Beyond the Variance Risk Premium:

Stock Market Index Return Predictability and Option-Implied Information

Marie-Hélène Gagnon†, Gabriel J. Power‡, and Dominique Toupin*

Abstract

This paper investigates international stock market index return predictability using option-implied information (SP500, DAX, FTSE, CAC, and SMI). We document at a daily frequency the predictive power of several forward-looking variables such as the variance risk premium and the Foster-Hart (FH) risk measure. Our results suggest that the variance risk premium is a significant predictor of returns at a daily frequency. Out-of-sample forecasts also show explanatory power for return horizons of less than one month, which could not be investigated using the typical monthly estimates. The FH measure also has significant forecasting power. Moreover, our results show that VRP and FH are complementary: the former best explains short term returns (daily to monthly) and the latter, longer horizon returns (monthly to yearly). While higher-order risk-neutral moments are often significant in-sample, they show less predictive power once VRP and FH are introduced in the model.

Keywords: Options; risk-neutral distribution; variance risk premium; return predictability; predictive regressions; international stock market returns; Foster-Hart risk.

JEL codes: C12, C22, G12, G13
EFMA codes : 350, 410, 450

† Gagnon is Associate Professor of Finance, Department of Finance, Insurance and Real Estate, FSA ULaval, and Research Fellow, CIRPEE, Université Laval, Quebec City, QC, Canada G1V 0A6. P: (418) 656-2131, # 4742. F: (418) 656-2624. E: Marie-Helene.Gagnon@fsa.ulaval.ca

‡ Power is Associate Professor of Finance, Research Fellow, CIRPEE, and Research Fellow, CRIB, Université Laval. P: (418) 656-2131, # 4619. F: (418) 656-2624. E: Gabriel.Power@fsa.ulaval.ca

* Toupin is a PhD candidate in Finance, Université Laval. E: dominique.toupin.1@ulaval.ca

Beyond the Variance Risk Premium:

Stock Market Index Return Predictability and Option-Implied Information

Abstract

This paper investigates international stock market index return predictability using option-implied information (SP500, DAX, FTSE, CAC, and SMI). We document at a daily frequency the predictive power of several forward-looking variables such as the variance risk premium and the Foster-Hart (FH) risk measure. Our results suggest that the variance risk premium is a significant predictor of returns at a daily frequency. Out-of-sample forecasts also show explanatory power for return horizons of less than one month, which could not be investigated using the typical monthly estimates. The FH measure also has significant forecasting power. Moreover, our results show that VRP and FH are complementary: the former best explains short term returns (daily to monthly) and the latter, longer horizon returns (monthly to yearly). While higher-order risk-neutral moments are often significant in-sample, they show less predictive power once VRP and FH are introduced in the model.

Keywords: Options; risk-neutral distribution; variance risk premium; return predictability; predictive regressions; international stock market returns; Foster-Hart risk.

JEL codes: C12, C22, G12, G13

1. Introduction

This paper investigates the predictive power of the information contained in the option-implied risk-neutral distributions of major international stock indexes on their future returns. In order to build a predictive linear model from a complete distribution, we must choose a set of measures in the aim of capturing the most of its informative content. In this paper, a particular focus is given to the variance risk premium (VRP), risk-neutral skewness, risk-neutral kurtosis and the Foster-Hart measure of riskiness (FH). We focus on those particular measures as they each have an economic interpretation that is related to financial risk and model we argue they capture a particular aspect of total risk. Also, as we will see, they represent distinct aspects of the risk-neutral distribution.

Risk-neutral distributions are often used by investors and researchers as indicators of the current expectations of the market. One of its more popular forms is the VIX index, a model-free way to compute the volatility of the option-implied risk-neutral distribution of the S&P500 index at a monthly horizon. In the financial literature, moments of the risk-neutral distribution in particular have been found to forecast realized returns [Chang, Christoffersen, Jacobs and Vainberg (2011), Duan and Wei (2009)] and to be related to investor sentiment (Han 2008) and firm-specific characteristics [Dennis and Mayhew (2002) and Hansis, Schlang and Vilkov (2010)]. Lately, higher-order moments of the risk-neutral distribution have received more attention. For example, the effect of risk-neutral skewness on the cross-section of expected returns has been studied by Chang, Christoffersen and Jacobs (2013). Higher-order moments by themselves have also been shown to have predictable dynamics as in Neumann and Skiadopoulos (2013) or Gagnon, Power & Toupin (2016). A large strand of the existing literature on option-implied moments focuses on the related problem of estimating risk-neutral densities and the associated pricing kernel [see e.g. Bakshi, Madan and Panayotov (2010), Jackwerth (2000), Melick and Thomas (1997)].

This paper is a natural extension to this research, as it investigates the predictive power of a more comprehensive set of measures of risk taken on the risk-neutral distribution. Indeed, while moments and derived measures such as the VRP are informative [Bollerslev, Tauchen and Zhou (2009), Bollerslev et al. (2014)], they might not reflect all the available information in the risk-neutral distribution. We argue that other measures could have supplemental predictive power over the popular VRP. For instance, Foster and Hart (2009) introduces a new risk measure representing the critical wealth level above which it becomes safe to accept a particular gamble. This corresponds to a strategy that always guarantees the absence of bankruptcy. While this risk measure was originally meant to be computed from the physical distribution underlying the gamble, it has been shown to have some predictive power over some forms of future returns when computed from a risk-neutral distribution. Bali, Cakici, & Chabi-Yo (2011) found a link to future adjusted returns while Leiss & Nax (2015) found a significant relation to future large negative returns.

The paper makes several contributions which are, to the best of our knowledge, novel to the literature. The first contribution of this paper provides comprehensive evidence of the performance of the variance-risk premium at a higher frequency than the monthly analysis previously found in the literature [e.g. Bollerslev et al. 2014]. This higher frequency (daily) analysis is particularly interesting because the length of time needed in order to observe a significant link between a risk premium and future returns could be very informative about the nature of this risk premium. For example, the predictive power observed could be interpreted as the reward that an investor receives to bear a risk caused by some specific form of uncertainty about future returns (e.g. “jump risk”). This would mean that the observation of a significant link with future returns after a specific time horizon represents a partial compensation for this investor and that part of the uncertainty has been resolved. The second contribution of this paper is to study the direct link between FH and the future returns of an index (not only the risk-adjusted returns or the large negative returns). It is also the first paper to study FH in an international setting and to look at its daily frequency predictive power. The third contribution of this paper is to examine, in a combined framework, the predictive power of those variables. Thus, the paper provides a novel look and characterization of the role each variable plays in predicting future returns.

This paper extends the literature on variables that have predictive power over equity index returns by investigating at a daily frequency and for several major international stock indexes the in-sample and out-of-sample predictive power of different measures based on the option-implied risk-neutral distribution. This study begins by estimating the risk-neutral distribution at a constant horizon for several major international stock indexes. This is done daily using data on options for those indexes. The variance risk-premium (VRP), risk-neutral skewness, risk-neutral kurtosis and the Foster-Hart measure of riskiness (FH) are measured on each of those daily risk-neutral distributions. The predictive power of the resulting time series is studied by introducing them in a linear model forecasting the future returns of an index. The estimations are performed both in and out-of-sample and for various return horizons.

The choice of those four particular measures is based on the existing literature that documents the forecasting abilities of the VRP [Bollerslev, Tauchen and Zhou (2009), Bollerslev et al. (2014)]. Since the VRP is defined as the difference between the risk-neutral variance of an asset and its historical variance, it is a simple transformation of the second moment of the risk-neutral distribution. Introducing similar measures for higher-order moments of this distribution like skewness and kurtosis is a logical extension of this research because such measures could provide additional information characterizing different dimensions of risk and additional information in the risk-neutral distribution above that contained in the VRP. While measures of downside risk like value at risk or expected shortfall would be logical additional candidates, they are redundant to the risk-neutral volatility and VRP and therefore inappropriate in our statistical context. For example, the correlation between the 5% expected shortfall on the S&P500 index and its risk-neutral volatility is -0.98. This high correlation is true across the studied indexes and also for different levels of expected shortfalls. In contrast, the FH measure of riskiness is related to a strategy that precludes the possibility of bankruptcy and can also be viewed as a measure of downside risk. As is shown in a later section of this paper, FH has the significant advantage of not being closely related to the moments of the risk-neutral distribution. The choice of the FH measure is further motivated by the fact that in theory it represents a different source of risk than the VRP. The VRP is commonly interpreted as being the risk coming from the volatility-of-volatility, as in Bollerslev, Tauchen and Zhou (2009), while the FH measure is related to the risk of bankruptcy (equivalent to a large negative cumulative return or more broadly, downside risk).

More specifically, this paper finds that the VRP has a significant predictive power on future returns. In particular, the predictive power is significant much sooner than what was previously found, even after only one day for the S&P500. Previous studies such as Bollerslev, Tauchen and Zhou (2009) and Bollerslev et al. (2014) only used monthly time series and therefore could not observe this characteristic of the VRP. We argue that this result illustrates that the uncertainty underlying the VRP is partially resolved after a short period of time and thus a characteristic that should be reflected in its economic interpretation. In addition, the FH measure of riskiness is found to have a significant explanatory power over future returns, but generally at a longer horizon than the VRP. The out-of-sample evidence confirms that those two measures are complementary as they are relevant to different investing horizons. While the VRP explains future returns well at a short horizon, the FH measure explains returns over longer periods. This would indicate that the Foster-Hart measure of riskiness is not a redundant measure of the risk-neutral distribution that was already accounted for by its moments. Also, while risk-neutral skewness and kurtosis are significant explanatory variables for future returns in-sample, the evidence regarding their explanatory power is mixed out-of-sample.

1.1. Literature review

The VRP is defined as the difference between the variance of the risk-neutral distribution and the historical variance of an asset. It can be interpreted as a fear gauge by the markets or as the price paid by investors to hold assets that will pay off higher returns in high-volatility states [Bollerslev, Tauchen and Zhou (2009), Carr and Wu (2009)]. The literature has found that the VRP is a concise and relevant way to relate option-implied information to the future returns of an asset. For example, there is a significant financial literature suggesting that the stock market VRP has explanatory power for US stock returns [Bali and Zhou (2015), Bollerslev, Tauchen and Zhou (2009), Carr and Wu (2009), Garcia et al. (2011)]. In particular, Bollerslev, Tauchen and Zhou (2009) looked at the returns on the S&P500 and found that "the magnitude of this predictability is particularly strong at the intermediate quarterly horizon, where it dominates that afforded by other popular predictor variables, such as the P/E ratio, the default spread and the consumption-wealth ratio."

However, much less is known regarding the VRP's relevance to future returns in international markets, with the exception of Bollerslev et al. (2014) and Londono (2015). While both of these papers document the performance of the VRP for in-sample forecast using overlapping monthly returns, it could be argued that the VRP has additional in and out-of-sample forecasting power for higher-frequency predictability. Chen, Shen, Wang & Zuo (2015) look at out-of-sample predictability but again only at a monthly frequency and only for this one measure on the risk-neutral distribution, the VRP. Hattori, Shim, & Sugihara (2016) look at the correlation of VRPs across different international markets.

The existing literature on the link between the Foster-Hart measure of riskiness and future returns is somewhat less comprehensive. Bali, Cakici, & Chabi-Yo (2011) find that stocks with a higher option-implied Foster-Hart measure of riskiness had lower future risk-adjusted returns, even after controlling for a measure of risk-neutral skewness, value at risk and expected shortfall. Notably absent from this study is the comparison to the variance risk premium. In contrast to the methodology of this paper, that study neglects to explicitly complete the tails of the risk-neutral distribution before computing the FH measure, making it more susceptible to variations in the available strike prices for an asset on different days. It also adjusts future returns for risk, making its results dependent on a specific form a risk-adjustment while this paper looks directly at future excess returns.

Bali, Cakici, & Chabi-Yo (2012) compute an aggregate riskiness index based on the option-implied Foster-Hart measure and find that it successfully predicts future economic downturns. Those economic downturns are measured by an aggregate of different monthly indicators of national economic activity. Leiss & Nax (2015) uses a logistic regression with only directly available option maturities to find that the Foster-Hart measure is related to future large negative returns for the S&P500 index. Anand, Kurosaki & Kim (2016) find that a Foster-Hart measure of riskiness based on historical returns of the Dow Jones Industrial Average constituents leads to a better risk-adjusted portfolio performance than commonly-used selection techniques. This literature related the FH measure points to potentially unexplored and yet unexplained attributes of future returns relating to a different source of risk than the variance risk premium and other moment-based measures.

2. Methodology

2.1. Regression on future returns

The interest of this paper is the predictive power of measures on the risk-neutral distribution for the future returns of an index. This is done with a simple linear model where the index returns[1] from time t to $t+h$ are explained by the measures on the risk-neutral distribution at time t:

[1] Index returns above the risk-free rate and including dividends.

(1) $$R_{t,t+h} = a + b \cdot VRP_t + c \cdot FH_t + d \cdot SK_t + e \cdot Kurt_t + u_t$$

where *VRP* is the variance risk premium defined as the risk-neutral variance minus the historical variance. *SK* and *Kurt* are respectively the third and fourth moment of the risk-neutral distribution. *FH* is the Foster-Hart measure of riskiness computed from the risk-neutral distribution.

2.2. Out-of-sample forecast of returns

The linear model described in equation (1) is used to perform out-of-sample forecasting for the last 30% of index returns. This model is calibrated using the first 70% of the sample of index returns in order to make one h-days ahead prediction. The following day, the model is recalibrated using the same length of past observations (70%) and one new h-days ahead prediction is made. This is done until the end of the sample. The prediction errors are then compared to those of more parsimonious variations of (1) in order to better identify the contribution of its different explanatory variables. The series of predicted returns are compared using a modified Diebold-Mariano test on the squared prediction errors that takes into account the effect of the overlapping periods of returns.

As developed in Harvey, Leybourne and Newbold (1997), this tests the null hypothesis of equal forecast performance for two models. It is based on the series d_t of differences between the squared errors of two models at each time t . The null hypothesis is that the expected value of d_t is zero. The observed sample mean $\bar{d}$ of this series of n observations is used to make the test statistic.

(2) $$\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$$

The variance of $\bar{d}$ is given by:

(3) $$V(\bar{d}) \approx \frac{1}{n}\left[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right]$$

where γ_k is the k^{th} autocovariance of d_t and h is the forecasting horizon. The autocovariance is estimated by:

(4) $$\gamma_k = \frac{1}{n}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d})$$

The null hypothesis can then be tested with this statistic:

(5) $$S_1^* = \frac{\bar{d}}{V(\bar{d})^{1/2}}\left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{1/2}$$

where critical values are taken from a Student t-distribution with $(n-1)$ degrees of freedom.

3. Data

3.1. Realized variance, index returns and risk-free rates

Index returns are computed including dividends using data from Bloomberg. Risk-free rates corresponding to the currency of each index are obtained from the FRED website of the Federal Reserve Bank of St. Louis. Excess returns are defined as the difference between a simple return and the risk-free rate. [2] Realized volatilities for each index are obtained from the Oxford-Man Institute's realized library's website (Gerd et al. 2009). The historical variance used in the VRP is defined as the sum of the 5-minute realized variance and squared overnight returns of the past month (21 business days), as in Bollerslev, Tauchen and Zhou (2009).

3.2. Option Data

Data for options traded on the underlying indexes are needed to recover the risk-neutral distribution. Index option data is obtained from OptionMetrics Ivy DB US and Europe. The indexes chosen are the S&P500 index for the United States, the DAX index of Germany, the SMI index of Switzerland, the CAC index of France and the FTSE index of the United Kingdom. Strike prices, maturities, and implied

[2] The constant maturity risk-free interest rates are computed from "the most recently auctioned 4-, 13-, 26-, and 52-week bills, plus the most recently auctioned 2-, 3-, 5-, 7-, and 10-year notes and the most recently auctioned 30-year bond, plus the composite rate in the 20-year maturity range."

volatilities are extracted at a daily frequency for all available options on the selected indexes. The beginning of the sample is restricted by the availability of realized volatility for the S&P500 and by the availability of option data for the European indexes. Table 1 presents descriptive statistics for the raw data on options. We see that the number of options available each day to construct a volatility surface varies between indexes. This is the main criteria used to choose which index to include in this paper since too few options could lead to an unreliable estimate of the continuous volatility surface resulting in a bad estimate of the risk-neutral distribution.

3.3. Empirical risk-neutral distributions

From Breeden and Litzenberger (1978), we know that the risk-neutral probability density function *f(K)* can be recovered by the second derivative of a continuum of call prices C with regard to strikes X where r is the risk-free rate for the maturity T:

$$f(X) = e^{rT} \frac{\partial^2 C}{\partial X^2} \tag{6}$$

In practice, this formula is difficult to use because there isn't a continuum of call prices from 0 to infinity. Each day, there are only option prices for a discrete set of strike prices and maturities. Interpolation can transform this set of points into a continuous surface, but assumptions still have to be made about what happens outside of the region where data is available.

If one is only interested in the moments of this empirical risk-neutral distribution, a method by Bakshi, Kapadia and Madan (2003) offers a way to calculate these moments directly from out-of-the-money call and put option prices. This method is well established in the literature, but unfortunately it limits how you can extrapolate outside of available strike prices. For this reason, this paper follows the method of Birru and Figlewski (2012) by first calculating the empirical RND and then fitting each missing tail to a Generalized Extreme Value (GEV) distribution. The use of this methodology to evaluate tails of the distribution more accurately is motivated by the goal of reliably measuring higher order risk-neutral moments (and other measures). What happens in the tails of the distribution might not have a large impact on the variance of a distribution if there is not a lot of mass there, but higher order moments are more sensitive to the tails since the distance to the center is elevated to the third or higher power.

The first step is to construct a continuous volatility surface. The usual filters used in the literature when working with options are applied to the raw data. Options are excluded if they are expiring in five days or less or having implied volatility that is negative or above 100%. Only out-of-the-money options are used to obtain the volatility surface. For all remaining options, implied volatilities are recovered from OptionMetrics. This provides, for each day, a surface of points in implied volatility-maturity-strike space (Figure 1 shows an example). For a chosen maturity, the next steps are to convert the corresponding IVs to call prices using Black-Scholes, fit the points to a continuous form like a spline and apply (6) to obtain the risk-neutral distribution for a given date and maturity.

Since the objective is to construct a daily frequency time series of measures on the risk-neutral distribution, a constant maturity has to be used on each daily observation. However, the actual option prices have maturities that change each day. For this reason, this set of points is transformed into a continuous surface. This is done using cubic spline interpolation where each node is an actual observation. After that, a “slice” of this surface is taken at a maturity of one month and is corrected to avoid arbitrage possibilities, following Aït-Sahalia and Duarte (2003). This can correct some pricing errors introduced by interpolating, but without having to make strong parametric assumptions about the form of the IV surface.

Following Birru and Figlewski (2012), a fourth-degree smoothing spline is used on the arbitrage-free interpolated IVs for a given maturity. This additional step of smoothing ensures that the derivatives of the price function will be continuous up to the third order. Since the risk-neutral distribution is calculated as the second derivative of the option prices with respect to moneyness, the extra smoothing will prevent the appearance of sharp spikes in the fitted density. Once the smoothed IV curve is converted back to a call option price function using Black-Scholes, the empirical risk-neutral distribution function is obtained from (6). Since there isn't a range of traded option strikes from 0 to infinity, the empirical distribution that is recovered is still missing its tails. To extend the IV surface, one approach is to assume that it is constant beyond the available data, as in Bliss and Panigirtzoglou (2004). Since the IV is calculated using Black-Scholes, this assumption implies that the tails of the risk-neutral distribution are normal. However, this is not an observed property for equity returns.

This paper instead follows Birru and Figlewski (2012) and uses a fitted Generalized Extreme Value (GEV) distribution for each missing tail. The motivation of those authors is that: "Similar to the way the Central Limit Theorem makes the Normal a natural model for the sample average from an unknown distribution, the Fisher-Tippett Theorem shows that the GEV distribution is a natural candidate for modeling the tails of an unknown density." This assumption is more appropriate and ensures that, for example, the risk-neutral moments computed for dates when the RND is missing a considerable part of its lower tail are closer to those computed on dates when this data is available. Imposing normality could lead to variation in the time series of moments that is due only to data availability and not option price information. The Generalized Extreme Value (GEV) cumulative distribution function used for each missing tail is given by:

$$F(X) = \exp\left[-\left(1 + \xi\,\frac{X-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] \tag{7}$$

For each tail, the three parameters (μ, σ, ξ) of the GEV distribution are fitted such that the total probability in the fitted tail is equal to the missing total probability in this tail of the empirical risk-neutral distribution and must connect with the empirical RND at the 2^{nd} and 5^{th} or 95^{th} and 98^{th} percentiles. Figure 2 shows this procedure. If the empirical RND does not extend to the 2nd or 98th percentiles, then the connection points for a tail are the last available percentile and another one that is 3 percentiles closer to the center of the distribution. The objective function is to minimize the sum of the squared distances between the empirical RND and the GEV distribution on the domain between each pair of connection points.

To show the impact of this approach relative to imposing normality, the completed distribution can be transformed back to an IV curve as is done in figure 3. In this example, we see that the normality assumption underestimates the prices of tail options compared to GEV. The GEV approach should then generate risk-neutral measures where the information contained in the options near the tails have more weight. As a result, these measures might be more informative, especially for measures that are sensitive to the tails of a distribution. Lastly, the GEV approach also precludes discontinuities in the IV curve. Such discontinuities are usually not present in observable option prices.

This method of completing the tails should also prevent a problem noted by Dennis and Mayhew (2002) where risk-neutral skewness is biased when the interval of strike prices available is not symmetric around the current price of the asset. Their solution is to use the largest available symmetric interval on each date and discard the data on options that are outside of it. However, their solution leaves out a lot of information, especially when the index studied has fewer traded options than the S&P500. With the GEV method, the result is the complete RND instead of a version that is cut off at the tails, as in Dennis and Mayhew (2002). That is why symmetry around the current asset price is not a problem here.

3.4. Risk-neutral moments and VRP

Once the risk-neutral distribution $f(x)$ is computed for a given day, its central moments are given by the usual mass function formulas:

(8)

$$M_n = \int_{-\infty}^{\infty} (x-c)^n f(x)\, dx$$

$$Var = \frac{M_2}{T}$$

$$Skew = \frac{M_3}{{M_2}^{3/2}}$$

$$Kurt = \frac{M_4}{{M_2}^{4/2}}$$

The variance risk premium (VRP) is obtained by subtracting the historical variance from the risk-neutral variance of equation (8).

3.5. Foster-Hart measure of riskiness

The Foster and Hart (2009) measure of riskiness ρ for a lottery represented by the random variable X is the solution to this equation:

(9) $$\mathrm{E}\left[\log\left(1+\frac{X}{\rho}\right)\right] = 0$$

The measure of riskiness ρ is the critical wealth level below which the investor will reject this gamble in order to avoid bankruptcy. Bali, Cakici & Chabi-Yo (2011) introduce a generalization of this measure that is the solution of the following equation:

(10) $$\mathrm{E}\left[\frac{\left(1+\frac{X}{\rho}\right)^{\delta}-1}{\delta}\right] = 0$$

In this version, the parameter δ is related to the risk aversion of the investor. This investor makes his decision to accept or reject the gamble based on the probability density function (pdf) that is distorted by

δ instead of the actual pdf of possible gains. Similarly to Leiss and Nax (2015), the last equation can be applied to financial returns by using the relative return r as the random variable:

$$\text{(11)} \qquad \mathrm{E}\left[\frac{(1+rW)^{\delta}-1}{\delta}\right]=0$$

where W is the fraction of wealth above which it becomes risky to invest in this asset. For example, at W=0.6 an investor can put up to 60% of his wealth in the asset without risking bankruptcy.

In contrast to other measures of the risk-neutral pdf, the Foster-Hart measure requires that the gamble has a positive expectation. This criterion is added to the GEV tail-fitting procedure, but since being numerically too close to a null return could cause problems the condition used is E[r] > 0.0005. The choice of δ is made to ensure that the Foster-Hart measure does not reach its ceiling of $W = 1$ during the period studied. Obviously, if the riskiness measure is allowed to stay at this maximum value for many consecutive days, it would not be a very informative signal.

To choose an appropriately low delta, a time series of δ corresponding to $W = 1$ is generated for each day. For the studied indexes, $\delta = -4$ is sufficiently low to ensure that $W < 1$ for the time period of interest. The transformation $FH_t = \log(1/W_t)$ is used in the regressions so that this measure of risk is appropriately scaled for a linear model. In this form it is a measure of riskiness since a high FH means that one risks bankruptcy even if only investing a small fraction of one's wealth.

3.6. Descriptive statistics of the time series

The descriptive statistics for the time series of explanatory variables considered are in table 2. From their low correlation coefficients, it would seem that the chosen measures represent relatively different components of the risk-neutral distribution. This is less true for the skewness and kurtosis with correlation coefficients up to 0.68. This comovement could mean that much of the variations in skewness are due to what happens in the tails of the distribution. Nevertheless, those higher moments are not very correlated to the other two measures and as such they might be a good way to check if the risk-neutral distribution contains relevant information above that contained in FH and VRP. Those time series are plotted in the appendix. There appears to be a common feature across indexes near the beginning of 2009, during the last financial crisis. At that point in time the VRPs of all studied indexes are more volatile and the Foster-Hart measure of riskiness visibly increases. This could indicate that FH and VRP both measure a kind of risk that is integrated across markets, at least in time of crisis.

4. Results

4.1. Linear model of future returns

As explained in the methodology, the in-sample predictive power of the explanatory variables is tested first. Future returns for all indexes and different horizons h (in days) are regressed on the explanatory variables:

$$R_{t,t+h} = a + b \cdot VRP_t + c \cdot FH_t + d \cdot SK_t + e \cdot Kurt_t + u_t$$

Figures 4 to 8 show the results for the Newey-West t-statistics of the regression coefficients. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

Previous research has found that the variance risk premium has predictive power for future returns [Bollerslev, Tauchen and Zhou (2009), Bollerslev et al. (2014)]. This kind of analysis has only been conducted at a monthly frequency. The results in figures 4 to 8 seem to indicate that those previous studies missed some interesting facts since the VRP is significantly related to future returns at short horizons for many indexes. This is the case for the S&P500 where the VRP is significant for all horizons under 175 days. For the DAX, the VRP is significant for some horizons around 20 days and longer horizons of 45 to 95 days. For the SMI, VRP is significant for most horizons except some around 35 days and those above 175 days. For the CAC, the VRP is significant for horizons between 2 and 175 days. For the FTSE, VRP is significant for some horizons around 18 days and for horizons between 50 and 135 days. We also observe that the VRP is more significant around the 2 to 4-month horizons (40 to 80 business days) for most indexes, as was reported by Bollerslev et al. (2014) for their monthly analysis of international stock indexes.

When an investor accepts to bear a certain risk premium measured by the VRP, he is compensated for this risk by having higher excess returns in the future. By looking only at monthly returns, previous studies missed an important clue pertinent to the speed at which this investor is compensated. This is relevant because it could be argued that if the investor has been partially compensated after a short amount of time, then it would mean that some of the uncertainty corresponding to this risk premium has been resolved. The stylized equilibrium model of Bollerslev, Tauchen and Zhou (2009) suggests that “the variance risk premium should serve as a useful predictor for the actual realized returns over horizons for which the volatility-of-volatility constitutes the dominant source of the variation in the equity premium”. This means that if the predictive power of the VRP really comes from uncertainty over the volatility-of-volatility, the results of this section indicate that a part of this uncertainty is very short lived because investors are rewarded in part for bearing this risk after only one day. This is a new result that has implications on the interpretation for the VRP that were not visible in the monthly analysis of previous studies.

The Foster-Hart measure is related to the risk of large negative returns (bankruptcy) and as such a positive relation with future returns would be expected. This is the classical risk-return relationship where an investor wants a higher expected return on days where the risk of large negative returns is higher. This is confirmed here since when FH is significant it is always positive. For the S&P500, FH is significant at horizons above 15 days (three weeks). For the DAX, FH is significant for all horizons between 15 and 220 days. The SMI is the only index where the FH measure is significant at short horizons. In fact, it is significant for all horizons under 220 days. For the CAC, the FH measure is significant for horizons between 10 and 165 days. For the FTSE, FH is significant for all horizons above 6 days. As explained in the next subsection, FH appears to not only be statistically significant but also economically relevant as it has an important effect on the R^2 of the linear model.

The skewness measure is significantly related to future returns for most indexes. For the S&P500 this factor is more significant around the daily and monthly (20 days) horizons. It is also significant at horizons above 60 days (about 3 months). The skewness is never significant for the DAX index. For the SMI, skewness is significant at horizons above 100 days. For the CAC it is significant at all horizons above 2 days. For the FTSE this factor is more significant around the weekly (5 days) and monthly (20 days) horizons but it is also significant at horizons above 40 days (about 2 months). The coefficient for the skewness measure is always positive when significant. This would mean that buying the index on a day when the risk-neutral skewness is high (more positive) would lead to higher returns in the future. This would not be coherent with an interpretation of skewness as “crash risk”.

Since kurtosis is usually interpreted as “tail risk”, we would expect a higher “tail risk” today to lead to higher future returns. This seems to be confirmed for most indexes since the coefficient for kurtosis is positive when significant for all indexes except the SMI. For the S&P500, kurtosis is significant for all horizons of 1 to 250 days. Kurtosis is never a significant factor for the DAX. For the SMI, kurtosis is negative and significant for some horizons around 40 days and above 175 days but only at a 10% level. For the CAC and FSTE, kurtosis is significant for all horizons above 20 days.

The R^2 of the regressions are presented in figures 9 to 13 and they follow a similar pattern for all indexes where it is lower at short horizons and higher at medium and longer horizons. This could be interpreted as an indication of the speed at which the uncertainty of the risk premiums corresponding to the explanatory variables are resolved. With this interpretation, the uncertainty corresponding to the studied variables is mostly resolved after about 80 business days (about 4 months). But, as noted by Bollerslev, Tauchen and Zhou (2009), this kind of regression on overlapping returns can lead to a R^2 that mechanically rises with the return horizon even there is no additional informative content and as such should be interpreted carefully.

4.2. Economic relevance of the measures on the risk-neutral distribution

To get an idea of the economic relevance of skewness and kurtosis compared to FH in the previous in-sample regression, we can compare the adjusted R^2 of the complete model with those of more parsimonious variations (figure 14). As noted by Bollerslev, Tauchen and Zhou (2009), we have to be careful when looking at the R squared of a regression on overlapping returns. The reason is that it can increase in value proportionally with the horizon without a true increase in predictability when predictor variables are highly persistent. Still, we can clearly see that the addition of the Foster-Hart measure of risk seems to have a much bigger impact on the fit of the regression than the addition of the higher moments of the risk-neutral distribution. This could explain why in the next section the out-of-sample evidence for the relevance of those higher moments is not as strong as that of FH. Another interesting aspect is that the marginal contribution of FH to the R^2 appears to be larger at longer horizons. This is in line with the out-of-sample results of the next section.

4.3. Out-of-sample forecasts

In this section, the out-of-sample forecasting performance of returns with different variations of the linear model in (1) is compared in order to better evaluate the importance of each measure. The squared prediction errors of a certain model are compared with those of a reference model. This is done with a modified Diebold-Mariano test that takes into account the effect of the overlapping periods of returns. If the test statistic is significantly negative, it means that the squared prediction errors of the forecasting model are significantly smaller than those of the reference model (the forecasting model outperforms the reference model). This method lets us compare the relative forecasting performance for different variations of the linear model described in equation (1). The base model only contains a constant which means it can be interpreted as a random walk where the expected return for the next period is the mean of the in-sample returns up to that day. The results are in table 3.

The first four tests compare the performance of a linear model that contains only one of the studied measures with a reference model that consist of just a constant. If the test statistic for those forecasting models is significantly negative, it would indicate that the corresponding measure has a significant predictive power by itself.

We observe that the variance risk premium has a significant predictive power by itself at some returns horizons for most indexes. It is the case at short and medium return horizons (1 to 50 business days) for the S&P500 and at some short horizons for the CAC (1, 2, 5 days) and FTSE (5 days). At longer return horizons (250 business days or about a year), the VRP by itself is significant for the DAX. This is not the case for the SMI where the VRP significantly underperforms the base model at 250 days. This is particularly interesting since most of the relevant horizons for the VRP are shorter than the monthly horizon (about 20 days) of the previous literature. Contrary to this previous literature, this out-of-sample evidence is also not dependent on the appropriate form for the standard error of the regression coefficients.[3]

The Foster-Hart measure of riskiness by itself does not seem to have a significant predictive power over the base model with only a constant. For longer horizon of the S&P500 it even significantly underperforms the base model. This could indicate that the component of risk measured by FH is not

[3] For example, if there is collinearity between kurtosis and skewness, this would make the interpretation of the significance of the coefficients for the in-sample regression unreliable while the out-of-sample test would not be affected.

independent of the risk measured by the VRP but it is rather complementary to the risk premium corresponding to the VRP.

The skewness measure by itself has significant predictive power for the CAC at 50 and 100 days. For the S&P500 the results are somewhat mixed, the skewness measure significantly outperforms the base model for some horizons (2 and 20 days) but it underperforms for a 5-day horizon. But, it seems to mirror the results we got from the in-sample regression on the S&P500 where skewness was more significant at 2 and 20 days and not significant at 5 days. The skewness outperforms the base model at all horizons for the DAX but never significantly so. There is also some significant underperformance for the FTSE (250 days) and SMI (50 days). For all indexes, kurtosis by itself never significantly outperforms the base model. Sometimes it even significantly underperforms compared to the base model.

The remaining comparisons let us evaluate the predictive performance of the linear model with all the variables with some variations that are more parsimonious. The main comparison is between the 4 variables model and the base model with only a constant. If the model composed of 4 measures taken from the option-implied risk-neutral distribution outperforms the base model it would strongly suggests that this distribution contains information about future returns. This seems to be the case for the S&P500 where the complete model significantly outperforms the base model for most horizons (5, 20, 50, 250 days). Since the test statistic is negative for the remaining horizons, we can say that the complete model also outperforms the base model for those horizons, but not significantly so. This is particularly interesting because for longer horizons (250 days), the complete model significantly outperforms the base model while none of the variables had a significant performance by themselves.

For the DAX index we observe that the complete model outperforms the base model for all horizons but not significantly so. For the SMI and the FTSE, the complete model does not seem to have any predictive power over the base model since the test statistic is not regularly negative. The complete model outperforms the base model for all horizons for the CAC but only significantly so for the shorter horizons (1,2 and 5 days). The last four comparisons give us some insight into where the predictive power of the complete model comes from. For the DAX we observe that the addition of kurtosis to a three-factor model seems to significantly degrade the predictive power of the model. This is made clear by the fact that the complete model significantly underperforms a model with all factors except for kurtosis at most horizons (2, 5, 20, 100 days). For the SMI we again find no evidence of predictive power for the models tested. For the FTSE we see that the addition of the VRP to a two-factor model significantly improve the forecasting performance at a 5-day horizon.

For the CAC we observe that the predictive power of the complete model seems to be mainly due to the VRP. This comes from the fact that the addition of the VRP to a three-factor model significantly improves the forecasting performance for short horizons of 1, 2 and 5 days. At longer horizons this is not the case with a significant underperformance at 250 days. For the S&P500, the source of the predictive power of the complete model seems to depend on the forecasting horizon. At shorter horizons the VRP seems to be more important. This is seen by the result that the addition of the VRP to a three-factor model significantly improves the forecasting performance for horizons of 1, 2, 5 and 20 business days. At longer horizons, the Foster-Hart measure of riskiness seems to be more important in explaining the predictive power of the complete model. This is seen in the result where adding FH to a three-factor model significantly improves the predictive performance of a two-factor model for horizons of 50 and 250 days. To better understand the predictive power of those two variables, additional tests for the S&P500 [4] are reported in table 4.

From those additional tests we can observe that the addition of the VRP to a model consisting only of the FH measure always improves the forecast and significantly so at most horizons (5 to 250

[4] For the other indexes, those tests do not give significant results, except for the CAC where again the predictive power of the VRP at short horizons is confirmed.

days). This is not surprising since we found that by itself the FH measure did not predict future returns. What is more surprising is the fact that adding FH to a model that consists only of the VRP significantly improves the forecast at longer horizons (20, 50 and 250 days). These additional tests confirm the result that the predictive power of the complete model for the S&P500 comes mainly from the VRP at shorter horizons and from the Foster-Hart measure of riskiness at longer horizons. This is coherent with the in-sample results for the S&P500 where for short and medium horizons the VRP was the more significant factor and at longer horizons it was the Foster-Hart measure.

For the VRP, this could mean that the uncertainty over the volatility-of-volatility of expected returns is mostly resolved over a relatively short horizon. This would make sense since it would be surprising if the market could take into account the characteristics of volatility at a very long horizon. This means that while the market could reasonably be worried about what will happen to volatility in the next 2-3 months, it might not have any idea of what the volatility will be like in a year. In contrast, for the risk premium measured by Foster-Hart measure of riskiness, it would make sense that an investor has to wait longer before being compensated. Since FH is related to the risk of bankruptcy of an investor, it represents the possibility of a large negative cumulative return. Since large negative returns can occur in one day, it makes sense that this uncertainty would not be resolved as quickly. In other words, it takes a longer time for an investor to be reassured and realize that he will not go bankrupt after all.

These out-of-sample tests do not seem to greatly support the role of the skewness and kurtosis in predicting future returns. This could mean that the higher moments of the risk-neutral distribution do not have an impact big enough on future returns to be detected by an out-of-sample test. Alternatively, the contradiction between the in-sample and out-of-sample results could indicate that the risk-neutral distribution contains additional information about future returns above the VRP and FH measures but that the higher moments are not the best way to measure it.

5. Conclusion

This paper contributes to the literature on the variance risk premium and the more general theme of index return predictability using measures on the option-implied risk neutral distribution of international stock indexes. It extends and expands a testing framework by considering several major international indexes in addition to the US S&P 500 and performs the analysis at a higher frequency and with more risk-neutral measures than what has been done before (e.g. Bollerslev et al. 2014). In this paper, we show that the variance risk premium is a significant predictor of returns using daily frequency and for several major international stock indexes. It is also true for return horizons under one month, which has not been shown in the literature before, because the previous literature used a monthly frequency. An out-of-sample forecast confirmed those results for several indices. This significant impact of the VRP on future returns for short horizons is an important clue to the exact nature of the uncertainty underlying this risk premium. This new result seems to indicate that this uncertainty is partially resolved after a very short amount of time. This could support the interpretation of the VRP as representing the risk of volatility-of-volatility since it would seem reasonable that the worries of the market about future volatility could be quickly resolved from one day to the next.

We have seen that in-sample, skewness and kurtosis, the higher moments of the option-implied risk-neutral distribution, have a significant relation to future returns for many indices. But, the fact that their coefficients were not always of the expected sign and that the relation was less evident out-of-sample suggest that although this indicates that the risk-neutral distribution contains relevant information about future returns above that of the VRP and Foster-Hart measure, the higher moments might not be the best way to measure it.

We have also seen that the Foster-Hart measure of riskiness is significantly related to future returns for all tested indices in-sample. The in-sample evidence pointed that this measure might be relevant at longer horizons than the VRP. Out-of-sample evidence on the S&P500 showed that the Foster-Hart measure has predictive power on future returns above that of the VRP. Those results suggest that while the VRP best explains short term returns (daily to monthly), the Foster-Hart measure has a complimentary predictability on a longer horizon (monthly to yearly). This evidence suggests that uncertainty underlying the FH measure cannot be resolved as quickly as that of the VRP. This would be consistent with an interpretation of the FH measure as the risk of future bankruptcy, or large losses. Since a large loss can happen in only a day, this kind of uncertainty would take longer to resolve.

References

Ait-Sahalia, Y. and J. Duarte (2003). Nonparametric option pricing under shape restrictions. Journal of Econometrics, 116, 9-47.

Anand, A., Li, T., Kurosaki, T., & Kim, Y. S. (2016). Foster-Hart optimal portfolios. Journal of Banking and Finance, 68, 117–130.

Bakshi, G., N. Kapadia and D. Madan (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. Review of Financial Studies, 16, 1, 101-143.

Bali, T. G., Cakici, N., & Chabi-Yo, F. (2011). A Generalized Measure of Riskiness. Management Science, 57(8), 1406–1423.

Bali, T., Cakici, N., & Chabi-Yo, F. (2012). Does Aggregate Riskiness Predict Future Economic Downturns? Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039078

Birru, J. and S. Figlewski (2012). Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008. Journal of Financial Markets 15, 2, 151-180.

Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock return predictability and variance risk premia: statistical inference and international evidence. Journal of Financial and Quantitative Analysis, 49(3), 633–661.

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

Chang, B. Y., Christoffersen, P., & Jacobs, K. (2013). Market skewness risk and the cross section of stock returns. Journal of Financial Economics, 107(1), 46–68.

Chang, B.Y., P. Christoffersen, K. Jacobs and G. Vainberg (2011). Option-implied measures of equity risk, Review of Finance, 16(2), 385-428.

Conrad, J., R. Dittmar and E. Ghysels (2013). Ex ante skewness and expected stock returns. Journal of Finance, 68, 1, 85-124.

Dennis, P. and S. Mayhew (2002). Risk-neutral skewness: evidence from stock options, Journal of Financial and Quantitative Analysis 37, 3, 471-493.

Duan, J.C. and J. Wei (2009). Systematic risk and the price structure of individual equity options. Review of Financial Studies, 22 (5), 1981-2006

Gerd, H., Lunde, A., Shephard, N. and Sheppard, K. (2009). Oxford-Man Institute's realized library v. 0.2, Oxford-Man Institute, University of Oxford.

Foster, D. P., & Hart, S. (2009). An Operational Measure of Riskiness. Journal of Political Economy, 117(5), 785–814.

Hansis, A., C. Schlag and G. Vilkov (2010). The dynamics of risk-neutral implied moments: evidence from individual options. Working paper, Goethe University.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281–291.

Jackwerth, J., (2000). Recovering risk aversion from option prices and realized returns. Review of Financial Studies, 13, 2, 433-451.

Leiss, M., & Nax, H. H. (2015). Option-Implied Objective Measures of Market Risk, (324247).

Londono, J. (2015). The variance risk premium around the world. Working Paper, Federal Reserve Board, Washington, DC

Melick, W., Thomas, C., (1997). Recovering an asset's implied SPD from option prices: An application to crude oil during the gulf crisis. Journal of Financial and Quantitative Analysis, 32, 91-115.

Neumann, M., and G. Skiadopoulos (2013). Predictable Dynamics in Higher Order Risk-Neutral Moments: Evidence from the S&P 500 Options. Journal of Financial and Quantitative Analysis 48(3), 947-977.

Table 1: Descriptive statistics for the raw option data. Source: Optionmetrics Ivy DB.

		S&P	DAX	CAC	FTSE	SMI
Dates		**2000/01/02 to 2013/12/31**	**2002/01/02 to 2013/12/31**	**2003/04/14 to 2013/12/31**	**2002/01/02 to 2013/12/31**	**2002/01/02 to 2013/12/31**
Number of options per day	**mean**	546.71	627.76	362.68	365.44	452.51
	min.	177	266	2	8	181
	max.	1664	1087	757	593	856
Implied volatility	**mean**	.26	.28	.24	.23	.22
	min.	.04	.01	.01	.01	.01
	max.	.99	1.00	.99	1.00	.99
Strike prices	**mean**	1172.42	5447.45	4044.97	5140.31	6388.85
	min.	50	500	800	1600	2000
	max.	3000	20000	8000	9600	13000
Number of strikes per day	**mean**	118.16	103.54	57.70	77.98	83.87
	min.	44	56	2	8	45
	max.	217	163	85	161	134
Time to maturity (days)	**mean**	202.59	358.45	449.53	233.18	335.56
	min.	6	6	7	7	7
	max.	1095	1824	1914	732	1824
Number of maturities per day	**mean**	10.98	13.59	13.26	9.85	11.74
	min.	7	7	1	4	7
	max.	19	17	22	11	17

Table 2: Descriptive statistics of the time series of measures on the risk-neutral distribution for the indexes studied. VRP is the variance risk premium defined as the difference between the risk-neutral variance and the historical realized variance. FH is the Foster-Hart measure of riskiness taken on the risk-neutral distribution each day of the sample. SK and Kurt are the skewness and kurtosis of the risk-neutral distribution. *Autocorr(1)* is the autocorrelation of the series at lag 1 and *P.Perron(10) p-value* is the p-value of a Phillips–Perron unit-root test with 10 lags.

	SPX			
	VRP	FH	SK	Kurt
Nb. Obs.	3356	3356	3356	3356
Mean	0.01	1.64	-1.05	11.73
Median	0.01	1.75	-0.91	7.73
Min	-0.40	-0.69	-7.03	2.33
Max	0.20	5.25	3.65	109.28
St. Err.	0.03	1.10	0.96	11.67
Skewness	-4.60	-0.13	-1.66	3.43
Kurtosis	58.84	2.25	11.41	19.10
Autocorr(1)	0.85	0.91	0.46	0.38
P.Perron(10) p-value	0.001	0.001	0.001	0.001
Correlations				
VRP	1	0.08	-0.05	0.06
FH		1	0.31	-0.22
SK			1	-0.51
Kurt				1

	DAX			
	VRP	FH	SK	Kurt
Nb. Obs.	2906	2906	2906	2906
Mean	0.00	1.82	-0.71	7.75
Median	0.01	1.84	-0.70	5.25
Min	-0.54	-0.93	-5.78	1.57
Max	0.29	9.01	14.00	562.90
St. Err.	0.05	0.98	0.80	16.06
Skewness	-4.94	0.40	6.79	24.32
Kurtosis	44.74	3.82	126.99	737.33
Autocorr(1)	0.93	0.87	0.34	0.18
P.Perron(10) p-value	0.001	0.001	0.001	0.001
Correlations				
VRP	1	0.01	-0.03	0.02
FH		1	0.03	-0.11
SK			1	0.45
Kurt				1

Table 2 (continued)

	SMI			
	VRP	FH	SK	Kurt
Nb. Obs.	2872	2872	2872	2872
Mean	0.00	1.89	-0.28	13.40
Median	0.00	1.82	-0.56	6.99
Min	-0.42	-0.68	-4.43	2.14
Max	0.12	8.72	24.33	1287.28
St. Err.	0.03	1.07	1.26	35.69
Skewness	-5.47	0.83	5.86	29.12
Kurtosis	57.66	5.07	85.99	1003.27
Autocorr(1)	0.91	0.74	0.41	0.11
P.Perron(10) p-value	0.001	0.001	0.001	0.001
Correlations				
VRP	1	0.00	0.10	0.05
FH		1	0.03	-0.03
SK			1	0.68
Kurt				1

	CAC			
	VRP	FH	SK	Kurt
Nb. Obs.	2523	2523	2523	2523
Mean	0.00	1.62	-0.74	10.27
Median	0.01	1.61	-0.68	5.44
Min	-0.55	-1.89	-6.06	2.19
Max	0.31	7.37	6.18	106.48
St. Err.	0.04	0.99	0.99	13.09
Skewness	-5.68	0.41	0.22	3.53
Kurtosis	59.44	3.42	12.90	17.23
Autocorr(1)	0.92	0.81	0.26	0.34
P.Perron(10) p-value	0.001	0.001	0.001	0.001
Correlations				
VRP	1	0.08	0.00	0.09
FH		1	0.12	-0.21
SK			1	-0.19
Kurt				1

Table 2 (continued)

	FTSE			
	VRP	FH	SK	Kurt
Nb. Obs.	2848	2848	2848	2848
Mean	0.01	1.19	-1.10	14.98
Median	0.01	1.16	-0.84	5.75
Min	-0.40	-0.69	-9.86	1.79
Max	0.13	7.57	11.90	224.13
St. Err.	0.03	1.16	1.61	23.51
Skewness	-5.11	0.39	-0.75	3.17
Kurtosis	56.60	2.93	12.40	14.08
Autocorr(1)	0.90	0.88	0.19	0.32
P.Perron(10) p-value	0.001	0.001	0.001	0.001
Correlations				
VRP	1	0.02	-0.11	0.14
FH		1	0.21	-0.29
SK			1	-0.46
Kurt				1

Table 3: Modified Diebold-Mariano test statistics on the difference between the squared prediction errors of returns for the forecasting model compared to the reference model. The usual levels of significance (1%,5%,10%) according to a Student distribution are shown as * for significantly negative numbers and ° for positive numbers. A significant test statistic that is negative indicates that the forecasting model outperforms the reference model.

SPX								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP	CST only	-1.69*	-1.76*	-2.31**	-2.12**	-1.66*	-1.45	-1.14
FH	CST only	-1.20	-1.54	1.16	2.07°°	2.09°°	1.71°	3.35°°°
SK	CST only	-1.51	-1.67*	1.79°	-2.29**	1.23	-0.02	-1.41
KUR	CST only	1.86°	1.68°	1.67°	3.50°°°	2.49°°	1.75°	1.39
All	CST only	-1.42	-1.29	-2.04**	-2.20**	-1.90*	-1.56	-2.77***
All	All ex. VRP	-1.76*	-1.77*	-2.15**	-1.77*	-1.23	-1.00	-1.01
All	All ex. FH	0.95	1.28	1.27	-1.59	-2.16**	-1.09	-9.78***
All	All ex. SK	0.03	0.02	0.07	-0.35	0.98	0.62	-0.04
All	All ex. KUR	0.94	0.75	0.45	0.55	0.35	0.32	0.50

DAX								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP	CST only	-1.15	-1.39	-1.33	-0.11	1.33	0.76	-3.39***
FH	CST only	-0.02	-0.82	-0.12	0.57	1.19	1.37	0.76
SK	CST only	-0.04	-1.52	-1.37	-1.24	-0.22	-0.29	-0.62
KUR	CST only	1.25	0.98	1.35	2.48°°	0.18	0.31	-0.33
All	CST only	-0.89	-1.03	-0.96	-0.75	-0.35	-0.58	-0.49
All	All ex. VRP	-1.20	-1.39	-1.34	-0.26	1.53	0.91	-1.12
All	All ex. FH	0.68	1.63	0.51	-0.46	-0.75	-0.77	-0.48
All	All ex. SK	-1.10	0.07	-0.31	-1.32	0.78	0.15	-0.80
All	All ex. KUR	0.93	1.97°°	2.08°°	2.59°°°	1.31	1.86°	1.35

Table 3 (continued)

SMI								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP	CST only	0.17	-0.16	-0.26	0.38	0.56	0.57	2.32°°
FH	CST only	1.59	1.25	1.37	1.18	1.12	1.16	0.71
SK	CST only	0.46	0.00	0.60	0.59	1.96°	0.50	0.05
KUR	CST only	0.24	0.31	0.13	0.05	0.52	-0.04	0.42
All	CST only	0.49	0.35	0.56	1.01	1.06	1.14	0.07
All	All ex. VRP	0.23	-0.13	-0.27	0.50	0.78	0.83	7.93°°°
All	All ex. FH	0.64	1.22	1.07	0.96	0.81	1.11	0.03
All	All ex. SK	-0.31	-0.43	-0.81	1.52	1.53	2.32°°	0.71
All	All ex. KUR	0.10	-0.28	-0.53	0.79	1.58	2.68°°°	1.05

CAC								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP	CST only	-1.81*	-1.86*	-2.12**	-1.01	0.17	0.12	1.87°
FH	CST only	1.23	1.13	1.47	1.86°	1.67°	1.35	0.80
SK	CST only	-0.18	-0.43	-1.23	-1.43	-3.51***	-2.12**	-0.90
KUR	CST only	-0.99	-1.27	0.73	1.16	1.49	1.89°	-0.17
All	CST only	-1.99**	-2.02**	-1.74*	-1.31	-1.36	-1.15	-0.54
All	All ex. VRP	-1.77*	-1.82*	-1.95*	-0.52	0.75	0.82	3.74°°°
All	All ex. FH	-0.19	0.45	0.00	-0.97	-1.45	-1.47	-0.54
All	All ex. SK	-0.01	-0.38	-1.12	-1.10	-1.32	-1.17	-0.65
All	All ex. KUR	-1.22	-1.55	0.51	0.26	0.26	-0.71	-0.96

Table 3 (continued)

FTSE								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP	CST only	-1.30	-1.61	-2.34**	-1.45	-0.60	-0.52	-0.71
FH	CST only	0.17	0.10	0.38	0.99	1.23	1.32	0.76
SK	CST only	1.26	0.78	-0.40	0.20	0.15	-0.19	2.18°°
KUR	CST only	0.76	-0.03	0.71	0.79	0.61	0.47	0.82
All	CST only	-0.90	-0.96	-0.19	0.10	0.01	0.12	0.76
All	All ex. VRP	-1.34	-1.63	-2.32**	-1.17	0.09	0.92	0.60
All	All ex. FH	1.42	1.07	0.84	0.54	0.02	0.14	0.73
All	All ex. SK	1.08	0.86	-0.42	-0.18	0.41	-1.03	0.60
All	All ex. KUR	-0.33	-0.22	0.66	0.77	0.39	-0.85	0.60

Table 4: Additional tests for the S&P500. Modified Diebold-Mariano test statistics on the difference between the squared prediction errors of returns for the forecasting model compared to the reference model. The usual levels of significance (1%,5%,10%) according to a Student distribution are shown as * for significantly negative numbers and ° for positive numbers. A significantly negative test statistics indicate that the forecasting model outperforms the reference model.

SPX								
Forecasting model	Reference model	Return horizon (in business days)						
		1	2	5	20	50	100	250
VRP,FH	FH	-1.31	-1.06	-2.23**	-2.62***	-2.46**	-1.89*	-5.19***
VRP,FH	VRP	1.23	1.47	1.63	-1.91*	-2.12**	-0.96	-6.96***
VRP,FH	CST only	-1.47	-1.35	-2.16**	-2.48**	-2.24**	-1.90*	-3.89***

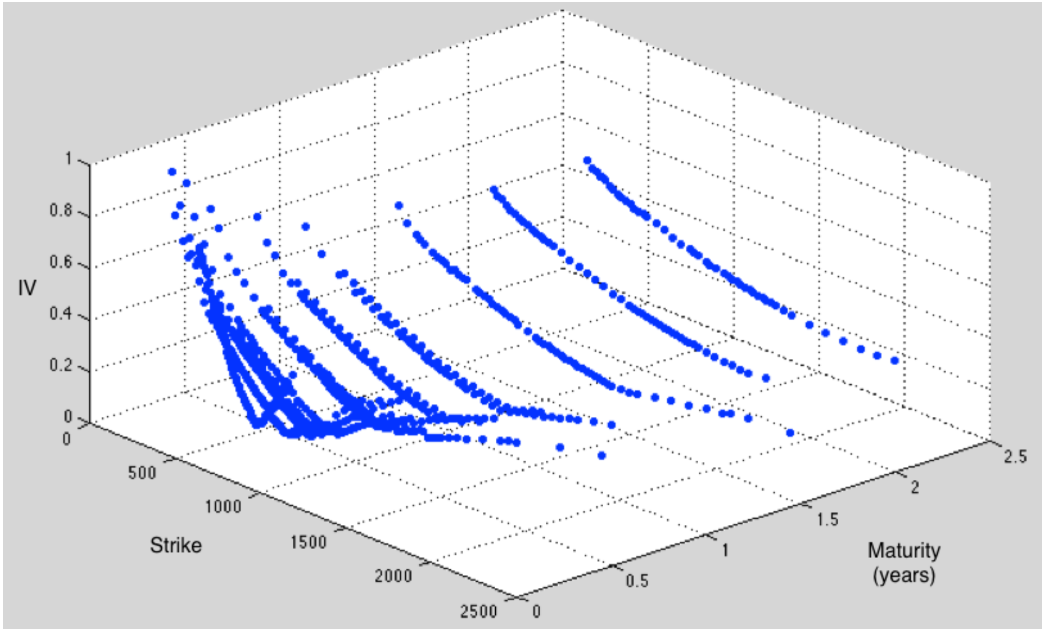

Figure 1: Example of an IV-maturity-strike surface of points.

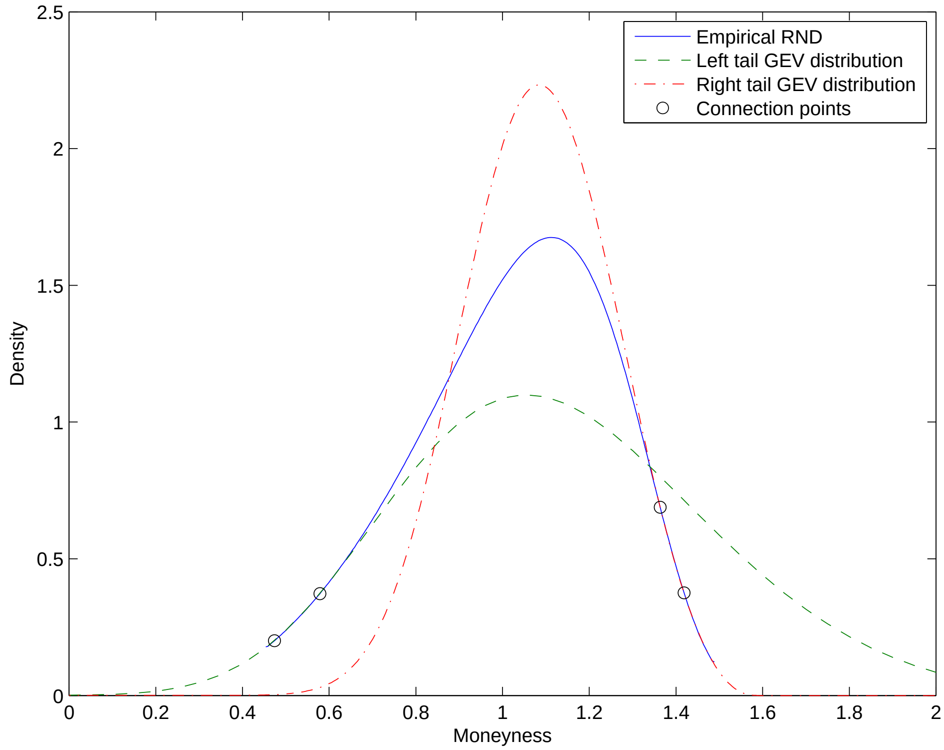

Figure 2: The empirical risk-neutral distribution and the GEV tails used to complete it.

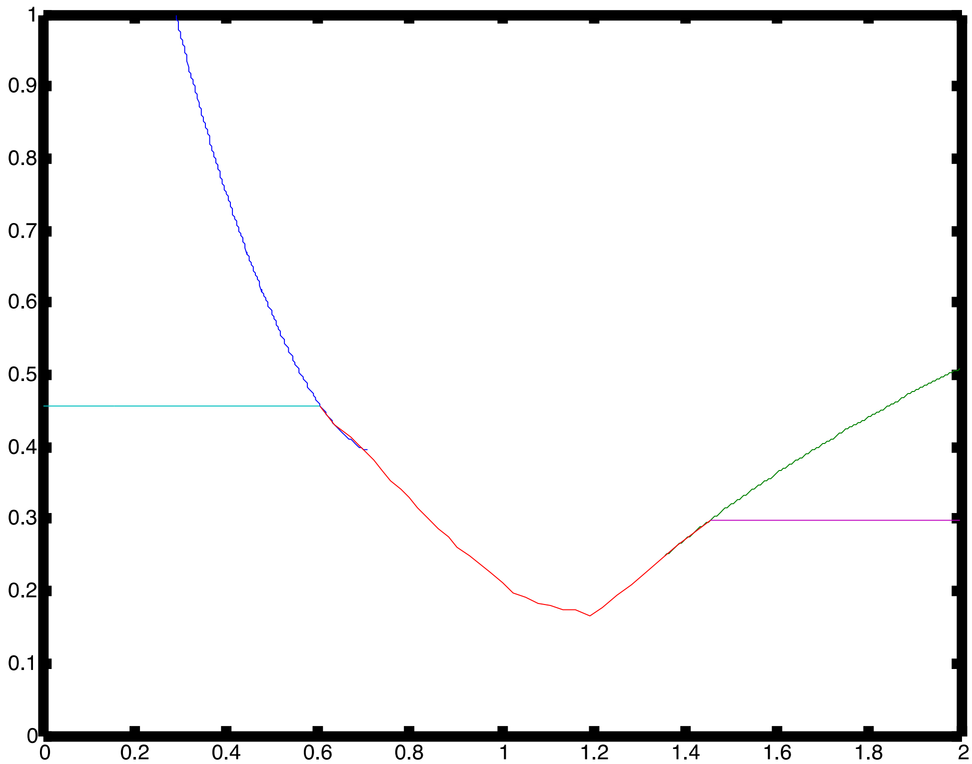

Figure 3: Completed IV curve with the two methods. The empirical IV curve for the S&P 500 options on 2002/01/02 is completed with the two different approaches. In teal and in purple the IV curve is assumed to be constant beyond the available data. In blue and in green a GEV distribution is fitted to each tail.

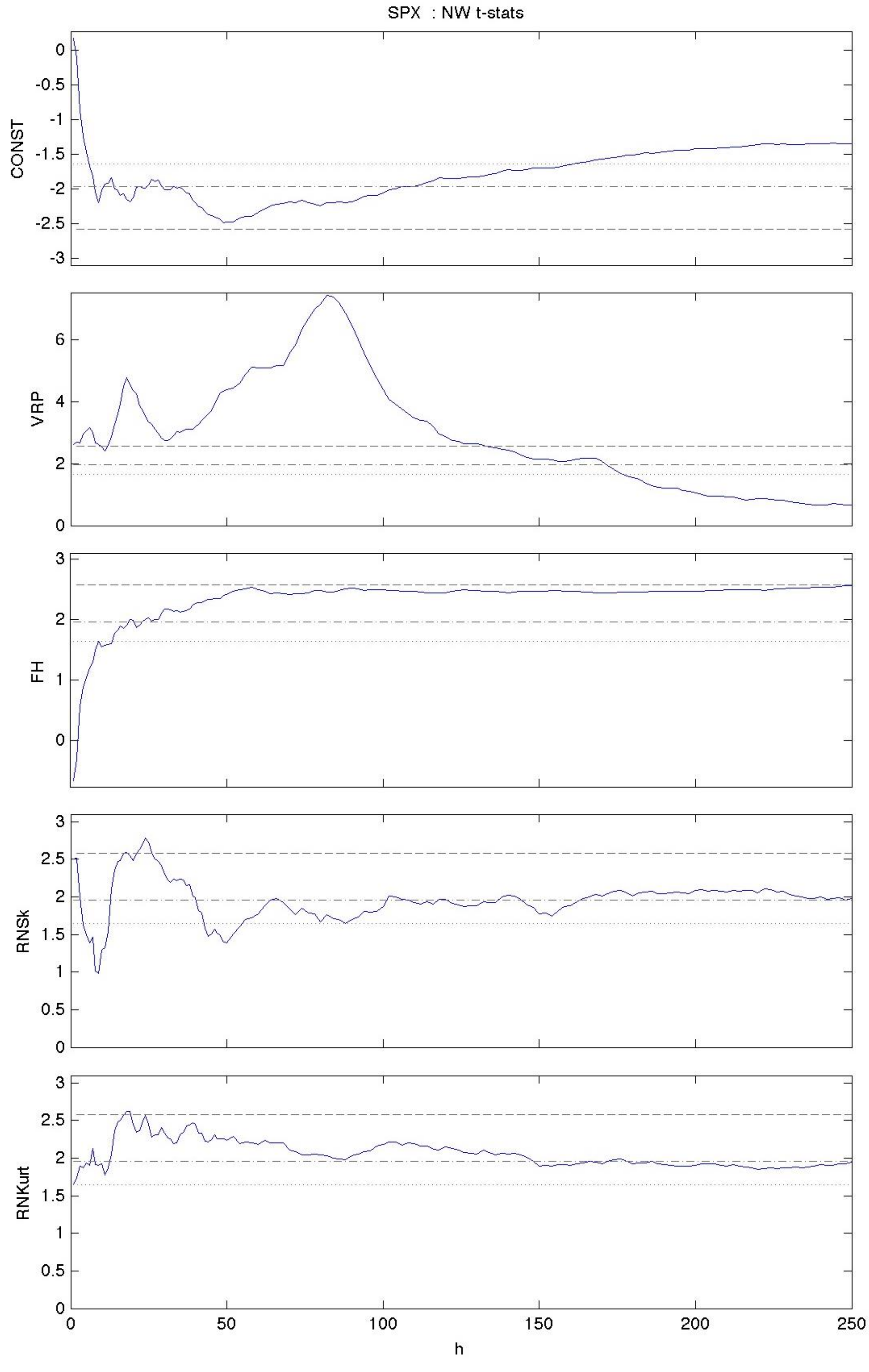

Figure 4: S&P 500. Newey-West adjusted t-statistics of the regression coefficients of equation (1) for future returns of the S&P 500 index at different horizons h in days. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

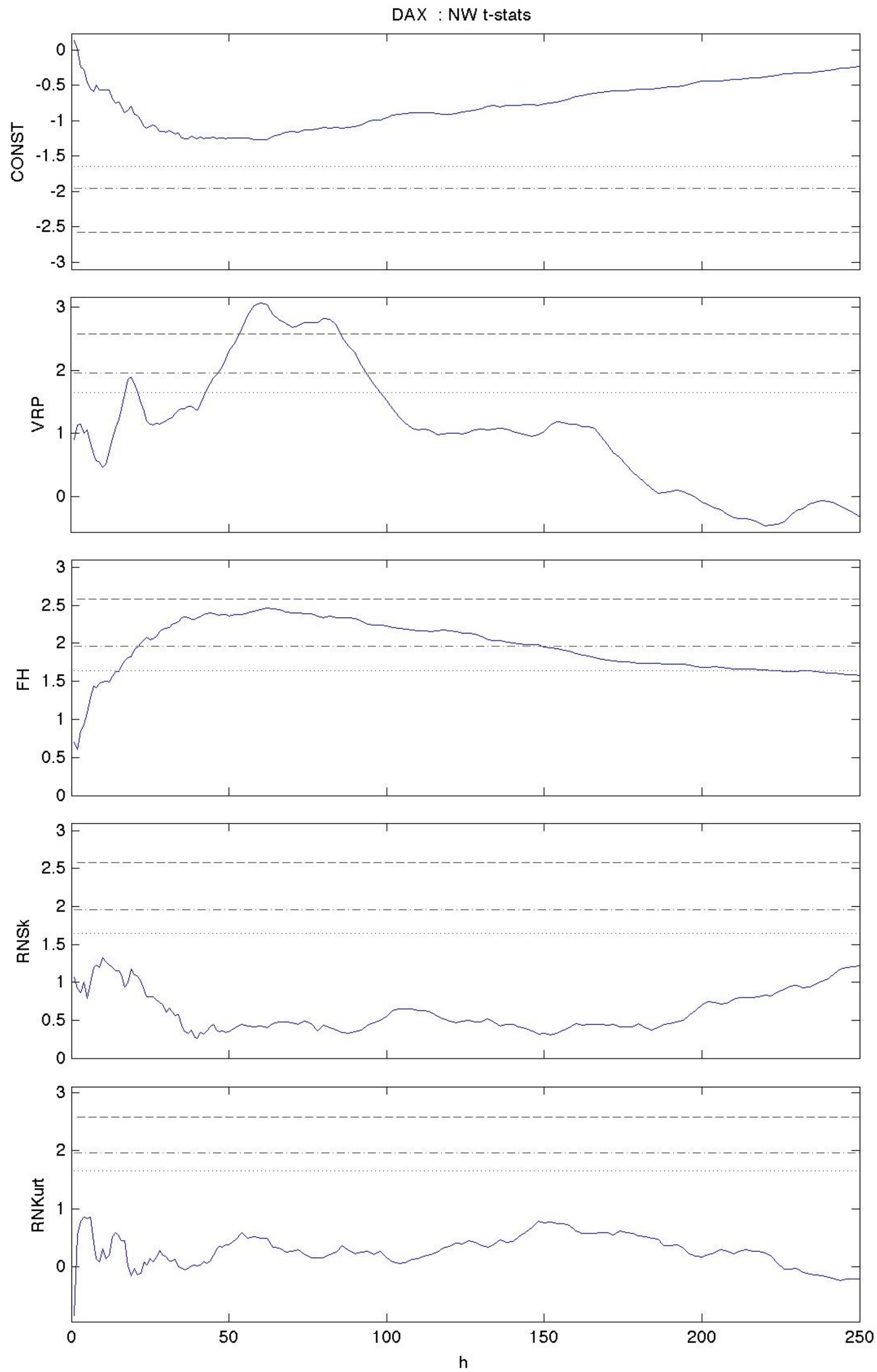

Figure 5: DAX. Newey-West adjusted t-statistics of the regression coefficients of equation (1) for future returns of the German DAX index at different horizons *h* in days. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

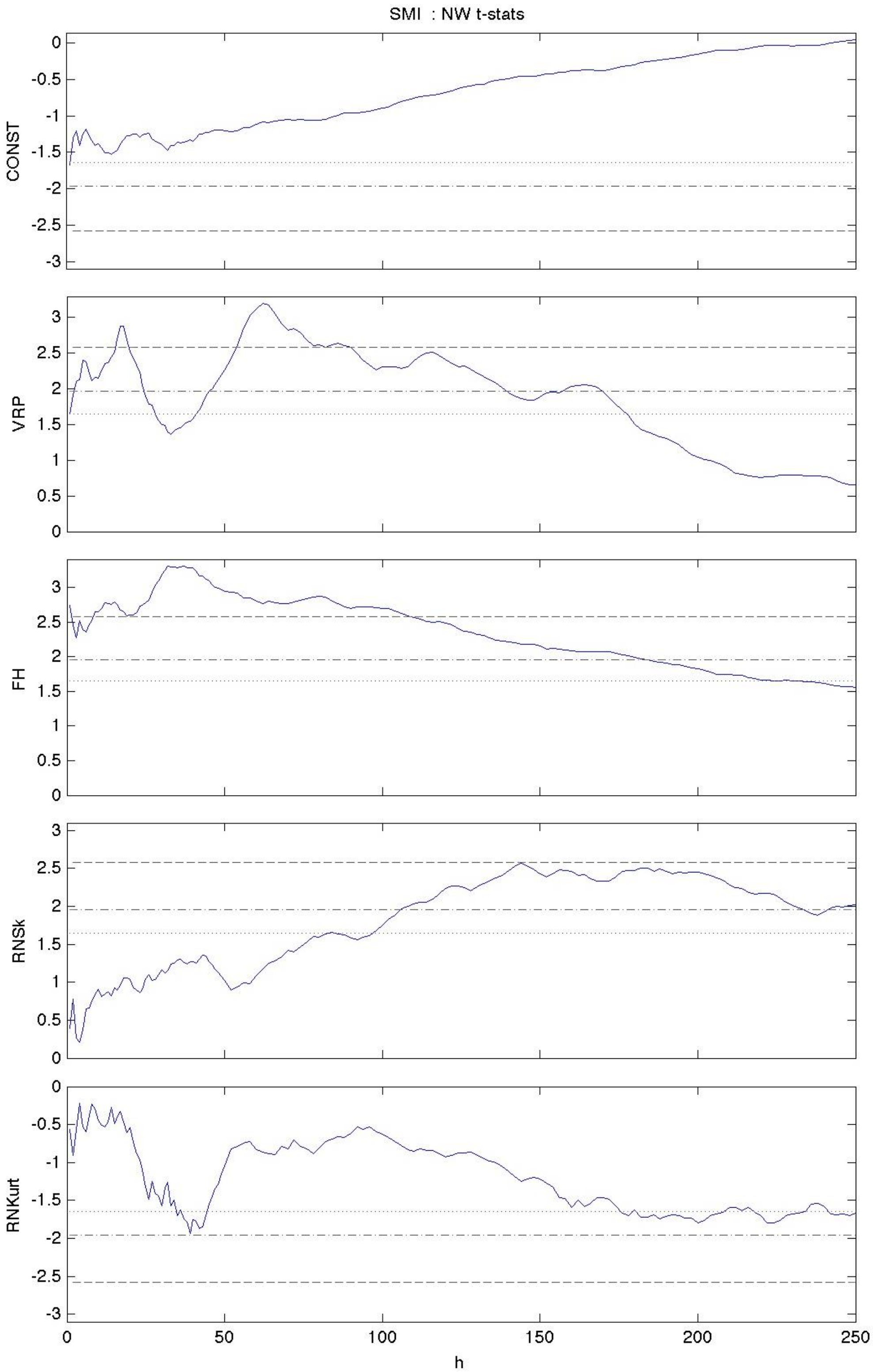

Figure 6: SMI. Newey-West adjusted t-statistics of the regression coefficients of equation (1) for future returns of the Swiss SMI index at different horizons *h* in days. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

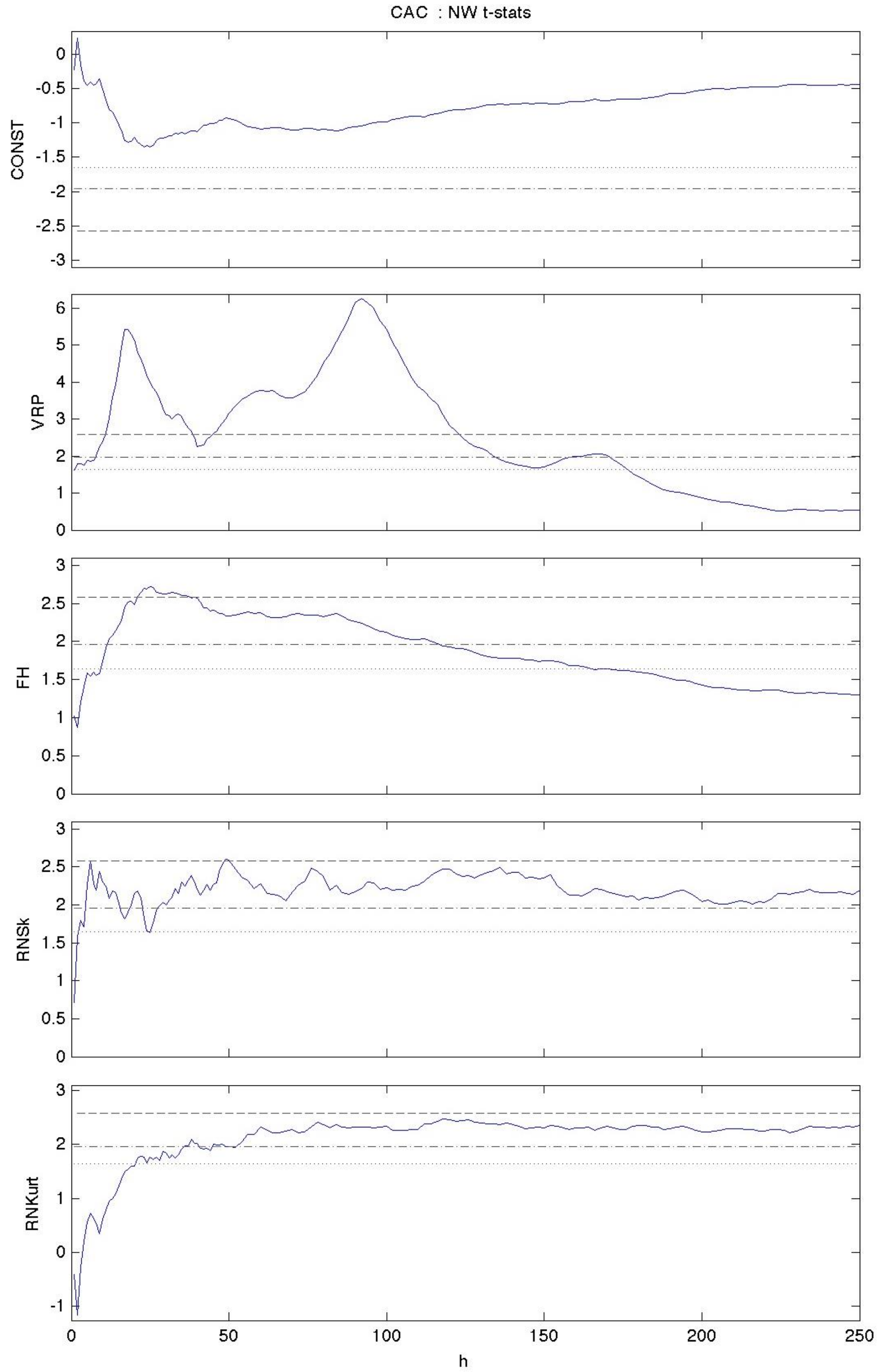

Figure 7: CAC. Newey-West adjusted t-statistics of the regression coefficients of equation (1) for future returns of the French CAC index at different horizons *h* in days. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

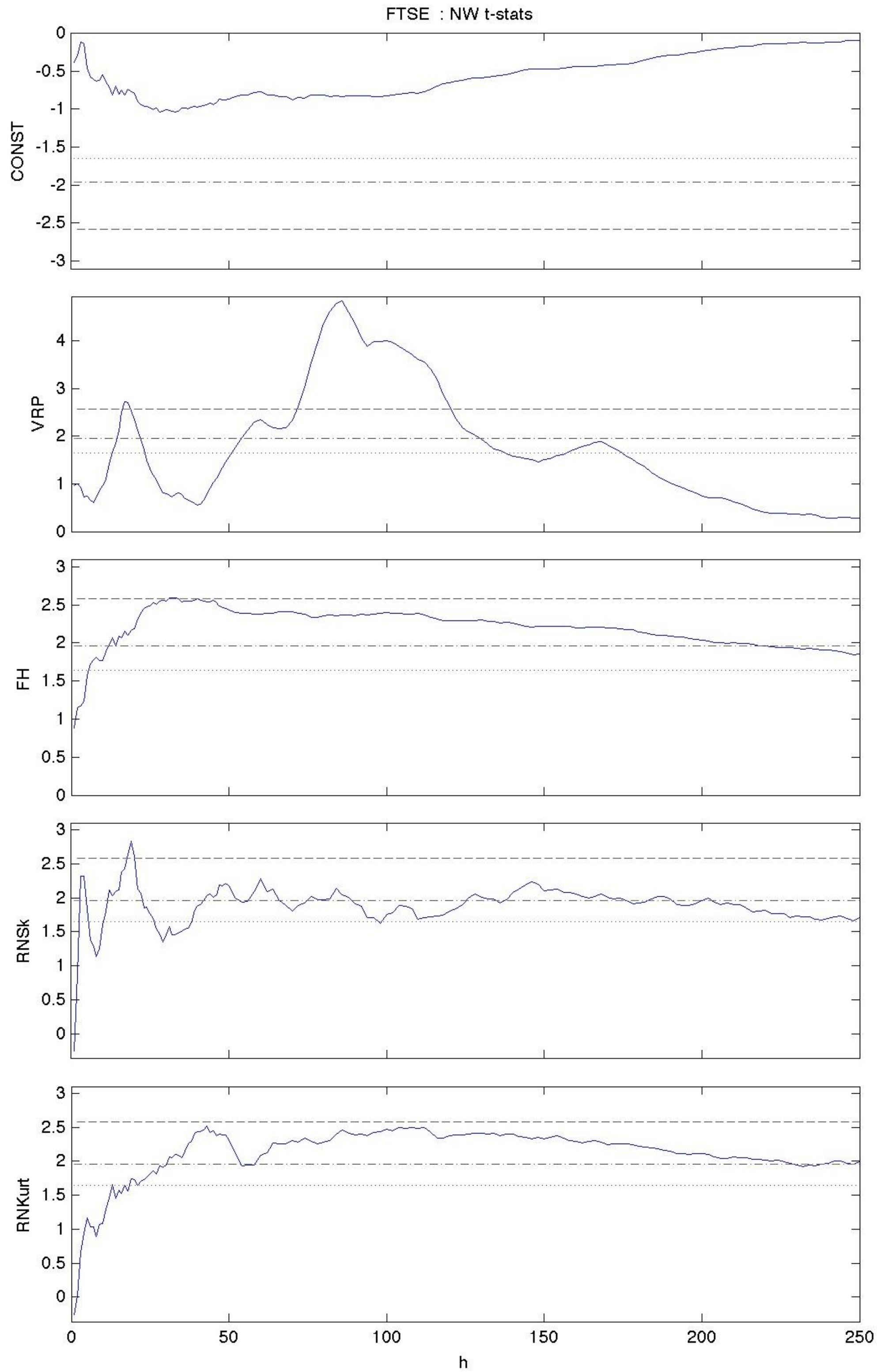

Figure 8: FTSE. Newey-West adjusted t-statistics of the regression coefficients of equation (1) for future returns of the UK FTSE index at different horizons *h* in days. The usual levels of significance are plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes).

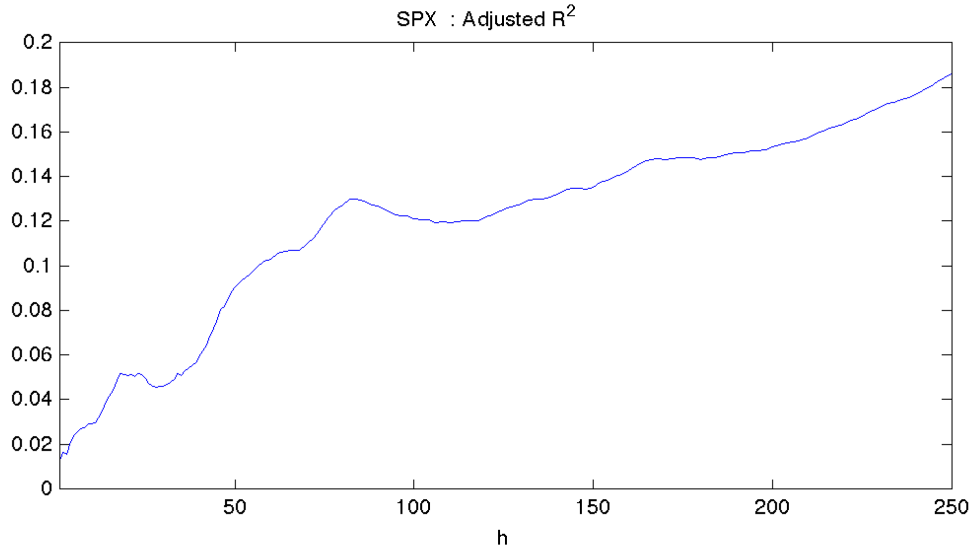

Figure 9: S&P500. Adjusted R^2 for the regression in equation (1) for future returns of the US S&P500 index at different horizons *h* in days.

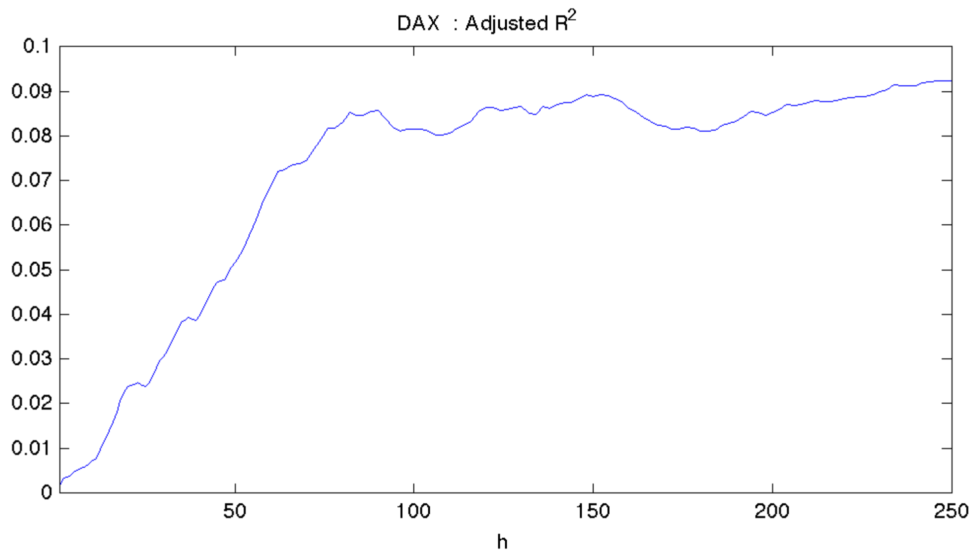

Figure 10: DAX. Adjusted R^2 for the regression in equation (1) for future returns of the German DAX index at different horizons *h* in days.

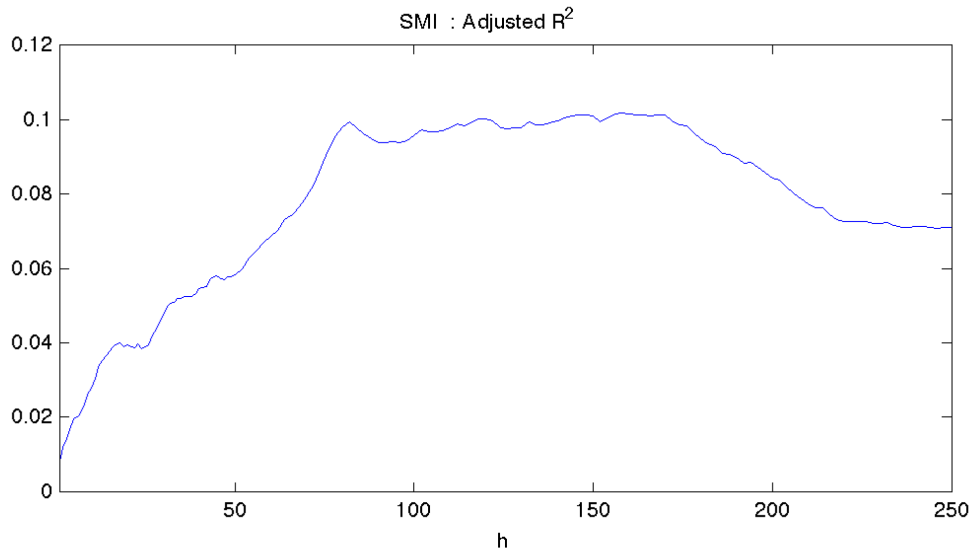

Figure 11: SMI. Adjusted R^2 for the regression in equation (1) for future returns of the Swiss SMI index at different horizons *h* in days.

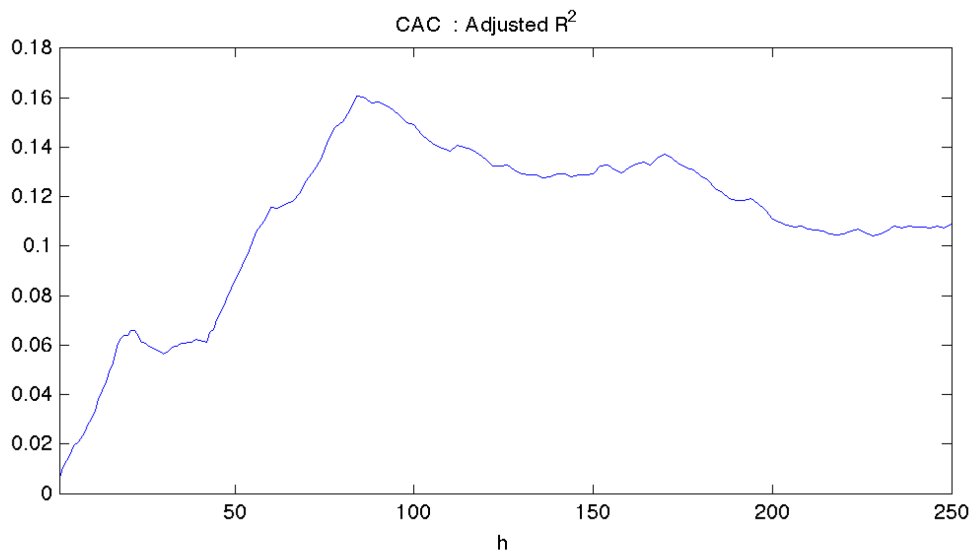

Figure 12: CAC. Adjusted R^2 for the regression in equation (1) for future returns of the French CAC index at different horizons *h* in days.

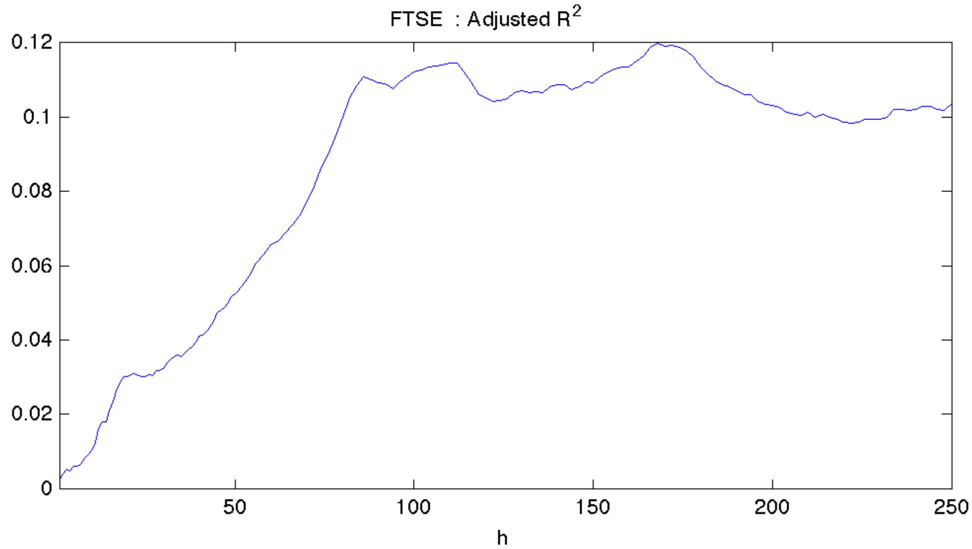

Figure 13: FTSE. Adjusted R^2 for the regression in equation (1) for future returns of the UK FTSE index at different horizons *h* in days.

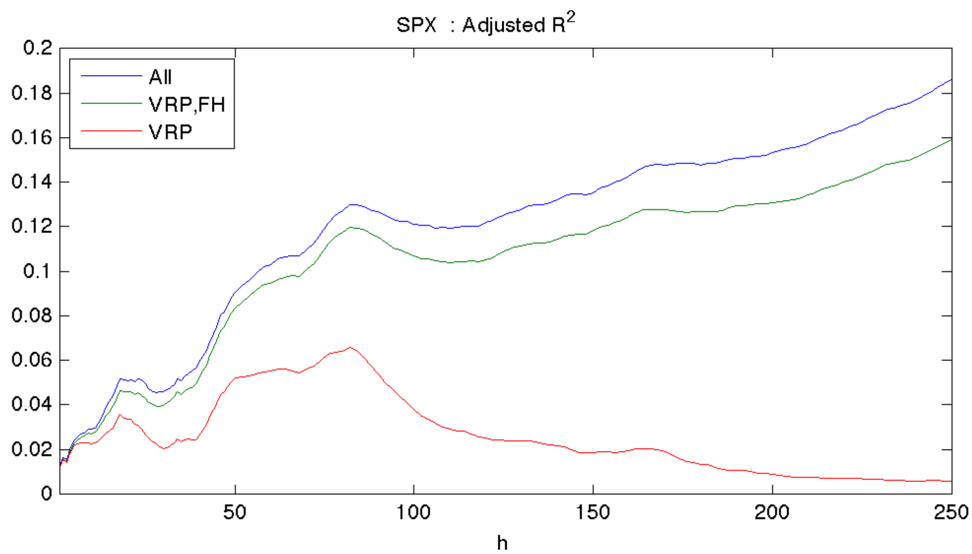

Figure 14: S&P500. Adjusted R^2 for the regression in equation (1) on future returns of the S&P 500 index at different horizons *h* in days. *All* is the original model with all the variables. *VRP* is a more parsimonious variation with only a constant and the VRP. *VRP,FH* is a variation with only a constant, the VRP and the FH measure.