

Modeling and predicting the CBOE market volatility index

Marcelo Fernandes

Sao Paulo School of Economics – FGV
and Queen Mary University of London
marcelo.fernandes@fgv.br

Marcelo C. Medeiros

Department of Economics,
Pontifical Catholic University of Rio de Janeiro
mcm@econ.puc-rio.br

Marcel Scharth

Australian School of Business,
University of New South Wales
m.scharth@unsw.edu.au

Abstract: This paper performs a thorough statistical examination of the time-series properties of the daily market volatility index (VIX) from the Chicago Board Options Exchange (CBOE). The motivation lies not only on the widespread consensus that the VIX is a barometer of the overall market sentiment as to what concerns investors' risk appetite, but also on the fact that there are many trading strategies that rely on the VIX index for hedging and speculative purposes. Preliminary analysis suggests that the VIX index displays long-range dependence. This is well in line with the strong empirical evidence in the literature supporting long memory in both options-implied and realized variances. We thus resort to both parametric and semiparametric heterogeneous autoregressive (HAR) processes for modeling and forecasting purposes. Our main findings are as follows. First, we confirm the evidence in the literature that there is a negative relationship between the VIX index and the S&P 500 index return as well as a positive contemporaneous link with the volume of the S&P 500 index. Second, the term spread has a slightly negative long-run impact in the VIX index, when possible multicollinearity and endogeneity are controlled for. Finally, we cannot reject the linearity of the above relationships, neither in sample nor out of sample. As for the latter, we actually show that it is pretty hard to beat the pure HAR process because of the very persistent nature of the VIX index.

JEL classification numbers: G12, C22, C53, E44

Keywords: heterogeneous autoregression, implied volatility, neural networks, VIX.

Acknowledgments: The authors are very grateful to financial support from the ESRC under the grant RES-062-23-0311 (Fernandes) and from CNPq-Brazil (Medeiros), respectively. The usual disclaimer applies.

1 Introduction

The Chicago Board Options Exchange (CBOE) computes since 1993 the volatility index VIX to measure market expectations of the near-term volatility implied by stock index option prices. Actually, as from September 2003, the CBOE reports two market volatility indices. The VXO represents the implied volatility of a hypothetical 30-calendar-day at-the-money S&P 100 index option, whereas the VIX hinges on the prices of a portfolio of 30-calendar-day S&P 500 calls and puts with weights being inversely proportional to the squared strike price. The latter thus gauges the expected market volatility by pooling the information from option prices over the whole volatility skew, not just at-the-money strikes as in the VXO index. Moreover, the VIX considers a model-free estimator of the implied volatility, and so it does not depend on any particular option pricing framework. The motivation for using options on the S&P 500 index rather than the S&P 100 lies on the fact that the S&P 500 is the main stock market benchmark in the US not only for derivative markets, but also for the hedge fund industry.

This means that the VIX essentially offers a market-determined, forward-looking estimate of one-month stock market volatility (Hentschel, 2003). Most studies in the literature that tackle the information content of implied market volatility employ either VIX or VXO time series. See, among others, Canina and Figlewski (1993), Christensen and Prabhala (1998), Fleming (1998), Poon and Taylor (2001a,b), Martens and Zein (2004) Koopman et al. (2005), and Bandi and Perron (2006). All in all, options-implied volatility is typically more informative than time-series volatility models based on stock market index returns for forecasting purposes, though the latter may sometimes carry incremental information; see, for example, Becker et al. (2007). Jorion (1995), Xu and Taylor (1995), Taylor and Xu (1997), and Martens and Zein (2004) provide similar evidence for foreign exchange markets.

This paper departs from this literature in that we do not attempt to compare the information content of VIX relative to volatility models based on the S&P 500 index returns. In some sense, we restrict attention to a much more basic task, namely, to understand the statistical behavior of the new VIX time series, though we carry out such exercise in a multivariate setting that controls for macroeconomic and financial market conditions. Our motivation lies on the widespread notion that the VIX stands for a barometer of the overall market conditions (Whaley, 2000). High VIX levels typically reflect pessimism, causing equity prices to overshoot on the downside and thus leading to subsequent rallies. In turn, low VIX levels would mirror complacency among market participants, setting up the market for disappointment and raising the likelihood of a market correction. In

addition, forecasting the VIX index is a necessary step of any trading strategy employing VIX futures and options at the CBOE Futures Exchange either for trading volatility or for hedging purposes (Konstantinidi et al., 2008; Carr and Lee, 2009; Clements and Fuller, 2012).

Our analysis complements well the evidence put forth by Fleming et al. (1995) in their examination of the statistical behavior of the VXO index. They conclude that the daily changes in the VXO index display a slightly first-order positive autocorrelation, whereas weekly changes exhibit significant mean reversion, even if there is no sign of either intra-day or -week seasonality. In addition, they also evince a strong negative and asymmetric association with contemporaneous stock market returns. In contrast, our findings suggest that the VIX index behaves in a somewhat different manner. First, the contemporaneous relationship between the VIX index and the S&P 500 index returns does not seem to feature nonlinear or asymmetric effects. Second, we uncover a strong long-range dependence in the data in line with the long memory that typically characterizes both options-implied and realized volatility measures (Koopman et al., 2005; Bandi and Perron, 2006; Corsi, 2009).

To capture the long-memory in the VIX index, we resort to the family of heterogeneous autoregressive (HAR) processes (Müller et al., 1997; Corsi, 2009; Hillebrand and Medeiros, 2010). Apart from the pure HAR model, we also set out parametric and semiparametric HAR-type processes with additional explanatory variables so as to account for the (contemporaneous and predictive) relationships between the VIX index and key financial and macroeconomic variables. Apart from the changes in the S&P 500 index and volume, we include multiperiod returns on the one-month oil futures contract, the change in the foreign value of the US dollar (as measured by the USD index), the term spread, the credit spread, and the difference between the effective and target Federal Fund rates. These are all linked to different dimensions of the overall market conditions in the US. Both oil prices and term spread convey information about the present and future real economic activity, whereas the credit spread relates to the amount of liquidity in the market. The strength of the dollar and the deviation in the Fed rates both reflect to some extent the macroeconomic context in the US.

The results we obtain are very robust in that the average partial effect of the macro-finance variables do not vary much across specifications. Even though accounting for nonlinear dependence seems to matter little. We estimate a very flexible nonlinear model based on neural networks, which have the ability to accurately approximate quite general nonlinear functions and can be interpreted as a semiparametric specification. The nonlinear models are estimated by Bayesian regularization

which avoids overfitting by automatically shrinking insignificant partial effects to zero. This turns out to entail a robust performance across different forecasting horizons. A careful analysis of the average partial effects within the full sample unveils some interesting relationships. As expected, we find a strong negative relationship with both contemporaneous and lagged S&P 500 index returns as well as a positive link with the contemporaneous S&P 500 volume. We also establish an inverse relationship of the VIX index with the term spread, when possible multicollinearity and endogeneity are controlled for. Finally, the VIX index does not seem to depend much on the USD index, the deviation in the Fed rates, the credit spread and the changes in oil prices.

The remainder of this paper is organized as follows. Section 2 provides some background and describes how the CBOE computes the market volatility index. Section 3 discusses the main features of the VIX data so as to shed some light on the specification of the econometric model. Section 4 then evaluates both the in-sample and out-of-sample performances of several augmented heterogeneous autoregressive models. Section 5 offers some concluding remarks.

2 Background for the VIX index

The idea of constructing a volatility index from option prices emerges soon after the introduction of exchange-traded index options in 1973. Gastineau (1977) proposes a volatility index that averages the volatilities implied by at-the-money call options of 14 stocks, whereas Cox and Rubinstein (1985) ameliorate the procedure by employing multiple call options on each stock and by weighting the volatilities in such a fashion that the index is at the money with a constant time to expiration. The CBOE volatility indices capture the spirit of these earlier efforts, extending the notion in two important directions. First, the VIX hinges on index options rather than stock options. Second, it depends on the implied volatilities of both call and put options. This not only increases the amount of information that the index pools, but also mitigates any eventual bias due to staleness in the observed index level and due to mismeasurement in the riskless rate.

In a nutshell, the VIX index measures the market expectations of the near-term volatility implied by stock index option prices. It features three main differences with respect to the VXO index. First, the VIX index relies on S&P 500 index options with a wide array of strike prices rather than restricting attention to at-the-money strike prices as in VXO. Second, the VIX does not assume the Black-Scholes-Merton option pricing framework, employing a model-free estimator of the implied volatility (Britten-Jones and Neuberger, 2000; Jiang and Tian, 2005). Third, the VIX calculation consider options on the S&P 500 index rather than on the S&P 100 index. This

seems much more natural for the S&P 500 is the primary stock market benchmark for both the hedge fund industry and derivative markets.

The model-free estimator of the implied volatility that CBOE employs to calculate the VIX index reads

$$\sigma_{\text{CBOE}}^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left(\frac{F}{K_0} - 1 \right)^2, \quad (1)$$

where T is time to expiration, F is the forward index level derived from the index options prices, K_i is the strike price of the i th out-of-the-money option (either a call if $K_i > F$ or a put if $K_i < F$), $\Delta K_i = (K_{i+1} - K_{i-1})/2$ (for the lowest/highest strike price is the difference between the lowest/highest strike and the next higher/lower strike, respectively), K_0 is the first strike below the forward index level, r is the risk-free interest rate to expiration, and $Q(K_i)$ is the mid-quote for the option with strike of K_i . The VIX index then equals 100 times the options-implied volatility given by σ_{CBOE} in (1). See discussion in Demeterfi et al. (1999).

The CBOE computes the VIX using primarily the put and call options in the two nearest-term expiration months so as to bracket a 30-day calendar period. At eight days to expiration, the VIX rolls to the second and third contract months to alleviate any sort of pricing anomaly that may occur due to the expiration proximity. For the sake of precision, the CBOE fixes the risk-free rate at $r = 1.162\%$ and measures the time to expiration T in minutes rather than days: $T = T_{SD}/T_Y$, where T_{SD} is the total number of minutes remaining until 8:30 on the settlement day and T_Y refers to the number of minutes in a year.

As for the forward index level, the CBOE assumes that $F = K_* + e^{rT_*} (C_* - P_*)$, where C_* and P_* are respectively the prices of the ‘at-the-money’ call and put options with a time to maturity of T_* and a strike price of K_* that minimizes the distance between the call and put prices. Finally, one determines the threshold strike K_0 as the strike price immediately below the forward index level F . The algorithm then sorts all options in ascending order by strike price so as to select only the call/put options with nonzero bid quote and strike price either at or below/above K_0 , respectively. To avoid double counting, one must average the mid-quote prices of the call and put options at K_0 . The CBOE executes the above calculations for both the near and next term options, resulting in a forward index level and a threshold strike for each term. This ultimately means that the algorithm will end up with estimates of the implied volatility in (1) for the near term options and for the next term options. The single VIX index then stems from a linear interpolation of these two estimates that ensures a constant maturity of 30 days to expiration. See Andersen and Bondarenko (2007) for an excellent discussion on the approximation errors of the VIX index.

3 Daily behavior of the market volatility index

We examine the daily VIX index for the period running from January 2, 1992 to January 15, 2013. The sample include altogether 5,807 daily observations. We use the full sample for the in-sample analyses, namely, descriptive statistics and contemporaneous modeling, whereas we employ a rolling window of 2,500 observations for the estimation of all predictive regressions. This means that the sample size for the out-of-sample performance evaluation amounts to about 3,240 observations after controlling for starting values.

Figure 1 illustrates the time evolution of the VIX index in the full-sample period. The VIX seems to oscillate in long swings between a quite volatile regime with high index values and a more stable regime with low index values. High volatility characterizes the periods ranging from January to December 1990, from July 1997 to April 2003, and from August 2007 onwards. In contrast, low volatility seems dominant from January 1991 to June 1997 and from April 2004 to July 2007. This is consistent with the claim in Whaley (2000) that one may interpret the VIX index as the investors' fear gauge. There are a series of financial crises in the periods featuring a high VIX index, e.g., Asian crisis in 1997, Russian crisis in 1998, Brazilian crisis in 1999, the internet bubble burst in 2000, the 9/11 terrorist attack in 2001, the corporate scandals in 2002, the quantitative long/short equity hedge funds meltdown in the first week of August 2007, the Lehmann Brother collapse in mid-September 2008, and the subsequent credit crunch and global financial crisis.

3.1 Statistical properties of the VIX time series

In this section, we attempt to characterize some of the statistical properties of the daily VIX index. Table 1 documents the results of our preliminary descriptive analyses. In particular, it reports the sample mean, standard deviation, minimum, first quartile, median, third quartile, maximum, and skewness for the VIX index time series (in logs) as well as the p-value of the Jarque-Bera test for normality. These descriptive statistics do not seem to change much according to the sample despite the seemingly different regimes in Figure 1. The only exception is the skewness coefficient, which substantially increases in the second half of the sample. As expected, the VIX time series is very skewed to the right, leptokurtic, and far from Gaussian. Applying a log transformation to the VIX index solves most of the excessive kurtosis, though a good deal of skewness (and hence nonnormality) remains.

Table 1 also evaluates the persistence of the VIX index through a battery of testing procedures. It reports the p-values of the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests for

unit root as well as the values of the KPSS test statistics for the null hypothesis of stationarity. We select the number of lags in the ADF test using the Bayesian information criterion, whereas we run the PP and KPSS tests using the quadratic spectral kernel with bandwidth choice as in Andrews (1991). Finally, Table 1 displays the value of the rescaled variance (V/S) test statistic for long memory (Giraitis et al., 2003).

We strongly reject the null hypothesis of a unit root for the VIX index with the ADF and PP tests in each half of the sample as well as in the full sample. Similarly, the KPSS test cannot reject the null of stationarity for both subsamples as well as for the full sample. However, there is strong evidence of long memory. The V/S test easily rejects the null of short memory for the VIX index. The sample autocorrelation and partial autocorrelation functions in Figure 2 corroborates our story in that the VIX series displays a highly persistent nature. The values of the sample autocorrelation function remain highly significant up to lag 500, though the partial correlation function seems to die out very fast. This explains the long swings in Figure 1.

3.2 Modeling the volatility index

Corsi (2009) argues that HAR specifications are particularly suitable to modeling and forecasting both realized and implied volatilities because they are able to capture the long-range dependence that arises from the asymmetric propagation of volatility between long and short horizons.¹ The HAR model implicitly assumes an additive cascade of different partial volatilities generated by the actions of distinct types of market participants (Müller et al., 1993). At each level of the cascade (or time scale), the corresponding unobserved partial volatility process is a function not only of its past value, but also of the expected values of the other partial volatilities. Corsi (2009) shows by straightforward recursive substitutions of the partial volatilities that this additive structure for the volatility cascade leads to a simple restricted linear autoregressive model featuring volatilities realized over different time horizons. The heterogeneous nature of the model derives from the fact that, at each time scale, the partial volatility relies on different autoregressive structures.

Let $\bar{y}_t^{(h)} = \frac{1}{h} \sum_{s=1}^h y_{t-s+1}$ and define $\mathbf{x}_t = \left(1, \bar{y}_t^{(\iota_1)}, \dots, \bar{y}_t^{(\iota_p)}\right)' \in \mathbb{R}^{p+1}$ for some vector of indexes $\boldsymbol{\iota} = (\iota_1, \dots, \iota_p)' \in \mathbb{Z}_+^p$. The time series $\{y_t, 1 \leq t \leq T\}$ then follows a HAR model if $y_t = \boldsymbol{\beta}' \mathbf{x}_{t-1} + \varepsilon_t$, where ε_t denotes a generic (weak) white noise. A typical choice in the literature for the index vector is $\boldsymbol{\iota} = (1, 5, 22)'$ so as to mirror the daily, weekly, and monthly components of the volatility process. In this paper, we augment the index vector by also including a biweekly and a quarterly component,

¹ We indeed find that ARFIMA models perform very poorly for the VIX index both in-sample and out-of-sample. The problem is that ARFIMA models impose a linear form of long memory that depends exclusively on a single parameter, i.e., the fractional integration order (Abadir and Talmain, 2002; Bhardwaj and Swanson, 2006).

so that $\iota = (1, 5, 10, 22, 66)'$.

We consider three variations of the HAR specification. The first includes a set of additional regressors \mathbf{z}_t such that

$$y_t = \beta' \mathbf{x}_{t-1} + \gamma' \mathbf{z}_t + \varepsilon_t, \quad (2)$$

where $\mathbf{z}_t = (z_{1t}, \dots, z_{Kt})$ is a K -dimensional vector of explanatory variables. Among the latter, we include the following macro-finance variables (both contemporaneously and with one lag): the κ -day continuously compounded return on the S&P 500 index for $\kappa = 1, 5, 10, 22, 66$ (S&P 500 κ -day return); the first difference of the logarithm of the volume of the S&P 500 index (S&P 500 volume change); the κ -day continuously compounded return on the one-month crude oil futures contract (oil κ -day return); the first difference of the logarithm of the trade-weighted average of the foreign exchange value of the US dollar index against the Australian dollar, Canadian dollar, Swiss franc, euro, British sterling pound, Japanese yen, and Swedish kroner (USD change); the excess yield of the Moody's seasoned Baa corporate bond over the Moody's seasoned Aaa corporate bond (credit spread); the difference between the 10-Year and 3-month treasury constant maturity rates (term spread); and the difference between the effective and target Federal Fund rates (FF deviation). We refer to (2) as the HARX specification.

The motivation for using S&P 500 returns is to take into account possible leverage and volatility feedback effects (Bollerslev and Zhou, 2006). Figure 3 evinces that there seems to exist a negative (virtually linear) link between changes in the VIX index and the S&P 500 index return, in line with the evidence in the literature (Fleming et al., 1995; Giot, 2005). Note that we include multiperiod returns on the S&P 500 index so as to comply with the HAR nature of the model. In addition, given the well-documented relationship between volume and volatility (Lamoureux and Lastrapes, 1990; Chan and Fong, 2000; Becker et al., 2007), we also add the S&P 500 volume change to the set of explanatory variables. The remaining regressors are all linked to different dimensions of the overall market conditions in the US. Both oil prices and term spread contain information about the future real economic activity (Estrella and Hardouvelis, 1991) as well as about future investment opportunities (Petkova, 2006). The credit spread gauges to some extent the amount of liquidity in the market, whereas USD change and FF deviation are both related to the macroeconomic conditions in the US.

The second variant is an HAR-type specification that controls for explanatory variables with asymmetric effects. The chief motivation is to capitalize on the seemingly asymmetric relation between the VIX index and the S&P 500 index returns. In particular, the AHARX model is given

by

$$y_t = \beta' \mathbf{x}_{t-1} + \gamma_0' \mathbf{z}_t^{(-)} + \gamma_1' \mathbf{z}_t^{(+)} + \varepsilon_t, \quad (3)$$

where $\mathbf{z}_t^{(-)} = \{z_{1t}^{(-)}, \dots, z_{Kt}^{(-)}\}$ and $\mathbf{z}_t^{(+)} = \{z_{1t}^{(+)}, \dots, z_{Kt}^{(+)}\}$, with

$$z_{kt}^{(-)} = \begin{cases} z_{kt} \mathbf{1}(z_{kt} < 0) & \text{if a return} \\ z_{kt} \mathbf{1}(\Delta z_{kt} < 0) & \text{if in levels} \end{cases} \quad z_{kt}^{(+)} = \begin{cases} z_{kt} \mathbf{1}(z_{kt} > 0) & \text{if a return} \\ z_{kt} \mathbf{1}(\Delta z_{kt} > 0) & \text{if in levels.} \end{cases}$$

To avoid an excessive number of parameters to estimate, we exclude the contemporaneous values of the multiperiod returns from the regression. Finally, we also consider a semiparametric specification that captures more general forms of nonlinear dependence through a neural network approximation. The motivation rests on the typical success that neural networks experience in the context of volatility modeling and forecasting (Donaldson and Kamstra, 1997; Hu and Tsoukalas, 1999; Hamid and Iqbal, 2004). As in Hillebrand and Medeiros (2010), we specify our neural network HARX (NNHARX) model as

$$y_t = \beta_0' \mathbf{x}_{t-1} + \gamma_0' \mathbf{z}_t + \sum_{m=1}^M \frac{\lambda_m}{1 + e^{-\beta_m' \mathbf{x}_{t-1} - \gamma_m' \mathbf{z}_t}} + \varepsilon_t. \quad (4)$$

We estimate the semiparametric NNHARX model using Bayesian regularization with M set either to 3 or 10. The results are very robust to changes in M and hence we report only those corresponding to the more parsimonious model with $M = 3$. The small number M of hidden units in the neural network is not surprising given that we find little evidence of nonlinear dependence in our empirical analysis.

Tables 2 and 3 report the least-square estimates of the (A)HARX coefficients as well as their heteroskedasticity-consistent standard errors. To conserve on space, we omit the coefficient estimates of the HAR components given that they are very stable across the different specifications. The first and last terms ($h = 1, 66$) are both significant at the 1% level, at about 0.90 and 0.03, respectively. The remaining terms are also jointly significant at the 1% level, even if not always individually insignificant despite their magnitudes (0.04 for $h = 5$, 0.05 for $h = 10$, and 0.003 for $h = 22$). Table 2 also documents the corresponding average partial effects and their 95% bootstrap-based confidence intervals for the semiparametric NNHARX model. To capture any sort of dependence structure in the data, including in the high-order moments, we employ a block bootstrap algorithm in which we form the artificial samples by resampling blocks of approximately $T^{1/3}$ observations. The results are both qualitatively and quantitatively very similar across specifications.

There is a strong negative link between the contemporaneous and lagged S&P 500 index 1-day return and the VIX index, whereas we find no significant influence from the S&P 500 index multiperiod returns unless we control for asymmetric effects. In addition, the positive volume effect is exclusively contemporaneous. This results is robust to the choice of the functional form of the model. The only oil-related regressor that seems to matter is the lagged one-day return on the one-month crude oil futures contract. All the other variables seem to have no impact on the VIX dynamics. It is worth stressing that the Lagrange multiplier tests for autocorrelation up to lag ℓ (with $\ell = 1, 5, 10, 22$) suggest no evidence of dynamic misspecification regardless of the specification we use. This is reassuring because the latter normally entails inconsistent coefficient estimates in the joint presence of residual autocorrelation and lagged values of the dependent variable among the regressors. Further analysis show that, as expected, there is some strong evidence of conditional heteroskedasticity. That is why we employ heteroskedasticity-consistent standard errors and a block bootstrap algorithm to compute the confidence intervals of the NNHARX partial average effects.

Including contemporaneous regressors obviously raises endogeneity issues. This is especially problematic for measuring the contemporaneous volume-volatility relation given that presumably both are jointly driven by the latent information arrival process (Lamoureux and Lastrapes, 1994).² We thus estimate the HARX model with only contemporaneous regressors by instrumental variables to see whether there is any change in the qualitative results. In particular, we instrument trading volume, credit spread, term spread, and FF deviation by their lagged values so as to ensure that instruments are strong enough. It is evident from Table 4 that the results do not change much qualitatively. All pointwise coefficient estimates are of similar magnitude to the sum of the least-squares estimates of the contemporaneous and lagged coefficients. The only difference is the negative effect of the term spread which becomes statistically significant. Note that we do not instrument the return-based regressors with their past values in order to avoid problems with weak instruments. Unreported results show that adding past returns as additional instruments yields insignificant coefficient estimates, even if very large in magnitudes, for the one-day returns on the S&P 500 index and on the one-month crude oil futures contract as well as for the changes in the USD index. There are no qualitative changes for the pointwise estimates of the other coefficients, but their standard errors increase in a substantial manner. This is not surprising given that past returns are very weak instruments for current returns. All in all, it seems that endogeneity is only a minor issue

² Nevertheless, there is a vast literature that employs trading volume as a natural proxy for the information arrival process, implicitly assuming that it is weakly exogenous. See, among others, Asmati and Pfleiderer (1988), Lamoureux and Lastrapes (1990), and He and Wang (1995).

here and hence we do not attempt to tackle the very daunting task of controlling for endogeneity in the estimation of the NNHARX model.

Another concern is multicollinearity. To check whether multicollinearity is responsible for the insignificant coefficient estimates we show the results of stepwise regression. Table 4 reports the results. They are very similar, corroborating the previous evidence of little endogeneity bias. As before, there is no qualitative change in the main results. The only exception is the significant and negative effect of the term spread when multicollinearity is controlled for. Finally, further analyses show that similar qualitative results ensue for the NNHARX specification as well as if we exclude insignificant regressors using a simple general-to-specific approach.

Given the amount of data persistence, it is not surprising that most of the explanatory power comes from the previous day value of the VIX index, with a partial R^2 around 50% across the different specifications. However, the one-day return on the S&P 500 index also entails a sizeable partial R^2 of about 18% if one considers both contemporaneous and lagged effects in the HARX specification. The next variable with most explanatory power is trading volume with a partial R^2 of 0.92%, whereas the term spread and the changes in the USD index add up to 0.42% and 0.12%, respectively. The partial coefficients of determination for the remaining regressors are all below 0.10%.

In the next section, we turn our attention to predictive regressions that exclude all contemporaneous terms for forecasting purposes. We employ a rolling window of 1,000 observations to estimate the regression coefficients of HARX, HARX, AHARX and NNHARX specifications and then assess their out-of-sample performance in the remainder of the sample by looking at ℓ -day ahead forecasts, with $\ell \in \{1, 5, 10, 22\}$.³ Given the above partial R^2 coefficients, it is evident that simply including additional explanatory variables will probably not suffice to improve the forecasting ability of the HAR model.

3.3 Forecasting the volatility index

Table 5 displays some descriptive results of the out-of-sample evaluation for forecasts 1, 5, 10, and 22 days ahead, respectively. In particular, we report the mean (MFE) and standard deviation (SDFE) of the forecast errors as well as the corresponding coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and the p-value of the test of predictive ability in terms of the MSE - SPE (SE) - and MAE - SPE (AE) - criteria (Hansen, 2005). The null

³ Note that to compute ℓ -day ahead forecasts, we employ a direct forecast approach in which we replace y_t with $y_{t+\ell-1}$ in the above models. This allows us to produce multi-step ahead forecasts without imposing any assumption about future realizations of the explanatory variables.

hypothesis of the latter is that the best alternative forecasting model does not outperform the benchmark model. As such, small p-values indicate that there is at least one alternative forecasting models with superior predictive ability and hence, ideally, we would like to find forecasting models with p-values exceeding the usual levels of significance at every horizon. Apart from the HAR-type models, we also include the forecasting results for a random walk with drift and ARX(1) specifications.⁴

In terms of bias, the results are mixed and there is no dominating model. The mean forecast errors are very close to zero for all specifications, implying virtually negligible contributions to the MSE. As for the standard deviation of the forecast errors, the pure HAR model seems to perform very well, confirming persistence as the most prominent feature of the VIX index. It consistently beats the random walk and the ARX models at every horizon. It also fares well relative to the HARX and AHARX specifications across every forecasting horizon as well as to the semiparametric NNHARX model for all but the 22-day ahead forecast. Comparing the NNHARX model against the linear HARX alternative, the former has a superior performance in terms of standard deviation for all horizons with the exception of five days ahead.

The relative success of the HAR model is not surprising. The VIX index measures the market expectations about the risk-neutral volatility 30 calendar days ahead, so that the overlapping implied by the daily frequency contributes to the strong persistence in the data. After 22 trading days (about 30 calendar days), the overlapping effect disappears, reducing persistence and increasing the relative contribution of the macro-finance factors. Accordingly, we also observe a drop of about 19% in the coefficient of determination once we move from 10-day to 22-day forecasts. This is a decline of dramatic proportion if one compares to the reduction of only about 7% from 1-day to 5-day forecasts and from 5-day to 10-day forecasts. As persistence subsides, the coefficient of determination is bound to decrease.

The MSE and MAE criteria tell virtually the same story. The HAR entails the best 1-day forecast results. The performances of the different models are again very similar for the 10-day-ahead forecasts, especially for the MAE criterion. The only striking difference resides in the 22-day forecasts, for which the NNHARX outperforms every other specification. However, according to the SPA test, apart from the random walk model, which is consistently outperformed in all horizons by the alternatives, all the other models seems to have a similar performance.

Although the NNHARX specification seemingly have many more parameters to estimate, the

⁴ Note that we compute the drift of the random walk recursively so as to capture the long swings in the VIX index. However, excluding the drift entails only a negligible (mostly negative) impact in the forecasting performance.

Bayesian regularization automatically shrinks the average partial effect of the insignificant coefficients to zero, therefore controlling the precision of the overall estimation/forecasting exercise. Figure 4 illustrates well this point. The average partial effects of the NNHARX model are typically within the confidence bands of the HARX coefficient estimates.

We complement the above results by running the unconditional Giacomini-White test for the mean absolute forecast error proposed by Giacomini and White (2006).⁵ Table 6 reports the p-values for testing the null hypothesis that the column and row models perform equally well in terms of mean absolute forecast error. The NNHARX model performs significantly better than the other models at the 1-day and 22-day horizons at the 10% level, though we cannot reject that most models forecast equally well the VIX index 5 and 10 days ahead.

Figure 5 shows the differences in the one-step-ahead cumulative squared errors for the out-of-sample period. The upper panel reports the difference between the linear HAR and the NNHARX models. Negative values indicate that the HAR model has lower cumulative squared errors. The lower panel reports the results concerning the HARX and the NNHARX models. Again, negative figures indicate that the HARX specification has lower squared errors. As can be seen, the differences are rather small. However, it seems that the NNHARX dominates the linear HARX. As the evidence of nonlinearity is not strong, we interpret this fact as a result of variable selection. Note that the Bayesian regularization technique is pruning the coefficients associated to irrelevant regressors towards zero.

Finally, it is reasonable to ask what happens if we exclude insignificant explanatory variables from the HARX, AHARX and NNHARX predictive regressions. After all, many of the coefficient estimates we report in Table 2 are not statistically different from zero. The answer is ‘not much’. The mean forecast errors of the (A)HARX models increase by about 40%, but their standard deviations decline as well. As a result, their mean squared error decreases by at most 10%. As for the NNHARX model, the net effect in the mean squared error is about zero. This is not very surprising given that the regularization method we employ automatically does model selection.

4 Conclusion

This paper examines the time-series properties of the CBOE’s market volatility index (VIX) at the daily frequency. As expected, preliminary analysis unearths strong evidence that the VIX time

⁵ We find no qualitative change in the results if we consider the conditional predictive ability test of Giacomini and White (2006) using the information set spanned by the lag values of the explanatory variables as well as of the loss-function difference. We employ Newey-West standard errors in both tests so as to account for possible heteroskedasticity in the VIX index.

series displays long-range dependence and so we employ HAR-type processes for modeling and forecasting purposes. In particular, we employ a pure HAR specification as well as both parametric and semiparametric HAR-type models that also use the information coming from several macro-finance variables. Among the latter, we include multiperiod returns on the S&P 500 index and on the one-month oil futures contract as well as the change in the volume of the S&P 500 index, the credit and term spreads, the change in the foreign value of the US dollar, and the difference between the effective and target Federal Fund rates.

The VIX index seems to depend neither on the deviation in the Fed rates nor on the credit spread. It however holds a very strong negative relationship with the S&P 500 index returns as well as a positive link with the contemporaneous S&P 500 volume change. In addition, both the 66-day return on the 1-month oil futures contract and the term spread entail a slightly negative long-run impact in the VIX index, whereas changes in the USD index significantly affects the VIX index in a positive linear fashion. However, the latter effect does not remain significant if one controls for nonlinear dependence of unknown form. Interestingly, this is the only link for which accounting for nonlinearity actually matters. All of the other relationships hold with similar magnitudes regardless of whether we take a semiparametric route.

As per the forecasting results, it turns out that it is very difficult to beat the forecasting performance of the pure HAR process because of the very persistent nature of the VIX index. This is partly due to the daily sampling frequency. Given that the VIX index reflects the market expectations about the stock market volatility 30 calendar days ahead, looking at daily figures implies a certain degree of overlapping that exacerbates data persistence. As a consequence, persistence becomes almost the only feature that matters for forecasting purposes at short horizons. In particular, this explains why exploiting the macro-finance information becomes relatively more valuable as the forecasting horizon approaches the 30 calendar days ahead threshold. Altogether, we find that the semiparametric NNHARX model performs as well as the linear HAR for all forecasting horizons.

References

- Abadir, K. M., Talmain, G., 2002, Aggregation, persistence and volatility in a macro model, *Review of Economic Studies* 69, 749–779.
- Andersen, T. G., Bondarenko, O., 2007, Construction and interpretation of model-free implied volatility, in: I. Nelken (ed.), *Volatility as an Asset Class*, Risk Publications, London.

- Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.
- Asmati, A., Pflereider, P., 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3–40.
- Bandi, F. M., Perron, B., 2006, Long memory and the relation between implied and realized volatility, *Journal of Financial Econometrics* 4, 636–670.
- Becker, R., Clements, A. C., White, S., 2007, Does implied volatility provide any information beyond that captured in model-based volatility forecasts?, *Journal of Banking and Finance* 31, 2535–2549.
- Bhardwaj, G., Swanson, N. R., 2006, An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series, *Journal of Econometrics* 131, 539–578.
- Blair, B. J., Poon, S.-H., Taylor, S. J., 2001a, Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns, *Journal of Econometrics* 105, 5–26.
- Blair, B. J., Poon, S.-H., Taylor, S. J., 2001b, Modelling S&P 100 volatility: The information content of stock returns, *Journal of Banking and Finance* 25, 1665–1679.
- Bollerslev, T., Zhou, H., 2006, Volatility puzzles: A simple framework for gauging return-volatility regression, *Journal of Econometrics* 131, 123–150.
- Britten-Jones, M., Neuberger, A., 2000, Option Prices, implied price processes, and stochastic volatility, *Journal of Finance* 55, 839–866.
- Canina, L., Figlewski, S., 1993, The informational content of implied volatility, *Review of Financial Studies* 6, 659–681.
- Carr, P., Lee, R., 2009, Volatility derivatives, *Annual Review of Financial Economics* 1, 319–339.
- Chan, K., Fong, W.-M., 2000, Trade size, order imbalance, and the volatility-volume relation, *Journal of Financial Economics* 57, 247–273.
- Christensen, B. J., Prabhala, N. R., 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125–150.

- Clements, A. C., Fuller, J., 2012, Forecasting increases in the VIX: A time-varying long volatility hedge for equities, NCER working paper series.
- Corsi, F., 2009, A simple approximate long memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.
- Cox, J. C., Rubinstein, M., 1985, *Options Markets*, Prentice Hall, New Jersey.
- Demeterfi, K., Derman, E., Kamal, M., Zou, J., 1999, More than you ever wanted to know about volatility swaps, *Journal of Derivatives* 6, 9–32.
- Donaldson, R. G., Kamstra, M., 1997, An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance* 4, 17–46.
- Estrella, A., Hardouvelis, G., 1991, The term structure as a predictor of real economic activity, *Journal of Finance* 46, 555–576.
- Fleming, J., 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317–345.
- Fleming, J., Ostdiek, B., Whaley, R. E., 1995, Predicting stock market volatility: A new measure, *Journal of Futures Markets* 15, 265–302.
- Gastineau, G. L., 1977, An index of listed option premiums, *Financial Analysts Journal* 30, 70–75.
- Giacomini, R., White, H., 2006, Tests of conditional predictive ability, *Econometrica* 74, 1545–1578.
- Giot, P., 2005, Relationships between implied volatility indices and stock index returns, *Journal of Portfolio Management* 31, 92–100.
- Giraitis, L., Kokoszka, P., Leipus, R., Teyssière, G., 2003, Rescaled variance and related tests for long memory in volatility and levels, *Journal of Econometrics* 112, 265–294.
- Hamid, S. A., Iqbal, Z., 2004, Using neural networks for forecasting volatility of S&P 500 Index futures prices, *Journal of Business Research* 57, 1116–1125.
- Hansen, P. R., 2005, A test for superior predictive ability, *Journal of Business and Economic Statistics* 23, 365–380.
- He, H., Wang, J., 1995, Differential information and the dynamic behavior of stock trading volume, *Review of Financial Studies* 8, 919–972.

- Hentschel, L., 2003, Errors in implied volatility estimation, *Journal of Financial and Quantitative Analysis* 38, 779–810.
- Hillebrand, E., Medeiros, M. C., 2010, The benefits of bagging for forecast models of realized volatility, *Econometric Reviews* 29, 571–593.
- Hu, M. Y., Tsoukalas, C., 1999, Combining conditional volatility forecasts using neural networks: An application to the EMS exchange rates, *Journal of International Financial Markets, Institutions and Money* 9, 407–422.
- Jiang, G., Tian, Y., 2005, Model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305–1342.
- Jorion, P., 1995, Predicting volatility in the foreign exchange market, *Journal of Finance* 50, 507–528.
- 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.
- Koopman, S. J., Jungbacker, B., Hol, E., 2005, Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements, *Journal of Empirical Finance* 12, 445–475.
- Lamoureux, C. G., Lastrapes, W. D., 1990, Heteroskedasticity in stock return data: Volume versus GARCH effects, *Journal of Finance* 45, 221–229.
- Lamoureux, C. G., Lastrapes, W. D., 1994, Endogenous trading volume and momentum in stock-return volatility, *Journal of Business and Economic Statistics* 12, 253–260.
- Martens, M., Zein, J., 2004, Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility, *Journal of Futures Markets* 24, 1005–1028.
- Müller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O., von Weizsacker, J., 1997, Volatilities of different time resolutions: Analysing the dynamics of market components, *Journal of Empirical Finance* 4, 213–239.
- Müller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O., Ward, J., 1993, Fractals and intrinsic time: A challenge to econometricians, *Proceedings of the XXXIX International AEA Conference on Real Time Econometrics*.

- Petkova, R., 2006, Do the Fama-French factors proxy for innovations in predictive variables?, *Journal of Finance* 61, 581–612.
- Taylor, S. J., Xu, X., 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317–340.
- Whaley, R. E., 2000, The investor fear gauge, *Journal of Portfolio Management* 26, 12–17.
- Xu, X., Taylor, S. J., 1995, Conditional volatility and the informational efficiency of the PHLX currency options markets, *Journal of Banking and Finance* 19, 803–821.

Table 1: **Descriptive statistics for the logarithm of the VIX index**

The sample period runs from January 2, 1990 to January 15, 2013, including altogether 5,807 time-series observations. We report the sample mean, median, minimum, maximum, standard deviation, skewness, and kurtosis for the logarithm of the VIX time series, as well as the p-values of the Jarque-Bera test for normality and of the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests for unit root. In addition, we also report the values of the KPSS test statistic for the null hypothesis of stationarity, whose critical values are 0.347, 0.463, and 0.739 at the 10%, 5%, and 1% significance levels, respectively. We select the number of lags in the ADF test using the Bayesian information criterion, whereas we carry out the PP and KPSS tests using the quadratic spectral kernel with bandwidth choice as in Andrews (1991). Finally, V/S refers to the value of the rescaled variance test statistic for long memory by Giraitis et al. (2003). The critical values of the V/S test are 1.36 and 1.63 at the 5% and 1% levels, respectively.

sample statistics	first half	second half	full sample
mean	2.9074	2.9984	2.9529
median	2.9096	2.9627	2.9370
minimum	2.2311	2.2915	2.2311
maximum	3.8230	4.3927	4.3927
standard deviation	0.3002	0.3847	0.3480
skewness	0.1703	0.5778	0.5385
kurtosis	2.3086	3.1782	3.2876
Jarque-Bera	0.0000	0.0000	0.0000
ADF	0.0001	0.0008	0.0000
PP	0.0000	0.0024	0.0000
KPSS	0.5060	0.1337	0.2098
V/S	8.2000	5.7399	5.1784

Table 2: Modeling the logarithm of the VIX index

The sample period runs from January 2, 1990 to January 15, 2013, including altogether 5,807 time-series observations. The first column lists the additional regressors we use apart from the day-of-the-week dummies and the average of the logarithm of the VIX index over the last $k \in \{1, 5, 10, 22, 66\}$ days. S&P500 k -day return is the k -day log-return on the S&P500 index; S&P500 volume change is the first difference of the logarithm of the volume of the S&P500 index; oil k -day return is the k -day log-return on the one-month crude oil futures contract; USD change is the first difference of the logarithm of the foreign exchange value of the U-S dollar index; credit spread is the excess yield of the Moody's seasoned Baa corporate bond over the Moody's seasoned Aaa corporate bond; term spread is the difference between the 10-Year and 3-month treasury constant maturity rates; and FF deviation is the difference between the effective and target Federal Funds rates. For the HARX specification, we provide the point estimates for the coefficients as well as their heteroskedasticity-consistent standard errors within parentheses, whereas we report average partial effects for the semiparametric NNHARX model with their corresponding 95% confidence intervals based on a block bootstrap algorithm.

	HARX		NNHARX	
	lag 0	lag 1	lag 0	lag 1
S&P500 1-day return	-3.658 (0.088)	-0.208 (0.076)	-3.6000 [-3.840,-2.170]	-0.199 [-0.315,0.0653]
S&P500 5-day return		-0.017 (0.047)		-0.015 [-0.118,0.112]
S&P500 10-day return		0.013 (0.036)		0.011 [-0.069,0.095]
S&P500 22-day return		-0.034 (0.021)		-0.035 [-0.083,0.008]
S&P500 66-day return		0.007 (0.010)		0.007 [-0.015,0.035]
S&P500 volume change	0.025 (0.004)	0.005 (0.003)	0.025 [0.010,0.035]	0.005 [-0.008,0.012]
oil 1-day return	0.043 (0.027)	0.047 (0.026)	0.042 [-0.044,0.097]	0.047 [0.001,0.123]
oil 5-day return		0.001 (0.015)		0.001 [-0.038,0.029]
oil 10-day return		-0.018 (0.012)		-0.018 [-0.043,0.010]
oil 22-day return		0.011 (0.008)		0.012 [-0.006,0.072]
oil 66-day return		-0.001 (0.004)		-0.001 [-0.012,0.008]
USD change	-0.044 (0.140)	0.118 (0.135)	-0.027 [-0.282,0.284]	0.124 [-0.107,0.386]
credit spread	0.022 (0.034)	-0.022 (0.034)	0.018 [-0.056,0.086]	-0.018 [-0.086,0.056]
term spread	0.014 (0.011)	-0.015 (0.011)	0.014 [-0.011,0.040]	-0.015 [-0.043,0.010]
FF deviation	0.001 (0.003)	-0.003 (0.003)	0.000 [-0.020,0.007]	-0.003 [-0.009,0.007]

Table 3: **Asymmetric effects in the VIX index**

The sample details are as in Table 2. We report the coefficient estimates of the AHARX model, with their heteroskedasticity-consistent standard errors within parentheses.

	lag 0		lag 1	
	positive	negative	positive	negative
S&P500 1-day return	-2.613 (0.137)	-4.659 (0.180)	0.163 (0.104)	-0.581 (0.138)
S&P500 5-day return			0.095 (0.075)	-0.114 (0.082)
S&P500 10-day return			0.073 (0.045)	0.016 (0.061)
S&P500 22-day return			0.025 (0.028)	-0.065 (0.039)
S&P500 66-day return			0.053 (0.013)	-0.017 (0.021)
S&P500 volume change	0.034 (0.007)	0.007 (0.005)	0.000 (0.004)	0.008 (0.006)
oil 1-day return	0.038 (0.052)	0.091 (0.042)	0.062 (0.052)	0.038 (0.040)
oil 5-day return			0.002 (0.023)	0.003 (0.024)
oil 10-day return			-0.035 (0.019)	0.000 (0.021)
oil 22-day return			0.008 (0.013)	0.013 (0.014)
oil 66-day return			0.000 (0.006)	0.011 (0.007)
USD change	0.208 (0.266)	-0.301 (0.233)	-0.187 (0.252)	0.303 (0.221)
credit spread	0.000 (0.001)	-0.002 (0.001)	-0.003 (0.001)	-0.001 (0.001)
term spread	0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)
FF deviation	0.006 (0.004)	-0.009 (0.006)	0.003 (0.004)	-0.003 (0.007)

Table 4: **Robustness checks: Endogeneity and multicollinearity issues**

The sample details are as in Table 2. We provide the IV and/or Stepwise-OLS coefficient estimates as well as their heteroskedasticity-consistent standard errors within parentheses for the different HARX specifications. The first specification considers only contemporaneous regressors using lagged values as instruments for the trading volume, credit spread, term spread, and FF deviation. To also account for multicollinearity, we estimate by Stepwise-OLS and the IV specification excluding the contemporaneous multi-day returns.

	IV	Stepwise		IV	
	lag 0	lag 0	lag 1	lag 0	lag 1
S&P500 1-day return	-3.583 (0.103)	-3.657 (0.046)	-0.217 (0.055)	-3.656 (0.089)	-0.201 (0.076)
S&P500 5-day return	-0.052 (0.051)				-0.022 (0.048)
S&P500 10-day return	0.024 (0.037)				0.012 (0.036)
S&P500 22-day return	-0.050 (0.021)				-0.034 (0.021)
S&P500 66-day return	0.010 (0.010)				0.008 (0.010)
S&P500 volume change	0.014 (0.007)	0.023 (0.002)		0.023 (0.003)	
oil 1-day return	0.036 (0.028)	0.052 (0.022)		0.049 (0.027)	0.045 (0.026)
oil 5-day return	0.022 (0.015)				0.001 (0.014)
oil 10-day return	-0.024 (0.012)				-0.019 (0.012)
oil 22-day return	0.011 (0.008)				0.011 (0.009)
oil 66-day return	0.000 (0.004)				-0.001 (0.004)
USD change	-0.010 (0.141)			-0.042 (0.142)	
credit spread	-0.001 (0.002)			-0.001 (0.002)	
term spread	-0.001 (0.000)		-0.001 (0.000)	-0.001 (0.000)	
FF deviation	-0.007 (0.008)			-0.005 (0.008)	

Table 5: **Forecasting performance at different horizons**

The sample period runs from January 2, 1990 to January 15, 2013, including altogether 5,807 observations. We use a rolling window of 2,500 time-series observations to estimate the different models and then perform out-of-sample forecasting evaluation in the remaining of the series. We consider the following specifications: random walk with drift (RW), heterogeneous autoregression (HAR), heterogeneous autoregression with exogenous variables (HARX), heterogeneous autoregression with exogenous variables and asymmetric effects (AHARX), and the neural-network heterogeneous autoregression with exogenous variables (NNHARX). We gauge forecasting performance by means of the mean forecast error (MFE), the standard deviation of the forecast error (SDFE), the mean squared forecast error (MSE), the mean absolute forecast error (MAE), and the Mincer-Zarnowitz coefficient of determination (R^2). We also report the p-value of the test of superior predictive ability in terms of the MSE - SPE (SE) - and MAE - SPE (AE) - criteria (Hansen, 2005). Rejection due to a low p-value means that there is at least one alternative forecasting model with superior predictive ability at that particular horizon.

	MFE	SDFE	MSE	SPA (SE)	MAE	SPE (AE)	R^2
ONE DAY AHEAD							
RW	-0.0002	0.0628	0.0039	0	0.0456	0	0.9715
ARX	-0.0003	0.0623	0.0039	0.0815	0.0447	0.5245	0.9718
HAR	-0.0003	0.0618	0.0038	0.8625	0.0445	0.8435	0.9722
HARX	0.0001	0.0621	0.0039	0.2380	0.0446	0.8275	0.9720
AHARX	0.0001	0.0623	0.0039	0.0325	0.0447	0.3780	0.9718
NNHARX	0.0000	0.0620	0.0038	0.2935	0.0446	0.7510	0.9720
FIVE DAYS AHEAD							
RW	-0.0007	0.1188	0.0141	0.0010	0.0891	0.0355	0.9000
ARX	-0.0000	0.1164	0.0135	0.2660	0.0876	0.3475	0.9017
HAR	-0.0011	0.1153	0.0133	0.8190	0.0873	0.6950	0.9034
HARX	0.0017	0.1158	0.0134	0.6665	0.0871	0.9275	0.9028
AHARX	-0.0002	0.1160	0.0135	0.4380	0.0872	0.7820	0.9024
NNHARX	0.0016	0.1160	0.0135	0.3695	0.0874	0.0595	0.9024
TEN DAYS AHEAD							
RW	-0.0006	0.1458	0.0212	0.3275	0.1105	0.2655	0.8276
ARX	0.0057	0.1455	0.0212	0.4700	0.1106	0.0985	0.8212
HAR	0.0017	0.1442	0.0208	0.8555	0.1098	0.3925	0.8237
HARX	0.0040	0.1454	0.0211	0.4940	0.1101	0.0600	0.8217
AHARX	0.0023	0.1453	0.0211	0.5250	0.1094	0.5685	0.8220
NNHARX	0.0066	0.1445	0.0209	0.7175	0.1088	0.8940	0.8229
TWENTY-TWO DAYS AHEAD							
RW	-0.0024	0.2072	0.0429	0.0060	0.1544	0.0310	0.7135
ARX	0.0004	0.2014	0.0406	0.5590	0.1500	0.3665	0.7092
HAR	-0.0033	0.2002	0.0401	0.7880	0.1502	0.2930	0.7108
HARX	0.0051	0.2015	0.0406	0.4705	0.1488	0.7935	0.7096
AHARX	0.0022	0.2030	0.0412	0.2265	0.1488	0.7520	0.7043
NNHARX	0.0059	0.1999	0.0400	0.8730	0.1484	0.8920	0.7119

Table 6: **Giacomini-White tests for the mean absolute forecast error**

The sample period runs from January 2, 1990 to January 15, 2013, including altogether 5,807 observations. We use a rolling window of 2,500 time-series observations to estimate the different models and then perform out-of-sample forecasting evaluation in the remaining of the series. We consider the following specifications: random walk with drift (RW), heterogeneous autoregression (HAR), heterogeneous autoregression with exogenous variables (HARX), heterogeneous autoregression with exogenous variables and asymmetric effects (AHARX), and the neural-network heterogeneous autoregression with exogenous variables (NNHARX). The p-values in each entry correspond to the modified Giacomini-White test for the null hypothesis that the column and row models perform equally well in terms of mean absolute forecast error.

	RW	ARX	HAR	HARX	AHARX
ONE DAY AHEAD					
ARX	0.0001				
HAR	0.0000	0.2253			
HARX	0.0000	0.2339	0.3488		
AHARX	0.0003	0.4108	0.1428	0.1402	
NNHARX	0.0001	0.2866	0.3244	0.4494	0.1852
FIVE DAYS AHEAD					
ARX	0.0699				
HAR	0.0314	0.3545			
HARX	0.0400	0.1630	0.3792		
AHARX	0.0428	0.2504	0.4155	0.4701	
NNHARX	0.0677	0.3928	0.4043	0.0173	0.2728
TEN DAYS AHEAD					
ARX	0.4937				
HAR	0.3177	0.3243			
HARX	0.4153	0.3150	0.4241		
AHARX	0.2960	0.2052	0.4124	0.2722	
NNHARX	0.2023	0.0801	0.2389	0.0314	0.3265
TWENTY-TWO DAYS AHEAD					
ARX	0.1389				
HAR	0.1224	0.4883			
HARX	0.1010	0.2534	0.3456		
AHARX	0.1134	0.3156	0.3274	0.4999	
NNHARX	0.0766	0.2549	0.2435	0.4156	0.4279

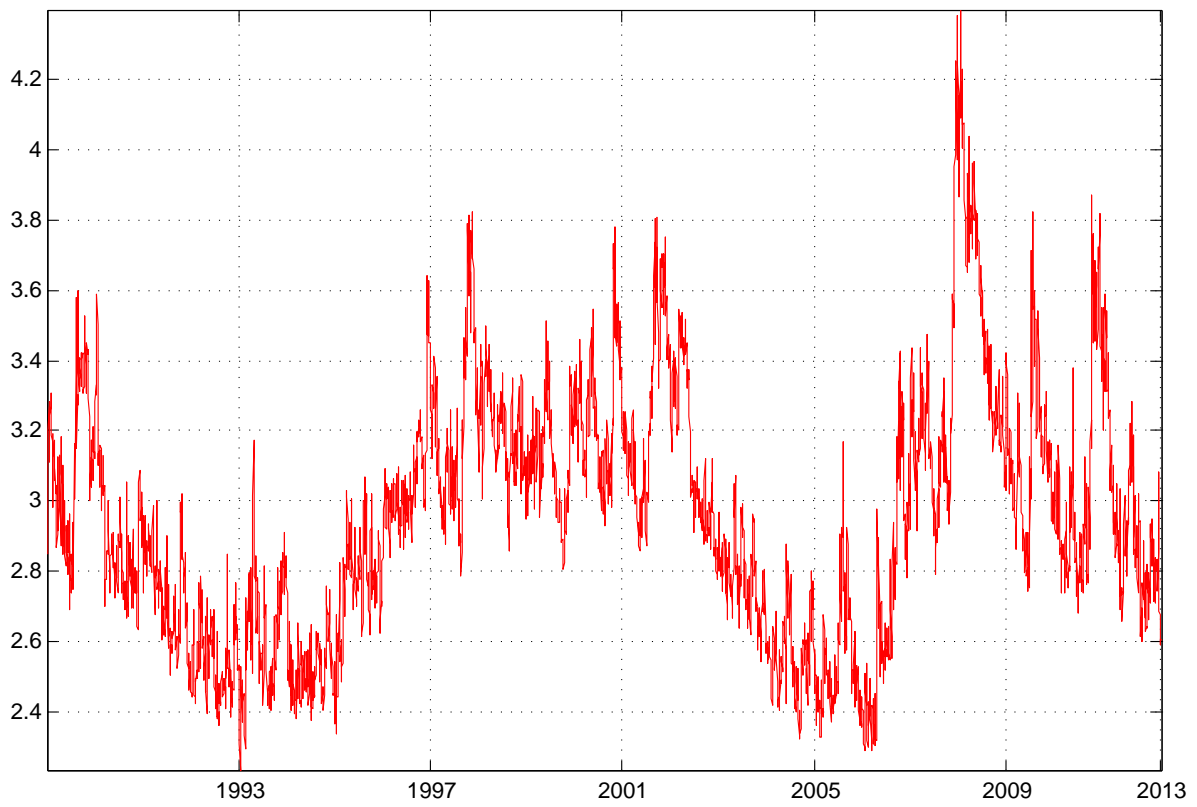


Figure 1: The daily VIX index from January 2, 1992 to January 15, 2013.

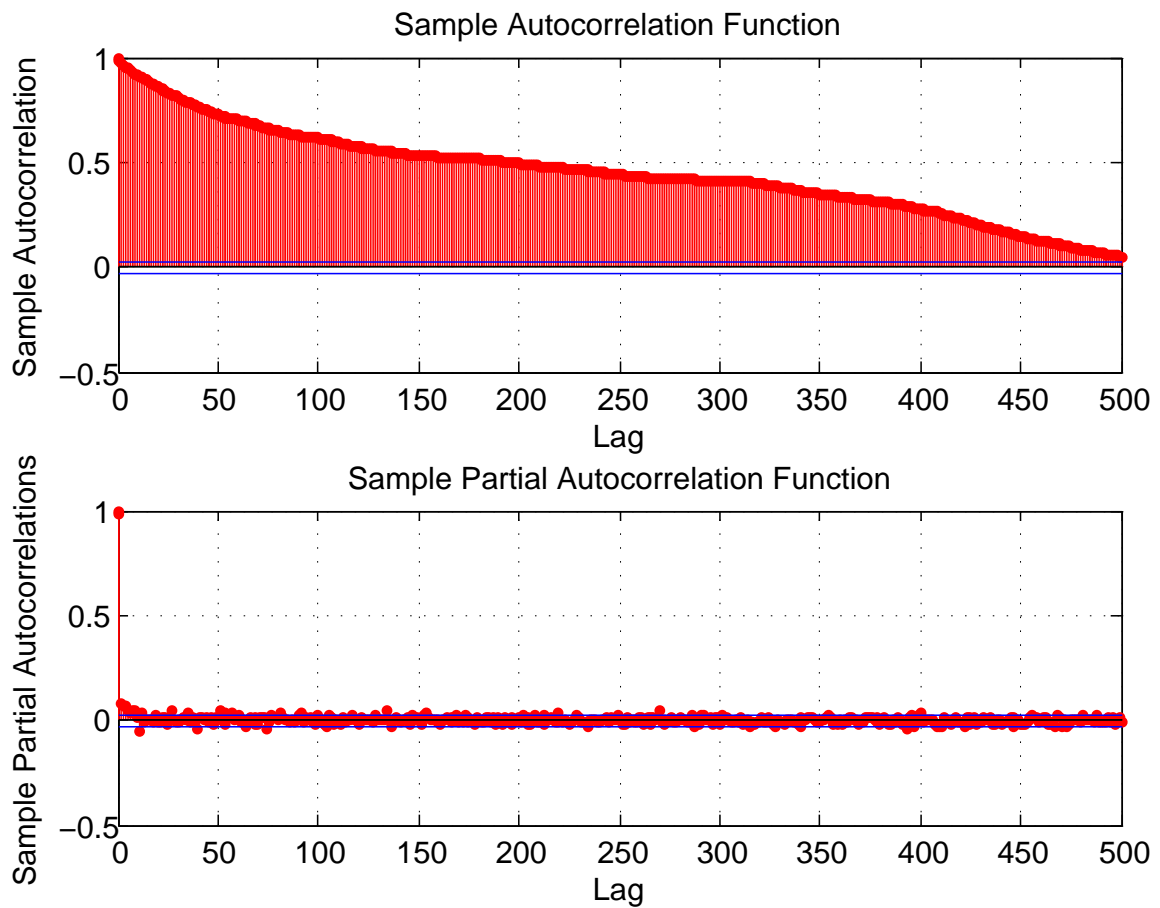


Figure 2: Sample autocorrelation and partial autocorrelation functions of the logarithm of the VIX index from January 2, 1990 to January 15, 2013. The blue line refers to the 95% confidence interval under the null of zero autocorrelation.

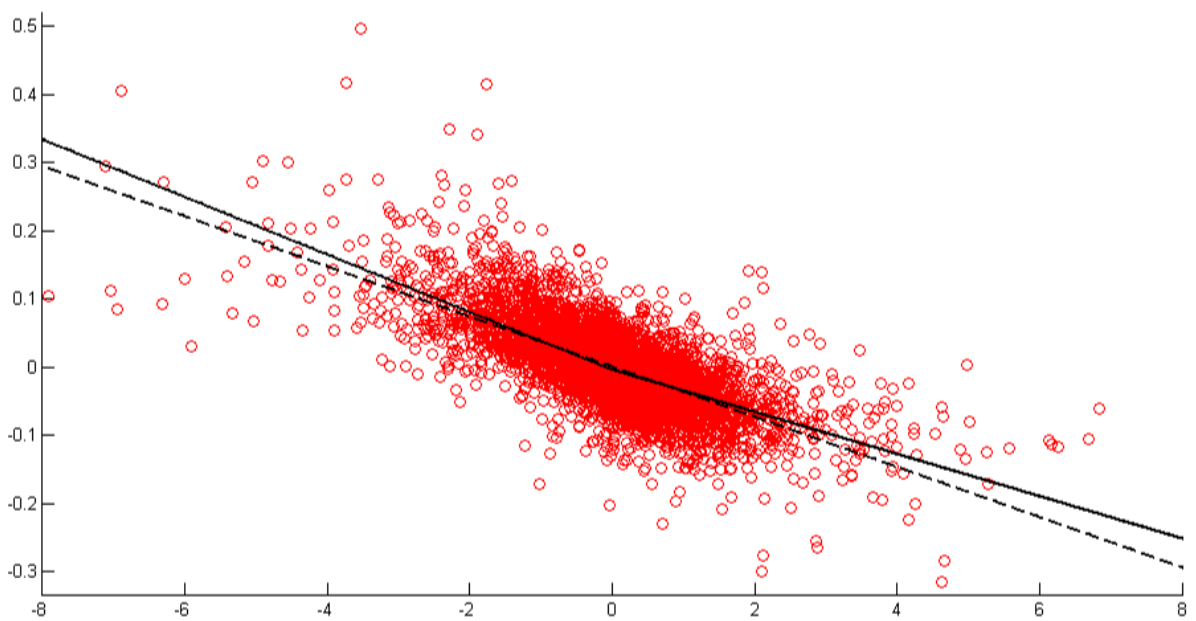


Figure 3: Scatter plot between the S&P 500 index returns and the changes in the VIX index (divided by 10) from January 2, 1990 to January 15, 2013. The dashed line results from a linear regression, whereas the solid line from a piecewise linear regression in which the VIX index responds differently to positive and negative S&P 500 index returns.

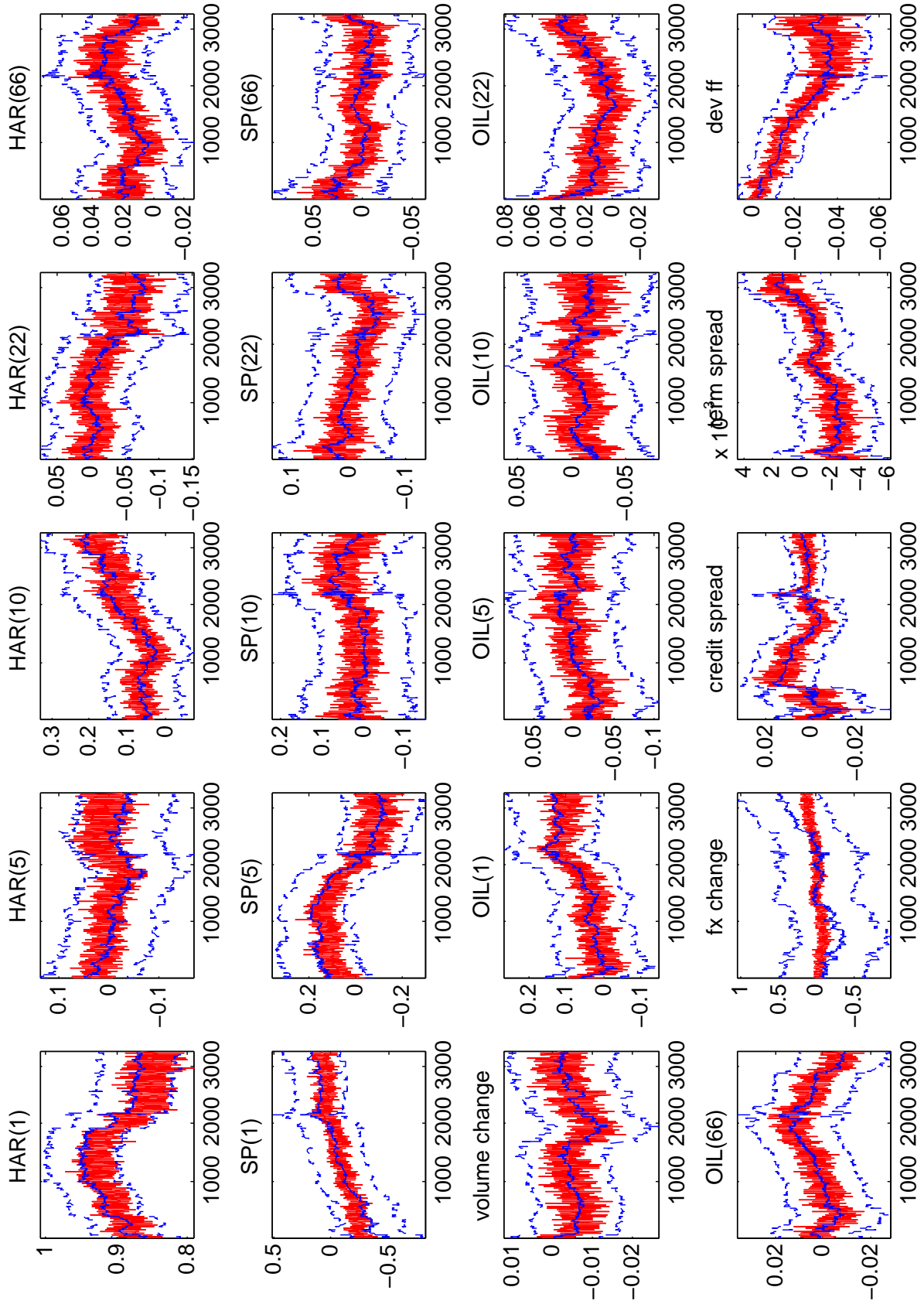


Figure 4: Average partial effects implied by the NNHARX model as compared to the corresponding confidence intervals of the HARX coefficient estimates (in shade).

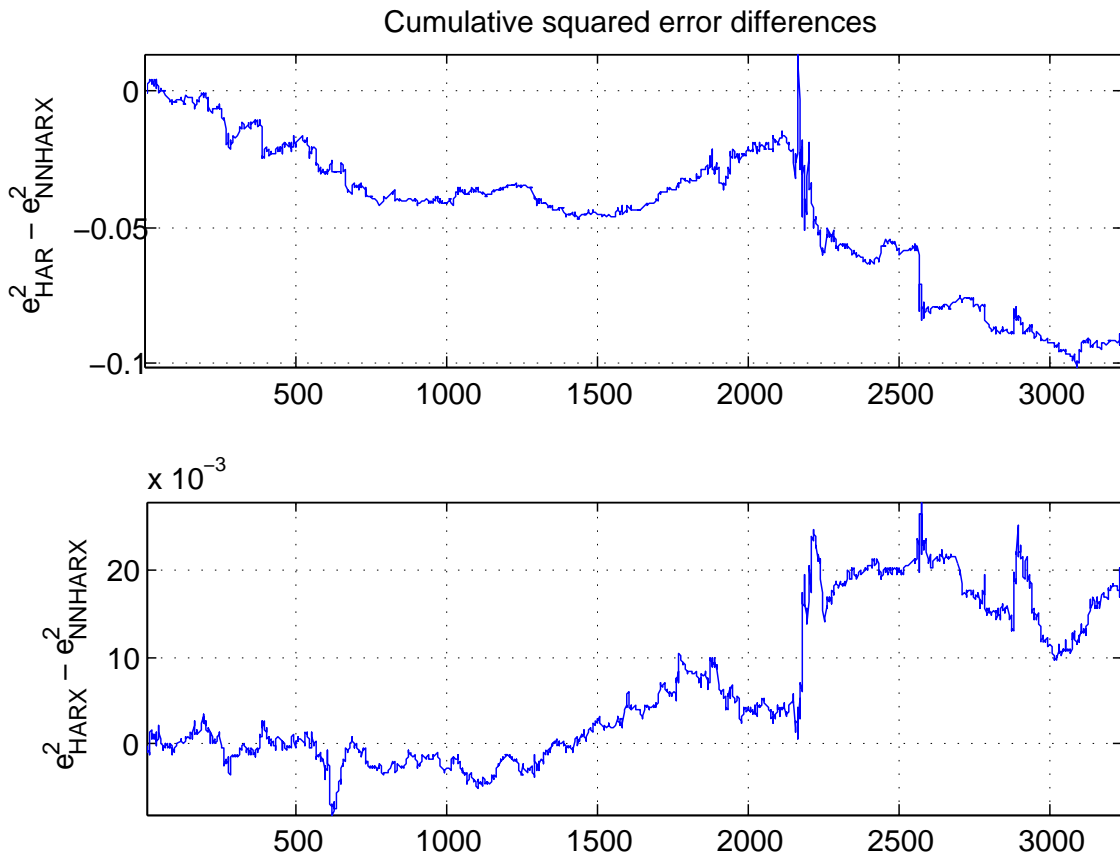


Figure 5: Differences in the cumulative squared errors. The upper panel illustrates the differences between the HAR and the NNHARX model and the lower panel reports the results concerning the HARX and the NNHARX models.