Taylor, N. (2019). Forecasting returns in the VIX futures market.
International Journal of Forecasting, 35(4), 1193-1210.
<https://doi.org/10.1016/j.ijforecast.2019.01.009>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.ijforecast.2019.01.009](https://doi.org/10.1016/j.ijforecast.2019.01.009)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.ijforecast.2019.01.009> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Forecasting returns in the VIX futures market [☆]

Nick Taylor^a

^a*School of Economics, Finance and Management, University of Bristol, Bristol, BS8 1TN, UK.*

Abstract

This paper introduces a new forecasting model of VIX futures returns. The model is structural in nature, parsimonious and contains parameters that are relatively easy to estimate. Forecasts of next day VIX futures returns based on this model are superior to those produced by a linear forecasting model that uses the same set of predictors. Moreover, the profits to a market timing model based on the proposed forecasts are statistically and economically significant, and are robust to the method used to adjust for risk and to transaction costs (up to around 15 basis points). By contrast, the forecasts generated by the linear forecasting model are not.

Keywords: VIX futures, forecasting model, market timing, volatility, economic significance.

JEL: C10, C22, C53, C58, G11, G17.

[☆]The comments of the two anonymous referees and associate editor are greatly appreciated. I also thank participants of the Workshop on Recent Developments in Financial Data Science and Econometrics at Loughborough University, and those who attended my finance seminar at Kent Business School for helpful comments.

Email address: `nick.taylor@bristol.ac.uk` (Nick Taylor)

1. Introduction

Predictability in financial markets has received considerable attention in the academic literature. Early studies focused on forecasting returns to conventional assets (that is, individual stocks, foreign exchange, stock indices, bonds and futures contracts). The results of this effort have been somewhat mixed, with the benefits appearing only at best marginal; see, e.g., Welch and Goyal (2008) for a seminal example in the context of stock indices. Within the context of futures contracts the vast majority of studies have examined the predictability of stock index or commodity futures; see Hamilton and Wu (2015), Narayan et al. (2015), Sun et al. (2015), Mishra and Smyth (2016) and Tse (2018) for recent examples. By contrast, futures contracts on the volatility index (VIX) have received less attention in this regard; see Konstantinidi and Skiadopoulos (2011) and Psaradellis and Sermpinis (2016). This is somewhat surprising given the importance of such derivatives in managing risk in the aftermath of the 2007/08 financial crisis. We add to this growing literature by introducing a new forecasting model of VIX futures price changes and returns. The statistical and economic significance of this model are investigated by conducting an out-of-sample forecasting exercise, with the latter significance assessed via the performance of an associated trading strategy.

The key theoretical result of Samuelson (1965) asserts that futures returns should be unpredictable. That is, irrespective of the dynamics of the underlying asset, futures prices should follow a martingale process. This is primarily because futures prices are themselves risk-adjusted measures of the expected value of the underlying at maturity. Updates to these expectations should be unpredictable – otherwise the expectations themselves could be improved by incorporating past unexpected changes (innovations). A key feature of this framework is that changes in the futures price at time t depend upon market-driven innovations in the value of the underlying asset (also) at time t . This fundamental setup forms the basis of the forecasting model proposed in this paper. We innovate by allowing expectations (futures prices) to be misspecified such that innovations are not orthogonal to the (sophisticated) investors/forecasters information set. In doing this, a novel structural forecasting model of futures price changes and returns is proposed, and is subjected to various tests of statistical and economic significance.

The virtues of the proposed forecasting model are considered within the context of the VIX futures market. The underlying in this market is the Chicago Board Options Exchange (CBOE)

volatility index (viz., the VIX). This index measures the expected volatility associated with the S&P 500 index returns over the subsequent 30 days, as implied by the prices of a basket of options contracts (on the S&P 500 index) with maturities from 23 to 37 days. Given the nature of the underlying, VIX futures contracts are cash settled, with final settlement taking place on the Wednesday that is 30 days prior to the third Friday of the calendar month following the expiry month. The primary purpose of these contracts is to enable hedgers and speculators to trade volatility in a low cost (high liquidity) environment.

There have been previous studies of the determinants of VIX futures prices. The majority of these pricing models derive a (non-arbitrage) theoretical expression for the futures prices based on a volatility model of S&P 500 index returns.¹ Popular volatility models used in this literature are the Heston model (Zhu and Lian, 2012), the Duffie-Pan-Singleton stochastic volatility model with jumps (Zhang et al., 2010), the Heston-Nandi GARCH model (Wang et al., 2017), and the HAR-based realised volatility model (Huang et al., 2017). The approach taken in the current paper is different in that we focus on deriving a discrete-time econometric model of futures price changes and returns based on a Wold decomposition of the VIX. A key driver of predictability in this model is the market's failure to adequately form appropriate conditional expectations of the VIX when determining VIX futures prices. In doing this it is possible to demonstrate that this model is capable of generating forecasts of futures returns that are statistically and economically significant even in the presence of moderate transaction costs. By contrast, a competing (and analogous) linear forecasting model is not capable of producing forecasts that are significant in such conditions.

The rest of the paper is organised as follows. The next section describes the proposed model. Section 3 provides an outline of the measures and tests used to examine the statistical significance of the proposed and competing models. Section 4 contains a description of the investment strategy used to test the economic significance of these models. Section 5 contains the application and testing of the models within the context of an out-of-sample forecasting exercise. The final section concludes.

2. Model Development

Let V_t be the value of the underlying at time t , where $t = 1, 2, \dots, T$. From Samuelson (1965), the *Axiom of Present-Discounted Expected Value* states that the futures price at time t (with λ

compensation for risk) for delivery at $t + h$ is given by

$$F_{t,t+h} = \lambda^{-h} \mathbb{E}[V_{t+h} | \mathcal{G}_t], \quad (1)$$

where $\mathcal{G}_t$ represents the information set of the representative market participant. Wold's decomposition can be applied to V_{t+h} such that it has the following infinite moving average, that is $\text{MA}(\infty)$, representation:

$$V_{t+h} = \mu_{t+h} + \sum_{j=0}^{\infty} \theta_j \epsilon_{t+h-j}, \quad (2)$$

where μ_{t+h} is a (linearly) deterministic series, θ_j is the j th MA coefficient associated with the underlying asset price (with $\theta_0 = 1$), and ϵ_t is the unexpected change in V_t , that is, the innovation.² A key assumption here is that the representative market participant assumes that future innovations are unpredictable such that

$$\mathbb{E}[\epsilon_{t+j} | \mathcal{G}_t] = 0, \quad (3)$$

where $j = 1, \dots, h$.

Substituting (2) into (1) and using (3) we obtain

$$F_{t,t+h} = \lambda^{-h} \mathbb{E}[V_{t+h} | \mathcal{G}_t] = \lambda^{-h} \left(\mu_{t+h} + \sum_{j=h}^{\infty} \theta_j \epsilon_{t+h-j} \right), \quad (4a)$$

$$F_{t+1,t+h} = \lambda^{-(h-1)} \mathbb{E}[V_{t+h} | \mathcal{G}_{t+1}] = \lambda^{-(h-1)} \left(\mu_{t+h} + \sum_{j=h-1}^{\infty} \theta_j \epsilon_{t+h-j} \right). \quad (4b)$$

Multiplying (4a) by λ , and subtracting from (4b), leads to the following expression for price changes:

$$\Delta F_{t+1,t+h} = (\lambda - 1) F_{t,t+h} + \lambda^{-(h-1)} \theta_{h-1} \epsilon_{t+1}, \quad (5a)$$

with an expression for proportional returns obtained by dividing by $F_{t,t+h}$. Taking conditional expectations at t and using (3) gives

$$\mathbb{E}[\Delta F_{t+1,t+h} | \mathcal{G}_t] = (\lambda - 1) F_{t,t+h}. \quad (6)$$

Thus implying that the conditional expectations of returns are constant (and hence returns are

unpredictable).³

It is possible, however, that the innovation process is predictable in the hands of an investor who possesses a sufficiently rich information set. To allow for this possibility, this process is assumed to evolve as follows:

$$\epsilon_{t+1} = c + \sum_{i=1}^k \alpha_i Z_{i,t} + \xi_{t+1}, \quad (7)$$

where $Z_{i,t}$ is a set of k predictor variables, and ξ_t is a pure white noise process. Representative market participants are assumed to form conditional expectations without knowledge of this relationship, that is, they implicitly assume that $c = 0$ and $\alpha_i = 0 \forall i$. By contrast, informed (sophisticated) investors/forecasters have knowledge of this relationship – highlighted by denoting their information set by $\mathcal{G}_t^*$. It follows that their conditional expectations of price changes are given by

$$\mathbb{E}[\Delta F_{t+1,t+h} | \mathcal{G}_t^*] = (\lambda - 1)F_{t,t+h} + \lambda^{-(h-1)}\theta_{h-1}(c + \sum_{i=1}^k \alpha_i Z_{i,t}). \quad (8)$$

The presence of the lagged predictors in (8) provides the necessary condition for predictability. This expression, and the analogous proportional return expression (obtained by dividing by $F_{t,t+h}$), are henceforth referred to as the *price change* and *proportional return* versions of the *structural forecasting* (SF) model.

For the purposes of comparison we consider analogous versions of a *linear forecasting* (LF) model given by

$$\mathbb{E}[\Delta F_{t+1,t+h} | \mathcal{G}_t^*] = c + \sum_{i=1}^k \alpha_i Z_{i,t}. \quad (9)$$

Again, we also consider a proportional return version of this model.⁴ The SF and LF models differ in that the latter implicitly assumes that λ and θ_{h-1} equal unity. The non-linear SF model of futures price changes and proportional returns differs from the LF model of VIX futures prices employed by Konstantinidi and Skiadopoulos (2011), and the complex non-linear (heterogeneous autoregressive, neural network, and support vector machine) models of VIX futures prices considered by Psaradellis and Sermpinis (2016).

As it stands, the SF model contains a large number of parameters. To create a more parsimo-

nious representation a predefined structure is imposed on the MA coefficients. In particular, θ_h is generated under the assumption that the underlying follows a fractionally integrated white noise (FIWN) process of the form:

$$(1 - L)^d(V_t - \mu) = \epsilon_t, \quad (10)$$

where d is the fractional difference operator, and μ is a constant term. The Wold representation of this series leads to the following expression for the decay associated with θ_h :

$$\text{FIWN decay: } \theta_h = \frac{\Gamma[h + d]}{\Gamma[d]\Gamma[h + 1]} \forall h, \quad (11)$$

where $\Gamma[\cdot]$ represents the gamma function. To examine the validity of this assumption it is compared with the alternative assumption of a random walk (RW) decay structure imposed on the underlying dynamics. Specifically,

$$\text{RW decay: } \theta_h = 1 \forall h. \quad (12)$$

This assumption corresponds to the restriction that d equals unity. It can be seen that the SF model represents a less restricted version of the LF model in which the restriction that λ equals unity and the above RW structure is imposed in the LF model.⁵

3. Statistical Significance

To assess the quality of the forecasts associated with the above models we firstly consider a number of statistical measures and tests of comparative performance. Let the realised value of the series that is to be forecasted be Y_{t+1} (that is, price changes or proportional returns to the VIX futures), and the 1-step ahead forecast associated with the i th model be denoted $\hat{Y}_{i,t+1}$. Two statistical loss functions are used to measure accuracy,

$$\text{Squared error loss function: } L_{i,t+1} \equiv L_{t+1}(Y_{t+1}, \hat{Y}_{i,t+1}) = (Y_{t+1} - \hat{Y}_{i,t+1})^2. \quad (13a)$$

$$\text{Absolute error loss function: } L_{i,t+1} \equiv L_{t+1}(Y_{t+1}, \hat{Y}_{i,t+1}) = |Y_{t+1} - \hat{Y}_{i,t+1}|. \quad (13b)$$

The mean values associated with these loss function represent our measures of statistical forecast accuracy, viz., the root mean squared error (RMSE) and mean absolute error (MAE) measures. In

addition, we consider the R^2 coefficient of determination associated with a regression of realisations on forecasts (that is, a Mincer-Zarnowitz regression).

A key requirement of the current paper is to compare the performance of the proposed SF model in comparison to the LF model. Thus we require a statistical test of comparative performance. To this end, the difference between the mean losses is tested using the *unconditional predictive ability* test suggested by Giacomini and White (2006), henceforth the GW test. Formally, we test the following null hypothesis of equal forecast accuracy:

$$H_0 : E[\Delta L_{i,j,t+1}] = 0, \quad (14)$$

where $\Delta L_{i,j,t+1} = L_{i,t+1} - L_{j,t+1}$ is the value of the loss function associated with forecasting model i minus the value of the loss function associated with forecasting model j . The null is tested using the following test statistic:

$$\mathcal{T}_{i,j} = \frac{\Delta \bar{L}_{i,j}}{\hat{\sigma}/\sqrt{n}}, \quad (15)$$

where $\Delta \bar{L}_{i,j}$ is the sample mean of $\Delta L_{i,j,t+1}$ and $\hat{\sigma}^2$ is a suitable HAC estimator of the asymptotic variance $\hat{\sigma}^2 = \text{var}[\sqrt{n}\Delta \bar{L}_{i,j}]$. Giacomini and White (2006) show that this test statistic is asymptotically Gaussian under a variety conditions (including rolling window estimation). We use the Newey-West estimator of the asymptotic variance.

As an alternative to the above pairwise testing, we also consider testing the null hypothesis that a set of models (for instance, all linear models) do not have superior predictive abilities to those based on a single benchmark model. The alternative hypothesis is that the former do have superior abilities. Formally, we use White's (2000) reality check approach to the null given by

$$H_0 : \max_{k=1,\dots,K} \{E[\Delta L_{0,k,t+1}]\} \leq 0, \quad (16)$$

where $\Delta L_{0,k,t+1} = L_{0,t+1} - L_{k,t+1}$, $L_{0,t+1}$ is the forecast loss associated with the benchmark model, and $L_{k,t+1}$ is the forecast loss associated with the k th competitor model within a particular set of models (denoted $\mathcal{S}$). The null hypothesis is that no model within set $\mathcal{S}$ outperforms the benchmark model (for a given loss function), against the alternative that at least one model within set $\mathcal{S}$

outperforms the benchmark model. The test statistic is given by

$$\mathcal{T}_{0,\mathcal{S}} = \max_{k=1,\dots,K} \{\sqrt{n} \times \Delta \bar{L}_{0,k}\}, \quad (17)$$

where $\Delta \bar{L}_{0,k}$ is the sample mean of $\Delta L_{0,k,t+1}$. We avoid a data snooping bias by using the block bootstrap procedure proposed by White (2000) to generate the distribution of this statistic under the null in order to test the above hypothesis.⁶

4. Economic Significance

The economic significance of the SF model (and the LF model) derived in section 2 is examined within the context of a simple market timing strategy. In particular, at each point in time the investor must decide whether to take a short position, a long position, or to stay out of the market (that is, no position). This decision is based on the conditional expectation of next period returns as described in (8), and so is myopic in nature. Let w_t denote this position such that $w_t = -1$ represents a short position, $w_t = 0$ represents no position, and $w_t = 1$ represents a long position, all of which take place at time t . The trading strategy return net of transaction costs is given by

$$f(R_{t+1}, w_t, \tau_t) = \frac{1 + w_t R_{t+1}}{1 + \tau_t}, \quad (18)$$

where R_{t+1} is the return to the futures contract at time t (with the contract maturity subscript notation dropped for convenience), $\tau_t = \kappa |\Delta w_t| = \kappa |w_t - w_{t-1}|$ is the total cost of trading, and κ is the proportional transaction cost associated with trading the futures contract. In the subsequent analysis, proportional transaction costs of up to 40 basis points are considered.

The strategy return depends, inter alia, on the position taken and the transaction cost incurred, which in turn depends on the difference in positions at $t - 1$ and t . A long position is taken if it (that is, the long position) generates a conditional expectation of strategy returns that is greater than the conditional expectation of strategy returns associated with no position or a short position. Similarly, no position or a short position is taken if it delivers the highest conditional expectation of strategies over those associated with the other positions. The presence of transaction costs mean that the optimisation problem seeks the highest conditional expectation of strategy returns

conditional on transaction costs. As these costs depend on the change in the weight, then this information should be considered when choosing the candidate set of strategy return conditional expectations which are to be maximised. For instance, if a long position is taken at $t - 1$ (that is, $w_{t-1} = 1$), then the transaction cost will be zero if a long position is selected at t , will equal κ if no position is selected at t , and 2κ if a short position is selected at t . In this case, the optimisation problem would be:

$$\max_{w_t \in \{-1, 0, 1\}} \{E[f(R_{t+1}, 1, 0)|\mathcal{G}_t^*], E[f(R_{t+1}, 0, \kappa)|\mathcal{G}_t^*], E[f(R_{t+1}, -1, 2\kappa)|\mathcal{G}_t^*]\}.$$

Similarly reasoning applies when a short or no position is taken at $t - 1$, with three strategy return conditional expectations in each candidate set.⁷ It follows that a total of nine conditional expectations are required at each point in time. As this approach explicitly incorporates transaction costs into the optimisation problem then it is referred to as the τ -informed strategy. Alternatively, if the investor does not incorporate transaction costs into the optimisation process (equivalently, incorrectly assumes that they equal zero) then the approach is referred to as the τ -ignorant strategy.

The performance of the above τ -ignorant and τ -informed strategies is assessed with respect to the following three different benchmark strategies. The *cash only strategy* assumes that no position is taken (that is, $w_t = 0 \forall t$); the *long only strategy* represents the buy and hold strategy (that is, $w_t = 1 \forall t$); and the *short only strategy* assumes that the position is continuously short sold (that is, $w_t = -1 \forall t$). The relative returns to the proposed trading strategy with respect to these timing strategies are given by

$$\text{Trading strategy w.r.t. cash only strategy: } R_{t+1}^c \equiv f(R_{t+1}, w_t^*, \tau_t^*) - f(R_{t+1}, 0, 0), \quad (19a)$$

$$\text{Trading strategy w.r.t. long only strategy: } R_{t+1}^l \equiv f(R_{t+1}, w_t^*, \tau_t^*) - f(R_{t+1}, 1, 0), \quad (19b)$$

$$\text{Trading strategy w.r.t. short only strategy: } R_{t+1}^s \equiv f(R_{t+1}, w_t^*, \tau_t^*) - f(R_{t+1}, -1, 0), \quad (19c)$$

where w_t^* denotes the weights derived under the τ -ignorant and τ -informed objective functions, and τ_t^* are the transaction costs associated with use of these weights.

To adjust for risk, Sharpe ratios associated with the above strategies are calculated. The statistical significance of this ratio is assessed by performing the two-sample test proposed by

Memmel (2003). In addition, to allow for alternative measures of risk, a Fama-French factor-based approach is adopted. Specifically, Jensen’s alpha is used in the following factor regression:

$$R_{t+1}^i = \alpha + \sum_{j=1}^M \beta_j X_{j,t} + v_t, \quad (20)$$

where α represents the mean risk-adjusted return to the i th strategy, β_j is the j th factor sensitivity, $X_{j,t}$ is the j th factor, and v_t is a suitably defined error term. This regression is estimated using ordinary least squares (OLS). It follows that performance can be subjected to statistical inference by applying t -tests to the constant term. In order to examine the sensitivity of the results to the test equation specification, various combinations of the five factors proposed by Fama and French (2015) are considered. The details of the factors used are described in the next section.

5. Results

This section contains the results associated with an assessment of the statistical and economic significance of the SF and LF models.

5.1. Data

VIX futures contracts were introduced by the CBOE in 2004. The daily closing prices of all of these futures contracts (CBOE S&P 500 Volatility Index with ticker VX) traded over the period from January 2004 to December 2017 were downloaded directly from the CBOE futures exchange (<http://cfe.cboe.com/market-data/historical-data>). A single ‘joined’ price series is constructed under the assumption that the next-to-mature (front) contract is rolled over five days prior to maturity.⁸ This continuous price series accounts for the roll yield in the series. Doing this ensures that price differences around roll dates are not based on different contracts being traded (and hence they do not contain discontinuities caused by the contango often observed in VIX futures), and more accurately reflect the profit and loss of a futures investment.⁹ Use of the front contract ensures that there is sufficient liquidity available, while the rollover date used ensures that the prices do not reflect periods of low liquidity specifically in the days just prior to contract maturity. In addition to above data, levels of the S&P 500 index and prices of the SPDR S&P 500 ETF (ticker SPY) are used. These prices are obtained using the CRSP database accessed via the WRDS

interface. The factors used to adjust returns for risk are obtained via the Kenneth French data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). In total, five different factors are considered: the market portfolio, the small-minus-big (SMB) size portfolio, the high-minus-low (HML) value portfolio, the robust-minus-weak (RMW) profitability portfolio, and the conservative-minus-aggressive (CMA) investment portfolio. See Fama and French (2015) for the virtues of using these factors. Finally, the underlying VIX index, prices of ten year US government bonds, and the one month Libor rate are downloaded from the Federal Bank of St. Louis FRED database (<https://fred.stlouisfed.org>).

5.2. *Summary statistics*

Summary statistics associated with various (related) series are provided in Table 1. A number of results are noteworthy. First, as expected, VIX futures prices are highly correlated with the VIX index. Second, as documented elsewhere the VIX futures market is in contango; see, e.g., Alexander and Korovilas (2012) for implications, and Eraker and Wu (2017) for explanations. This is evinced by the mean futures price being above the mean of the VIX. It can also be seen by the fact that VIX futures are characterised by a very large cost of carry. Indeed, the mean futures return is -49.503% per annum (cf. the mean S&P 500 index return is 8.177%). Consistent with this negative risk premium is the negative correlation between futures returns and market returns (as proxied by the S&P 500 index return) – a result that supports the popularity of VIX futures within a risk management context. Third, all series appear not to be normally distributed, with excess kurtosis the common characteristic. This non-normality is somewhat mitigated when the logarithmic transform is applied, though non-normality is still apparent.

Insert Table 1 here

Results associated with the (non)-stationarity properties of each series are also provided in Table 1. In particular, the augmented Dickey-Fuller (ADF) unit root test (Dickey and Fuller, 1979, 1981, Said and Fuller, 1984), and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for stationarity (Kwiatkowski et al., 1992) are performed.¹⁰ The latter test is particularly useful as it is shown to have the power to reject stationarity in the presence of a long memory alternative (Lee and Schmidt, 1996). The ADF test results indicate that the null of a unit root can be rejected for VIX

futures prices and the VIX. Interestingly, the KPSS test for stationarity is also rejected. Taken together these results imply that these series are long memory stationary series.

5.3. *The dynamics of the underlying*

As described in section 2, the assumed MA coefficients that describe the dynamics of the underlying (that is, the VIX index) can be summarised using a simple fractionally integrated model (see (10)). To examine the quality of this assumption, a general model is considered and its performance relative to the simple counterpart is examined. In particular, the following autoregressive fractionally integrated moving average (ARFIMA) model is estimated:

$$\Phi(L)(1-L)^d(V_t - \mu) = \Theta(L)\epsilon_t, \quad (21)$$

where $\Phi(L)$ and $\Theta(L)$ are autoregressive and moving average polynomials such that $\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$, and $\Theta(L) = 1 + \Theta_1 L + \dots + \Theta_p L^p$. See Ahoniemi (2006), Konstantinidi et al. (2008), and Psaradellis and Sermpinis (2016) for previous applications of this model to the VIX. The parameters in this model (and restrictions thereof) are estimated using maximum likelihood via the Constrained Maximum Likelihood MT application module in Gauss v17. For reasons of parsimony only a first-order ARFIMA model is considered. Two versions of the dependent variable are considered: the VIX index in its raw form and in its log form. Parameter estimates, associated standard errors, and measures of fit are provided in Table 2.

Insert Table 2 here

Restricting the model to the case where $d = 1$ and with no AR or MA coefficients, that is, the random walk model, appears to provide an inadequate representation of the dynamics of the VIX. In particular, the Akaike information criterion (AC), the Hannan-Quinn information criterion (HC), and the Bayesian information criterion (BC), all indicate that this version of the model provides the worst fit to the data. Adding AR and MA coefficients improves the fit. However, it is not until the restriction on the fractional difference parameter is lifted does the fit improve considerably; see Fernandes et al. (2014) for corroborating evidence of long memory dynamics in the VIX. Perhaps most noticeably it is the fractionally integrated (FI) model with no AR or MA coefficients that

provides the best fit according to the AC, HC and BC measures. Thus, it is reasonable to assume the simplified expression for the series of MA coefficients as given in (11).

It could be argued that the ARFIMA model is better suited to log transformed values of the VIX. The results in Table 2 contain the parameters of this version of the model, with measures of fit transformed such that they correspond to the raw form of the dependent variable.¹¹ Again, it is the FI model that provides the best fit to the data. Moreover, the log form version of the FI model provides a superior fit to the data than the raw form version of the FI model. However, the differences are small. Consequently, given the forecasting model derived in this paper makes use of the raw form version, the subsequent results are unlikely to be adversely affected from making this assumption. This is an appealing finding as the forecasting model becomes extremely complex if the log version is used to derive the forecasting model.

5.4. Estimating the forecasting model

The parameters in (8) are estimated using maximum likelihood estimation under the Gaussian error assumption. Six predictor variables (that is, six $Z_{i,t}$ variables) are considered. The first five of these correspond to the predictors used in Konstantinidi and Skiadopoulos (2011), viz. the change in the VIX futures price denoted $Z_{1,t}$; the log-return on the S&P 500 index denoted $Z_{2,t}$; the one month Libor rate (in log-differences) denoted $Z_{3,t}$; the slope of the yield curve (the difference between the ten year US government bond and the one month Libor rates) denoted $Z_{4,t}$; and the futures basis (the difference between the VIX and the VIX futures price) denoted $Z_{5,t}$. Following the extant literature we refer to these as the *economic predictors*.

In addition to these predictors, we use the exponentially weighted moving average (EWMA) estimate of stock market volatility given by

$$\hat{\sigma}_{t+1} = \sqrt{\varphi\sigma_t^2 + (1 - \varphi)R_{S,t}^2}, \quad (22)$$

where $R_{S,t}^2$ is the squared return to the SPDR S&P 500 ETF and represents the measure of realised variance. This measure coincides with the integrated generalised autoregressive conditional heteroscedasticity (IGARCH) model, albeit the IGARCH parameter φ is not estimated. The extant literature is followed such that $\varphi = 0.06$; see, e.g., the commonly employed Riskmetrics method-

ology (Mina and Xiao, 2001). This is denoted $Z_{6,t}$ and is henceforth referred to as the *volatility predictor*.

A number of versions of the proposed SF model are estimated. Two main versions of the model based on the MA coefficient structure are considered: viz. the FIWN and RW decay structures described in (11) and (12), respectively. In addition, various restrictions on these models with respect to the coefficients on the predictors (that is, α_i) are imposed. In particular, for each decay structure assumption we consider four different predictor sets: no predictors (except the intercept); the economic predictors only; the volatility predictor only; and all predictors. To keep the FIWN decay as simple as possible we assume that $d = 1/2$. In addition, we consider four versions of the LF model using these same sets of predictors. Parameter estimates (and standard errors) associated with these LF and SF models of futures price changes and proportional returns estimated over the full sample period are provided in Table 3 and 4, respectively. In addition, measures of model fit (the adjusted R^2 , AC, HC and BC measures) and likelihood ratio (LR) tests of the restriction that λ equals unity are included.

Insert Table 3 & 4 here

A number of interesting features of the dynamics of VIX futures price changes and returns are highlighted by the results. First, the LF almost always provides an inferior fit to the data in comparison to its SF counterpart. Second, when the discount parameter λ is freely estimated it is always less than unity. Moreover, it is significantly less than unity for the majority of models (particularly those that have the best fit to the data). This implies that the VIX futures market is generally in contango – a result commonly found in the literature; see, e.g., Alexander and Korovilas (2012) for implications, and Eraker and Wu (2017) for explanations. Third, of the two decay structures it is the FIWN decay structure that provides the best fit to the data (according to the AC, HC and BC measures). Fourth, the coefficients on the change in the futures price, the basis and the volatility predictors that are significant in the SF models. Of these, the volatility predictor appears strongest as the SF model (with FIWN decay) that includes this predictor has superior fit (based on the BC) in comparison to the counterpart models based only on the economic predictors.¹² As one would expect, it has a positive coefficient indicating that higher (lower) volatility has a positive (negative) impact on VIX futures returns.

5.5. Forecast accuracy

The SF model with FIWN decay provides a good (in-sample) representation of the data. Moreover, the restriction that λ equals unity is rejected under most variants of the SF model. As this restriction is inherent in the LF model, these results cast doubt on the usefulness of the LF model. To examine this in further detail, the out-of-sample performance of the LF and SF models is examined, with various versions of the models associated with futures price changes and proportional returns estimated using a 1000-day rolling window of observations. Starting with the 1000-day estimation sample from January 1, 2004 to December 31, 2007 this scheme adds one observation to the end of the estimation sample and removes one observation from the start of the sample until the December 31, 2017 end observation is reached. After each estimation of the models, out-of-sample 1-step ahead forecasts are constructed for each day in the sample January 1, 2008 to December 31, 2017. To ensure a level playing field, the same predictors are used in both the LF and SF models. In doing this we can assess the importance of model specification on forecast quality.

The RMSE, MAE and R^2 statistics from the Mincer-Zarnowitz regression are provided in Table 5. The results are consistent with the in-sample results. The SF model with FIWN decay based on the volatility predictor generally delivers the most accurate forecasts of price changes and proportional returns. It is also noticeable that the LF model with no predictors (equivalently the random walk model) performs remarkable well, often delivering more accurate forecasts than the models with predictors. Indeed, it is only the SF model with FIWN decay based on the volatility predictor that is consistently superior.

Insert Table 5 here

The relative accuracy of the SF versus LF forecasts is assessed using the RMSE and MAE loss functions, with the difference between the mean losses tested using the GW test described in section 3. The results associated with forecasts of price changes and proportional returns generated by the LF and SF models are provided in Tables 6 and 7, respectively. The proportional differences in losses (in percentage terms) are presented for each possible pairwise comparison with the GW test performed on each difference. A positive value indicates that the model listed in the column delivers more accurate forecasts than those listed in the rows of the table. Similarly, rejections of

the null indicate that the model listed in the column delivers significantly more accurate forecasts than those listed in the row.

Insert Tables 6 & 7 here

The results indicate that the forecasts associated with the SF model are generally more accurate than the LF model forecasts. Perhaps most noticeable are the results associated with the SF model with FIWN decay based on the volatility predictor. This version of the model delivers the most accurate forecasts as evinced by the relative mean losses. Moreover, these forecasts are almost always significantly more accurate than all versions of the LF model and the SF model with RW decay particularly when the RMSE is used. Under MAE loss, similar results hold, albeit now the SF model with FIWN decay based on the economic predictors is most successful in beating the other models. The results provide compelling evidence that the proposed structural specification underlying the SF model is preferred to the conventional linear specification.

As an alternative to the above pairwise testing, we also consider the reality check approach described in section 3. We examine the predictive performances of three different sets of models against that of a single benchmark model. The three sets are: all LF models (denoted model set $\mathcal{A}$), all SF models with RW decay (denoted model set $\mathcal{B}$), and all SF models with FIWN decay (denoted model set $\mathcal{C}$). The benchmark model is assumed to be each of the individual models in turn. The null hypothesis is that no model within a set outperforms the benchmark model (for a given loss function), against the alternative that at least one model within a set outperforms the benchmark model. Results are provided in Table 8.

Insert Table 8 here

The results indicate that under the RMSE assumption, the null that no LF model is better than the benchmark cannot be rejected. By contrast, there is at least one SF model with FIWN decay that outperforms the benchmark. Moreover, in terms of the benchmark models, it is the SF model with FIWN decay based on the volatility predictor that appears to be the most difficult to beat. Indeed, the null that no model within a model set outperforms this particular benchmark is never rejected. Under the MAE assumption, performances differences are less exaggerated. These

findings are consistent with those presented in the previous tables and highlight the virtues of the SF model from a statistical significance point of view.

5.6. Trading strategy performance

The excellent relative performance of the SF model does not necessarily imply that it has economic value. To assess this value the LF and SF models are used in conjunction with the trading strategy described in the section 4. In particular, out-of-sample conditional expectations of next step ahead returns are estimated and positions are taken accordingly. To ensure that the strategy is time consistent (that is, positions are taken only using information that is available at that time), the models are estimated using a 1000-day rolling window of observations. At each point in time out-of-sample conditional expectations of VIX futures returns, positions, and strategy returns are constructed. These data cover the period January 1, 2008 to December 31, 2017. Having constructed a set of strategy returns over this sample, the performance of the strategy is assessed by calculating Sharpe ratios and risk-adjusted mean returns based on the factor model in (20). These (annualised) ratios and returns are provided in Table 9 for all versions of the LF and SF models, and with respect to the cash only, long only and short only benchmark strategies.¹³ Statistical inference is based on the p-values associated with the Memmel (2003) test (Sharpe ratios) and t-tests applied to the intercept in the factor regression (risk-adjusted returns).

Insert Table 9 here

The results are clear-cut. The LF models and the SF models with RW decay show only limited evidence of abnormal risk-adjusted performance. By contrast, the SF models with FIWN decay (based on economic and volatility predictors) appear to deliver Sharpe ratios and risk-adjusted returns that are large and significantly above zero. This result holds for all benchmarks. For instance, the SF model with FIWN decay based on the economic predictors yields (5-factor) risk-adjusted returns of 71.7% (against the cash only benchmark), 85.9% (against the long only benchmark) and 57.4% (against the short only benchmark) – all of which are significant at the 5% level.¹⁴ Similar results hold when the volatility predictor is used within the context of the SF model with FIWN decay albeit with slightly lower levels of performance achieved. This result lends further support to notion that VIX futures returns are unpredictable when LF models are used, and predictable

when SF models with FIWN decay are used. Thus model structure appears more important than which particular predictors are used.

The encouraging results in Table 9 assume that it is costless to trade. To address this shortcoming, the performance of the trading strategies in the presence of positive transaction costs is considered. The risk-adjusted mean returns to the SF model with FIWN decay based on the volatility predictor are calculated in the presence of transaction costs ranging from zero to 40 basis points. Moreover, the results associated with analogous strategies based on the LF model that employs the volatility predictor are also presented. To examine the importance of incorporating transaction costs into the objective function, we consider the performance of both the τ -ignorant and τ -informed strategies. The results based on using the short only benchmark and various factor models are provided in Figure 1 (SF model) and Figure 2 (LF model).

Insert Figures 1 & 2 here

As one would expect, the results indicate that as transaction costs increase, the performance of the proposed SF model strategies decrease. This deterioration is less acute for the τ -informed investors than for the τ -ignorant investors, both in terms of the size of the risk-adjusted returns and in terms of their statistical significance. For the former investors, performance is positive and significant for transaction cost levels around 15 basis points or less (when the 1-factor, 3-factor and 5-factor models are used). Therefore, it would appear that use of the proposed SF model delivers performance that has both statistical and economical significance in the presence of transaction costs most likely encountered in the VIX futures market. By contrast, the results in Figure 2 indicate that the LF model strategies yield returns that are considerably lower than those associated with the SF model strategies. Crucially, these returns are always insignificantly different from zero in the presence of transaction costs. This lends further support to the use of the proposed SF model over the conventional LF model.

5.7. Robustness checks

It is quite possible that the factor models employed do not sufficiently account for the risk inherent in the trading strategy. To investigate whether the variance risk premium is playing a role, the relationship between risk-adjusted returns to the proposed SF model trading strategy

and this premium is investigated. This is achieved by performing a simple OLS regression of risk-adjusted returns on an estimate of the variance risk premium. This premium is constructed via the methodology used in Bekaert and Hoerova (2014). In particular, the variance risk premium is given by the difference between the VIX and the conditional expectation of realised variance. The latter conditional expectations are constructed using a 1000-day rolling window of observations in which 22-day realised variances are regressed on a 22-day lagged values of the VIX (squared), 22-day lagged values of 22-day realised variances, 22-day lagged values of 5-day realised variances, and 22-day lagged values of 1-day realised variances. This corresponds to Model 8 in Bekaert and Hoerova (2014). The realised variance measure is based on summing 5-minute frequency squared S&P 500 index returns each day. These were obtained from the Oxford-Man Institute’s “realised library” (<http://realized.oxford-man.ox.ac.uk>).

The risk-adjusted strategy returns associated with the SF model with FIWN decay based on the volatility predictor and the short only benchmark are assumed. Results are provided in Table 10. When strategy returns are not risk adjusted there is a significant relationship between strategy performance and the variance risk premium. However, when the returns are risk adjusted this relationship disappears. This implies that the Fama-French factors are sufficiently accounting for the risk inherent in the proposed strategy returns, and that the variance risk premium plays no role in this regard. The central result of this paper therefore still holds that VIX futures returns are predictable.

Insert Table 10 here

The results demonstrate that volatility is a significant predictor. It is therefore natural to consider whether the results depend on the way this predictor is constructed. To this end we consider alternative measures of volatility. Recall that the volatility predictor is given by (22). The preceding analysis uses the squared return to the SPDR S&P 500 ETF as the measure of realised variance (that is, $R_{S,t}^2$). To examine the sensitivity of this assumption we consider two alternative measures of realised variance: squared returns to the S&P 500 index, and the daily sum of 5-minute frequency squared S&P 500 index returns.¹⁵ Moreover, it is also possible that the performance of this predictor is dependent on the value of the IGARCH parameter φ , which has been set equal to 0.06 in the preceding analysis. To investigate this issue we consider a variety of values of this

parameter: specifically, $\varphi \in [0, 0.1]$. The risk-adjusted returns (with respect to the short only position) associated with the LF model and the SF model with FIWN decay both based on the new volatility predictor are presented in Figure 3 (based on daily SPDR S&P 500 ETF returns), Figure 4 (based on daily S&P 500 index returns), and Figure 5 (based on intraday S&P 500 index returns).

Insert Figures 3, 4 & 5 here

The risk-adjusted returns to the LF model are consistently below those associated with the SF model. Moreover, the returns to the latter model are statistically significant for all volatility definitions and for the majority of φ values. Interestingly, there exists a positive relationship between the IGARCH parameter and performance, with risk-adjusted returns peaking around φ equal to 0.1. However, the SF model performance is consistently above the LF performance over this parameter space, and therefore does not alter the central conclusion that relative forecasting performance is highly dependent on model specification. Rather, the use of an IGARCH parameter equal to 0.06 delivers a conservative (and not exaggerated) estimate of the performance of the proposed model.

6. Conclusions

In contrast to previous studies (see, e.g., Konstantinidi and Skiadopoulos, 2011), the proposed SF model appears to have statistical and economic significance in the VIX futures market. The predictability of returns in this market is sufficiently large to deliver an associated market timing strategy that is highly profitable. Moreover, this profitability is robust to transaction costs and the factor model used to account for risk. Transaction costs in excess of 15 basis points would have to be encountered to render the trading strategy valueless. Whether such transaction costs are attainable largely depends on the nature of the investor (individual versus institutional).

The key conclusion of this paper is that the ability to forecast VIX futures returns is more a function of model specification rather than the use of a particular set of predictors. This finding is robust to a number of changes in the design of the study. The proposed model is superior (both in a statistical and economic sense) to standard linear forecasting models over a wide range of performance measures. Variation in the predictor set (for instance, economic to volatility predictors),

and in the way in which the predictors are constructed causes minor changes in performance. By contrast, changes in the specification of the forecasting model (for instance, linear to structural) leads to large (and significant) changes in performance.

It could be argued that speculators should employ the proposed SF model. However, some caution is required. It is quite possible that the factor models employed do not sufficiently account for the risk inherent in the trading strategy. Previous studies provide a rationale for the positive returns to a short position in VIX futures by arguing that this reflects a variance risk premia (above and beyond the traditional Fama-French factor premia); see, e.g., Eraker and Wu (2017). However, in the current context it is the returns to the timing strategy with respect to such a passive benchmark that delivers significant risk-adjusted returns. It, therefore, does not necessarily follow that the predictability observed in the current context reflects a variance risk premia. Indeed, the evidence presented in Table 10 would suggest otherwise.

References

- Ahoniemi, K. 2006. Modeling and forecasting implied volatility: An econometric analysis of the VIX index. Working paper, Helsinki School of Economics.
- Alexander, C., & Korovilas, D. 2012. Diversification of equity with VIX futures: Personal views and skewness preference. ICMA Centre Discussion Paper.
- Bekaert, G., & Hoerova, M. 2014. The VIX, the variance premium and stock market volatility. *Journal of Econometrics* 183, 181–192.
- Dickey, D., & Fuller, W. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Dickey, D., & Fuller, W. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
- Eraker, B., & Wu, Y. 2017. Explaining the negative returns to volatility claims: An equilibrium approach. *Journal of Financial Economics* 125, 72–98.
- Fama, E., & French, K. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.
- Fernandes, M., Medeiros, M., & Scharth, M. 2014. Modeling and predicting the CBOE market volatility index. *Journal of Banking and Finance* 40, 1–10.
- Frijns, B., Tourani-Rad, A., & Webb, R. 2016. On the intraday relation between the VIX and its futures. *Journal of Futures Markets* 36, 870–886.
- Giacomini, R., & White, H. 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Hamilton, J., & Wu, J. 2015. Effects of index-fund investing on commodity futures prices. *International Economic Review* 56, 187–205.
- Huang, Z., Tong, C., & Wang, T. 2017. Pricing the CBOE VIX term structure and VIX futures with realized volatility. Downloaded from https://acfr.aut.ac.nz/_data/assets/pdf_file/0008/

77525/Pricing-the-CBOE-VIX-Term-Structure-and-VIX-Futures-with- Realized-Volatility.pdf
on 16/01/2018.

- Konstantinidi, E., & Skiadopoulos, G. 2011. Are VIX futures prices predictable? An empirical investigation. *International Journal of Forecasting* 27, 543–560.
- Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. 2008. Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking and Finance* 32, 2401–2411.
- Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–178.
- Lee, D., & Schmidt, P. 1996. On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics* 73, 285–302.
- Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. *Finance Letters* 1, 21–23.
- Mina, J., & Xiao, J. 2001. Return to riskmetrics: The evolution of a standard. Technical report, Risk Metrics Group. available at: <http://www.riskmetrics.com/pdf/rrmfinal.pdf>.
- Mishra, V., & Smyth R. 2016. Are natural gas spot and futures prices predictable? *Economic Modelling* 54, 178–186
- Narayan, P., Ahmed, H., & Narayan, S. 2015. Do momentum-based trading strategies work in the commodity futures markets? *Journal of Futures Markets* 35, 868–891.
- Politis, D., & Romano, J. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Psaradellis, I., & Sermpinis, G. 2016. Modelling and trading the U.S. implied volatility indices. Evidence from the VIX, VXN and VXD indices. *International Journal of Forecasting* 32, 1268–1283.
- Said, S., & Dickey, D. 1984. Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika* 71, 599–607.

- Samuelson, P. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6, 41–49.
- Shu, J., & Zhang, J. 2012. Causality in the VIX futures market. *Journal of Futures Markets* 32, 24–46.
- Sun, B., Guo, H., Karim, H., Ge, Y., & Xiong, S. 2015. Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing* 151, 1528–1536.
- Tse, Y. 2018. Return predictability and contrarian profits of international index futures, *Journal of Futures Markets*, forthcoming.
- Wang, T., Shen, Y., Jiang, Y., & Huang, Z. 2016. Pricing the CBOE VIX futures with the Heston-Nandi GARCH model. *Journal of Futures Markets* 37, 641–659.
- Welch, I., & Goyal, A. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–1508.
- White, H. 2000. A reality check for data snooping. *Econometrica* 68, 1097–1126.
- Zhang, J., Shu, J., & Brenner, M. 2010. The new market for volatility trading. *Journal of Futures Markets* 30, 809–833.
- Zhu, S., & Lian, G. 2012. An analytical formula for VIX futures and its applications. *Journal of Futures Markets* 32, 1096–9934.

Notes

¹A smaller tranche of the literature exists which investigates causality amongst VIX spot and futures markets (Shu and Zhang, 2012, and Frijns et al., 2016).

²This decomposition allows *every* covariance-stationary time series to be expressed as the sum of a deterministic and stochastic process. Thus it enables possibly complex non-linear processes to be expressed in linear form.

³This expression forms the basis of Samuelson’s *Theorem of Mean-Percentage Price Drift*, whereby $F_{t,t+h}$ follows a martingale process such that the futures market represents a ‘fair game’ to investors after accounting for the risk inherent in the price of the futures contract.

⁴The proposed model is structural in the sense that it is based on economic theory. In the absence of such theory a natural (and often used) simple alternative is the LF model with parameters estimated by OLS.

⁵An alternative assumption based on a pure white noise (PWN) decay imposed on the underlying dynamics is also possible. Specifically, $\theta_h = 0 \forall h$, corresponding to the assumption that $d = 0$. However, within the subsequent empirical analysis this particular model was found to provide a universally inferior fit to models based on the RW and FIWN decay structures. Consequently, it is omitted from the analysis.

⁶We follow White (2000) and make use of the stationary bootstrap proposed by Politis and Romano (1994), where the bootstrap data consist of sub-sample blocks with lengths drawn from a geometric distribution with the mean equal to a pre-selected value. We assume that 1000 replications of the data are drawn, with the mean block length being set equal to five observations.

⁷When a short position is taken at $t - 1$ the optimisation problem is

$$\max_{w_t \in \{-1, 0, 1\}} \{E[f(R_{t+1}, 1, 2\kappa)|\mathcal{G}_t^*], E[f(R_{t+1}, 0, \kappa)|\mathcal{G}_t^*], E[f(R_{t+1}, -1, 0)|\mathcal{G}_t^*]\}.$$

When no position is taken at $t - 1$ the optimisation problem is

$$\max_{w_t \in \{-1, 0, 1\}} \{E[f(R_{t+1}, 1, \kappa)|\mathcal{G}_t^*], E[f(R_{t+1}, 0, 0)|\mathcal{G}_t^*], E[f(R_{t+1}, -1, \kappa)|\mathcal{G}_t^*]\}.$$

Note here that the candidate set differ in terms of the transaction costs incurred by changing the position.

⁸This choice of rollover date is also used in Konstantinidi and Skiadopoulos (2011).

⁹For further details of the construction of such series see *Campbell White Paper Series* (2014) accessed via the Chicago Mercantile Exchange (CME) website:

<https://www.cmegroup.com/education/files/deconstructing-futures-returns-the-role-of-roll-yield.pdf>.

¹⁰The lag structure in the ADF test is determined using the Bayesian information criterion (BC), and the bandwidth in the KPSS test is based on Newey-West lag truncation with a Bartlett window.

¹¹This is achieved by transforming the fitted values of the model from log form to raw form under the Gaussian approximation.

¹²The BC measure is the most penal fit measure and places more emphasis on model parsimony – an important consideration when conducting an out-of-sample forecasting exercise.

¹³The benchmarks are used by subtracting their returns (ratios) from the trading strategy returns (ratios).

¹⁴Note that the returns are smallest against the short only benchmark because this benchmark is the best amongst the benchmark strategies. In turn, this reflects the negative returns to VIX futures over the period as these contracts are in contango. This contango also means that VIX futures returns are highly negatively correlated with market returns – hence the importance of appropriately adjusting for the (negative systematic) risk within these returns.

¹⁵These were obtained from the Oxford-Man Institute’s “realised library”.

Table 1 – Summary Statistics

Series	Moment				Correlation			(Non)-Stationarity	
	MN	SD	SK	KT	ρ_1	ρ_2	ρ_3	ADF	KPSS
Panel A: Futures data									
VIX futures prices	19.340	8.077	2.199	6.037	0.990	0.977	-0.083	-2.961**	0.979***
VIX futures prices (log)	2.897	0.339	1.052	0.942	0.990	0.935	-0.076	-3.660***	1.131***
VIX futures proportional returns	-49.503	67.595	1.147	6.164	-0.033	0.167	-0.743	-60.712***	0.131
Panel B: VIX index data									
VIX index	18.592	9.072	2.599	9.059	0.982	1.000	-0.125	-3.929***	0.960***
VIX index (log)	2.840	0.378	1.108	1.235	0.982	0.958	-0.114	-4.612***	1.175***
Panel C: S&P 500 index data									
S&P 500 index proportional returns	8.177	18.521	-0.106	12.567	-0.099	-0.126	1.000	-46.763***	0.146

Notes: This table contains the sample moments and various correlations associated with VIX futures and index data. The following moments are included: mean (MN), standard deviation (SD), skewness (SK), and excess kurtosis (KT). The mean and standard deviation values associated with the proportional returns are given in annualised percentage form. Three correlations are included: first-order serial correlation (ρ_1), contemporaneous correlation with the VIX index (ρ_2), and contemporaneous correlation with S&P 500 index returns (ρ_3). The ADF unit root test is based on BC determined lag length, and the KPSS test for stationarity is based on Newey-West/Bartlett kernel determined bandwidth. The ADF test critical values are -3.432 (1%), -2.862 (5%) and -2.567 (10%); and the KPSS test critical values are 0.739 (1%), 0.463 (5%) and 0.347 (10%). Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 2 – Modelling the VIX

Parameter	Model							
	I	ARI	IMA	ARIMA	FI	ARFI	FIMA	ARFIMA
Panel A: Raw VIX index dynamics								
d	1	1	1	1	0.832*** (0.014)	0.809*** (0.022)	0.810*** (0.021)	0.818*** (0.016)
Φ_1		-0.126*** (0.017)		0.626*** (0.042)		0.036* (0.028)		-0.555*** (0.212)
Θ_1			-0.155*** (0.019)	-0.778*** (0.033)			0.036* (0.027)	0.583*** (0.204)
-LL/10	680.064	677.278	676.681	673.731	673.804	673.716	673.713	673.657
AC	680.164	677.478	676.881	674.031	674.004	674.016	674.013	674.057
HC	680.273	677.698	677.100	674.360	674.223	674.345	674.343	674.496
BC	680.471	678.093	677.495	674.953	674.618	674.938	674.935	675.286
Panel B: Log VIX index dynamics								
d	1	1	1	1	0.855*** (0.014)	0.798*** (0.025)	0.809*** (0.021)	0.753*** (0.039)
Φ_1		-0.096*** (0.017)		0.571* (0.356)		0.084*** (0.031)		0.431*** (0.135)
Θ_1			-0.110*** (0.018)	-0.699** (0.343)			0.071*** (0.026)	-0.303*** (0.118)
-LL/10	680.652	677.704	677.175	673.960	673.621	673.727	673.650	674.113
AC	680.752	677.904	677.375	674.260	673.821	674.027	673.950	674.513
HC	680.861	678.124	677.594	674.589	674.040	674.356	674.279	674.952
BC	681.059	678.519	677.989	675.182	674.435	674.949	674.872	675.742

Notes: This table contains the parameter estimates and measures of fit associated with the ARFIMA(1,d,1) model (and restricted versions thereof) applied to the VIX (in raw and log form). Standard errors are given in parentheses. Measures of model fit are the log-likelihood value (LL), the Akaike information criterion (AC), the Hannan-Quinn information criterion (HC), and the Bayesian information criterion (BC). The LL value is negated such that a lower value indicates superior fit. Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 3 – Modelling VIX futures price changes

Parameter	Linear				Structural with RW decay				Structural with FIWN decay			
	LF1	LF2	LF3	LF4	SF1	SF2	SF3	SF4	SF5	SF6	SF7	SF8
λ					0.999*** (0.002)	0.993*** (0.002)	0.983*** (0.002)	0.985*** (0.003)	0.997*** (0.002)	0.993*** (0.002)	0.967*** (0.003)	0.968*** (0.003)
c	-0.040** (0.018)	0.012 (0.037)	-0.117*** (0.031)	0.021 (0.047)	-0.011 (0.042)	0.109** (0.044)	0.021 (0.027)	0.087*** (0.037)	0.142 (0.237)	0.791*** (0.272)	0.609*** (0.132)	0.816*** (0.185)
α_1		-0.062* (0.033)		-0.042 (0.031)		-0.056* (0.030)		-0.025 (0.024)		-0.426*** (0.100)		-0.198*** (0.073)
α_2		0.761 (2.888)		2.923 (2.668)		0.668 (2.680)		2.265 (2.061)		0.005 (0.000)		0.060 (0.000)
α_3		0.011 (0.021)		0.012 (0.020)		0.014 (0.019)		0.010 (0.015)		0.067 (0.124)		0.045 (0.081)
α_4		-0.004 (0.024)		-0.002 (0.023)		0.013 (0.022)		0.012 (0.017)		0.062 (0.141)		0.106 (0.091)
α_5		0.057*** (0.009)		0.059*** (0.011)		0.061*** (0.008)		0.039*** (0.009)		0.437*** (0.056)		0.170*** (0.045)
α_6			0.005*** (0.002)	-0.001 (0.002)			0.013*** (0.001)	0.008*** (0.002)			0.115*** (0.007)	0.096*** (0.009)
LR Test ($\lambda = 1$)					0.462	0.004	0.000	0.000	0.115	0.001	0.000	0.000
Adj. R^2 (%)	0.000	1.277	0.231	1.272	-0.014	1.477	1.110	1.784	-0.001	1.678	2.411	2.773
LL	500.038	497.568	499.588	497.527	500.012	497.167	498.010	496.579	499.990	496.815	495.723	494.831
AC	500.138	498.168	499.788	498.227	500.212	497.867	498.310	497.379	500.190	497.515	496.023	495.631
HC	500.248	498.827	500.008	498.996	500.432	498.636	498.640	498.257	500.409	498.283	496.352	496.509
BC	500.445	500.012	500.403	500.379	500.827	500.019	499.232	499.838	500.804	499.666	496.945	498.090

Notes: This table contains the parameter estimates and measures of fit associated with the proposed LF and SF models (and restricted versions thereof) applied to VIX futures price changes. Standard errors are given in parentheses (non-restricted parameters only). The likelihood ratio (LR) test p-values associated with the null that λ equals unity are also provided. Measures of model fit are the adjusted R^2 statistic, log-likelihood value (LL), the Akaike information criterion (AC), the Hannan-Quinn information criterion (HC), and the Bayesian information criterion (BC). The LL value is negated such that a lower value indicates superior fit. All measures of fit are given in terms of the raw data. Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 4 – Modelling VIX futures proportional returns

Parameter	Linear				Structural with RW decay				Structural with FIWN decay			
	LF1	LF2	LF3	LF4	SF1	SF2	SF3	SF4	SF5	SF6	SF7	SF8
λ					0.996*** (0.002)	0.994*** (0.003)	0.987*** (0.003)	0.988*** (0.003)	0.997*** (0.002)	0.996*** (0.002)	0.979*** (0.029)	0.972*** (0.004)
c	-0.002*** (0.001)	0.000 (0.001)	-0.005*** (0.001)	0.000 (0.002)	0.028 (0.038)	0.096** (0.041)	0.027 (0.028)	0.082*** (0.035)	0.152 (0.197)	0.482** (0.224)	0.200 (1.049)	0.712*** (0.178)
α_1		-0.002 (0.001)		-0.002 (0.001)		-0.048* (0.025)		-0.026 (0.024)		-0.350*** (0.093)		-0.102 (0.123)
α_2		0.091 (0.105)		0.091 (0.110)		0.658 (2.031)		1.893 (2.057)		0.468 (0.000)		6.659 (10.753)
α_3		0.000 (0.001)		0.000 (0.001)		-0.007 (0.013)		-0.005 (0.012)		-0.080 (0.092)		-0.027 (0.061)
α_4		0.000 (0.001)		0.000 (0.001)		0.000 (0.015)		0.003 (0.014)		-0.042 (0.106)		0.056 (0.074)
α_5		0.002*** (0.000)		0.002*** (0.000)		0.043*** (0.009)		0.032*** (0.009)		0.314*** (0.065)		0.114** (0.047)
α_6			0.017** (0.007)	-0.001 (0.009)			0.010*** (0.002)	0.007*** (0.002)			0.104** (0.052)	0.094*** (0.012)
LR Test ($\lambda = 1$)					0.139	0.036	0.000	0.000	0.093	0.078	0.474	0.000
Adj. R^2 (%)	0.000	0.734	0.142	0.706	-0.015	0.628	0.413	0.784	0.005	0.696	0.742	1.005
LL	-598.725	-600.248	-599.020	-600.248	-598.750	-600.114	-599.539	-600.435	-598.785	-600.231	-600.111	-600.820
AC	-598.625	-599.648	-598.820	-599.548	-598.550	-599.414	-599.239	-599.635	-598.585	-599.531	-599.811	-600.020
HC	-598.515	-598.989	-598.600	-598.780	-598.330	-598.646	-598.910	-598.756	-598.365	-598.762	-599.482	-599.142
BC	-598.318	-597.803	-598.205	-597.397	-597.935	-597.263	-598.317	-597.176	-597.970	-597.379	-598.889	-597.561

Notes: This table contains the parameter estimates and measures of fit associated with the proposed LF and SF models (and restricted versions thereof) applied to VIX futures proportional returns. Standard errors are given in parentheses (non-restricted parameters only). The likelihood ratio (LR) test p-values associated with the null that λ equals unity are also provided. Measures of model fit are the adjusted R^2 statistic, log-likelihood value (LL), the Akaike information criterion (AC), the Hannan-Quinn information criterion (HC), and the Bayesian information criterion (BC). The LL value is negated such that a lower value indicates superior fit. All measures of fit are given in terms of the raw data. Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 5 – Out-of-sample forecast accuracy

Model	Price changes			Proportional returns		
	Metric			Metric		
	RMSE	MAE	R^2 (%)	RMSE	MAE	R^2 (%)
Panel A: Linear forecasting model						
LF1: no predictors (random walk)	100.000	100.000	0.025	100.000	100.000	0.075
LF2: economic predictors	100.465	99.357	0.344	100.257	99.484	0.148
LF3: volatility predictor	100.290	100.279	0.305	100.078	100.177	0.146
LF4: all predictors	100.928	99.962	0.366	100.458	99.985	0.191
Panel B: Structural forecasting model with RW decay						
SF1: no predictors	100.202	100.510	0.018	100.050	100.465	0.001
SF2: economic predictors	100.198	99.404	0.555	100.093	99.742	0.170
SF3: volatility predictor	99.932	99.945	0.589	99.861	100.211	0.206
SF4: all predictors	100.658	99.906	0.502	100.181	99.976	0.176
Panel C: Structural forecasting model with FIWN decay						
SF5: no predictors	100.051	100.180	0.005	99.973	100.032	0.020
SF6: economic predictors	100.035	99.041	0.771	99.962	99.446	0.342
SF7: volatility predictor	99.220	99.218	1.645	99.578	99.643	0.692
SF8: all predictors	100.190	99.987	0.952	99.932	100.029	0.516

Notes: This table contains the root mean squared error (RMSE), mean absolute error (MAE), and the R^2 statistics from the Mincer-Zarnowitz regression for various versions of the LF and SF models. A value of RMSE and MAE above 100 indicates inferior forecast accuracy with respect to the linear model with no predictors (that is, LF1). The models are estimated using a rolling window of 1000 days, with out-of-sample forecasts generated at each point in time.

Table 6 – Comparing out-of-sample forecast accuracy (proportional mean losses associated with forecasts of price changes)

Model	Linear				Structural with RW decay				Structural with FIWN decay			
	LF1	LF2	LF3	LF4	SF1	SF2	SF3	SF4	SF5	SF6	SF7	SF8
Panel A: Relative RMSE values												
LF1	-0.463	-0.289	-0.919	-0.919	-0.202	-0.198	0.068	-0.654	-0.051	-0.035	0.786	-0.190
LF2	0.465	0.175	-0.458	-0.458	0.262	0.266**	0.533	-0.192	0.414	0.430	1.255**	0.275
LF3	0.290	-0.174	-0.632	-0.632	0.087	0.091	0.358***	-0.366	0.239	0.255	1.078***	0.100
LF4	0.928	0.460*	0.636	0.636	0.724	0.728***	0.996**	0.268**	0.876	0.892**	1.721***	0.736***
SF1	0.202	-0.262	-0.087	-0.719	-0.004	0.004	0.270	-0.453	0.151	0.167	0.990**	0.012
SF2	0.198	-0.266	-0.091	-0.723	-0.004	-0.266	0.266	-0.457	0.147	0.163	0.986**	0.008
SF3	-0.068	-0.531	-0.357	-0.986	-0.270	-0.266	0.726	-0.721	-0.119	-0.103	0.718***	-0.257
SF4	0.658	0.192	0.367	-0.267	0.455	0.459**	0.726	-0.465	0.607	0.623*	1.449**	0.467**
SF5	0.051	-0.412	-0.238	-0.869	-0.151	-0.147	0.119	-0.603	0.016	0.016	0.837*	-0.139
SF6	0.035	-0.428	-0.254	-0.885	-0.167	-0.163	0.103	-0.619	-0.016	0.016	0.821*	-0.155
SF7	-0.780	-1.239	-1.067	-1.692	-0.980	-0.976	-0.713	-1.429	-0.830	-0.815	0.978**	-0.968
SF8	0.190	-0.274	-0.099	-0.731	-0.012	-0.008	0.258	-0.465	0.139	0.155	0.978**	-0.968
# Rejections	0	1	0	0	0	3	2	1	0	2	10	2
Panel B: Relative MAE values												
LF1	0.647	-0.278	0.038	0.038	-0.507	0.599	0.055	0.094	-0.180	0.968**	0.788*	0.013
LF2	-0.643	-0.919	-0.605	-0.605	-1.147	-0.048	-0.589	-0.550	-0.822	0.319	0.140	-0.630
LF3	0.279	0.928**	0.317	0.317	-0.230	0.880**	0.334**	0.373	0.099	1.250***	1.070***	0.292
LF4	-0.038	0.609***	-0.316	-0.316	-0.545	0.561***	0.017	0.056	-0.218	0.929***	0.750*	-0.025
SF1	0.510**	0.230	0.548	0.548	-1.100	1.112***	0.565**	0.604	0.329***	1.483***	1.302***	0.523
SF2	-0.596	0.048	-0.872	-0.558	-0.562	0.544	-0.541	-0.502	-0.774	0.366*	0.188	-0.583
SF3	-0.055	0.592	-0.333	-0.017	-0.601	0.505**	-0.039	0.039	-0.234	0.913**	0.733***	-0.042
SF4	-0.094	0.553**	-0.372	-0.056	-0.601	0.505**	-0.039	0.039	-0.274	0.873***	0.694**	-0.081
SF5	0.180	0.828*	-0.099	0.218	-0.328	0.780*	0.235	0.274	-1.137	1.150**	0.970**	0.193
SF6	-0.959	-0.318	-1.234	-0.921	-1.461	-0.365	-0.904	-0.865	-0.961	0.178	-0.178	-0.946
SF7	-0.782	-0.140	-1.058	-0.745	-1.286	-0.188	-0.728	-0.689	-0.961	0.178	-0.178	-0.770
SF8	-0.013	0.634**	-0.291	0.025	-0.520	0.587**	0.042	0.081	-0.192	0.955***	0.776**	-0.770
# Rejections	1	6	0	0	0	6	2	0	1	9	8	0

Notes: This table contains the proportional mean losses associated with forecasts of price changes for various versions of the LF and SF models, and are given in percentage points. A positive value indicates that the model listed in the column delivers more accurate forecasts than those listed in the rows of the table. Similarly, rejections of the null (based on the GW test) indicate that the model listed in the column delivers significantly more accurate forecasts than those listed in the row. The # Rejections rows contain the number of instances in which the null is rejected at the 10% (or lower) level. Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 7 – Comparing out-of-sample forecast accuracy (proportional mean losses associated with forecasts of proportional returns)

Model	Linear				Structural with RW decay				Structural with FIWN decay			
	LF1	LF2	LF3	LF4	SF1	SF2	SF3	SF4	SF5	SF6	SF7	SF8
Panel A: Relative RMSE values												
LF1		-0.286	-0.078	-0.483	-0.050	-0.092	0.139	-0.181	0.027	0.038	0.424**	0.068
LF2	0.287		0.209	-0.197	0.237	0.195	0.427*	0.106	0.315	0.325**	0.712***	0.355*
LF3	0.078	-0.209		-0.406	0.028	-0.015	0.217*	-0.103	0.105	0.116	0.502***	0.146
LF4	0.486*	0.198*	0.407*		0.435	0.393**	0.625**	0.304*	0.513*	0.524**	0.911***	0.554**
SF1	0.050	-0.237	-0.028	-0.434		-0.043	0.189*	-0.131	0.077	0.088	0.474**	0.118
SF2	0.093	-0.194	0.015	-0.391	0.043		0.232	-0.088	0.120	0.131	0.517**	0.161
SF3	-0.139	-0.425	-0.217	-0.622	-0.189	-0.231		-0.319	-0.112	-0.101	0.284**	-0.071
SF4	0.181	-0.106	0.103	-0.303	0.131	0.088	0.320*		0.208	0.219	0.605**	0.249*
SF5	-0.027	-0.314	-0.105	-0.511	-0.077	-0.120	0.112	-0.208		0.011	0.396**	0.040
SF6	-0.038	-0.324	-0.116	-0.521	-0.088	-0.131	0.101	-0.219	-0.011		0.385*	0.030
SF7	-0.422	-0.707	-0.499	-0.903	-0.471	-0.514	-0.283	-0.602	-0.394	-0.384		-0.354
SF8	-0.068	-0.354	-0.146	-0.551	-0.118	-0.160	0.071	-0.248	-0.040	-0.030	0.355*	
# Rejections	1	1	1	0	0	1	5	1	1	2	11	3
Panel B: Relative MAE values												
LF1		0.429	-0.177	-0.033	-0.462	0.258	-0.211	0.024	-0.032	0.557*	0.359	-0.029
LF2	-0.427		-0.603	-0.460	-0.888	-0.170	-0.637	-0.403	-0.459	0.128	-0.070	-0.456
LF3	0.177	0.607**		0.144	-0.286	0.436	-0.034	0.201	0.145	0.735**	0.536**	0.147
LF4	0.033	0.462***	-0.144		-0.430	0.291	-0.178	0.057	0.001	0.590**	0.392	0.003
SF1	0.465***	0.896***	0.287	0.432		0.724**	0.253*	0.489*	0.433***	1.024***	0.825***	0.435
SF2	-0.258	0.170	-0.434	-0.290	-0.719		-0.468	-0.233	-0.289	0.298*	0.100	-0.287
SF3	0.211	0.641**	0.034	0.178	-0.252	0.470*		0.235	0.179	0.770**	0.570***	0.182
SF4	-0.024	0.405**	-0.201	-0.057	-0.487	0.234**	-0.235		-0.056	0.533**	0.334	-0.054
SF5	0.032	0.461*	-0.145	-0.001	-0.431	0.290	-0.179	0.056		0.589**	0.390*	0.002
SF6	-0.554	-0.128	-0.730	-0.587	-1.014	-0.297	-0.764	-0.530	-0.586		-0.198	-0.583
SF7	-0.357	0.070	-0.533	-0.390	-0.818	-0.100	-0.567	-0.333	-0.389	0.198		-0.387
SF8	0.029	0.459*	-0.147	-0.003	-0.433	0.288	-0.181	0.054	-0.002	0.587**	0.388*	
# Rejections	1	7	0	0	0	3	1	1	1	9	5	0

Notes: This table contains the proportional mean losses associated with forecasts of proportional returns for various versions of the LF and SF models, and are given in percentage points. A positive value indicates that the model listed in the column delivers more accurate forecasts than those listed in the rows of the table. Similarly, rejections of the null (based on the GW test) indicate that the model listed in the column delivers significantly more accurate forecasts than those listed in the row. The # Rejections rows contain the number of instances in which the null is rejected at the 10% (or lower) level. Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 8 – Reality check tests of forecast accuracy

Benchmark	Price changes			Proportional returns		
	Model set			Model set		
	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$
Panel A: p-values under RMSE						
LF1: no predictors (random walk)	0.998	0.860	0.022	0.983	0.219	0.002
LF2: economic predictors	0.038	0.008	0.000	0.014	0.000	0.000
LF3: volatility predictor	0.325	0.092	0.000	0.512	0.042	0.000
LF4: all predictors	0.007	0.001	0.000	0.002	0.000	0.000
SF1: no predictors	0.459	0.258	0.000	0.711	0.058	0.000
SF2: economic predictors	0.390	0.178	0.000	0.500	0.013	0.000
SF3: volatility predictor	0.994	1.000	0.000	1.000	1.000	0.011
SF4: all predictors	0.031	0.005	0.000	0.119	0.004	0.000
SF5: no predictors	0.856	0.695	0.003	1.000	0.321	0.007
SF6: economic predictors	0.892	0.689	0.001	0.983	0.491	0.001
SF7: volatility predictor	1.000	1.000	1.000	1.000	1.000	1.000
SF8: all predictors	0.528	0.270	0.001	0.997	0.617	0.008
# Rejections	3	4	11	2	6	11
Panel B: p-values under MAE						
LF1: no predictors (random walk)	0.006	0.009	0.000	0.001	0.078	0.000
LF2: economic predictors	1.000	0.985	0.108	1.000	1.000	0.827
LF3: volatility predictor	0.000	0.000	0.000	0.000	0.002	0.000
LF4: all predictors	0.006	0.006	0.000	0.000	0.114	0.001
SF1: no predictors	0.000	0.000	0.000	0.000	0.000	0.000
SF2: economic predictors	0.818	1.000	0.049	0.061	1.000	0.032
SF3: volatility predictor	0.007	0.006	0.000	0.000	0.000	0.000
SF4: all predictors	0.010	0.008	0.000	0.001	0.095	0.000
SF5: no predictors	0.000	0.000	0.000	0.000	0.036	0.001
SF6: economic predictors	1.000	1.000	0.997	0.986	1.000	1.000
SF7: volatility predictor	0.998	1.000	0.390	0.339	0.999	0.135
SF8: all predictors	0.009	0.007	0.000	0.001	0.085	0.001
# Rejections	8	8	9	9	7	9

Notes: This table contains the reality check p-values associated with the null hypothesis that the best model in the model set is not superior to an individual benchmark model. The following model sets are considered: all linear models (denoted model set $\mathcal{A}$), all structural models with RW decay (denoted model set $\mathcal{B}$), and all structural models with FIWN decay (denoted model set $\mathcal{C}$). The benchmark model is excluded from the model set when calculating the test p-values. The # Rejections rows contain the number of instances in which the null is rejected at the 10% (or lower) level.

Table 9 – Trading strategy performance

Benchmark	Linear				Structural with RW decay				Structural with FIWN decay			
	LF1	LF2	LF3	LF4	SF1	SF2	SF3	SF4	SF5	SF6	SF7	SF8
Panel A: Sharpe ratios												
Cash only benchmark	0.012	0.501	0.680*	0.326	0.085	0.446	0.540	0.488	0.182	0.982**	0.926**	0.900**
Long only benchmark	0.726*	1.029**	1.208***	0.854**	0.613	0.974*	1.068**	1.016**	0.710	1.510***	1.454***	1.428***
Short only benchmark	-0.330	-0.027	0.152	-0.202	-0.444	-0.082	0.012	-0.040	-0.346	0.454	0.398	0.372
Panel B: Mean returns (unadjusted)												
Cash only benchmark	14.579	36.851*	50.015**	23.975	6.226	32.822	39.716*	35.878	13.404	72.141***	68.026***	66.171***
Long only benchmark	53.433	75.706**	88.869**	62.829*	45.080	71.677**	78.570**	74.733**	52.258	110.995***	106.881***	105.025***
Short only benchmark	-24.276	-2.003	11.161	-14.879	-32.629	-6.032	0.862	-2.976	-25.451	33.286	29.172	27.316
Panel C: Mean returns (1-factor adjusted)												
Cash only benchmark	0.240	29.137	46.969**	20.980	-0.505	29.961	33.656	34.289	-0.192	71.002***	62.407***	65.878***
Long only benchmark	12.843	41.740	59.573*	33.583	12.099	42.565	46.259	46.892	12.411	83.606***	75.010**	78.482***
Short only benchmark	-12.363	16.533	34.366*	8.377	-13.108	17.358	21.052	21.685	-12.796	58.399**	49.804**	53.275**
Panel D: Mean returns (3-factor adjusted)												
Cash only benchmark	-4.713	29.282	44.477**	21.448	-4.658	29.482	30.914	34.171	-4.725	70.386***	60.160***	65.959***
Long only benchmark	6.070	40.064	55.260*	32.230	6.125	40.265	41.697	44.953	6.057	81.168***	70.942**	76.741***
Short only benchmark	-15.495	18.500	33.695*	10.666	-15.440	18.700	20.132	23.388	-15.508	59.603***	49.378**	55.177**
Panel E: Mean returns (5-factor adjusted)												
Cash only benchmark	0.590	28.502	46.979**	21.828	-0.428	28.809	32.241	33.991	0.730	71.651***	59.204***	65.675***
Long only benchmark	14.880	42.792	61.270*	36.119	13.863	43.100	46.532	48.282*	15.021	85.942***	73.495**	79.965***
Short only benchmark	-13.701	14.211	32.689*	7.538	-14.718	14.519	17.951	19.701	-13.560	57.361**	44.914**	51.384**

Notes: This table contains the annualised Sharpe ratios and risk-adjusted returns to the trading strategy based on various versions of the proposed LF and SF models. Returns are given in percentage terms. These are provided against three benchmarks (cash only, long only and short only benchmarks), with adjusted returns calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 10 – Strategy performance and the variance risk premium

Risk adjustment method	Λ_0	Λ_1	Adj. R^2 (%)
None	-0.003** (0.001)	2.661*** (0.455)	1.183
Fama-French 1-factor model	0.002 (0.001)	0.244 (0.382)	0.015
Fama-French 3-factor model	0.002 (0.001)	0.245 (0.382)	0.015
Fama-French 5-factor model	0.001 (0.001)	0.293 (0.382)	0.021

Notes: This table contains the parameter estimates and a measure of fit associated with an OLS regression of risk-adjusted trading strategy returns on an estimate of the variance risk premium. The constant term is denoted by Λ_0 and the slope coefficient is denoted by Λ_1 . The trading strategy assumes the SF model with FIWN decay based on the volatility predictor and the short only benchmark. Returns are adjusted using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). Significance is indicated as follows: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Figure 1 – Strategy performance and transaction costs (SF model)

This figure contains plots of the annualised risk-adjusted returns to the trading strategy (in percentage terms) that uses the SF model with FIWN decay based on the volatility predictor against transaction costs (in basis points). These are provided against the short only benchmark and are calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). The investors are assumed to take into account transaction costs (τ -informed) or not (τ -ignorant). Significance at the 5% level is indicated with a solid red symbol.

Figure 2 – Strategy performance and transaction costs (LF model)

This figure contains plots of the annualised risk-adjusted returns to the trading strategy (in percentage terms) that uses the LF model based on the volatility predictor against transaction costs (in basis points). These are provided against the short only benchmark and are calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). The investors are assumed to take into account transaction costs (τ -informed) or not (τ -ignorant). Significance at the 5% level is indicated with a solid red symbol.

Figure 3 – Strategy performance and volatility measurement (based on S&P 500 ETF returns)

This figure contains plots of the annualised risk-adjusted returns to the trading strategy (in percentage terms) based on various versions of the LF model and the SF model with FIWN decay (based on the volatility predictor) against the IGARCH parameter (φ) used in the construction of the volatility predictor. These are provided against the short only benchmark and are calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). Transaction costs are set equal to zero. Significance at the 5% level is indicated with a solid red symbol.

Figure 4 – Strategy performance and volatility measurement (based on S&P 500 index returns)

This figure contains plots of the annualised risk-adjusted returns to the trading strategy (in percentage terms) based on various versions of the LF model and the SF model with FIWN decay (based on the volatility predictor) against the IGARCH parameter (φ) used in the construction of the volatility predictor. These are provided against the short only benchmark and are calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). Transaction costs are set equal to zero. Significance at the 5% level is indicated with a solid red symbol.

Figure 5 – Strategy performance and volatility measurement (based on intraday S&P 500 index returns)

This figure contains plots of the annualised risk-adjusted returns to the trading strategy (in percentage terms) various versions of the LF model and the SF model with FIWN decay (based on the volatility predictor) against the IGARCH parameter (φ) used in the construction of the volatility predictor. These are provided against the short only benchmark and are calculated using various factor models (no factors, and the Fama-French 1-factor, 3-factor and 5-factor models). Transaction costs are set equal to zero. Significance at the 5% level is indicated with a solid red symbol.