Volatility-related exchange traded assets: an econometric investigation*

Javier Mencía
Bank of Spain
<javier.mencia@bde.es>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

February 2015
revised April 2016

Abstract

We develop a theoretical framework for covariance stationary but persistent positively valued processes which combines a semi-nonparametric expansion of the Gamma distribution with a component version of the Multiplicative Error Model. Our conditional mean assumption allows for slow, possibly non-monotonic mean-reversion while our distribution assumption provides more flexibility than a traditional Laguerre expansion while preserving positivity of the density. We apply our framework to a dynamic portfolio allocation for Exchange Traded Notes tracking short- and mid-term VIX futures indices, which are increasingly popular but risky financial instruments. We show the superior performance of the strategies based on our econometric model.

Keywords: Density Expansions, Exchange Traded Notes, Multiplicative Error Model, Volatility Index Futures.

JEL: G13, C16.

*We are grateful to Dante Amengual, Nour Meddahi, Michael Rockinger, Genaro Sucarrat and participants at the XXIII Finance Forum for very useful comments and suggestions. We have also benefited from the comments of the associate editor and two anonymous referees. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Economy and Competitiveness through grant ECO 2014-59262 (Sentana) is also gratefully acknowledged. The views expressed in this paper are those of the authors, and do not reflect those of the Bank of Spain or the Eurosystem. Address for correspondence: Casado del Alisal 5, E-28014, Madrid, Spain, tel: +34 91 429 05 51, fax: +34 91 429 1056.

1 Introduction

The introduction of the new VIX index by the Chicago Board Options Exchange (CBOE) in 2003 meant that volatility became widely regarded as an asset class on its own. As is well known, VIX captures the volatility of the Standard & Poor's 500 (S&P500) over the next month implicit in stock index option prices, and for that reason it has become a widely accepted measure of stock volatility and a market fear gauge. In addition, since March 26, 2004 it is possible to directly invest in volatility through futures contracts on the VIX negotiated at the CBOE Futures Exchange (CFE). More recently, several volatility related Exchange Traded Notes (ETNs) have provided investors with equity-like long and short exposure to constant maturity futures on the VIX, and even dynamic combinations of long-short exposures to different maturities (see Rhoads, 2011). Although the poor performance of some of these assets during decreasing volatility periods have raised some concerns about their risks, by 2013 there were already about 30 ETNs with a market cap of around \$3 billion and a trading volume on some of them of close to \$5 billion per day (see Alexander and Korovilas, 2013, for further details).

Volatility indices such as the VIX provide important examples of financial time series where the original data is always positive but stationary in levels, with a slow reversion to their long run mean. Many discrete and continuous time models have been proposed to capture this strong persistence. An increasingly popular example is the discrete-time Multiplicative Error Model (MEM) proposed by Engle (2002), which has been applied not just to volatility modelling but also to trading volumes and durations (see Brownlees, Cipollini, and Gallo, 2012, for a recent review). In this model, a positive random variable is treated as the product of a time varying, recursive mean times a positive random error with unit conditional mean. In this regard, Engle and Gallo (2006) show on the basis of earlier results by Gourieroux, Monfort, and Trognon (1984) that the mean parameters can be consistently estimated assuming a Gamma distribution for the error term even when the true distribution is not Gamma, as long as the conditional mean is correctly specified. Unfortunately, this pseudo-likelihood approach is insufficient when the interest goes beyond the first conditional moment. For that reason, some authors have proposed more flexible distributional assumptions (see e.g. De Luca and Gallo (2004, 2009) and Lanne (2006)).

One particularly important situation where the entire conditional distribution matters arises in assessing trading strategies involving VIX ETNs. Intuitively, risk averse investors must take into account not only the expected value of the resulting payoffs, which can be obtained from the mean forecasts generated by the MEM, but also some suitable measures of the risks involved, which necessarily depend on features of the conditional distribution beyond its mean. In this context, we develop a comprehensive dynamic asset allocation framework to invest in precisely those financial instruments.

We begin by modelling the mean-reverting features of the VIX with a two component MEM specification analogous to the GARCH model proposed by Engle and Lee (1999). Then, we make use of a semi-nonparametric expansion of the Gamma density (Gamma SNP or GSNP for short). SNP expansions were introduced by Gallant and Nychka (1987) for nonparametric estimation purposes as a way to ensure by construction the positivity of the resulting density (see also Fenton and Gallant, 1996; Gallant and Tauchen, 1999). In our case, though, we follow León, Mencía, and Sentana (2009) in treating the SNP distribution parametrically as if it reflected the actual data generating process instead of an approximating kernel. Next, we specify a stochastic discount factor (SDF) with which we derive an equivalent risk-neutral measure that allows us to obtain closed-form expressions for the prices of VIX futures ETNs. Using those three ingredients, we study asset allocation strategies in ETNs tracking the VIX futures short and mid-term indices. We compare our strategy with buy and hold positions on existing ETNs, some of which are already dynamic combinations of the VIX futures indices, as well as other strategies that have been previously proposed in the literature. Finally, we assess the sensitivity of our results to the evaluation criterion, and compare our model with two alternative approaches: (i) a reduced form model and (ii) the autoregressive Gamma process proposed by Gourieroux and Jasiak (2006).

The rest of the paper is organised as follows. In the next section, we study the statistical properties of the GSNP density. In Section 3, we describe our pricing framework, relate the real and risk-neutral measures, and obtain futures prices. Section 4 presents the empirical application. Finally, we conclude in Section 5. Proofs and auxiliary results can be found in the appendices.

2 Semi-nonparametric Gamma density expansions

Consider the Gamma distribution, whose probability density function (pdf) can be expressed as

$$f_G(x,\nu,\psi) = \frac{1}{\Gamma(\nu)\psi^{\nu}} x^{\nu-1} \exp(-x/\psi), \tag{1}$$

where $\Gamma(\cdot)$ denotes the Gamma function, ν are the degrees of freedom and ψ the scale parameter. For the sake of brevity, we will denote this density as $G(\nu,\psi)$. Following Gallant and Nychka (1987), we consider SNP expansions of this density (GSNP for short):

$$f_{GSNP}(x,\nu,\psi,\boldsymbol{\delta}) = f_G(x,\nu,\psi)\left[\sum_{j=0}^{m}\delta_j\left(\frac{x}{\psi}\right)^j\right]^2\frac{1}{d}, \tag{2}$$

where $\boldsymbol{\delta} = (\delta_0, \delta_1, \cdots, \delta_m)'$, and d is a constant that ensures that the density integrates to 1.

In order to interpret (2), it is convenient to expand the squared term. This yields the following result:

Proposition 1 *Let x be a $GSNP_m(\nu,\psi,\boldsymbol{\delta})$ variable with density $f_{GSNP}(x,\nu,\psi,\boldsymbol{\delta})$ given by (2). Then*

$$\begin{aligned} f_{GSNP}(x,\nu,\psi,\boldsymbol{\delta}) &= f_G(x,\nu,\psi)\frac{1}{d}\sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\left(\frac{x}{\psi}\right)^j, & (3)\\ &= \frac{1}{d}\sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\frac{\Gamma(\nu+j)}{\Gamma(\nu)}f_G(x,\nu+j,\psi), & (4)\end{aligned}$$

where

$$\gamma_j(\boldsymbol{\delta}) = \sum_{k=\max\{j-m,0\}}^{\min\{j,m\}}\delta_j\delta_{j-k}.$$

Using Proposition 1, it is straightforward to show that the constant of integration can be expressed as

$$d = \sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\frac{\Gamma(\nu+j)}{\Gamma(\nu)}.$$

But since (2) is homogeneous of degree zero in $\boldsymbol{\delta}$, there is a scale indeterminacy that we must solve by imposing a single normalising restriction on these parameters, such as $\delta_0 = 1$, or preferably $\boldsymbol{\delta}'\boldsymbol{\delta} = 1$, which we can ensure by working with hyperspherical coordinates.[1]

[1]In particular, $\nu_0 = \cos\theta_1$; $\nu_i = (\prod_{k=1}^{i}\sin\theta_k)\cos\theta_{i+1}$ for $0 < i \leq m-1$; and $\nu_m = \prod_{k=1}^{m}\sin\theta_k$, where $\theta_k \in [0,\pi)$, for $1 < k \leq m-1$, and $\theta_m \in [0, 2\pi)$.

Proposition 1 allows us to interpret the GSNP distribution as a mixture of $2m + 1$ Gamma distributions.[2] We can exploit the mixture interpretation together with the results in Appendix A.1 to write the moment generating function of a GSNP variable x as

$$E\left[\exp(nx)\right] = \frac{1}{d} \sum_{j=0}^{2m} \gamma_j(\boldsymbol{\delta}) \frac{\Gamma(\nu + j)}{\Gamma(\nu)} (1 - \psi n)^{-(\nu+j)}.$$

Similarly, its characteristic function can be expressed as

$$\psi_{GSNP}(i\tau) = \frac{1}{d} \sum_{j=0}^{2m} \gamma_j(\boldsymbol{\delta}) \frac{\Gamma(\nu + j)}{\Gamma(\nu)} (1 - i\psi n)^{-(\nu+j)},$$

where i is the usual imaginary unit. As a result, we can write the moments of x as

$$E(x^n) = \frac{\psi^n}{d} \sum_{j=0}^{2m} \gamma_j(\boldsymbol{\delta}) \frac{\Gamma(\nu + j + n)}{\Gamma(\nu)}.$$

Hence, it is straightforward to show that the condition

$$\psi = d \left[\sum_{j=0}^{2m} \gamma_j(\boldsymbol{\delta}) \frac{\Gamma(\nu + j + 1)}{\Gamma(\nu)} \right]^{-1} \tag{5}$$

ensures that $E(x) = 1$. Since we plan to use the GSNP distribution to model the residual in MEM models, we assume in what follows that (5) holds to fix its scale.

By reordering the terms in (3) appropriately, we can also interpret the GSNP distribution as a finite order version of the Laguerre expansion of the Gamma distribution discussed in appendix A. We can formally express this relationship as follows:

Proposition 2 *Let x be a $GSNP_m(\nu, \psi, \boldsymbol{\delta})$ variable with density $f_{GSNP}(x, \nu, \psi, \boldsymbol{\delta})$ given by (2). Then, this density can be expressed as a Laguerre expansion (A2) of order $2m$ with coefficients*

$$c_n = \frac{(-1)^n}{d} \sqrt{\frac{\Gamma(\nu)n!}{\Gamma(\nu + n)}} \sum_{i=0}^{n} \sum_{j=0}^{2m} \frac{(-1)^i}{i!} \binom{n + \nu - 1}{n - i} \frac{\Gamma(\nu + i + j)\psi^i}{\Gamma(\nu)\bar{\psi}^i} \gamma_j(\boldsymbol{\delta})$$

for $n = 0, \cdots, 2m$.

Importantly, we systematically treat the GSNP distribution as a flexible parametric distribution which remains non-negative for all possible values of x by construction.[3]

[2] This interpretation is consistent with Bowers (1966), who expands general density functions for positive random variables using sums of Gamma densities. Interestingly, the mixing variable of the equivalent mixture might have some negative weights, as in Steutel (1967) and Bartholomew (1969). However, this causes no inconsistencies because by construction the GSNP density is positive for all values of the parameters.

[3] The GSNP satisfies sufficient conditions for positivity. See Meddahi (2001) and León, Mencía, and Sentana (2009) for a discussion of necessary and sufficient conditions.

3 Component MEM applied to the valuation of volatility futures

3.1 Real measure

Consider a non-traded volatility index whose value at time t is $V_t \geqslant 0$. We model this variable using the Multiplicative Error Model (MEM) proposed by Engle (2002). Specifically, we model the volatility index under the real measure $\mathbb{P}$ as

$$V_t = \mu_t(\boldsymbol{\theta})\varepsilon_t, \quad \mu_t(\boldsymbol{\theta}) = E(V_t|I_{t-1}), \tag{6}$$

where I_{t-1} denotes the information observed at $t-1$, $\boldsymbol{\theta}$ is a vector of parameters and ε_t is a unit mean *iid* non-negative variable. Engle and Gallo (2006) show that we can obtain a consistent estimator of $\boldsymbol{\theta}$ using the Gamma distribution even though the true distribution is not Gamma as long as $\mu_t(\boldsymbol{\theta})$ is correctly specified. However, in our case we are also interested in higher order moments because we want to study asset allocation strategies. Therefore, we will assume that ε_t follows a $GSNP_m(\nu, \psi, \boldsymbol{\delta})$ as a natural flexible generalisation of the Gamma distribution. As we mentioned before, we will use the scale restriction (5) to ensure that ε_t has unit mean.

Figure 1a shows that historically the VIX has mean reverted, but experiencing highly persistent swings. Figure 1b shows the more recent evolution of the VIX together with that of the CBOE S&P500 3-month volatility index, or VXV for short. Both series display similar mean reverting features, which is natural given that they measure volatility on the same variable at different horizons, but they do not coincide. For example, the VIX reached a maximum value of 80.86 on November 20, 2008, which was around 10 points higher than the VXV. As highlighted by Schwert (2011), this indicates that during the financial crisis the market did not expect the volatility of the S&P500 to remain at such high levels forever.

In an earlier paper (Mencía and Sentana, 2013), we modelled the VIX index in a continuous time framework, finding that it is crucial to allow for mean reversion to a time-varying long run mean, which in turn mean reverts more slowly (see also Amengual and Xiu, 2013, Bardgett, Gourier, and Leippold, 2014, and Song and Xiu, 2016, for other related continuous time models that explicitly look at the evolution of the VIX under the real and risk neutral measures). In this paper, though, we prefer to use a discrete time model because it allows us to uncouple the specification of the mean process from the shape of the conditional distribution. Thus, we are able to easily modify the distribution while keeping the autocorrelation structure of the model fixed.

In order to incorporate the aforementioned mean-reverting features in a discrete time setting, we use the MEM analogue to the component GARCH model proposed by Engle and Lee (1999). In particular, we model the conditional mean as the sum of two components $\mu_t(\boldsymbol{\theta}) = \varsigma_t(\boldsymbol{\theta}) + s_t(\boldsymbol{\theta})$. We parametrise the first component as

$$\varsigma_t(\boldsymbol{\theta}) = \omega + \rho\varsigma_{t-1}(\boldsymbol{\theta}) + \varphi(V_{t-1} - \mu_{t-1}(\boldsymbol{\theta})),$$

while

$$s_t(\boldsymbol{\theta}) = (\alpha + \beta)s_{t-1}(\boldsymbol{\theta}) + \alpha(V_{t-1} - \mu_{t-1}(\boldsymbol{\theta})).$$

Hence, the second component mean reverts to zero, while the first one mean reverts to $\omega/(1-\rho)$. In turn, the coefficients ρ and $(\alpha+\beta)$ indicate the corresponding mean reversion speeds, so that if $\rho > \alpha+\beta$ then $\varsigma_{t+n|t}(\boldsymbol{\theta})$ will be more persistent than $s_{t+n|t}(\boldsymbol{\theta})$. The unconditional mean implied by this model is $E[\mu_t(\boldsymbol{\theta})] = \omega/(1-\rho)$. Using the results in Engle and Lee (1999), we can show that the n-period ahead forecast can be easily obtained in closed form as $E(V_{t+n}|I_t) = \varsigma_{t+n|t}(\boldsymbol{\theta}) + s_{t+n|t}(\boldsymbol{\theta})$, where

$$\begin{aligned}
\varsigma_{t+n|t}(\boldsymbol{\theta}) &= \omega\frac{1}{1-\rho} + \rho^{n-1}\left[\varsigma_{t+1}(\boldsymbol{\theta}) - \frac{\omega}{1-\rho}\right], \\
s_{t+n|t}(\boldsymbol{\theta}) &= (\alpha+\beta)^{n-1}s_{t+1}(\boldsymbol{\theta}).
\end{aligned}$$

As a result, the convergence of $E(V_{t+n}|I_t)$ to its long-run value $\omega/(1-\rho)$ can be non-monotonic.

3.2 Risk-neutral measure

We solve the problem of pricing derivatives on V_t by defining a stochastic discount factor with an exponentially affine form

$$M_{t-1,t} \propto \exp(-\alpha\varepsilon_t). \tag{7}$$

Such a specification corresponds to the Esscher transform used in insurance (see Esscher, 1932). In option pricing applications, this approach was pioneered by Gerber and Shiu (1994), and has also been followed by Buhlman, Delbaen, Embrechts, and Shyraev (1996, 1998), Gourieroux and Monfort (2006, 2007) and Bertholon, Monfort, and Pegoraro (2003) among others. On this basis, we can easily characterise the risk-neutral measure as follows:

Proposition 3 *Assume that the volatility index V_t follows the process given by (6) under the real measure $\mathbb{P}$, where the distribution of ε_t is a $GSNP_m(\nu, \psi, \boldsymbol{\delta})$ and (5) holds. Then, if the stochastic discount factor is defined by (7), under the equivalent risk-neutral measure $\mathbb{Q}$ we will have that $V_t = \mu_t(\boldsymbol{\theta})\varepsilon_t$, where $\varepsilon_t \sim^{\mathbb{Q}}$ iid $GSNP_m(\nu, \psi^{\mathbb{Q}}, \boldsymbol{\delta}^{\mathbb{Q}})$, with $\psi^{\mathbb{Q}} = \psi/(1+\alpha\psi)$, $\boldsymbol{\delta}^{\mathbb{Q}} = (\delta_0^{\mathbb{Q}}, \ldots, \delta_m^{\mathbb{Q}})'$ and $\delta_i^{\mathbb{Q}} = \delta_i(1+\alpha\psi)^i$.*

Hence, if we model $\mu_t(\boldsymbol{\theta})$ as a Component-MEM process under $\mathbb{P}$, the process under $\mathbb{Q}$ will be another Component-MEM. However, the residual $\varepsilon_t^{\mathbb{Q}}$ will no longer have unit mean because

$$E^{\mathbb{Q}}[\varepsilon_t] = \varkappa = \frac{\psi}{d(1+\alpha\psi)} \sum_{j=0}^{2m} \gamma_j(\boldsymbol{\delta}^{\mathbb{Q}}) \frac{\Gamma(\nu+j+1)}{\Gamma(\nu)} \tag{8}$$

will be generally different from 1. We can exploit this feature to extract from VIX futures prices relevant economic information about the risk premia implicit in the CBOE market.

In order to price futures defined on V_t it is important to keep in mind that since V_t is not a directly traded asset, there is no cost of carry relationship between the price of the futures and V_t (see Grünbichler and Longstaff, 1996, for more details). Therefore, absent any other market information, the price at time t of a futures contract maturing at $t+n$ must be priced according to its risk-neutral expectation, i.e.

$$F_{t,t+n} = E^{\mathbb{Q}}(V_{t+n}|I_t). \tag{9}$$

On this basis, we can obtain the following analytical formula for (9):

Proposition 4 *The price at time t of a future written on the volatility index V_{t+n} under the risk-neutral measure defined in Proposition 3 can be written as*

$$F_{t,t+n} = \varkappa E^{\mathbb{Q}}[\varsigma_{t+n}(\boldsymbol{\theta}) + s_{t+n}(\boldsymbol{\theta})|I_t],$$

where

$$E^{\mathbb{Q}}\left[\begin{array}{c} \varsigma_{t+n}(\boldsymbol{\theta}) \\ s_{t+n}(\boldsymbol{\theta}) \end{array} \middle| I_t\right] = (\mathbf{I}_2 - \mathbf{A}_1)^{-1}\left[\mathbf{I}_2 - \mathbf{A}_1^{n-1}\right]\mathbf{A}_0 + \mathbf{A}_1^{n-1}\left[\begin{array}{c} \varsigma_{t+1}(\boldsymbol{\theta}) \\ s_{t+1}(\boldsymbol{\theta}) \end{array}\right],$$

$\mathbf{I}_2$ *is the identity matrix of order* 2, $\mathbf{A}_0 = (\begin{array}{cc} \omega & 0 \end{array})'$ *and*

$$\mathbf{A}_1 = \left[\begin{array}{cc} \rho + \varphi(\varkappa - 1) & \varphi(\varkappa - 1) \\ \alpha(\varkappa - 1) & \alpha\varkappa + \beta \end{array}\right].$$

Thus, the futures price is an affine function of the two components of the MEM process, whose coefficients depend on the time to maturity. Proposition 4 also shows that the change of measure not only affects the mean of the residual, but also the term structure of the forecasts of V_{t+n} for $n > 1$.

4 Empirical application

4.1 Estimation

As we mentioned in the introduction, nowadays volatility is widely regarded as an asset class on its own. But although the VIX is a not a directly tradeable asset, S&P created VIX futures indices that are themselves tradeable. Their short term index measures the return from a daily rolling long position in the first and second VIX futures contracts that replicates the evolution of a one-month constant-maturity VIX futures. In turn, the mid term index takes long positions in the fourth, fifth, sixth and seventh month VIX futures contracts (see Standard & Poor's, 2012, and Appendix B for further details). As can be seen from Figure 2, both indices experienced large gains from the beginning of their history until the peak of the financial crisis in the Autumn of 2008. From then on, though, they have lost most of their value due to the reversion of the VIX to lower volatility levels. In the same figure we also display the contrarian strategies, which would yield losses of value in the first half of the sample, and substantial gains after volatility started to decrease. Given that a comparison of the original futures indices with their tracking ETNs shows that the counterparty risk implicit in the latter is negligible, in what follows we will ignore such tracking errors and directly model the two S&P500 VIX futures indices.

We will also model the VIX directly, and infer the distribution of the futures index returns conditional on the values of this volatility index. In this way, we can exploit the much larger historical information available on the VIX[4] (see Figure 1a). Specifically, let $\mathbf{y}_t$ denote the two dimensional vector which contains the VIX futures index returns at time t. Using the results from Section 3.2, we assume the following pricing structure,

$$\mathbf{y}_t = E^{\mathbb{Q}}(\mathbf{y}_t | V_t, I_{t-1}) + \boldsymbol{\epsilon}_t, \tag{10}$$

where $E^{\mathbb{Q}}(\mathbf{y}_t | V_t, I_{t-1})$ denotes the expected value of the index returns at time t given V_t (the VIX) and the information available at time $t-1$, and $\boldsymbol{\epsilon}_t$ the corresponding pricing errors, which simply reflect the fact that no model will be able to fit actual market futures prices perfectly. In addition, given that Bates (2000) and Eraker (2004)

[4] Another advantage is that we could value other indices different from the ones used in the estimation. In addition, by modeling the contract daily returns instead of the indices, we can avoid the distortions that compounding errors from different days would create.

convincingly argue that if an asset is mispriced at time t, then it is likely to be mispriced at $t+1$, we assume that $\boldsymbol{\epsilon}_t|V_t, I_{t-1} \sim N(\rho_f \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\Sigma}_f)$.

We obtain the model prices by exploiting the fact that the two futures index returns are portfolios of n_f VIX futures contracts maturing at $T_1, T_2, \cdots T_{n_f}$. Hence, we can express the price of the i^{th} element in $\mathbf{y}_t$ as

$$E^{\mathbb{Q}}(y_{it}|V_t, I_{t-1}) = \sum_{j=1}^{n_f} \zeta_{i,T_j-t} F_{t,T_j}(\boldsymbol{\theta}),$$

where $F_{t,T_j}(\boldsymbol{\theta})$ are the model-based futures prices and the loadings ζ_{i,T_j-t} deterministically depend on the time to maturity $T_j - t$ (see Standard & Poor's, 2012, and Appendix B for further details).

Although there is no sequential cut in parameters, we can nevertheless decompose the joint log-likelihood as

$$l(\mathbf{y}_t, V_t|I_{t-1}) = l(\mathbf{y}_t|V_t, I_{t-1}) + l(V_t|I_{t-1}), \tag{11}$$

where $l(\mathbf{y}_t|V_t, I_{t-1})$ denotes the Gaussian log-likelihood of the two futures index returns given the current value of the VIX and I_{t-1}, and $l(V_t|I_{t-1})$ the marginal likelihood of the VIX given I_{t-1}. We model $l(V_t|I_{t-1})$ by assuming that $V_t - \Delta$ follows a Component-MEM process with a $GSNP_m(\nu, \psi, \boldsymbol{\delta})$ conditional distribution given I_{t-1}. We introduce the constant shift Δ because the VIX cannot take values close to zero as they would imply constant equity prices over one month for all the constituents of the S&P500.[5] Thus, we can obtain large gains in fit by assigning zero probability to those events in which $V_t < \Delta$. Importantly, the assumed Gaussianity of $l(\mathbf{y}_t|V_t, I_{t-1})$ means that its optimal value depends exclusively on the second moment matrix of the differences between the actual returns on the two indices and the returns predicted by the different models. Given that financial market participants are mostly interested in the forecasting ability of a model to predict the returns of portfolios of those indices, which depends on a quadratic form in that matrix evaluated at the portfolio weights, the conditional component of the log-likelihood function has a direct economic interpretation.

We use 5,847 daily VIX index observations from December 11, 1990, until February 28, 2014. In turn, we look at 2,060 observations on the S&P 500 VIX short and mid-term futures indices starting December 20, 2005 until the same final date. We have

[5] The minimum historical end-of-day value of the VIX has been 9.31 on December 22, 1993.

carried out all the calculations using Matlab in a desktop PC. Table 1 compares the parameter estimates and asymptotic standard errors that we obtain with the Gamma distribution and a symmetrically normalised GSNP(2) density in which we fix the scale of $\boldsymbol{\delta}$ using hyperspherical coordinates, so that the number of free shape parameters is 3. The parameters of the conditional mean are similar for both distributions. This is reasonable given that the Gamma distribution, which only uses a single distributional parameter, yields consistent estimates of the conditional mean under misspecification (once again, see Engle and Gallo, 2006). However, likelihood ratio tests show that the additional shape parameters of the GSNP densities provide hugely significant gains. In addition, those gains are confirmed by the values of the Bayesian information criterion (BIC) despite the penalty for the added complexity of the GSNP densities.

Table 1 also reports one-component versions of the MEM model in (6) in which the conditional mean evolves according to the recursive equation

$$\mu_t(\boldsymbol{\theta}) = \omega + \rho\mu_{t-1}(\boldsymbol{\theta}) + \varphi(V_{t-1} - \mu_{t-1}(\boldsymbol{\theta}))$$

using both a Gamma and a GNSP distribution for the innovations. As expected, one component models provide a poor fit, particularly for the futures contracts.

As we mentioned in the introduction, though, we are not the first to suggest distributions for the standard innovations ε_t in the MEM model (6) other than the Gamma. In particular, De Luca and Gallo (2004) consider a mixture of two exponentials for high frequency intra-trade durations while Lanne (2006) proposed a two component Gamma mixture multiplicative error model for realised volatility. Another important difference between those models is that while the former fully disentagles mean dynamics and distributional features, the latter allows for different conditional means for each of the Gamma mixture components. We have estimated both these models with our data augmented with an Esscher transform for the purposes of mapping the physical to the risk neutral measure necessary for pricing the underlying futures contracts. The results that we obtain suggest that Lanne (2006) model also provides a better fit than either the one- or two-component Gamma models for the VIX series but its ability for pricing futures is clearly below par. Given that this model has three parameters more than our preferred choice, the differences in the log-likelihood function with the GSNP model in Table 1 are exacerbated when looking at the BIC values (-6610.20 vs -6688.14). In contrast, De Luca and Gallo (2004) model offers a very poor match (BIC=-12067.48), which reflects the fact that density of an exponential mixture is necessarily monotonically decreasing while the distribution of ε_t in the VIX has a clear interior mode.

Figure 3 contains a qq-type plot that assesses the ability of the different models that we have considered so far to fit the VIX series by looking at the cumulative distribution function of the probability integral transforms (PITs) of the one-period ahead forecast errors. The advantage of this plot is that the PITs should be uniform for the true model regardless of its distribution. Although the differences are not huge, except for the mixture of exponentials, the plot confirms that the Gamma distribution provides the second worst fit. Overall, our results suggest that having one or two components does not make much of a difference for the fit of the VIX, which seems to depend more on the assumed distribution. In contrast, the fit of the futures prices seems to depend mostly on the dynamics of the mean. In view of this evidence, in what follows we will focus more on a component MEM with a GSNP density for the innovations.

Figure 4 displays the temporal evolution of the conditional mean of the VIX, $\mu_t(\boldsymbol{\theta})$, and its two components, $\varsigma_t(\boldsymbol{\theta})$ and $s_t(\boldsymbol{\theta})$ that this model generates. As expected from the results in Table 1, the first component, which can be interpreted as a "moving mean", is substantially more persistent than the second component, whose path mostly reflects oscillations around the "moving mean". Those deviations, though, can be substantially positive in crisis periods, such as in the fourth quarter of 2008 or coinciding with the most severe episodes of the European sovereign debt crisis in 2010 and 2012.

Table 1 also shows that we obtain a negative and significant risk premium parameter with the two-component GSNP model. To analyse its implications, we use the results from Proposition 4 to plot in Figure 5 the coefficients of the affine prediction formulas of the VIX at different horizons under both the real and risk-neutral measures. We can observe that the loadings on the short term factor decrease very quickly, whereas the long run component has a strong effect even at very long horizons. In other words, the VIX mean-reverts more slowly towards a higher mean under $\mathbb{Q}$ than under $\mathbb{P}$. Thus, we can conclude that it incorporates investors' risk-aversion by introducing more harmful prospects for the evolution of the VIX. Our results are consistent with the parameter estimates of the continuous time model in Mencía and Sentana (2013), and therefore confirm earlier findings by Andersen and Bondarenko (2007), among others, who show that the VIX almost uniformly exceeds realised volatility because investors are on average willing to pay a sizeable premium to acquire a positive exposure to future equity-index volatility.

4.2 Asset allocation

The surge in interest on volatility futures ETNs might seem surprising on the basis of the evolution of the iPath S&P 500 VIX short term futures ETN (VXX), which, introduced on January 29, 2009, was the first VIX related equity-like ETN. The VXX, which is a 1-month constant-maturity VIX futures tracker, yielded an 8.6% profit during its first month of existence, but from then on until January 2013 it experienced losses of close to 100% due to the fall in volatility over this period. Its poor performance led some commentators to question the potential benefits of VIX futures ETNs (see e.g. Dizard, 2012). However, a short position on a 1-month constant maturity VIX futures has been available since December 2010 through the XIV ETN. Not surprisingly, by January 2013 this inverse ETN had yielded 95% accumulated profits, which confirms that volatility derivatives might give rise to significant but risky returns. The real problem, though, is how to choose the most appropriate investment strategy using only the information available at each point in time. For that reason, in this section we study asset allocation strategies for investors seeking exposure to the two VIX futures indices based on the model we have estimated in the previous section. Although a more precise model in the statistical sense does not always lead to better financial performance (see Engle and Colacito (2006) and Sentana (2005) for a theoretical example and counterexample, respectively), we believe the exercise can still shed light on the usefulness of our model in a relevant real life application.

Consider an investor whose wealth at $t-1$ is A_{t-1}, and denote by $\mathbf{w}_t$ the 2×1 vector of portfolio weights chosen with information known at $t-1$. Then, the investor's wealth at t will be $A_t = A_{t-1}(1 + \mathbf{w}_t'\mathbf{y}_t)$, where $\mathbf{w}_t'\mathbf{y}_t$ is the return of the portfolio. We set $\sum_{j=1}^{n_f} |w_{jt}| = 1$ to fix the leverage of the portfolio, which implies that the investor allocates all her initial wealth in the two assets. Importantly, we consider the sum of the absolute value of the weights instead of the sum of the signed values because a short position is in practice a long position on the inverse ETN. Geometrically, this means that in $\mathbb{R}^2$ the weights lie on a rhombus centred at the origin with vertices $(\pm 1, 0)$ and $(0, \pm 1)$. Subject to this scaling restriction, we consider an investor who chooses $\mathbf{w}_{t-1}$ to maximise the conditional Sharpe Ratio (SR):

$$SR = \frac{E(\mathbf{w}_t'\mathbf{y}_t|I_{t-1})}{\sqrt{Var(\mathbf{w}_t'\mathbf{y}_t|I_{t-1})}}. \tag{12}$$

Unfortunately, the conditional distribution of $\mathbf{y}_t$ given I_{t-1} alone that appears in (12) is not directly available in our setting. In contrast, we know the distribution of $\mathbf{y}_t$ conditional on V_t and I_{t-1}. For that reason, we compute the moments of any given function $g(\cdot)$ of $\mathbf{w}_t'\mathbf{y}_t$ via the law of iterated expectations as follows

$$E[g(\mathbf{w}_t'\mathbf{y}_t)|I_{t-1}] = \int_{\Delta}^{\infty} E[g(\mathbf{w}_t'\mathbf{y}_t)|V_t, I_{t-1}]f(V_t|I_{t-1})dV_t, \tag{13}$$

where we exploit that

$$\mathbf{w}_t'\mathbf{y}_t \sim N[\rho_f \mathbf{w}_t'\boldsymbol{\epsilon}_{t-1} + \mathbf{w}_t'E^{\mathbb{Q}}(\mathbf{y}_t|V_t, I_{t-1}), \mathbf{w}_t'\boldsymbol{\Sigma}_f\mathbf{w}_t]$$

conditional on V_t and I_{t-1} to obtain the expectation in the integrand.[6] Importantly, (13) confirms that the SR depends on the entire conditional distribution of the VIX given its past history even though it only involves the first two moments of $\mathbf{y}_t$.

Given that the parameters reported in Table 1 have been obtained using the whole sample, we avoid any look-ahead bias by considering a feasible allocation procedure which re-estimates the parameters of the Component MEM - GSNP(2) distribution at each day in the sample using prior historical data only. Thus, we rebalance our investment strategies each day using feasible parameter estimates. In order to have sufficient data at the beginning of the sample, we only consider trading days from January 2, 2008, until the end of the sample. Nevertheless, our sample includes the bulk of the financial crisis.

Figure 6a shows the accumulated value of the SR maximising strategy (GSNP-SR for short) assuming that the initial wealth on January 2, 2008, was \$100. The gains from this strategy are vastly superior to those obtained from just investing in either the direct or inverse indices. As we mentioned before, the original short and mid indices performed better until December 2008, mainly because the VIX consistently grew during 2008. However, as the VIX started to reverse to lower levels in 2009, the short and mid-term indices rapidly lost value. In contrast, our dynamic strategy automatically rebalances the portfolio to deal with mean reversion.

[6] In practice, we compute the required integrals with Matlab's adaptive Gauss-Kronrod numerical quadrature procedure.

To assess the extent to which the performance of the GSNP-SR strategy is driven by the use of a more flexible distribution, we have repeated the exercise using a Gamma distribution. The results show a slight deterioration in performance, which confirms that for the purposes of predicting futures prices, the conditional distribution of the VIX plays a non-negligible but secondary role.

In Figure 6b we consider the strategies of two different ETNs that combine long and short positions on the indices: XVIX and XVZ. The XVIX, launched by UBS, follows a long-short static strategy that allocates -0.5 to the short term VIX futures index and 1 to the mid term index. Barclays XVZ follows a more sophisticated dynamic strategy that rebalances the investment weights on the short and mid-term indices depending on whether the S&P500 volatility term structure is in contango or backwardation (see Standard & Poor's, 2011; UBS, 2012, for further details).[7] In addition, we consider the CVIX and CVZ strategies, which are two artificial indices proposed by Alexander and Korovilas (2013). The CVIX allocates 75% of capital to the XVIX and 25% of capital to the XVZ. Alexander and Korovilas (2013) choose these weights arguing that 75% (25%) is the proportion of days that the S&P500 volatility term structure is in contango (backwardation). The CVZ index follows a dynamic strategy which holds the XVIX when the S&P500 volatility term structure is in contango, and the XVZ when it is in backwardation. Figure 6b shows that these long-short strategies perform better than the pure long strategies, at least until April 2012. Moreover, the accumulated gains from the CVZ index were slightly superior to those of the GSNP-SR strategy until the summer of 2010. However, at this point the VIX, which had been growing steadily in response to the European sovereign crisis, started a downward trend that lasted until the spring of 2012, when it stabilised. Interestingly, this change of trend deteriorated the performance of the CVZ index without affecting the GSNP-SR strategy. As a result, the accumulated gains at the end of the sample are more than twice as big for the GSNP-SR strategy than for the CVZ index.

It is also illustrative to look at the temporal evolution of the positions on the short and long contracts implied by all those strategies. The scaling constraint $\sum_{j=1}^{2} |w_{jt}| = 1$ allows us to do this by using a single series for each trading strategy which reflects the ratio of the two weights. Figure 7 contains a rolling monthly window of the arctangent

[7] On a given day there is contango if the VIX (or one-month volatility) is below the VXV index. Backwardation occurs when the VXV is higher than the VIX.

of that ratio. In that graph 100% long positions on the short- and mid-term contracts correspond to angles of 0 and 90°, respectively, while the corresponding short positions are represented by angles of 180° and -90°. More generally, negative angles indicate short positions in the mid term contract combined with long positions in the short term contract while the opposite is true for values above 90°. Although we never observed them, short positions in both contracts would correspond to angles between -90° and -180°. As can be seen, our optimal strategy is somewhat similar to a short version of the XVZ fund. In particular, it tended to be long in the long term contract and short in the short-term one during the global financial crisis, but it has often taken the contrarian position afterwards.

Figure 6 is useful to compare investments beginning on the first day of the sample. However, it does not reliably rank investments initiated at other points in the sample because accumulated gains are sensitive to the starting point. For that reason, we also compare the realised daily returns, which do not suffer from this problem. Table 2 shows descriptive statistics of the different strategies over the whole sample. The first column shows that in terms of annualised ex-post SR, the GSNP-SR strategy yields the highest values, followed by the two-component Gamma and the CVZ, which is another dynamic strategy. In turn, the second column shows the low proportion of days with positive returns that would result from directly investing in the futures indices. Finally, the last columns of Table 2 show some quantiles of realised returns. The numbers indicate that the main benefit offered by the GSNP-SR strategy is that it substantially reduces the left tail. Specifically, we can see that the left-tail quantiles of the SR maximising strategy are higher than in the competing models. Not surprisingly, though, this result is achieved at the cost of giving away part of the benefits offered by some of the other strategies in the right tail.

Figures 8 and 9 show the sample SR and the proportion of positive returns over one-year rolling moving windows. Those figures confirm that the aggregate results observed in Table 2 for the whole sample are relatively stable across different subperiods. For example, Figure 8 shows that the GSNP-SR strategy is consistently among the strategies with highest SR's. The specific values, though, experience substantial swings over the sample, which partly reflect the difficulties in precisely estimating Sharpe ratios with such short sample spans. The rolling SR from the GSNP-SR strategy reached peak levels during the second halves of 2010 and 2013. In contrast, Figure 8a shows that although going short on the original indices was a good strategy during the last year of the sample, such a strategy performed very poorly in 2010 and 2011. Similarly, CVZ yields high SR's in 2010, but negative values afterwards (Figure 8b). Finally, Figure 9a once again shows that long positions on the indices yield too many negative returns, with only a high proportion of days with positive returns at the very beginning of the sample, when the VIX was still at its highest historical values. The long-short static and dynamic strategies shown in Figure 9b perform better, but they still suffer very large swings over the sample.

4.3 Additional comparisons

In this subsection, we consider three alternative modifications of our asset allocation procedure. In the first one, we maintain the GSNP distributional assumption, but change the investor's preferences for an alternative profitability measure known as the Upside Potential Ratio (UPR). For a given return threshold r, the GSNP-UPR approach involves choosing the portfolio weights that maximise the conditional UPR, defined as

$$UPR(r) = \frac{E[\max(0, \mathbf{w}_t'\mathbf{y}_t - r)|I_{t-1}]}{\sqrt{E[\min(0, \mathbf{w}_t'\mathbf{y}_t - r)^2|I_{t-1}]}}. \tag{14}$$

Intuitively, the preferences implied by (14) penalise more heavily than the SR the uncertainty coming from the left tail.

The second robustness check that we consider consists of maximising the conditional SR, but based on a reduced form model that disregards the risk neutral valuation approach developed in Section 3.2. In particular, we directly estimate a bivariate Gaussian ARMA(2,1)-GARCH(1,1) with constant conditional correlation on the short and mid VIX futures return indices.

Lastly, we consider an alternative maximisation of the SR using another model not based on the MEM structure. In particular, we model $V_t - \Delta$ using a first order Autoregressive Gamma process (ARG). This discrete time process, which was originally proposed by Gourieroux and Jasiak (2006), can be interpreted as the discrete time counterpart to the popular square root process (see Cox, Ingersoll, and Ross, 1985). Specifically, in this model the conditional distribution of the VIX is a non-central chi-square. We show in Appendix C that we can easily price futures on the VIX in this setting using another Esscher transform.

Table 3 compares the performance of the realised returns of these three alternative approaches with those of the GSNP-SR strategy. We can observe that the GSNP-UPR strategy is able to yield a higher realised SR and UPR, and a very similar proportion of days with positive returns. In contrast, the strategy based on the bivariate ARMA-GARCH model yields much smaller values for the SR and UPR, although the proportion of days with positive returns is slightly higher in this case. Finally, the ARG process, estimated with the pricing error structure in (10), yields a slightly higher SR and UPR than the ARMA-GARCH model, but they are still noticeably smaller than those obtained with the GSNP framework.

Figure 10a shows that investing \$100 on January 2, 2008, would have yielded similar gains at the end of the sample under both the GSNP-SR and GSNP-UPR strategies. However, the ARMA-GARCH bivariate model and the ARG process would have yielded much smaller gains. In the ARMA-GARCH case, it is mainly due to its bad performance in 2008. In the ARG case, the restrictive AR(1) time series structure does not seem to adapt well to the decreasing futures prices over the last year of the sample. Figures 10b and 10c show the evolution of the realised SR and UPR, respectively, computed over one-year moving windows. We can observe that the GSNP-SR and GSNP-UPR strategies are very similar in terms of the SR, while the GSNP-UPR strategy is slightly superior in terms of the UPR. Once again, the strategy based on the bivariate ARMA-GARCH model clearly underperforms in 2008, while the ARG framework performs poorly in 2013. The ARMA-GARCH model works better over the following years, but it systematically yields lower performance statistics than the strategies based on the GSNP distribution.

5 Conclusions

We develop a theoretical framework for covariance stationary but persistent positively-valued processes combining a semi-nonparametric expansions of the Gamma distribution with a component version of the Multiplicative Error Model. In addition, we define an exponentially affine stochastic discount factor that allows us to price futures on the VIX index in closed form. Our estimated parameters indicate a short run component that mean-reverts to zero and a long run component, which mean-reverts more slowly towards a long run mean. We also find that the GSNP expansion yields significant improvements in fit relative to the original Gamma distribution, and it also performs better than other distributions suggested in the literature. Overall, our results suggest that having one or two components does not make much of a difference for the fit of the VIX, which seems to depend more on the assumed distribution. In contrast, the fit of the futures prices seems to depend more on the dynamics of the mean.

We then apply our framework to a dynamic portfolio allocation for VIX Exchange Traded Notes that maximises the one-day ahead conditional Sharpe Ratio, which depends on the GSNP expansion through a convolution formula. Our results show that the GSNP strategy yields realised returns with the highest ex-post SRs over the whole sample. In effect, our strategy manages to increase the left tail quantiles of the return distribution at the cost of having a somewhat thinner right tail than other strategies. We also observe that we generally obtain a superior performance with our GSNP strategy when we assess performance over rolling one-year sample sub-periods.

Finally, we investigate the extent to which our results are related to our choice of performance measure and modelling approach. To do so, we consider the Upside Potential Ratio (UPR) as an alternative performance measure, maintaining the GSNP distributional assumption. In addition, we check the impact of the GSNP distribution by keeping the SR preferences but considering either a bivariate ARMA-GARCH model that we directly estimate on the VIX futures index returns, or an Autoregressive Gamma process. We find that the alternative preferences yield minor improvements in performance, but the elimination of our flexible distributional assumption clearly leads to underperformance relative to GSNP-based strategies.

Monte Carlo simulations looking at the reliability of the ML parameter estimators and standard errors in finite samples would consitute a valuable addition. Another fruitful avenue for future research would be to consider multivariate expansions, which could be used to invest simultaneously in ETNs on different volatility indices. It would also be interesting to explore time varying specifications of the shape parameters, as well as long memory alternatives to the MEM.

References

Abramowitz, M. and A. Stegun (1965). *Handbook of mathematical functions.* New York: Dover Publications.

Alexander, C. and D. Korovilas (2013). Volatility exchange-traded notes: curse or cure? *The Journal of Alternative Investments 16*, 52–70.

Amengual, D., G. Fiorentini, and E. Sentana (2013). Sequential estimators of shape parameters in multivariate dynamic models. *Journal of Econometrics 177*, 233–249.

Amengual, D. and D. Xiu (2013). Resolution of policy uncertainty and sudden declines in volatility. Discussion Paper, University of Chicago Booth School of Business.

Andersen, T. G. and O. Bondarenko (2007). Construction and interpretation of model-free implied volatility. NBER Working Paper No. W13449.

Bardgett, C., E. Gourier, and M. Leippold (2014). Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. Discussion Paper, University of Zurich and Swiss Finance Institute.

Bartholomew, D. J. (1969). Sufficient conditions for a mixture of exponentials to be a probability density function. *The Annals of Mathematical Statistics 40*(6), 2183–2188.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Bertholon, H., A. Monfort, and F. Pegoraro (2003). Pricing and inference with mixtures of conditionally normal processes. Mimeo CREST.

Bontemps, C. and N. Meddahi (2012). Testing distributional assumptions: a GMM approach. *Journal of Applied Econometrics 27*, 978–1012.

Bowers, N. L. (1966). Expansion of probability density functions as a sum of Gamma densities with applications in risk theory. *Transactions of Society of Actuaries 18*, 125–147.

Brownlees, C., F. Cipollini, and G. M. Gallo (2012). Multiplicative error models. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook in Financial Engineering and Econometrics: volatility models and their applications*, pp. 225–248. Wiley.

Buhlman, H., F. Delbaen, P. Embrechts, and A. Shyraev (1996). No arbitrage, change of measure and conditional Esscher transforms in a semi-martingale model of stock processes. *CWI Quarterly 9*, 291–317.

Buhlman, H., F. Delbaen, P. Embrechts, and A. Shyraev (1998). On Esscher transforms in discrete finance models. *ASTIN Bulletin 28*, 171–186.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

De Luca, G. and G. Gallo (2004). Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics and Econometrics 8*, 1–20.

De Luca, G. and G. Gallo (2009). Time-varying mixing weights in mixture autoregressive conditional duration models. *Econometric Reviews 28*, 102–120.

Dizard, J. (2012). Vix products are not for rational investors. Financial Times. April 15, 2012.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*, 425–446.

Engle, R. F. and R. Colacito (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics 24*, 238–253.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*, 3–27.

Engle, R. F. and G. Lee (1999). A long-run and short-run component model of stock return volatility. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, 475–497.

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance 59*, 1367–1404.

Esscher, F. (1932). On the probability function in the collective theory of risk. *Skandinavisk Aktuariedskrift 15*, 165–195.

Fenton, V. and A. R. Gallant (1996). Qualitative and asymptotic performance of SNP density estimators. *Journal of Econometrics 74*, 77–118.

Gallant, A. and G. Tauchen (1999). The relative efficiency of method of moment estimators. *Journal of Econometrics 92*, 149–172.

Gallant, A. R. and D. W. Nychka (1987). Seminonparametric maximum likelihood estimation. *Econometrica 55*, 363–390.

Gerber, H. and E. Shiu (1994). Option pricing by Esscher transforms. *Transactions of Society of Actuaries 46*, 99–140.

Gourieroux, C. and J. Jasiak (2006). Autoregressive Gamma processes. *Journal of Forecasting 25*, 129–152.

Gourieroux, C. and A. Monfort (2006). Pricing with splines. *Annales de Economie et de Statistique 82*, 3–33.

Gourieroux, C. and A. Monfort (2007). Econometric specification of stochastic discount factor models. *Journal of Econometrics 136*, 509–530.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica 52*(3), 681–700.

Grünbichler, A. and F. A. Longstaff (1996). Valuing futures and options on volatility. *Journal of Banking & Finance 20*, 985–1001.

Johnson, N., S. Kotz, and N. Balakrishnan (1994). *Continuous univariate distributions*, Volume 1. New York: John Wiley and Sons.

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics 4*, 594–616.

León, A., J. Mencía, and E. Sentana (2009). Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business & Economic Statistics 27*, 176–192.

Meddahi, N. (2001). An eigenfunction approach for volatility modeling. Université de Montréal Working Paper 29-2001.

Mencía, J. and E. Sentana (2013). Valuation of VIX derivatives. *Journal of Financial Economics 108*, 367–391.

Rhoads, R. (2011). *Trading VIX derivatives: trading and hedging strategies using VIX futures, options and exchange traded notes*. New York: John Wiley and Sons.

Schwert, G. W. (2011). Stock volatility during the recent financial crisis. *European Financial Management 17*, 789–805.

Sentana, E. (2005). Least squares predictions and mean-variance analysis. *Journal of Financial Econometrics 3*(1), 56–78.

Song, Z. and D. Xiu (2016). A tail of two option markets: state-price densities and volatility risk. *Journal of Econometrics 190*, 176–196.

Standard & Poor's (2011). S&P 500 dynamic VIX futures methodology.

Standard & Poor's (2012). S&P Dow Jones indices: S&P 500 VIX futures indices methodology.

Steutel, F. W. (1967). Note on the infinite divisibility of exponential mixtures. *The Annals of Mathematical Statistics 38*(4), 1303–1305.

Stuart, A. and K. Ord (1977). *Kendall's Advanced Theory of Statistics* (6th ed.), Volume 1. London: Griffin.

UBS (2012). UBS AG exchange traded access securities (E-TRACS) daily long-short

VIX ETN due November 30, 2040. Amendment No. 2 dated January 11, 2012 to prospectus supplement dated November 30, 2010.

A Further distributional results

A.1 Properties of the Gamma distribution

Assume that x is a Gamma random variable whose pdf is given by (1). We summarise here the main properties of this distribution, as described in Johnson, Kotz, and Balakrishnan (1994). Its moment generating function is

$$E\left[\exp(nx)\right] = (1-\psi n)^{-\nu},$$

for $n < \psi^{-1}$, while its characteristic function is $\psi_G(i\tau) = (1-i\psi n)^{-\nu}$. Therefore, we can express the moments of x as

$$E(x^n) = \psi^n \frac{\Gamma(\nu+n)}{\Gamma(\nu)}. \tag{A1}$$

A.2 Gamma-based Gram Charlier expansions

The Gamma distribution can be used in place of the normal distribution as the parent distribution in a Gram Charlier expansion. In particular, if we consider a non-negative random variable y, under certain assumptions its density function $h(y)$ can be expressed as the product of a Gamma density times an infinite series of polynomials,

$$h(y) = f_G(y,\nu,\bar{\psi}) \sum_{j=0}^{\infty} c_j P_j(y,\nu,\bar{\psi}), \tag{A2}$$

where $P_j(y,\nu,\bar{\psi})$ denotes the polynomial of order j that forms an orthonormal basis with respect to the Gamma distribution, so that $E[P_j(y,\nu,\bar{\psi})] = 0$, $V[P_j(y,\nu,\bar{\psi})] = 1$ and $E[P_j(y,\nu,\bar{\psi})P_k(y,\nu,\bar{\psi})] = 0$, for all $j,k \geq 0$ and $j \neq k$ (see Johnson, Kotz, and Balakrishnan, 1994).

Following Bontemps and Meddahi (2012), we can express those polynomials as

$$\begin{aligned}
P_0(y,\nu,\bar{\psi}) &= 1, \\
P_1(y,\nu,\bar{\psi}) &= \frac{\bar{\psi}^{-1}y - \nu}{\sqrt{\nu}}, \\
P_2(y,\nu,\bar{\psi}) &= \frac{[(\bar{\psi}^{-1}y)^2 - 2(\nu+1)\bar{\psi}^{-1}y + \nu(\nu+1)]}{\sqrt{2\nu(\nu+1)}},
\end{aligned}$$

and in general

$$P_n(y,\nu,\bar{\psi}) = \frac{(\bar{\psi}^{-1}y - \nu - 2n - 2)P_{n-1}(y,\nu,\bar{\psi}) - \sqrt{(n-1)(\nu+n-2)}P_{n-2}(y,\nu,\bar{\psi})}{\sqrt{n(\nu+n-1)}}.$$

Given that

$$P_n(y, \nu, \bar{\psi}) = (-1)^n L_n(\bar{\psi}^{-1} y, \nu - 1) \sqrt{\frac{\Gamma(\nu) n!}{\Gamma(\nu + n)}}, \tag{A3}$$

where $L_n(\cdot, \cdot)$ is the generalised Laguerre polynomial of order n, we will refer to (A2) as the Laguerre expansion of the density of y. The orthonormal properties of these polynomials imply that we can obtain the coefficients of the expansion as

$$c_n = \int_0^\infty P_n(y, \nu, \bar{\psi}) h(y) dy. \tag{A4}$$

A.3 Truncated Laguerre expansions

A truncated Laguerre expansion is another finite version of (A2) which treats the $c_j' s$ as free parameters. Specifically,

$$h(x) = f_G(x, \nu, \nu^{-1}) \left[1 + \sum_{j=2}^{k} c_j P_j(x, \nu, \nu^{-1}) \right], \tag{A5}$$

where we have imposed that $c_1 = 0$ and $\bar{\psi} = 1/\nu$ so that this distribution has unit mean too. Unfortunately, this approach does not ensure the non-negativity of the resulting density function, a property that is satisfied by construction by the GSNP distribution. In this sense, Amengual, Fiorentini, and Sentana (2013) have studied the parametric restrictions that the c_j coefficients must satisfy to ensure positivity in second and third-order Laguerre expansions.

Since both the GSNP distribution and the truncated Laguerre expansion have unit mean, one may ask which of them can generate a wider range of higher order moments. We address this question by comparing the coefficients of variation, skewness and kurtosis of the two distributions, which we will denote as τ, ϕ and λ, respectively. In particular, we compare (A5) for $k = 3$ with a GSNP distribution of order $m = 2$ since both have the same number of free parameters. Figures A1a to A1c show the regions generated by both distributions on the $\tau - \phi$, $\tau - \lambda$ and $\phi - \lambda$ spaces. We have computed these regions using numerical methods. Specifically, for the GSNP, we simulate values for $\boldsymbol{\delta}$ in the unit sphere for a dense grid of values for ν, and compute the envelope of the coefficients on the $\tau - \phi$, $\tau - \kappa$ and $\phi - \kappa$ spaces. For the Laguerre expansion we obtain the envelopes by combining a dense grid for ν with another dense grid for the frontier, as parametrised by Amengual, Fiorentini, and Sentana (2013). We also use the results in Appendix A.1 to derive the values generated by the Gamma distribution, which are

$\tau_G = \sqrt{1/\nu}$, $\phi_G = \sqrt{4/\nu}$ and $\kappa_G = 3 + 6\nu^{-1}$. Finally, we also include the lower bounds that no properly-defined density can exceed derived in Appendix A.4. As can be observed, both distributions provide similar flexibility for coefficients of variation smaller than 0.5. For larger coefficients of variation, the GSNP turns out to be superior in terms of feasible values of skewness and kurtosis. Interestingly, the flexibility of the Laguerre distribution relative to the Gamma distribution decreases drastically for coefficients of variation larger than around 1.8. In contrast, we do not observe this phenomenon in the GSNP distribution. In terms of skewness and kurtosis, the Laguerre expansion remains less flexible than the GSNP, but the differences are smaller.

Finally, another way of adding flexibility would be to shift the expanded distribution by a constant amount Δ, as in our empirical application. This shift would affect τ, but not ϕ or λ.

A.4 Feasible moments of distributions

Stuart and Ord (1977) explain that regardless of the shape of the distribution, the skewness-kurtosis relationship

$$\kappa \geq 1 + \phi^2 \tag{A6}$$

must hold. In a similar spirit, we can apply the Cauchy-Schwarz inequality to show that for a positive random variable x:

$$[E(x^{3/2}x^{1/2})]^2 \leq E(x^3)E(x),$$

so that $\mu_2'^2 \leq \mu_1'\mu_3'$. If we introduce in this expression the relationships between the central and non-central moments, $\mu_2' = \mu_2 + \mu_1'^2$ and $\mu_3' = \mu_3 + 3\mu_1'\mu_2 + \mu_1'^3$, we can show that

$$\phi \geq \tau - \tau^{-1}. \tag{A7}$$

Finally, if we combine (A7) with (A6), we can show that $\kappa \geq 1 + [\max\{\tau - \tau^{-1}, 0\}]^2$.

B S&P 500 VIX indices

The volatility indices developed by S&P are based on a time series that they define as index excess return. Let ξ_t denote this index, which can be expressed as

$$\xi_t = \xi_{t-1}(1 + CDR_t),$$

where CDR_t is the contract daily return given by

$$CDR_t = \frac{TDWO_t}{TDWI_{t-1}} - 1,$$

where $TDWO_t$ denotes the total dollar weight obtained on t. For simplicity, we focus on the short term index, which is determined as follows:

$$TDWO_t = CRW_{m,t-1}F_{m,t} + CRW_{n,t-1}F_{n,t},$$

where $F_{m,t}$ and $F_{n,t}$ are the Futures prices of the m and n contracts at t. For the sake of concreteness, assume that m is the shortest maturity of the two. The contract roll weights of the VIX contracts at t are defined as

$$\begin{aligned} CRW_{m,t} &= 100\frac{dr}{dt}, \\ CRW_{n,t} &= 100\frac{dt - dr}{dt}, \end{aligned}$$

where dr is the number of business days remaining until the maturity of contract m, including t but not the day on which m matures, dt is the number of business days between the contract immediately prior to m (i.e. $m-1$) and m. This second period includes the day in which $m-1$ matures, but not the day in which m matures.

Similarly, the total dollar weight invested on $t-1$ is defined as

$$TDWI_{t-1} = CRW_{m,t-1}F_{m,t-1} + CRW_{n,t-1}F_{n,t-1}.$$

Entirely analogous derivations apply to the mid-term index, the only difference being that they involve four different futures contracts (see Standard & Poor's, 2012, for more details).

C Futures pricing based on the ARG process

Let V_t follow an Autoregressive Gamma process of order 1 under the real measure, or $ARG(1)$ for short. Then, it can be shown that the distribution of $2V_t/c$ conditional on I_{t-1} is a non-central chi-square with noncentrality parameter $2\beta V_{t-1}$ and degrees of freedom 2δ. If we consider the exponentially affine stochastic discount factor

$$M_{t-1,t} = \exp(-\alpha V_t),$$

then it can be easily shown that $2(1+2\alpha)V_t/c$ will be, under the risk-neutral measure, a non-central chi-square with degrees of freedom 2δ and non-centrality parameter

$2\beta V_{t-1}/(1+2\alpha)$. In practice, this process can be reinterpreted as an ARG(1) process with parameters $\delta_{\mathbb{Q}} = \delta$,

$$c_{\mathbb{Q}} = \frac{c}{1+2\alpha}, \quad \beta_{\mathbb{Q}} = \frac{\beta}{1+2\alpha}.$$

Hence, the futures price can be written as

$$F_{t,t+n} = E^{\mathbb{Q}}[V_{t+n}|I_t] = c_{\mathbb{Q},n}\delta + c_{\mathbb{Q},n}\beta_{\mathbb{Q},n}V_t,$$

where

$$c_{\mathbb{Q},n} = \frac{1-c_{\mathbb{Q}}^n\beta_{\mathbb{Q}}^n}{1-c_{\mathbb{Q}}\beta_{\mathbb{Q}}}c_{\mathbb{Q}}, \quad \beta_{\mathbb{Q},n} = \frac{c_{\mathbb{Q}}^{n-1}\beta_{\mathbb{Q}}^n(1-c_{\mathbb{Q}}\beta_{\mathbb{Q}})}{1-c_{\mathbb{Q}}^n\beta_{\mathbb{Q}}^n}.$$

D Proofs of propositions

D.1 Proposition 1

We can show through tedious but straightforward algebra that

$$\left[\sum_{j=0}^{m}\delta_j\left(\frac{x}{\psi}\right)^j\right]^2 = \sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\left(\frac{x}{\psi}\right)^j.$$

Then, we can use (1) to show that

$$\begin{aligned}\left(\frac{x}{\psi}\right)^j f_G(x,\nu,\psi) &= \frac{1}{\Gamma(\nu)\psi^{\nu+j}}x^{\nu+j-1}\exp(-x/\psi) \\ &= \frac{\Gamma(\nu+j)}{\Gamma(\nu)}f_G(x,\nu+j,\psi).\end{aligned}$$

D.2 Proposition 2

Introducing (4) in (A4), we can express the coefficients of the Laguerre expansion as

$$c_n = \frac{1}{d}\sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\frac{\Gamma(\nu+j)}{\Gamma(\nu)}\int_0^\infty P_n(y,\nu,\bar{\psi})f_G(y,\nu+j,\psi)dy. \tag{D8}$$

If we write $P_n(y,\nu,\bar{\psi})$ in terms of the n-order Laguerre polynomial, as in (A3), we obtain

$$c_n = (-1)^n\sqrt{\frac{\Gamma(\nu)n!}{\Gamma(\nu+n)}}\frac{1}{d}\sum_{j=0}^{2m}\gamma_j(\boldsymbol{\delta})\frac{\Gamma(\nu+j)}{\Gamma(\nu)}\int_0^\infty L_n(\bar{\psi}^{-1}y,\nu-1)f_G(y,\nu+j,\psi)dy.$$

Then, if we use the following property

$$L_n(\bar{\psi}^{-1}y,\nu-1) = \sum_{i=0}^{n}\frac{(-1)^i}{i!}\binom{n+\nu-1}{n-i}(\bar{\psi}^{-1}y)^i.$$

from Abramowitz and Stegun (1965) (page 775), then we obtain

$$\int_0^\infty L_n(\bar{\psi}^{-1}y,\nu-1)f_G(y,\nu+j,\psi)dy=\sum_{i=0}^{n}\frac{(-1)^i}{i!}\binom{n+\nu-1}{n-i}\bar{\psi}^{-i}\int_0^\infty y^i f_G(y,\nu+j,\psi)dy, \tag{D9}$$

where

$$\int_0^\infty y^i f_G(y,\nu+j,\psi)dy=\psi^i\frac{\Gamma(\nu+i+j)}{\Gamma(\nu+j)} \tag{D10}$$

from (A1). Introducing (D9) and (D10) in (D8), we obtain the final result.

D.3 Proposition 3

The risk-neutral density of ε_t will be proportional to

$$\begin{aligned} & f_{GSNP}(\varepsilon_t,\nu,\psi,\boldsymbol{\delta})\exp(-\alpha\varepsilon_t), \\ = \; & f_G(\varepsilon_t,\nu,\psi)\exp(-\alpha\varepsilon_t)\left[\sum_{j=0}^{m}\delta_j\left(\frac{\varepsilon_t}{\psi}\right)^j\right]^2 \end{aligned}$$

It can be easily shown that $f_G(\varepsilon_t,\nu,\psi)\exp(-\alpha\varepsilon)\propto\varepsilon^{\nu-1}\exp\left(-\varepsilon/\psi^{\mathbb{Q}}\right)$, where $\psi^{\mathbb{Q}}=\psi/(1+\alpha\psi)$. Similarly, we can write

$$\sum_{j=0}^{m}\delta_j\left(\frac{\varepsilon_t}{\psi}\right)^j=\sum_{j=0}^{m}\delta_j(1+\alpha\psi)^j\left(\frac{\varepsilon_t}{\psi^{\mathbb{Q}}}\right)^j.$$

Hence, we can always define $\delta_j^{\mathbb{Q}}=\delta_j(1+\alpha\psi)^j$. This proves that the resulting density is a $GSNP_m(\nu,\psi^{\mathbb{Q}},\boldsymbol{\delta}^{\mathbb{Q}})$.

D.4 Proposition 4

If we use (8), we can show that

$$F_{t,t+n}=E^{\mathbb{Q}}[V_{t+n}|I(t)]=\varkappa E^{\mathbb{Q}}[\varsigma_{t+n}(\boldsymbol{\theta})+s_{t+n}(\boldsymbol{\theta})|I_t]$$

and

$$E^{\mathbb{Q}}[\varsigma_{t+n}(\boldsymbol{\theta})|I_{t+n-2}]=\omega+[\rho+\varphi(\varkappa-1)]\varsigma_{t+n-1}(\boldsymbol{\theta})+\varphi(\varkappa-1)s_{t+n-1}(\boldsymbol{\theta}).$$

Similarly, we can obtain

$$E^{\mathbb{Q}}[s_{t+n}(\boldsymbol{\theta})|I_{t+n-2}]=\alpha(\varkappa-1)\varsigma_{t+n-1}(\boldsymbol{\theta})+[\alpha\varkappa+\beta]s_{t+n-1}(\boldsymbol{\theta}).$$

Hence, we have

$$E^{\mathbb{Q}}\left[\begin{array}{c}\varsigma_{t+n}(\boldsymbol{\theta})\\ s_{t+n}(\boldsymbol{\theta})\end{array}\middle| I_{t+n-2}\right]=\mathbf{A}_0+\mathbf{A}_1\left[\begin{array}{c}\varsigma_{t+n-1}(\boldsymbol{\theta})\\ s_{t+n-1}(\boldsymbol{\theta})\end{array}\right].$$

By applying the law of iterated expectations recursively to condition on $I_{t+n-3},I_{t+n-4},\cdots,I_t$, we can obtain the final result after some straightforward algebraic manipulations.

Table 1
Maximum likelihood estimates of Component-MEM models

	1c-Gamma	s.e.	1c-GSNP(2)	s.e.	2c-Gamma	s.e.	2c-GSNP(2)	s.e.
α					0.662	0.010	0.666	0.010
β					0.286	0.011	0.282	0.011
ω	0.125	0.002	0.124	0.002	0.025	0.002	0.025	0.002
ρ	0.992	0.001	0.992	0.001	0.998	0.000	0.998	0.000
φ	0.706	0.004	0.711	0.004	0.221	0.009	0.224	0.009
Δ	5.064	0.136	5.089	0.173	5.179	0.134	5.386	0.160
ν	137.314	4.275	116.934	4.995	139.903	4.113	115.187	4.556
θ_1			0.018	0.001			0.019	0.001
θ_2			3.137	0.000			3.137	0.000
Risk premium	0.338	0.147	0.292	0.106	-0.387	0.159	-0.248	0.113
σ_{short}	0.022	0.000	0.022	0.000	0.019	0.000	0.019	0.000
σ_{mid}	0.014	0.000	0.014	0.000	0.012	0.000	0.012	0.000
$\rho_{short,mid}$	0.748	0.008	0.748	0.008	0.689	0.009	0.689	0.009
ρ_f	0.997	0.001	0.997	0.001	0.99	0.002	0.99	0.002
Likelihood	2615.329		2864.242		3183.954		3465.495	
LR test		497.827				563.083		
BIC	-5057.185		-5520.316		-6159.732		-6688.137	

Notes: The estimation uses VIX data from December 11, 1990, until February 28, 2014, as well as data on the S&P 500 VIX short and mid-term futures indices from December 20, 2005 until the same final date. "Gamma" denotes those MEM models whose conditional distribution given the information known at $t-1$ is Gamma, while in "GSNP(2)" the conditional distribution is a SNP expansion of order 2 of the Gamma distribution. 1c and 2c denote one and two component-MEM models, respectively. Standard errors have been computed from the outer product of the analytical score. LR tests show the likelihood ratio tests of the Gamma distribution vs. the GSNP model for the 1c and 2c versions. BIC denotes the Bayesian Information Criterion.

Table 2

Descriptive statistics of the realised returns of different asset allocation strategies.

	SR	Ret>0(%)	Mean	Std. Dev.	Skewness	Kurtosis	5% Perc.	25% Perc.	Median	75% Perc.	95% Perc.
Short	-0.594	42.6	-0.146	3.974	0.844	6.663	-5.943	-2.277	-0.581	1.541	6.935
Mid	-0.281	45.1	-0.035	1.995	0.632	6.619	-3.042	-1.052	-0.186	0.875	3.307
-1×Short	0.594	57.8	0.146	3.974	-0.844	6.663	-6.935	-1.541	0.581	2.277	5.943
-1×Mid	0.281	55.1	0.035	1.995	-0.632	6.619	-3.307	-0.875	0.186	1.052	3.042
XVIX	0.586	53.1	0.032	0.880	-0.139	5.888	-1.346	-0.443	0.058	0.518	1.378
XVZ	0.398	49.1	0.034	1.384	0.885	11.260	-1.796	-0.558	-0.012	0.507	2.127
CVIX	0.611	52.1	0.032	0.860	0.210	6.461	-1.312	-0.433	0.047	0.480	1.357
CVZ	0.911	53.3	0.074	1.324	0.764	11.641	-1.655	-0.505	0.073	0.568	1.964
GAM-SR	1.755	58.0	0.142	1.311	-0.502	12.086	-1.685	-0.357	0.130	0.679	2.118
GSNP-SR	1.868	57.7	0.147	1.274	-0.249	14.944	-1.613	-0.339	0.128	0.631	2.036

Notes: The sample used is 1-Jan-2008 to 27-Feb-2014. SR denotes the Sharpe Ratio, expressed in annualised terms. The column labelled "Ret>0 (%) indicates the proportion of days with positive returns. "Short" and "Mid" denote the S&P 500 VIX short and mid futures indices. "-1×" denote short sales on those indices. XVIX is a UBS ETN following a long-short static strategy on the VIX futures indices, while XVZ is a Barclays ETN following a dynamic strategy. CVIX and CVZ are investment strategies proposed by Alexander and Korovilas (2013). GSNP-SR and GAM-SR denote the returns obtained by maximising the conditional Sharpe Ratio, based on the parameters obtained from a two-component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution, and a Gamma distribution, respectively. The parameters are estimated each day using the information available at that point.

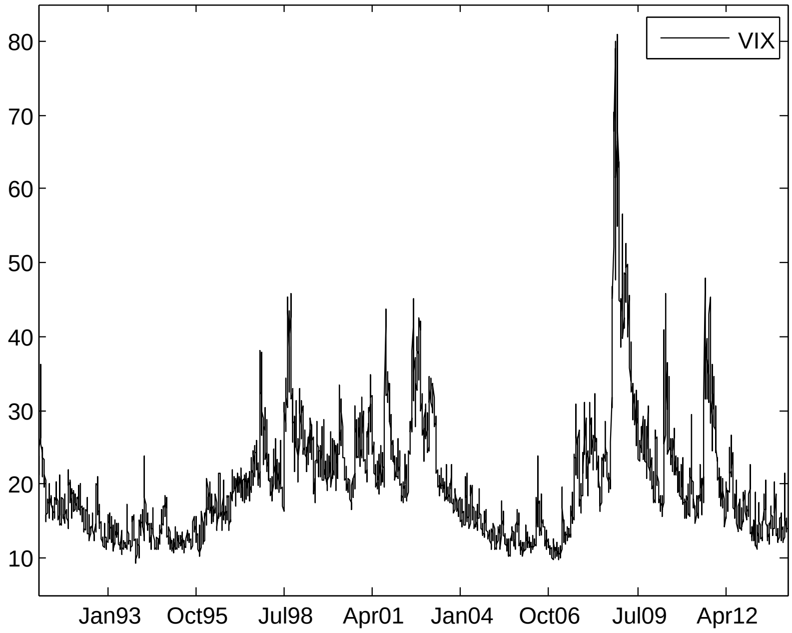
Table 3
Profitability measures of the realised returns of alternative dynamic asset allocation strategies.

	SR	Ret>0(%)	UPR
GSNP-SR	1.868	57.7	9.027
GSNP-UPR	1.917	57.2	9.582
ARMA-GARCH	0.781	58.5	7.288
ARG-SR	1.063	56.0	8.235

Notes: The sample used is 1-Jan-2008 to 27-Feb-2014. SR denotes the Sharpe Ratio, while UPR denotes the Upside Potential Ratio with zero as the return threshold. Both the SR and UPR are expressed in annualised terms. The column labelled "Ret> 0 (%) indicates the proportion of days with positive returns. GSNP-SR (GSNP-UPR) denotes the returns obtained by maximising the conditional SR (UPR), based on the parameters obtained from a Component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution. ARMA-GARCH denotes the returns obtained by maximising the conditional SR, based on the parameters obtained from a bivariate ARMA(2,1)-GARCH(1,1) with constant conditional correlation, estimated on the short and mid VIX future index returns. ARG-SR denotes the returns obtained by maximising the conditional SR, based on the parameters obtained from a first order Autoregressive Gamma process. The parameters are estimated each day using the information available at that point.

Figure 1: Historical evolution of the VIX index

(a) Dec 1990- Jan 2013



(b) Comparison with VXV (Dec 2007- Jan 2013)

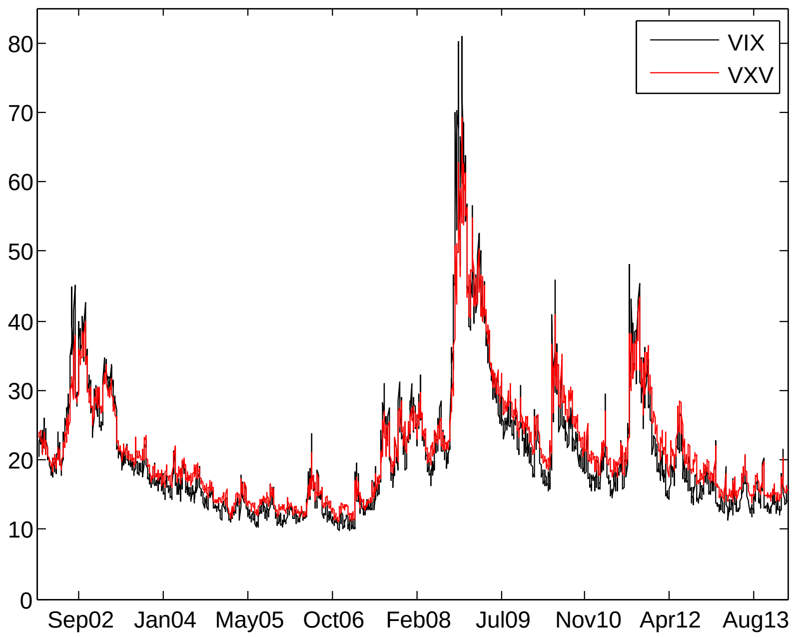
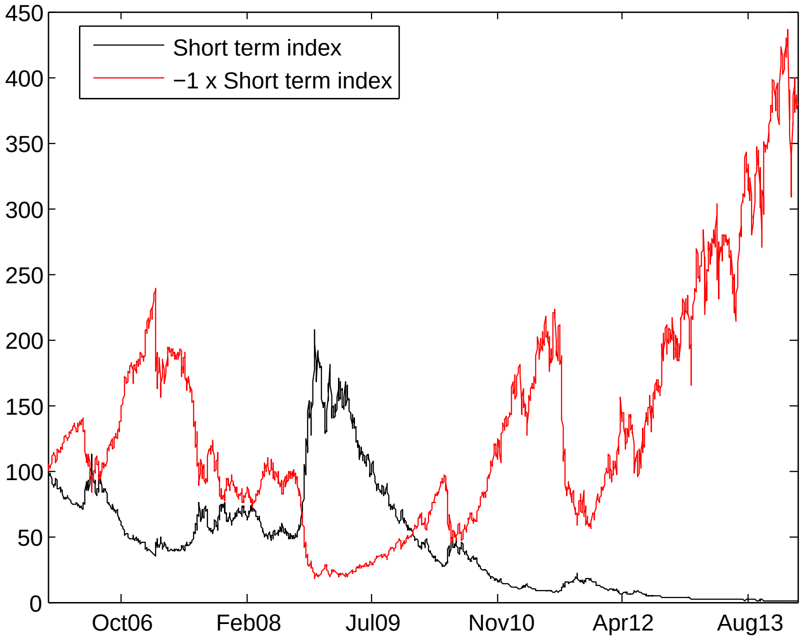
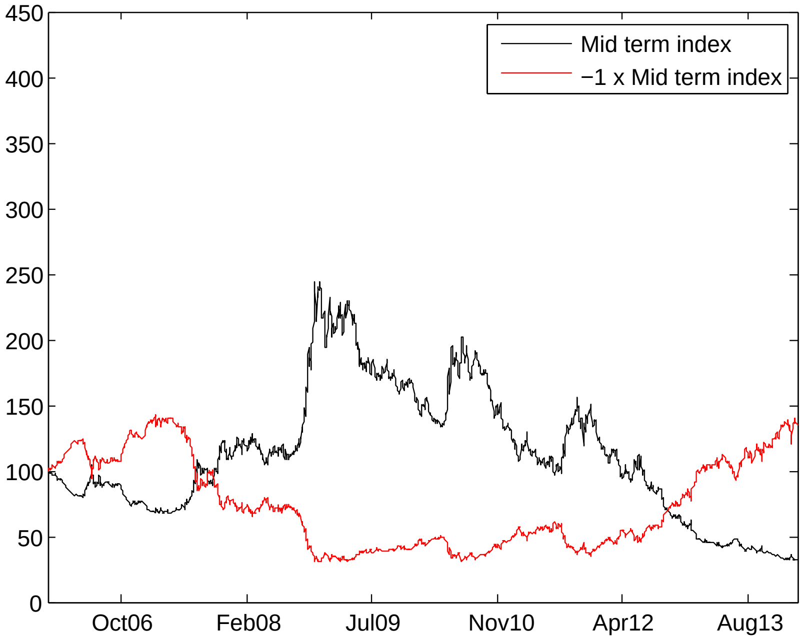


Figure 2: Historical evolution of S&P 500 VIX futures indices

(a) Short term index

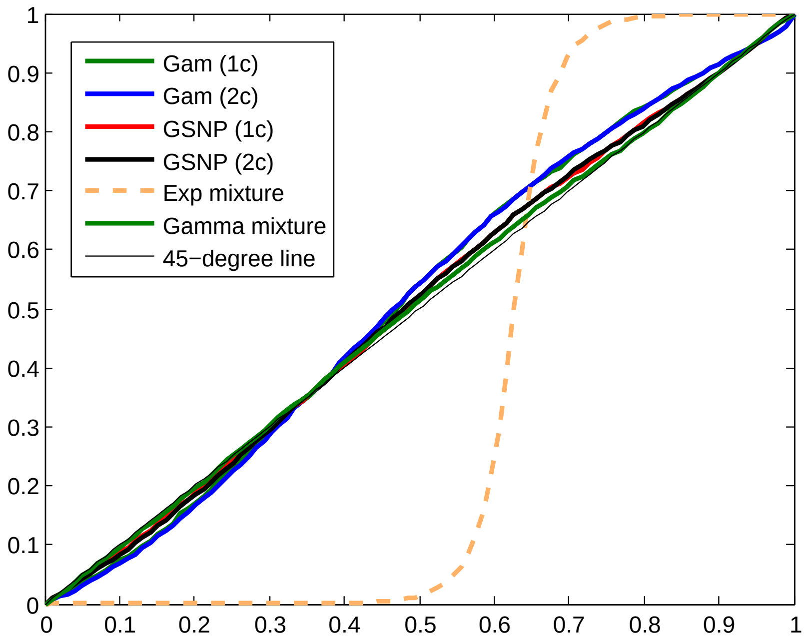


(b) Mid-term index



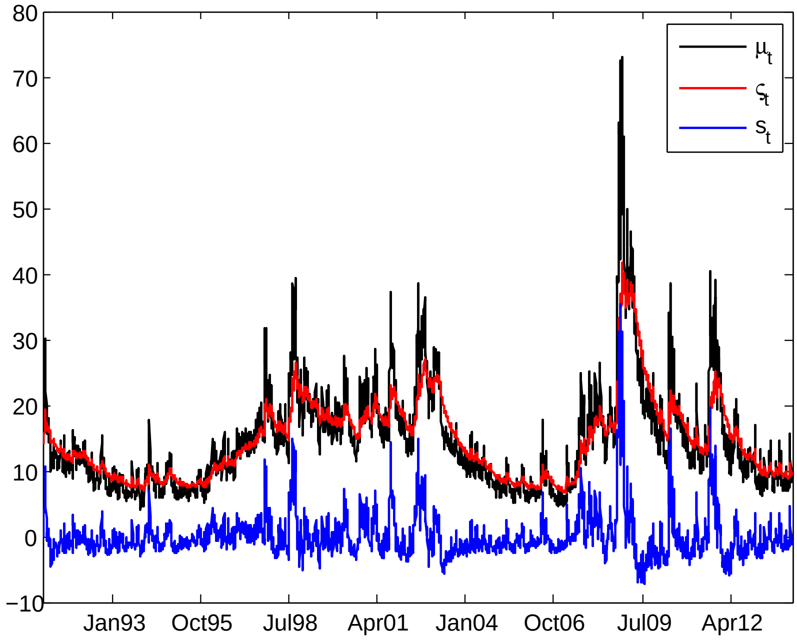
Note: The black lines show the evolution of the original S&P 500 VIX futures indices, while the red lines show the evolution of indices with exactly the opposite returns from the original ones.

Figure 3: Probability integral transforms of different VIX conditional distributions



Note: Gam, GSNP and "Exp Mixture" denote a MEM for the VIX with a Gamma distribution, a GSNP(2) expansion of the Gamma distribution, and a mixture of two exponential distributions, respectively. 1c and 2c denote one and two component-MEM models, respectively. "Gamma Mixture" denotes the mixture of two Gamma distributions in which each component of the mixture follows a different two-component MEM.

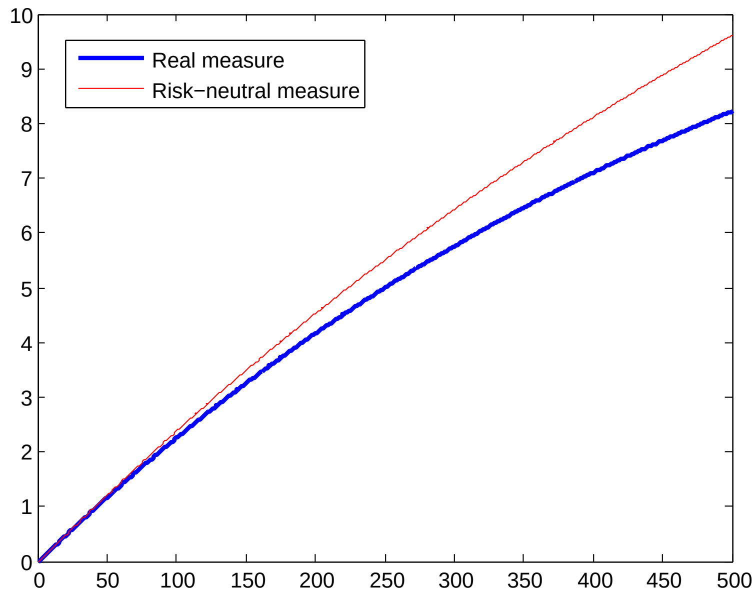
Figure 4: Conditional mean decomposition under the two-component GSNP model



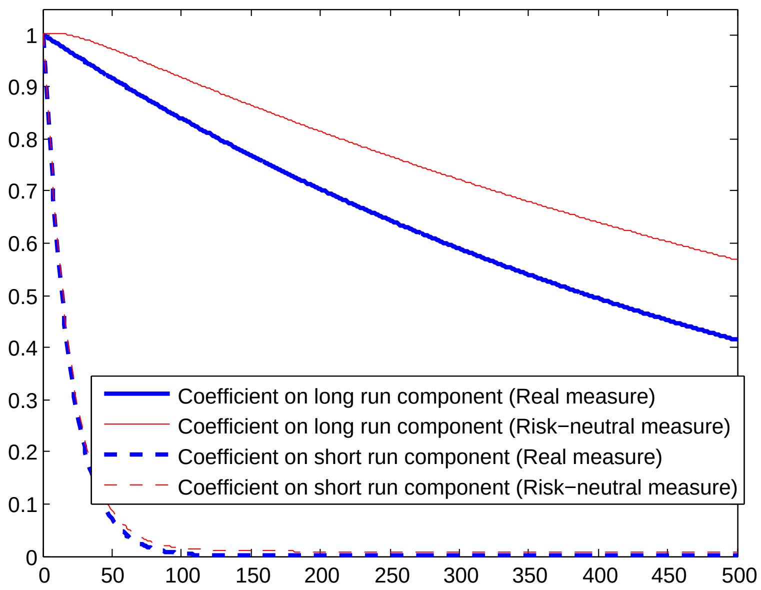
Note: The estimated model is a two-component GSNP MEM, which has a short-run component s_t and a long run component ς_t.

Figure 5: Coefficients of the affine prediction formulas of the VIX at different horizons under the real and risk-neutral densities

(a) Intercept



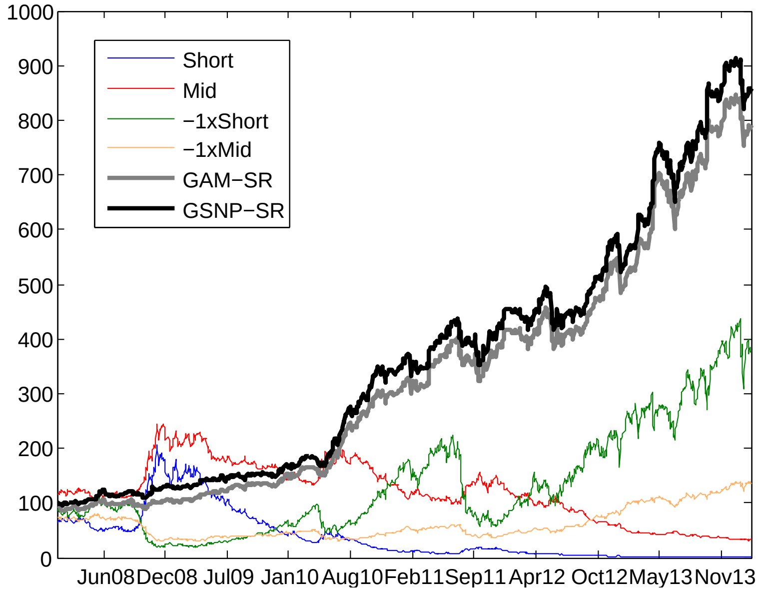
(b) Coefficients on the short and long-run components



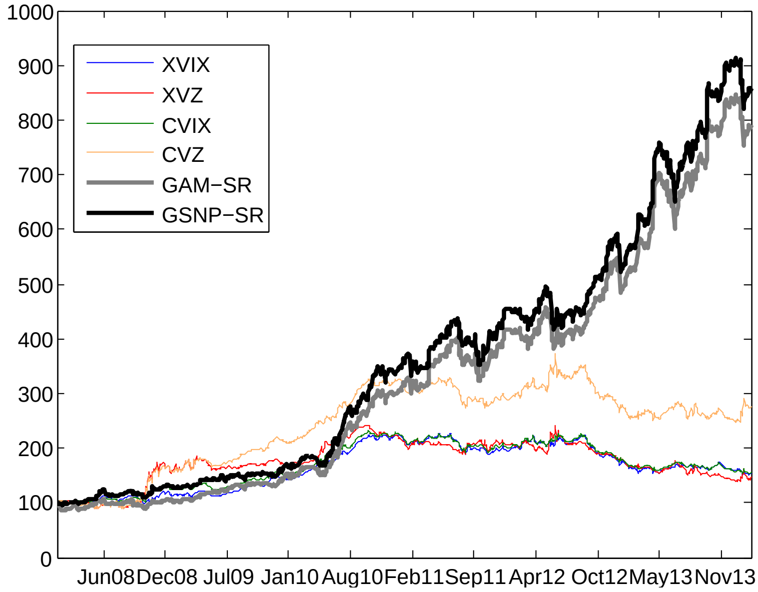
Note: The estimated model is a two-component GSNP MEM, which has a short-run component s_t and a long run component ς_t. The x-axis represents the time horizon in days.

Figure 6: Evolution of investment strategies accumulated gains

(a) GSNP vs. buy and hold strategies

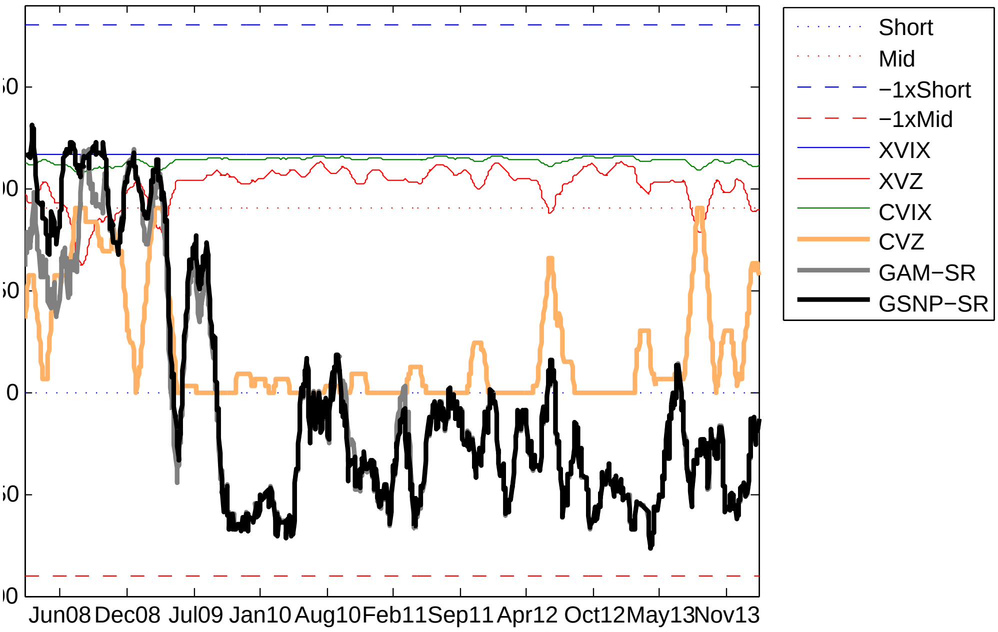


(b) GSNP vs. long-short static and dynamic strategies



Note: All the strategies start from an initial investment of \$100. "Short" and "Mid" denote the S&P 500 VIX short and mid futures indices. "-1×" denote short sales on those indices. XVIX is a UBS ETN following a long-short static strategy on the VIX futures indices, while XVZ is a Barclays ETN following a dynamic strategy. CVIX and CVZ are investment strategies proposed by Alexander and Korovilas (2013). GSNP-SR and GAM-SR denote the returns obtained by maximising the conditional Sharpe Ratio, based on the parameters obtained from a two-component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution, and a Gamma distribution, respectively. The parameters are estimated each day using the information available at each point.

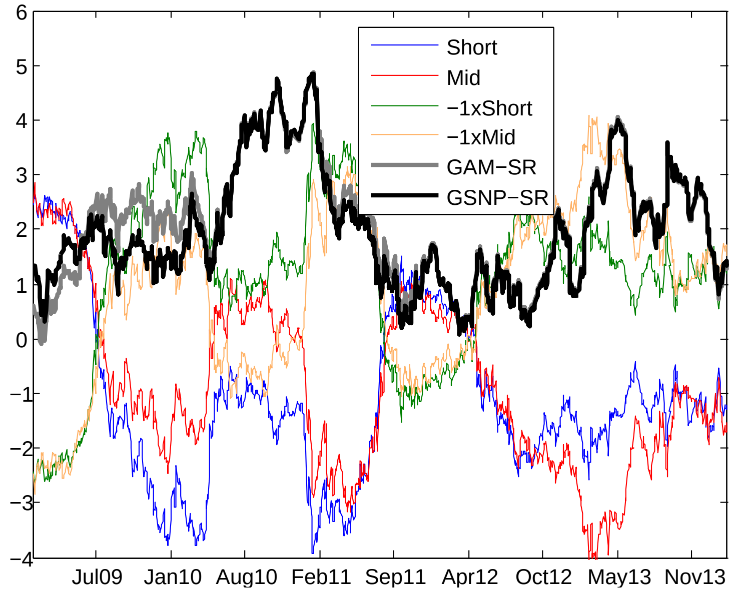
Figure 7: Evolution of portfolio weights



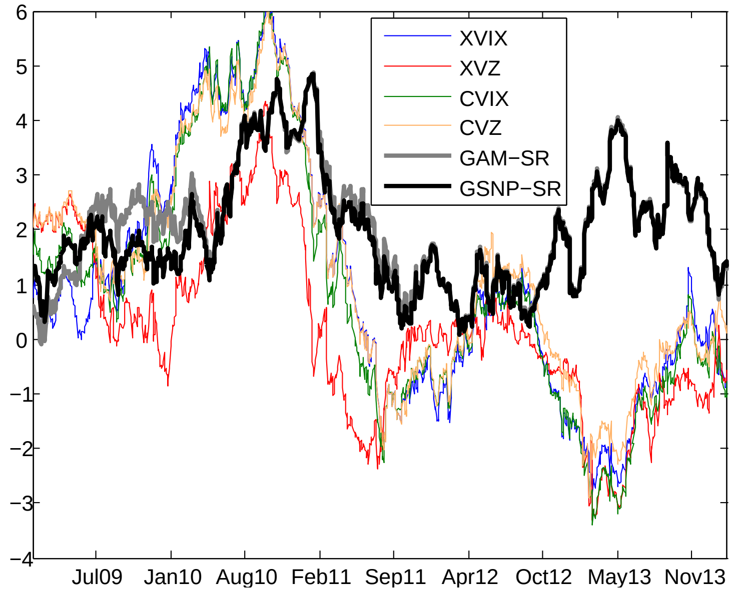
Note: 30-day moving averages of daily weights. “Short” and “Mid” denote the S&P 500 VIX short and mid futures indices. “-1×” denote short sales on those indices. XVIX is a UBS ETN following a long-short static strategy on the VIX futures indices, while XVZ is a Barclays ETN following a dynamic strategy. CVIX and CVZ are investment strategies proposed by Alexander and Korovilas (2013). GSNP-SR and GAM-SR denote the returns obtained by maximising the conditional Sharpe Ratio, based on the parameters obtained from a two-component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution, and a Gamma distribution, respectively. The parameters are estimated each day using the information available at each point.

Figure 8: Sharpe Ratio of realised returns over a one-year moving window

(a) GSNP vs. buy and hold strategies



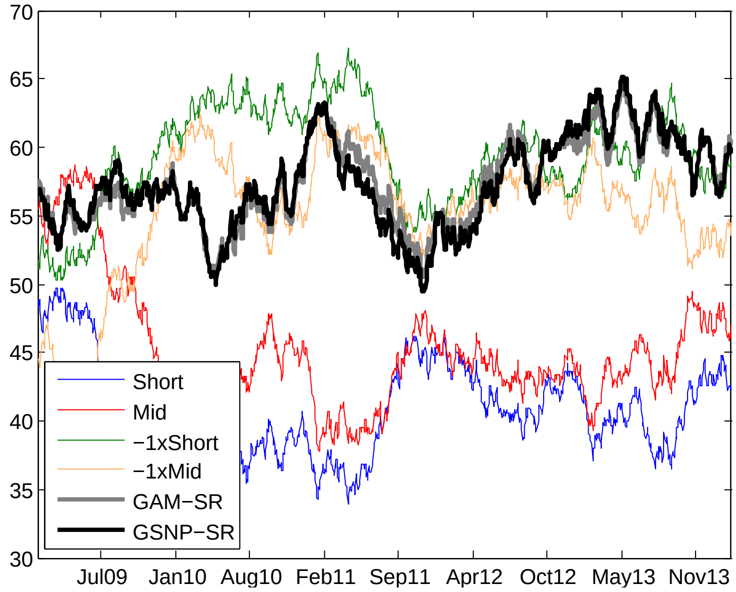
(b) GSNP vs. long-short static and dynamic strategies



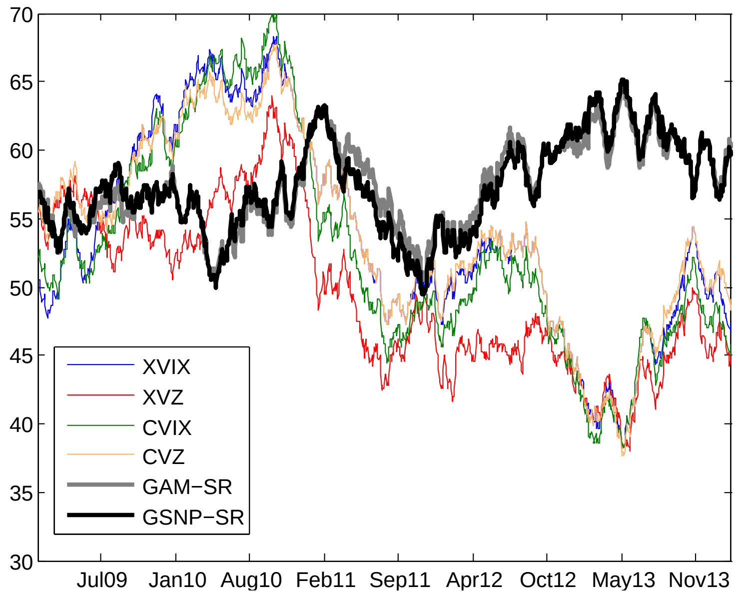
Note: "Short" and "Mid" denote the S&P 500 VIX short and mid futures indices. "-1×" denote short sales on those indices. XVIX is a UBS ETN following a long-short static strategy on the VIX futures indices, while XVZ is a Barclays ETN following a dynamic strategy. CVIX and CVZ are investment strategies proposed by Alexander and Korovilas (2013). GSNP-SR and GAM-SR denote the returns obtained by maximising the conditional Sharpe Ratio, based on the parameters obtained from a two-component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution, and a Gamma distribution, respectively. The parameters are estimated each day using the information available at each point.

Figure 9: Proportion of days with positive realised returns over a one-year moving window (%)

(a) GSNP vs. buy and hold strategies



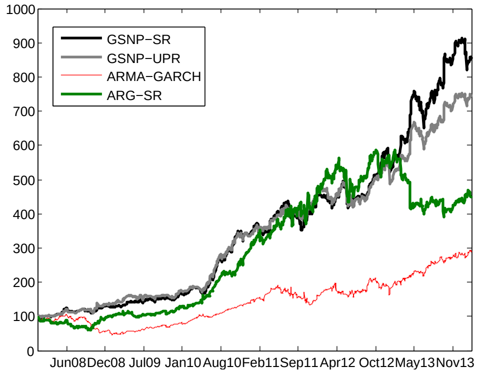
(b) GSNP vs. long-short static and dynamic strategies



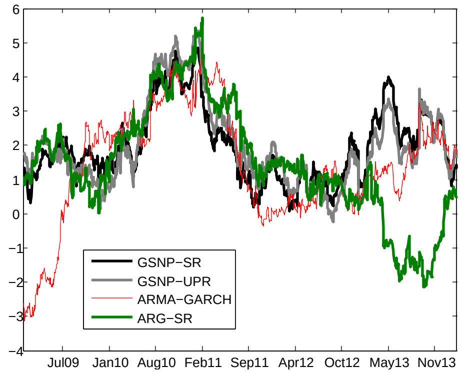
Note: "Short" and "Mid" denote the S&P 500 VIX short and mid futures indices. "-1×" denote short sales on those indices. XVIX is a UBS ETN following a long-short static strategy on the VIX futures indices, while XVZ is a Barclays ETN following a dynamic strategy. CVIX and CVZ are investment strategies proposed by Alexander and Korovilas (2013). GSNP-SR and GAM-SR denote the returns obtained by maximising the conditional Sharpe Ratio, based on the parameters obtained from a two-component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution, and a Gamma distribution, respectively. The parameters are estimated each day using the information available at each point.

Figure 10: Profitability measures of the realised returns of alternative dynamic asset allocation strategies.

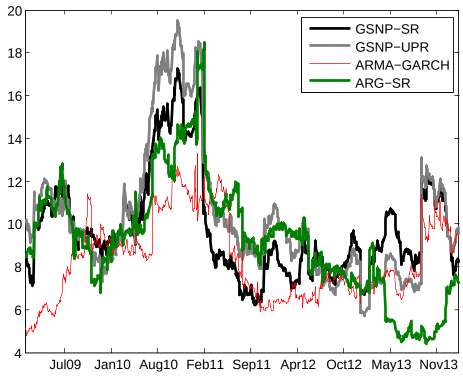
(a) Accumulated gains since Jan-2008



(b) Realised Sharpe Ratio over one-year moving windows



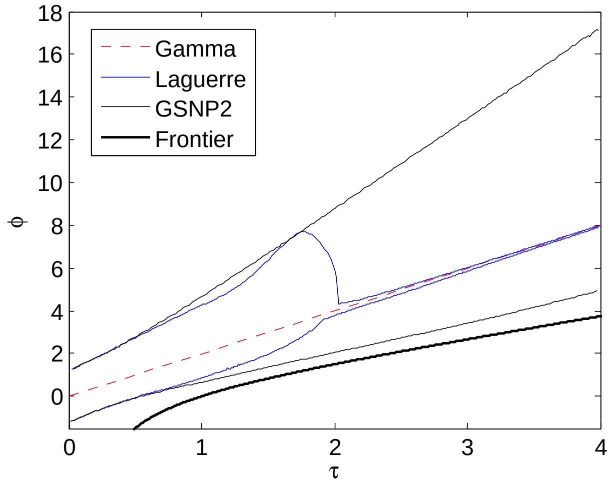
(c) Realised Upside Potential Ratio over one-year moving windows



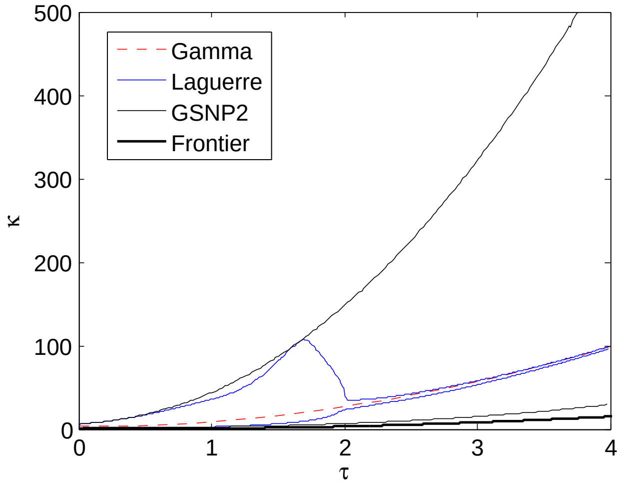
Note: Both the Sharpe Ratio (SR) and Upside Potential Ratio (UPR) are expressed in annualised terms. GSNP-SR (GSNP-UPR) denotes the returns obtained by maximising the conditional SR (UPR), based on the parameters obtained from a Component MEM for the VIX with a GSNP(2) expansion of the Gamma distribution. ARMA-GARCH denotes the returns obtained by maximising the conditional SR, based on the parameters obtained from a bivariate ARMA(2,1)-GARCH(1,1) with constant conditional correlation, estimated on the short and mid VIX future index returns. ARG-SR denotes the returns obtained by maximising the conditional SR, based on the parameters obtained from a first order Autoregressive Gamma process. The parameters are estimated each day using the information available at each point.

Figure A1: Regions of the coefficients of variation, skewness and kurtosis credit institutions

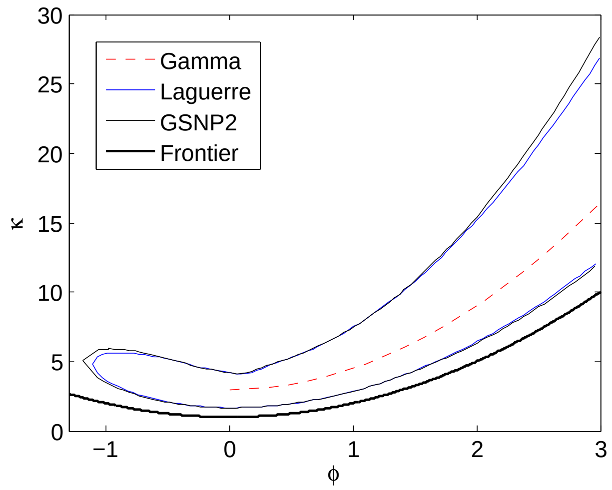
(a) Variation vs. Skewness



(b) Variation vs. Kurtosis



(c) Skewness vs. Kurtosis



Notes: τ, ϕ and κ denote the coefficients of variation, skewness and kurtosis, respectively. The lines labelled "Frontier" denote the limits that no density can surpass. "Laguerre" denotes a truncated third order Laguerre expansion of the Gamma distribution, while "GSNP2" denotes a second order SNP expansion of the Gamma distribution.