0138]			

RA	-0.0246	128	2/3/2000	8/31/2010
	[-0.0672, 0.1869]			

Figures



(i) ACF: RV_t (ii) ACF: BV_t

(iii) ACF: VIX_t^2 (iv) ACF: $r_t^{(e)}$

Figure 1: ACF estimates for the three variance series and returns - The figure plots the estimates of the autocorrelation of the realized variance, RV_t, the bipower variation, BV_t, the volatility index, VIX_t^2, and daily close to close excess log returns on the the S&P 500, $r_t^{(e)}$. The x-axis measures lags in daily units.

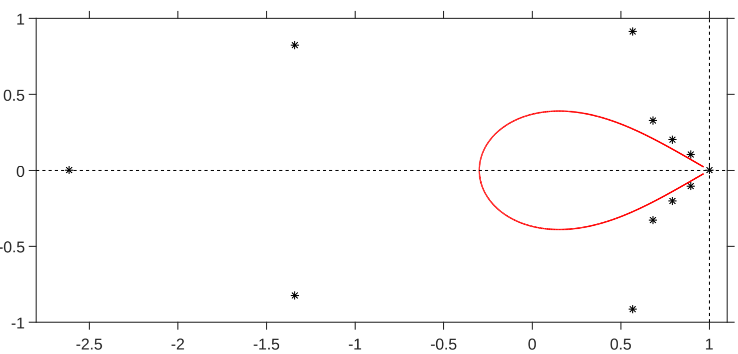
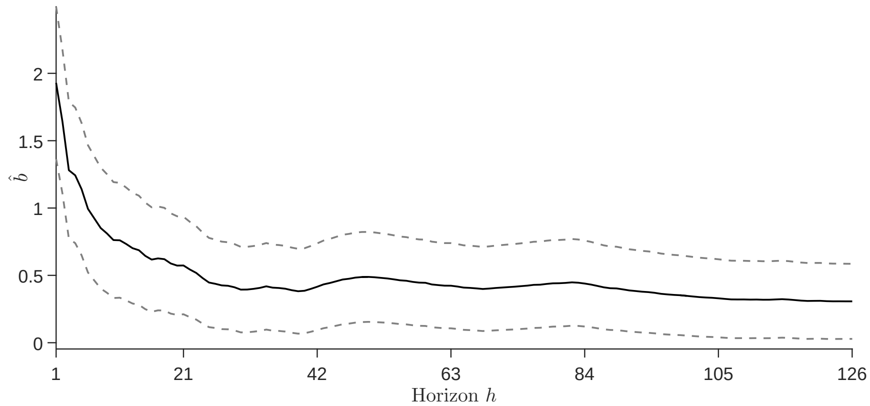
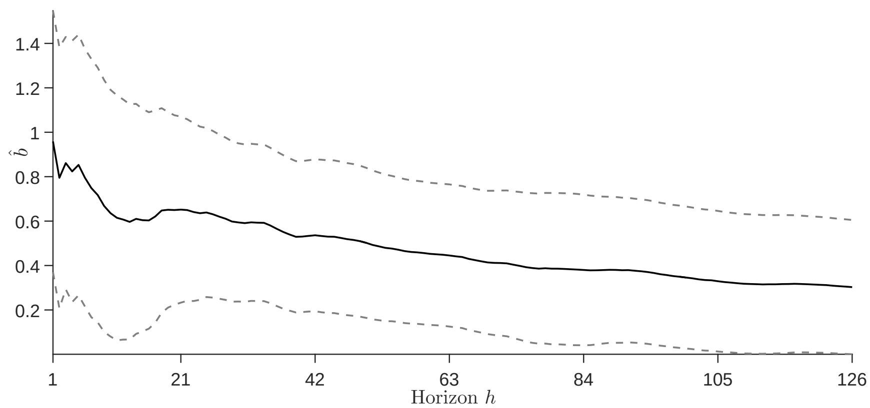


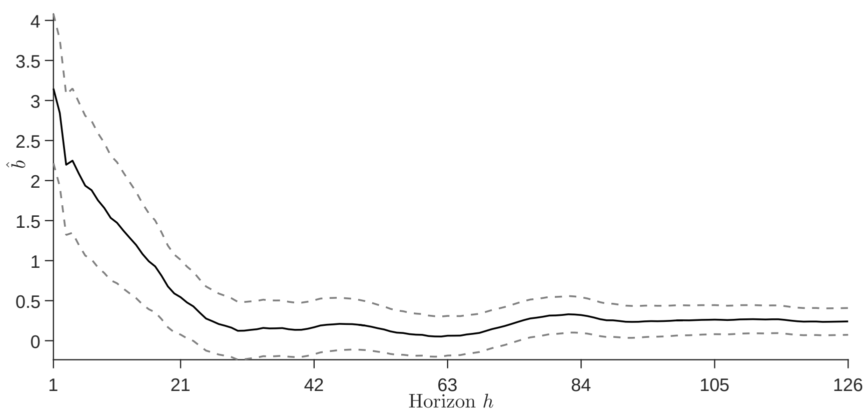
Figure 2: Roots of the characteristic polynomial of the co-fractional VAR - The figure plots the roots of the characteristic equation $|(1-c)I_{3\times3} - \varphi\theta' c - (1-c)\sum_{i=1}^{n}\Gamma_i c^i| = 0$, indicated by the black stars. The red line is the image of the complex disk $\mathbb{C}_d$, for $\hat{d} = 0.3775$. For $\theta'\tilde{X}_t$ to be $I(0)$, all roots must be equal to one or lie outside the disk.



(i) $\hat{b}_h$ - Entire Sample Period (2/3/2000-6/30/2014)



(ii) $\hat{b}_h$ - 'Normal' Periods (2/3/2000-2/26/2007 & 3/3/2009-6/30/2014)



(iii) $\hat{b}_h$ - Financial Crisis (2/27/2007-3/2/2009)

Figure 3: Estimated risk-return trade-off parameter - The figure plots the estimated risk-return trade-off parameter $\hat{b}_h$ over different horizons measured in days. We estimate the unbalanced misspecified and endogeneous predictive regression for cumulative returns by GMM, using the instruments q_t. The dashed lines represent 95% confidence intervals.

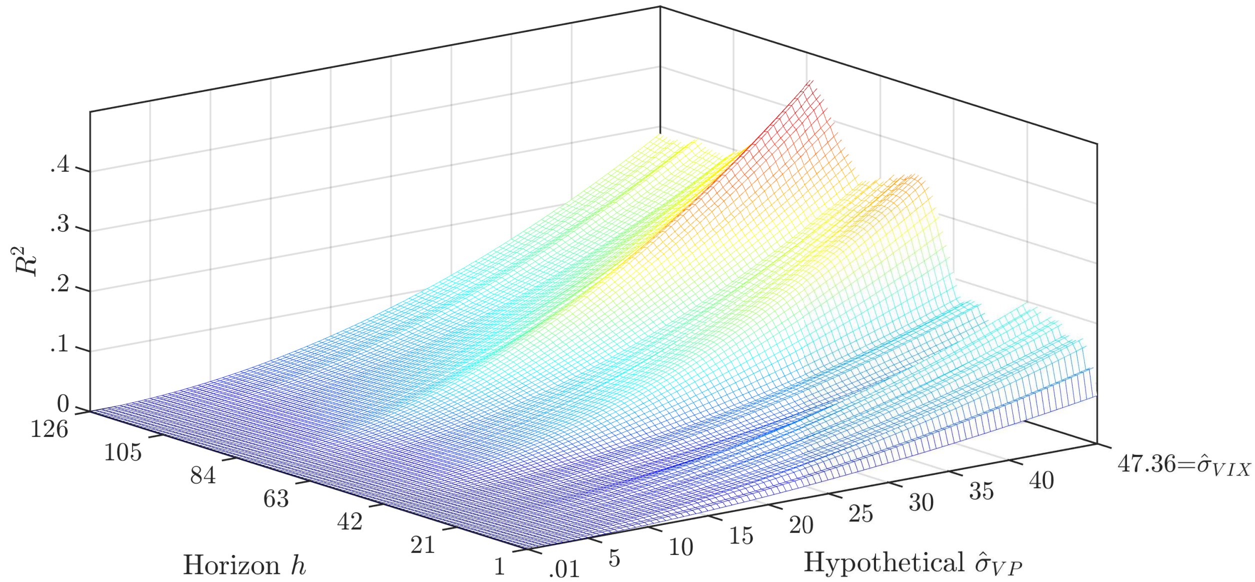


Figure 4: Predictive R^2 - The figure plots the implied predictability of the GMM regression over different horizons h. The R^2 changes with different hypothetical values considered for the sample standard deviation of the latent variance premium, $\hat{\sigma}_{VP}$.

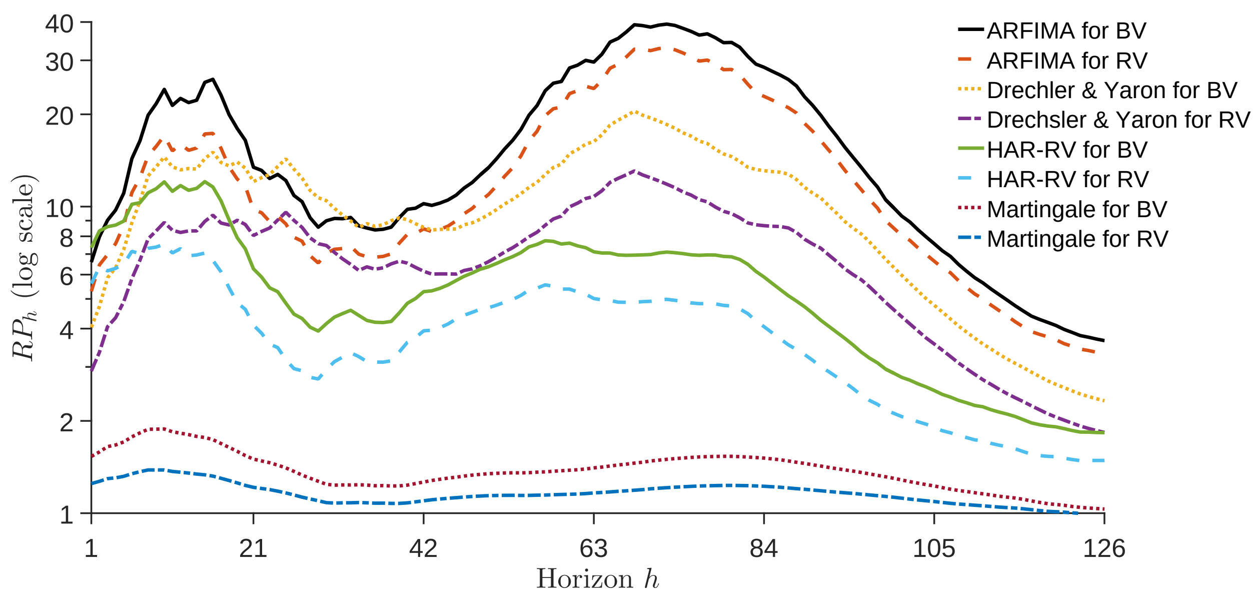


Figure 5: Relative Predictability RP_h - The figure plots the relative predictive R^2, RP_h, for different models. The numerator of the ratio is the squared slope estimate from the GMM regression. The denominator is the squared slope estimate from an OLS regression, where the latent VP_t is replaced by different estimates. The y-axis has a logarithmic scale.

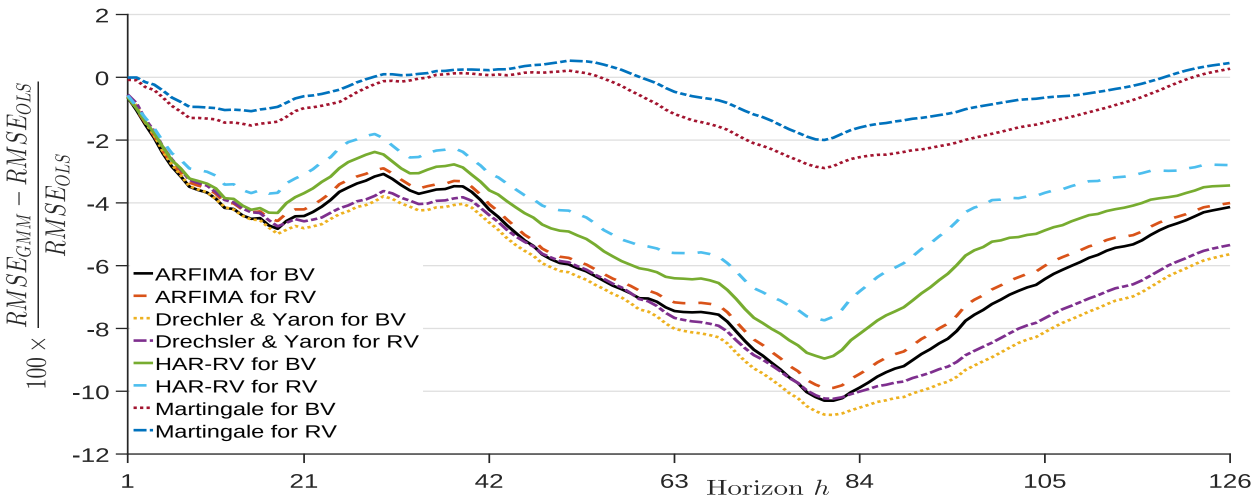


Figure 6: Percentage Difference in Out-of-Sample Forecasting Efficiency - The figure plots the percentage difference in OOS forecasting efficiency for different forecasting horizons. RMSE_{GMM} is the root mean squared error resulting from predictions using the proposed GMM approach and replacing the unknown $\mathsf{E}_t^P(IV_{t,t+\tau})$ by $\tilde{BV}_t$ in the forecast. RMSE_{OLS} is the same measure for forecasts from the OLS predictions using different estimates for the unobserved VP_t.

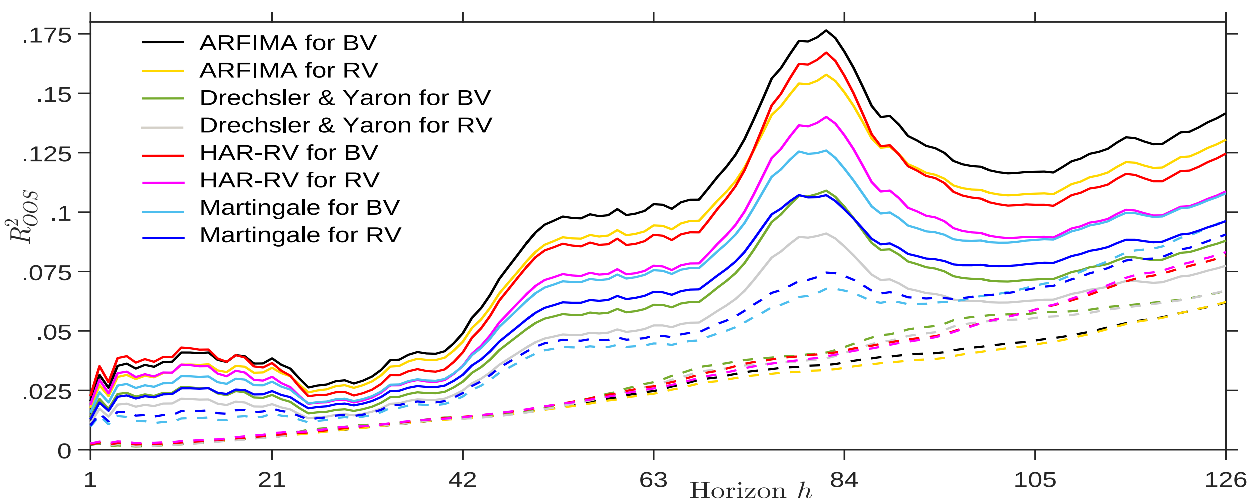


Figure 7: Out-of-Sample R-squared, R^2_{OOS} - The figure plots the OOS R-squared, $R^2_{OOS} = \mathsf{Var}(\hat{r}^{(e)}_{T_{IS}+h})/\mathsf{Var}(r^{(e)}_{T_{IS}+h})$. The solid lines refer to forecasts that use the proposed GMM approach for in-sample estimation; the out-of-sample prediction is made by multiplying the slope estimate by different estimates for VP_t. The dashed lines refer to forecasts that use the standard OLS approach for in-sample estimation; the out-of-sample prediction is made by multiplying the slope estimate by different estimates for VP_t.