A Stochastic Dominance Approach to Financial Risk Management Strategies*

Chia-Lin Chang
Department of Applied Economics
Department of Finance
National Chung Hsing University
Taichung, Taiwan

Juan-Ángel Jiménez-Martín
Department of Quantitative Economics
Complutense University of Madrid, Spain

Esfandiar Maasoumi
Department of Economics
Emory University, USA

Teodosio Pérez-Amaral
Department of Quantitative Economics
Complutense University of Madrid, Spain

Revised: April 2014

* The authors are most grateful to Michael McAleer for numerous insightful comments guidance. For financial support, the first author wishes to thank the National Science Council, Taiwan, and the second and fourth authors acknowledge the Ministerio de Economía y Competitividad and Comunidad de Madrid, Spain.
Abstract

The Basel III Accord requires that banks and other Authorized Deposit-taking Institutions (ADIs) communicate their daily risk forecasts to the appropriate monetary authorities at the beginning of each trading day, using one of a range of alternative risk models to forecast Value-at-Risk (VaR). The risk estimates from these models are used to determine the daily capital charges (DCC) and associated capital costs of ADIs, depending in part on the number of previous violations, whereby realized losses exceed the estimated VaR. In this paper we define risk management in terms of choosing sensibly from a variety of risk models and discuss the optimal selection of the risk models. Previous approaches to model selection for predicting VaR proposed combining alternative risk models and ranking such models on the basis of average DCC, or other quantiles of its distribution. These methods are based on the first moment, or specific quantiles of the DCC distribution, and supported by restrictive evaluation functions. In this paper, we consider robust uniform rankings of models over large classes of loss functions that may reflect different weights and concerns over different intervals of the distribution of losses and DCC. The uniform rankings are based on recently developed statistical tests of stochastic dominance (SD). The SD tests are illustrated using the prices and returns of VIX futures. The empirical findings show that the tests of SD can rank different pairs of models to a statistical degree of confidence, and that the alternative (recentered) SD tests are in general agreement.

Key words and phrases: Stochastic dominance, Value-at-Risk, daily capital charges, violation penalties, optimizing strategy, Basel III Accord, VIX futures, global financial crisis.

JEL Classifications: G32, G11, G17, C53, C22.
1. Introduction

The Basel III Accord requires that banks and other Authorized Deposit-taking Institutions (ADIs) communicate their daily risk forecasts to the appropriate monetary authorities at the beginning of each trading day, using one of a range of alternative financial risk models to forecast Value-at-Risk (VaR). The risk estimates from these models are used to determine the daily capital charges (DCC) and associated capital costs of ADIs, depending in part on the number of previous violations, whereby realized losses exceed the estimated VaR (for further details see, for example, Chang et al. (2011)).

In 1993 the Chicago Board Options Exchange (CBOE) introduced a volatility index, VIX (Whaley, 1993), which was originally designed to measure the market expectation of 30-day volatility implied by at-the-money S&P100 option prices. In 2003, together with Goldman Sachs, CBOE updated VIX to reflect a new way of measuring expected volatility, one that continues to be widely used by financial theorists.

The new VIX is based on the S&P500 Index, and estimates expected volatility by averaging the weighted prices of S&P500 puts and calls over a wide range of strike prices. Although many market participants considered the index to be a good predictor of short term volatility, namely daily or intraday, it took several years for the market to introduce volatility products, starting with over-the-counter products, such as variance swaps and other financial derivatives. The first exchange-traded product, VIX futures, was introduced in March 2004, and was followed by VIX options in February 2006. Both of these volatility derivatives are based on the VIX index as the underlying asset.

McAleer et al. (2013a, b, c) analyse, from a practical perspective, how the new market risk management strategies performed during the 2008-09 global financial crisis (GFC), and evaluate how the GFC affected the best risk management practices. These papers define risk management in terms of choosing, using appropriate financial targets, from a variety of financial risk models, and discuss the selection of optimal risk models. They forecast VaR using ten univariate conditional volatility models with different error distributions. Additionally, they analyze twelve new strategies based on combinations of the previous standard univariate model forecasts of VaR, namely: Infinum (0th percentile), Supremum
(100th percentile), Average, Median and nine additional strategies based on the 10th through to the 90th percentiles.

Such an approach is intended to select a robust VaR forecast, irrespective of the time period, that provides reasonable daily capital charges and number of violation penalties under the Basel II Accord. They found that the Median is a GFC-robust strategy, in the sense that maintaining the same risk management strategy before, during and after the GFC leads to comparatively low daily capital charges and violation penalties under the Basel II Accord. Chang et al. (2011) apply a similar methodology for choosing the best strategy to forecast VaR for a portfolio based on VIX futures.

These prior methods focus on the first moment, or certain quantiles of the DCC distribution. Alternative criteria may consider mean-variance trade-offs, as in substantial areas of financial research, or general evaluation criteria that incorporate higher moments and quantiles of the underlying probability distributions. These will all provide appropriate “cardinal” and “complete” rankings of models and strategies, based on subjective valuations of different aspects, or parts of the DCC distribution. For instance, were DCC to be a Gaussian variate, mean-variance assessments would be strongly justified. This is not likely, however, and consensus on appropriate weighting and assessment functionals of non-Gaussian distributions has been elusive. It is of some importance to point out, that mean-variance type assessments are justified by a joint consideration of quadratic risk function, as well as the full characterization of the Gaussian case by the second moment. Absent a Gaussian setting, justification of a quadratic loss function itself becomes questionable. Why would we not be concerned with higher moments (when they exist), and often asymmetrical tail area behavior, especially when tail functions such as VaR are of interest?

A complementary robust alternative, is to seek weak uniform rankings over entire classes of evaluation functions, and based on nonparametric distributions of DCC. In this respect, Stochastic Dominance (SD) tests have been developed to test for statistically significant rankings of prospects. Assuming $F$ and $G$ are the distribution functions of DCC produced by model 1 and model 2, respectively, model 1 first order SD model 2, over the support of DCC value $dcc$, iff $F(dcc) \preceq G(dcc)$, with strict inequality over some values of DCC. This means that the model that produces $G$ is dominant over all merely increasing evaluation functions.
since, at all quantiles, the probability that capital charges are smaller under G is greater than under F. In particular, the distribution F will have a higher median DCC than G. Similarly, each and every (quantile) percentile of the F distribution will be at a higher DCC level than the corresponding percentile of the G distribution. Consequently, model 2 will be preferred to model 1, to a statistical degree of confidence, on the basis of lower capital charges.

Higher order SD rankings reference further subclasses of evaluation functions, those that are increasing and concave, reflecting increasing risk aversion; see sections 4-5 below.

In this paper we examine several standard models for forecasting VaRs, including GARCH, EGARCH, and, GJR, paired with Gaussian and Student-t distributions. The results show that the Gaussian distribution is preferred to Student-t in forecasting DCC. With the Student-t distribution, the EGARCH model provides a greater likelihood of higher DCC in comparison with GARCH and GJR. Using the Gaussian distribution to forecast DCC does not lead to either first or second order stochastic dominance. In respect of the CDF and integrated CDF, the basis of first and Second Order Stochastic dominance testing, it seems that the higher expected DCC of GJR or GARCH may be compensated by lower risk compared with EGARCH.

The remainder of the paper is organized as follows. In Section 2 the definition, notation and properties of stochastic dominance are presented. Section 3 describes briefly the Basel II Accord for computing daily capital charges. Section 4 presents alternative GARCH models to produce daily capital charges. Section 5 introduces the data, describes the block bootstrapping method to simulate time series, and illustrates the application of stochastic dominance to enhance financial risk management strategies of banks. Section 6 presents the main results. Section 7 gives some concluding comments.

2. Forecasting Value-at-Risk and Daily Capital Charges

In this section we introduce the calculation of daily capital charges (DCC) as a basic criterion for choosing between risk models. The Basel II Accord stipulates that daily capital charges (DCC) must be set at the higher of the previous day’s VaR or the average VaR over the last 60 business days, multiplied by a factor \((3+k)\) for a violation penalty, where a violation occurs when the actual negative returns exceed the VaR forecast negative returns for a given day:
where

\[ DCC_t = \sup \left\{ -3 + k, \frac{VaR_{t-1} - VaR_{60}}{\hat{\sigma}_t} \right\} \]  \hspace{1cm} (1) \]

\[ VaR_t = \hat{Y}_t - z_t \cdot \hat{\sigma}_t \], the Value-at-Risk for day \( t \),

\[ \overline{VaR}_{60} = \text{mean VaR over the previous 60 working days}, \]

\[ \hat{Y}_t = \text{estimated return at time } t, \]

\[ z_t = 1\% \text{ critical value of the distribution of returns at time } t, \]

\[ \hat{\sigma}_t = \text{estimated risk (or square root of volatility) at time } t, \]

\[ 0 \leq k \leq 1 \] is the Basel II violation penalty (see Table 1).

The formula given in equation (1) is contained in the 1995 amendment to Basel I, while Table 1 appears for the first time in the Basel II Accord in 2004. The multiplication factor (or penalty), \( k \), depends on the central authority’s assessment of the ADI’s risk management practices and the results of a simple backtest. It is determined by the number of times actual losses exceed a particular day’s VaR forecast (see Basel Committee on Banking Supervision (2004)). The minimum multiplication factor of 3 is intended to compensate for various errors that can arise in model implementation, such as simplifying assumptions, analytical approximations, small sample biases and numerical errors that tend to reduce the true risk coverage of the model (see Stahl (1997)). Increases in the multiplication factor are designed to increase the confidence level that is implied by the observed number of violations at the 99% confidence level, as required by regulators (for a detailed discussion of VaR, as well as exogenous and endogenous violations, see McAleer (2009), and McAleer et al. (2010)).
In calculating the number of violations, ADIs are required to compare the forecasts of VaR with realised profit and loss figures for the previous 250 trading days. In 1995, the 1988 Basel Accord (Basel Committee on Banking Supervision (1988)) was amended to allow ADIs to use internal models to determine their VaR thresholds (Basel Committee on Banking Supervision (1995)). However, ADIs that propose using internal models are required to demonstrate that their models are sound. Movement from the green zone to the red zone arises through an excessive number of violations. Although this will lead to a higher value of $k$, and hence a higher penalty, violations will also tend to be associated with lower daily capital charges. It should be noted that the number of violations in a given period is an important, though not the only, guide for regulators to approve a given VaR model.

VaR refers to the lower bound of a confidence interval for a (conditional) mean, that is, a “worst case scenario on a typical day”. If interest lies in modelling the random variable, $Y_t$, it could be decomposed as follows:

$$Y_t = E(Y_t | F_{t-1}) + \varepsilon_t.$$  \hspace{1cm} (2)

This decomposition states that $Y_t$ comprises a predictable component, $E(Y_t | F_{t-1})$, which is the conditional mean, and a random component, $\varepsilon_t$. The variability of $Y_t$, and hence its distribution, is determined by the variability of $\varepsilon_t$. If it is assumed that $\varepsilon_t$ follows a conditional distribution, $\varepsilon_t \sim D(\mu_t, \sigma_t^2)$, where $\mu_t$ and $\sigma_t$ are the time-varying conditional mean and standard deviation of $\varepsilon_t$, respectively, these can be estimated using a variety of parametric, semi-parametric or non-parametric methods.

The VaR threshold for $Y_t$ can be calculated as:

$$VaR_t = E(Y_t | F_{t-1}) - \alpha \sigma_t,$$  \hspace{1cm} (3)

where $\alpha$ is the critical value from the distribution of $\varepsilon_t$ to obtain the appropriate confidence level. It is possible for $\sigma_t$ to be replaced by alternative estimates of the conditional standard deviation in order to obtain an appropriate VaR (for useful reviews of theoretical results for
conditional volatility models, see Li et al. (2002) and McAleer (2005), where several univariate and multivariate, conditional, stochastic and realized volatility models are discussed.

Some recent empirical studies (see, for example, Berkowitz and O’Brien (2001), Gizycki and Hereford (1998), and Pérignon et al. (2008)) have indicated that financial institutions tend to overestimate their market risks in disclosures to the appropriate regulatory authorities, which can imply a costly restriction to the banks trading activity. ADIs may prefer to report high VaR numbers to avoid the possibility of regulatory intrusion. This conservative risk reporting suggests that efficiency gains may be feasible. In particular, as ADIs have effective tools for the measurement of market risk, while satisfying the qualitative requirements, ADIs could conceivably reduce daily capital charges by implementing a context-dependent market risk disclosure policy. McAleer (2009) and McAleer et al. (2010) discuss alternative approaches to optimize VaR and daily capital charges.

The next section describes several volatility models that are widely used to forecast the 1-day ahead conditional variances and VaR thresholds.

3. Models for Forecasting VaR

ADIs can use internal models to determine their VaR thresholds. There are alternative time series models for estimating conditional volatility. In what follows, we present several well-known conditional volatility models that can be used to evaluate strategic market risk disclosure, namely GARCH, GJR and EGARCH, with Gaussian and Student-\(t\) distributions. These models are chosen as they are widely used in the literature. For an extensive discussion of the theoretical properties of several of these models, see Ling and McAleer (2002a, 2002b, 2003a), McAleer (2005), and Caporin and McAleer (2012).

3.1 GARCH

For a wide range of financial data series, time-varying conditional variances can be explained empirically through the autoregressive conditional heteroskedasticity (ARCH) model, which
was proposed by Engle (1982). When the time-varying conditional variance has both autoregressive and moving average components, this leads to the generalized ARCH(\(p,q\)), or GARCH(\(p,q\)), model of Bollerslev (1986). It is very common in practice to impose the widely estimated GARCH(1,1) specification in advance.

Consider the stationary AR(1)-GARCH(1,1) model for daily returns, \(y_t\):

\[
y_t = \varphi_1 + \varphi_2 y_{t-1} + \varepsilon_t, \quad |\varphi_2| < 1
\]

(4)

for \(t = 1, \ldots, n\), where the shocks to returns are given by:

\[
\varepsilon_t = \eta_t \sqrt{h_t}, \quad \eta_t \sim iid(0,1) \\
h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1},
\]

(5)

and \(\omega > 0, \alpha \geq 0, \beta \geq 0\) are sufficient conditions to ensure that the conditional variance \(h_t > 0\). The stationary AR(1)-GARCH(1,1) model can be modified to incorporate a non-stationary ARMA(\(p,q\)) conditional mean and a stationary GARCH(\(r,s\)) conditional variance, as in Ling and McAleer (2003b).

### 3.2 EGARCH

An alternative model to capture asymmetric behaviour in the conditional variance is the Exponential GARCH, or EGARCH(1,1), model of Nelson (1991), namely:

\[
\log h_t = \omega + \alpha \left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}} + \beta \log h_{t-1}, \quad |\beta| < 1
\]

(6)

where the parameters \(\alpha, \beta, \gamma\) have different interpretations from those in the GARCH(1,1) and GJR(1,1) models.

EGARCH captures asymmetries differently from GJR. The parameters \(\alpha\) and \(\gamma\) in EGARCH(1,1) represent the magnitude (or size) and sign effects of the standardized residuals, respectively, on the conditional variance, whereas \(\alpha + \gamma\) represent the
effects of positive and negative shocks, respectively, on the conditional variance in GJR(1,1). Unlike GJR, EGARCH can accommodate leverage, depending on the restrictions imposed on the size and sign parameters, though leverage is not guaranteed.

As noted in McAleer et al. (2007), there are some important differences between EGARCH and the previous two models, as follows: (i) EGARCH is a model of the logarithm of the conditional variance, which implies that no restrictions on the parameters are required to ensure $h_t > 0$; (ii) moment conditions are required for the GARCH and GJR models as they are dependent on lagged unconditional shocks, whereas EGARCH does not require moment conditions to be established as it depends on lagged conditional shocks (or standardized residuals); (iii) Shephard (1996) observed that $|\beta| < 1$ is likely to be a sufficient condition for consistency of QMLE for EGARCH(1,1); (iv) as the standardized residuals appear in equation (6), $|\beta| < 1$ would seem to be a sufficient condition for the existence of moments; and (v) in addition to being a sufficient condition for consistency, $|\beta| < 1$ is also likely to be sufficient for asymptotic normality of the QMLE of EGARCH(1,1).

The three conditional volatility models given above are estimated under the following distributional assumptions on the conditional shocks: (1) Gaussian and (2) Student-$t$, with estimated degrees of freedom. As the models that incorporate the $t$ distributed errors are estimated by QMLE, the resulting estimators are consistent and asymptotically normal, so they can be used for estimation, inference and forecasting.

3.3 GJR

In the symmetric GARCH model, the effects of positive shocks (or upward movements in daily returns) on the conditional variance, $h_t$, are assumed to be the same as the effect of negative shocks (or downward movements in daily returns) of equal magnitude. In order to accommodate asymmetric behaviour, Glosten, Jagannathan and Runkle (1992) proposed a model (hereafter GJR), for which GJR(1,1) is defined as follows:

$$h_t = \omega + (\alpha + \gamma I(\eta_{t-1}))\varepsilon_{t-1}^2 + \beta h_{t-1},$$

(7)
where $\omega > 0, \alpha \geq 0, \alpha + \gamma \geq 0, \beta \geq 0$ are sufficient conditions for $h_t > 0$, and $I(\eta_t)$ is an indicator variable defined by:

$$I(\eta_t) = \begin{cases} 1, & \varepsilon_t < 0 \\ 0, & \varepsilon_t \geq 0 \end{cases}$$

(8)

as $\eta_t$ has the same sign as $\varepsilon_t$. The indicator variable differentiates between positive and negative shocks, so that asymmetric effects in the data are captured by the coefficient $\gamma$. For financial data, it is expected that $\gamma \geq 0$ because negative shocks have a greater impact on risk than do positive shocks of similar magnitude. The asymmetric effect, $\gamma$, measures the contribution of shocks to both short run persistence, $\alpha + \gamma/2$, and to long run persistence, $\alpha + \beta + \gamma/2$.

Although GJR permits asymmetric effects of positive and negative shocks of equal magnitude on conditional volatility, the special case of leverage, whereby negative shocks increase volatility while positive shocks decrease volatility (see Black (1976) for an argument using the debt/equity ratio), cannot be accommodated, in practice (for further details on asymmetry versus leverage in the GJR model, see Caporin and McAleer (2012)).

4. **Stochastic Dominance**

The objective is to evaluate each of the alternative conditional volatility models with respect to the DCC function. Observe that each model will yield different values of DCC because they will produce different VaR forecasts. The stochastic dominance concept is applied to determine which model should be used to produce the lowest DCC for a given investment, while not taking into account the number of violations since the primary purpose of the analysis is to assist risk managers in choosing among alternative models. Below we briefly describe the SD tests that are used in this paper.

4.1 Definitions and Hypothesis Formulation

---

In this section we follow heavily the exposition in Donald and Hsu (2013), which is basic for the subsequent
Let $X$ and $Y$ be two random variables with cumulative distribution functions (CDF) $F_X$ and $F_Y$, respectively. For first order stochastic dominance (SD1), $Y$ SD1 $X$, if $F_Y(z) \leq F_X(z)$ for all $z \in \mathbb{R}$. Let $W_U(F)$ denote an evaluation function of the form $W_U(F) = \int U(z) dF(z)$, where $F$ is the distribution of an underlying variable, and $U$ is any “utility” function. SD1 is defined over monotonically increasing utility functions, that is, $W_U(F_Y) \geq W_U(F_X)$ for all $U(z)$ such that $U'(z) \geq 0$.

Technical assumptions for the underlying statistical theory include the following:

**Assumption 4.1:**
1. $Z = [0, \bar{z}]$ where $\bar{z} < \infty$.
2. $F_X$ and $F_Y$ are continuous functions on $Z$ such that $F_X(z) = F_Y(z) = 0$ iff $z = 0$, and $F_X(z) = F_Y(z) = 1$ iff $z = \bar{z}$.

(see Linton, Maasumi and Whang, (2005) (hereafter LMW), Linton, Song and Whang (2010) and Donald and Hsu (2013) for further details).

In order to test if $Y$ SD1 $X$, Donald and Hsu (2013) formulate their hypotheses as:

$$
H_0 : F_Y(z) \leq F_X(z) \quad \text{for all } z \in Z, \quad (9)
$$

$$
H_1 : F_Y(z) > F_X(z) \quad \text{for some } z \in Z. \quad (10)
$$

This is different from LMW, who provide a simultaneous test of either $Y$ SD1 $X$ or $X$ SD1 $Y$.

**Assumption 4.2:**
1. $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^M$ are samples from distributions with CDF’s $F_X$ and $F_Y$, respectively. Some authors deal with independent samples and observations. LMW allow dependent time series and possibly dependent $X$ and $Y$.
2. $M$ is a function of $N$ satisfying that $M(N) \rightarrow \infty$ and $N/(N + M(N)) \rightarrow \lambda$ $(0, 1)$ when $N \rightarrow \infty$.

The CDF’s $F_X$ and $F_Y$ are estimated by empirical CDFs:

$$
\hat{F}_{X,N}(z) = \frac{1}{N} \sum_{i=1}^N 1(X_i \leq z), \quad \hat{F}_{Y,M}(z) = \frac{1}{M} \sum_{i=1}^M 1(Y_i \leq z),
$$

where $1(\cdot)$ denotes the indicator function. The Kolmogorov-Smirnov test statistic is given by:
\[
\hat{S}_N = \sqrt{\frac{NM}{N+M}} \sup_{z \in \mathbb{Z}} (\hat{F}_{Y,N}(z) - \hat{F}_{X,M}(z)).
\]

Let \(\Psi_{h_2}\) denote a mean zero Gaussian process with covariance kernel equal to \(h_2 \in H_z\), where \(H_z\) denotes the collection of all covariance kernels on \(Z \times Z\). For \(F_X\) and \(F_Y\) satisfying Assumption 4.1, and let \(h_2^{X,Y}\) denote the covariance kernel on \(Z \times Z\) such that:

\[
h_2^{X,Y}(z_1, z_2) = \lambda \cdot F_X(z_1)(1 - F_X(z_2)) + (1 - \lambda) \cdot F_Y(z_1)(1 - F_Y(z_2)) \quad \text{for } z_1 \leq z_2
\]

with \(\lambda\) defined in Assumption 4.2. Then,

\[
\sqrt{NM/(N+M)} (\hat{F}_{Y,M}(z) - \hat{F}_{X,N}(z) - (F_Y(z) - F_X(z))) \Rightarrow \Psi_{h_{2,X,Y}}
\]

See, for example, Linton et al (2005), or Donald and Hsu (2013). A typical limiting result is:

1. Under \(H_0\), \(\hat{S}_N \xrightarrow{D} \sup_{z \in \mathbb{Z}} \Psi_{h_{2,X,Y}}\)

2. Under \(H_1\), \(\hat{S}_N \xrightarrow{D} \infty\)

Several approaches for resampling and subsampling implementation of SD tests have been proposed. Some methods simulate the limiting Gaussian process, in the spirit of Barret and Donald (2003), using the Multiplier method, bootstrap with separate samples, or bootstrap with combined samples. Simulated processes weakly converge to the same process as the limit process, conditional on the sample path with probability approaching 1.

4.2 Re-centering Functions.

These simulation methods do not work when the data are weakly dependent, as for time series samples in this paper. In these cases one has to appeal to either the subsampling technique of Linton, Maasoumi and Whang (2005), or a variant of the block bootstrap. Donald and Hsu (2013) provide a comparative examination of these alternatives.

Donald and Hsu (2013) and LMW (2005) propose re-centering methods introduced by Hansen (2005) to construct critical values for the Kolmogorov-Smirnov type tests. This approach provides a test with improved size and power properties compared with the unadjusted test mounted at the composite boundary of the null and alternative spaces, the so
called Least Favorable Case (LFC). The re-centering function proposed by Donald and Hsu (2013) is conditioned to apply on the null space:

\[ \hat{\mu}_N(z) = \left( \hat{F}_{Y,N}(z) - \hat{F}_{X,N}(z) \right) \cdot 1 \left( \sqrt{N} \left( F_{Y,N}(z) - F_{X,N}(z) \right) < aN \right). \]

For \( \alpha < 1/2 \), let

\[ \check{c}_{bN} = \max \{ \hat{c}_{bN}, \eta \}, \]
\[ \hat{c}_{N} = \sup \left\{ c \mid P^u \left( \sup_{z \in Z} \sqrt{N} \left( \hat{D}_N^b(z) + \hat{\mu}_N(z) \right) \leq c \right) \leq 1 - \alpha \} \]

where \( \hat{D}_N^b(z) \) is the “S” statistic computed with the b-th resample/subsample block.

If the decision rule is to reject the null hypothesis, \( H_0 : F_Y(z) \leq F_X(z) \) for all \( z \in Z \) when \( \hat{S}_{N} > \check{c}_{bN} \), then the corresponding test has the same size properties as in the independent random samples case.

### 4.3 More on Weakly Dependent Data

Let \( \{(X_i, Y_i)\}_{i=1}^N \) be a strictly stationary time series sequence with joint distribution function \( F_{XY} \) on \( \mathbb{Z}^2 \) and marginal CDF’s \( F_X \) and \( F_Y \), respectively. Suppose that Assumption 1 of LMW holds. Then under the null hypothesis that \( H_0 : F_Y(z) \leq F_X(z) \) for all \( z \in Z \), \( \hat{S}_N \overset{D}{\rightarrow} \sup_{z \in Z} \Psi_{h_2}(z) \), where

\[ h_2(z_1, z_2) = \lim_{N \to \infty} \text{Cov} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N (1(Y_i \leq z_1) - 1(X_i \leq z_1) - F_Y(z_1) + F_X(z_1)), \right. \]
\[ \left. \frac{1}{\sqrt{N}} \sum_{i=1}^N (1(Y_i \leq z_2) - 1(X_i \leq z_2) - F_Y(z_2) + F_X(z_2)) \right) \leq (12) \]

which is the long-run covariance kernel function. In order to simulate \( \Psi_{h_2} \), Donald and Hsu (2013) propose the blockwise bootstrap as in LMW (2005) because the multiplier method and the bootstrap methods do not account for the weak dependence of the data. Then under the same conditions as in LMW, we have \( \sqrt{N}D_N^b(\cdot) \overset{p}{\Rightarrow} \Psi_{h_2}(\cdot) \) where \( h_2 \) is defined in (12).

We adopt the Donald and Hsu recentering procedure and critical values as they are less conservative under the null and at least as powerful under the alternative. To appreciate
the role played by the choice of critical values here, consider the critical value in either of the multiplier method (mp), bootstrap with separate samples (bs), or bootstrap with combined samples (bc), given below for \( k = mp, bs \) and \( bc \) is as follows:

\[
\hat{q}_N^k = \sup \{ q \mid P_N^N \left( \frac{\sqrt{N M}}{N + M} \sup_{z \in \mathbb{Z}} \hat{D}_N^k (z) \leq q \right) \leq 1 - \alpha \}. \tag{13}
\]

The critical value \( \hat{q}_N^k \) is bounded away from zero in probability. Since \( \eta \) can be chosen to be arbitrarily small, we can assume that \( \eta < \hat{q}_N^k \), which implies that \( \hat{c}_{n,N}^k = \max\{ \hat{c}_{n,N}^k, \eta \} \leq \hat{q}_N^k \) given that \( \hat{c}_{n,N}^k \leq \hat{q}_N^k \). Thus, Donald and Hsu (2013) are able to show that, given Assumptions 4.1, 4.2, and \( \alpha < \frac{1}{2} \),

\[
P(\hat{S} > \hat{q}_N^k) \leq P(\hat{S} > \hat{c}_{n,N}^k), \text{ for } k = mp, bs \text{ and } bc.
\]

This would imply that the power and size of these tests are never smaller than those of BD (2003).

4.4 Linton, Maasoumi and Whang's Subsampling Test

LMW (2005) estimate the critical value by the subsampling method proposed in Politis and Romano (1994). Donald and Hsu (2013) introduce LMW’s test with a modification that allows for different sample sizes. For \( s \geq 1 \), let \( X_s \) denote the collection of all of the subsets of size \( s \) of \( \{X_1, ..., X_N\} \):

\[
X_s = \left\{ \{X_{i_1}, ..., X_{i_s}\} \mid \{i_1, ..., r_s\} \subseteq \{1, ..., N\} \right\}.
\]

A random draw denoted by \( \{X_{1}^b, ..., X_{s}^b\} \) from \( X_s \) would be a random sample of size \( s \) without replacement from the original data. Let \( \hat{F}_{X,s}^b \) be the empirical CDF based on the random draw, \( \{X_{1}^b, ..., X_{s}^b\} \). Define \( \hat{F}_{Y,s}^b \) similarly. Let \( s_N \) and \( s_M \) denote the subsampling sizes for the \( X \) and \( Y \) samples, respectively. The subsampling critical value \( \hat{c}_N^k \) is given by
\[ \hat{c}_N^{t} = \sup \left\{ c \mid P^{\mu} \left( \frac{S_N S_M}{s_N + s_M} \sup_{z \in \mathbb{Z}} (\hat{\mathcal{F}}^{b}_{\text{Y},s_N}(z) - \hat{\mathcal{F}}^{b}_{\text{Y},s_M}(z)) \leq c \right) \leq 1 - \alpha \right\}. \]

Assume that:

1. \( s_N \to \infty, s_M \to \infty, s_N/M \to 0 \) and \( s_M/M \to 0 \) as \( N \to \infty \).
2. \( s_N/(s_N + s_M) \to \lambda, \) where \( \lambda \) is defined in Assumption 2.2.

The first part is standard for the subsampling method. The second part requires that the subsample sizes from the two samples grow at the same rate and that the limit of the ratio of the subsample sizes be the same as that of the original samples. This condition is important if, for example, \( s_N/(s_N + s_M) \to \lambda \neq \hat{\lambda}, \) so that, under the null hypothesis:

\[ \sqrt{\frac{S_N S_M}{s_N + s_M}} \sup_{z \in \mathbb{Z}} (\hat{\mathcal{G}}^{b}_{\text{Y},s_N}(z) - \hat{\mathcal{G}}^{b}_{\text{Y},s_M}(z)) \rightarrow_{D} \sup_{z \in \mathbb{Z}} \sqrt{\hat{\lambda}_s B_G} \ oG(z) - \sqrt{1 - \hat{\lambda}} B_F \ oF(z) \quad (14) \]

conditional on the sample(s) with probability one (denoted as D\( \rightarrow \) p). When \( \lambda \neq \hat{\lambda}, \) in the limit the left-hand side of (14) may not be a good approximation to the limiting null distribution of the original test statistic.

The limiting distribution theory in LMW (2005) covers weakly stationary, dependent samples, with certain mixing conditions, such as we deal with in our application. In addition, they allow for the prospects being ranked to be estimated functions, rather than original series described above. As long as estimators involved in these estimated functions permit certain expansions, as described in LMW (2005), assumption 2, section 3.1, the limiting distribution theory is preserved with re-centering. The DCC functions being ranked here are certainly estimated functions of the data. The conditional means and variance estimators in the VaR forecasts are generally consistent, and admit the type of expansions assumed in LMW (2005). But, DCC is a maximal function of two functions of past VaRs, and to the extent that this may lead to discontinuities, or non-smoothness, in the estimated empirical CDFs, the accuracy of the underlying limiting distributions may be affected. It is known that the subsampling method is valid for many non-standard cases of this kind. But the same may not be true of bootstrap methods. A more careful technical examination of this issue is beyond the scope of this paper.
5. Data and Implementation of Tests

5.1. Data description

The data used for estimation and forecasting are closing daily prices (settlement prices) for the 30-day maturity CBOE VIX volatility index futures (ticker name VX). They were obtained from the Thomson Reuters-Data Stream Database for the period 26 March 2006 to 29 November 2013 (2526 observations). The settlement price is calculated by the CBOE as the average of the closing bid and ask quote so as to reduce the noise due to any microstructure effects. The contracts are cash settled on the Wednesday 30 days prior to the third Friday on the calendar month immediately following the month in which the contract expires. The underlying asset is the VIX index that was originally introduced by Whaley (1993) as an index of implied volatility on the S&P100. In 2003 the new VIX based on the S&P500 index was introduced.

VIX is a measure of the implied volatility of 30-day S&P500 options. It is independent of an option pricing model and is calculated from the prices of the front month and next-to-front month S&P500 at-the-money and out-the-money call and put options. The level of VIX represents a measure of the implied volatilities of the entire smile for a constant 30-day to maturity option chain. VIX is quoted in percentage points (for example, 30.0 VIX represents an implied volatility of 30.0%). In order to invest in VIX, an investor can take a position in VIX futures or VIX options.

Although VIX represents a measure of the expected volatility of the S&P500 over the next 30-days, the prices of VIX futures are based on the current expectation of what the expected 30-day volatility will be at a particular time in the future (on the expiration date). Although the VIX futures should converge to the spot at expiration, it is possible to have significant disparities between the spot VIX and VIX futures prior to expiration. Figure 1 shows the daily VIX futures index together with the 30-day maturity VIX futures closing prices. VIX has a correlation (0.96) with the 30-day maturity VIX futures. VIX futures prices tend to show significantly lower volatility than VIX, which can be explained by the fact that VIX futures
must be priced in a manner that reflects the mean reverting nature of VIX. For the whole sample, the standard deviation is 9.99 for VIX and 8.58 for VIX futures prices.

[Insert Figures 1-2 here]

If \( P_t \) denotes the closing prices of the VIX futures contract at time \( t \), the returns at time \( t \) (\( R_t \)) are defined as:

\[
R_t = 100 \times \log \left( \frac{P_t}{P_{t-1}} \right).
\]

(15)

Figure 2 shows the daily VIX futures returns, and the descriptive statistics for the daily returns are given in Table 2. The returns to the VIX futures are driven by changes in expectations of implied volatility. Figure 3 shows the histograms for the daily returns, together with the theoretical Gaussian and Student-t probability density functions and a kernel density estimator. The Student-t density fits the returns distributions better than does its Gaussian counterpart.

[Insert Figure 3]

Regarding the returns volatility, several measures of volatility are available in the literature. In order to gain some intuition, we adopt the measure proposed in Franses and van Dijk (1999), who define the true volatility of returns as:

\[
V_t = \left[ \left( R_t - E \left( R_t | F_{t-1} \right) \right)^2 \right]^{1/2},
\]

(16)

[Insert Figure 4 here]

where \( F_{t-1} \) is the information set at time \( t-1 \). Figure 4 presents \( V_t \) in equation (16) as “volatilities”. The series exhibit clustering that should be captured by an appropriate time series model. Until January 2007, a month before the first reports of subprime losses, the volatility of the series seems to be stable. The volatility reached an all time peak on 27 February 2007, when it climbed to 26\% (the mean (median) for the entire sample is 2.62(1.78)), as the US equity market had its worst day in four years. Then it remained above historic levels, but the VIX futures volatility increases again after August 2008, due in large
part to the worsening global credit environment, with a maximum again on 3 November 2008. Then the volatility remained low until the news about the sovereign debt crisis in the Eurozone created another spike in volatility in the first week of May 2010, when the VIX futures reached 35 with a high volatility in returns. Finally, at the end of September 2011 we observe another maximum of 42 because of the 2011 US debt ceiling episode.

5.2. The Block Bootstrapping and subsampling

In order to test for SD rankings between risk models using different conditional volatility models for forecasting VaR, we implement the Circular Block Bootstrapping (CBB) method developed in Politis and Romano (1992) for resampling the VIX futures through the MFE toolbox of Kevin Sheppard. The block bootstrap is widely used for implementing the bootstrap with time series data. It consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement.

In the CBB, let the data consist of observations \( \{X_i : i = 1, \ldots, n\} \), and let \( l \in \{1, \ldots, n\} \) and \( b \geq 1 \) denote the length and the number of blocks, respectively, such that \( l \times b \leq n \). Let \( n \) and \( m \) be the initial data size and the bootstrap sample size, \( m \leq n \) and \( k \) the number of blocks chosen. CBB consists of dividing the time series into \( b \) blocks of consecutive observations denoted by:\n
\[
B_i = \left( X_{(i-1)l+1}, \ldots, X_{il} \right) \quad i = 1, \ldots, n
\]

A random sample of \( k \) blocks, \( k \geq 1 \), \( B'_1, \ldots, B'_k \) is selected with replacement from \( B'_1, \ldots, B'_k \). Joining the \( k \) blocks with \( m = k \times l \) observations, the bootstrap sample is given as:\n
\[
\left( X'_1, \ldots, X'_l, \ldots, X'_{(k-1)l+1}, \ldots, X'_l \right)
\]

The CBB procedure is based on wrapping the data around a circle and forming additional blocks using the “circularly defined” observations. For \( i \geq n \), it is defined that \( X_i = X_{i_n} \), where \( i_n = i \mod n \) and \( X_0 = X_n \). The CBB method resamples overlapping and periodically extended blocks of length \( l \). Notice that each \( X_i \) appears exactly \( l \) times in the collection of blocks and, as the CBB resamples the blocks from this collection with equal probability, each of the original observations \( X_1, \ldots, X_n \) receives equal weight under the CBB. This property
distinguishes the CBB from previous methods, such as the non-overlapping block bootstrap of Carlstein (1992).
Note that DCC (VaR) is estimated for each drawn sample, thus generating the bootstrap (subsample) distribution of the test statistics.

5.3. Daily Capital Charges (DCC) and Evaluation Framework: Stochastic Dominance

The primary objective is to evaluate each of the alternative conditional volatility models with respect to the DCC function. Each model will imply different values of DCC as they will produce different VaR forecasts. The stochastic dominance concept is used to determine which model should be used for producing the lowest DCC for a given investment, not taking into account the number of violations. The main purpose of the analysis is to assist risk managers in choosing among alternative financial risk models.

Analysing risk management, McAleer et al. (2013b) forecast VaR using ten univariate conditional volatility models with different error distributions. Additionally, they analyze twelve new strategies based on combinations of the previous standard univariate forecasts of VaR, namely: Infimum (0th percentile), Supremum (100th percentile), Average, Median and nine additional strategies based on the 10th through to the 90th percentiles. This was intended to select a robust VaR forecast, irrespective of the time period, that provides reasonable daily capital charges and number of violation penalties under the Basel Accord. They found that the Median is a GFC-robust strategy, in the sense that maintaining the same risk management strategy before, during and after the GFC leads to comparatively low daily capital charges and violation penalties under the Basel Accord. Chang et al. (2011) apply this model selection criterion for choosing the best risk management strategy when dealing with VIX futures.

For each criterion above there corresponds an implied evaluation, or weighting function, over the quantiles of the distribution. SD ranking may provide “uniform” evaluations based on large classes of loss functions, and/or distribution functions. To a statistical degree of confidence, SD rankings may provide for robust risk model selection and reporting. This method offers the advantage of always being consistent with expected utility maximization, while not requiring an actual specification of a loss function.

As an example, consider the probability density functions of the DCC from two alternative models: model 1 and model 2 as shown in Figure 5. In the left panel of Figure 5, the bold line
represents model 1, which has a Gaussian distribution function with mean 4.5 and standard deviation 1. The dotted line is for model 2 and depicts the density of a Chi-square with 4 degrees of freedom. If we made the choice according to previous criteria based only on first moments, this would imply choosing model 2, since the mean DCC from model 2 is lower than that from model 1. However, looking at the right panel in Figure 5, which depicts the CDF for models 1 (Gaussian) and 2 (Chi-square), we can see that DCC values of model 2 are more uncertain, with a greater probability of either small (the CDF for Model 2 is initially above the CDF for Model 1) or large DCC (the CDF for model 2 lies below model 1 for high DCC values). Because these CDF functions cross, first order stochastic dominance cannot be established. But given the high crossing point, Second Order dominance is prospective, and we can test for it based on the cumulative areas under the CDFs (ICDF), as shown in the right panel of Figure 6. Because the difference in areas (ICDF model 2 - ICDF model 1) is always positive, this implies that model 1 second order dominates model 2, and will be preferred by a risk averse risk manager, whatever his/her cardinal (concave) loss function. Choosing model 2 would imply a greater risk during periods of turmoil, that is, greater uncertainty.

This is quite important since safe practice based on a specific risk assessment function, such as mean-variance, or any particular quantile, may be an artefact of the particular functional employed, and how that functional places different weights on different quantiles, as well as how different quantiles are compared. This is a question of increasing risk aversion, embodied by, or absent from, different loss functions.

The way that SD may be used to choose the best risk management strategy is as follows. For notational simplicity, we write \( FSD \) and \( SSD \) whenever Model 1 dominates Model 2 according to FSD and SSD, respectively. Let Y and X be the DCC produced using model 1 and model 2, respectively. Based on the previous definition, if Y first-order stochastically dominates X, then Y will involve higher DCC than X in the sense that it has a higher probability of producing higher values of DCC. Therefore, Y must be associated with a higher DCC than X if both X and Y require the same initial investment. In this context, \( FSD \) would mean that the risk manager would prefer Model 2 because the probability is higher that the bank will have to set aside less money for covering losses.

In summary, the decision rule would be as follows:
### Decision Rule

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>Y FSD X</th>
<th>If do not reject $H_0$: Y dominates X</th>
<th>DCC of Model 1 is likely to be higher than of Model 2</th>
<th>Risk manager prefers Model 2 to Model 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y=\text{DCC of Model 1}$</td>
<td>$X=\text{DCC of Model 2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graphically, FSD exists when the cumulative distribution functions do not intersect; if they do cross, then the FSD is statistically indeterminate. This is the case shown in Figure 5, where the right box includes CDFs for models 1 and 2 from the previous example. As the two CDFs cross, first order stochastic dominance cannot be established. From this figure, a lot of information in the probability distribution would be lost examining only the first or second moments. For example, for low values of DCC, model 2 provides a higher probability of having higher values of DCC, while for high values of DCC, a higher probability would be more likely under model 1.

We can test for second order stochastic dominance (SSD) in order to account for risk aversion, over increasing and concave utility functions. Model 2 SSD model 1 when the area under the cumulative distribution for model 2 is less than the corresponding area for model 1, at every point on the support. In Figure 6, model 2 SDD model 1, with the graph on the right showing the difference in areas, which is always greater than zero over the entire range of possible values of DCC. This means that model 2 yields higher DCC for all but one set of circumstances. If this case is not too severe, SSD indicates that model 1 is better than model 2. This testing strategy may also be extended to consider cases where a limited part of the support is of interest.

6.- Results

Table 3A describes the different alternatives analysed in the paper. Tables 3B and 3C present rejection rates from three different tests, namely Donald and Hsu (2013) (BB), Barret and Donald (BD) (2003), and Linton, Maasoumi and Whang (2005) (LMW) for the null hypothesis: $H_0$: Y SD X. For example, in alternative 1, failing to reject $H_0$ would imply that DCC produced by GARCH with a Student-t distribution (Y) FSD the DCC produced by GARCH with a Gaussian distribution (X). Following Donald and Hsu (2013), when implementing the blockwise bootstrap, the block sizes are set to 12 and 24. The subsample size is set at 25. The p-values for the blockwise bootstraps method are approximated based on
200 replications, and the p-values for the subsampling method are approximated based on 176 possible subsamples. The significance level is set at 5%.

[Insert Tables 3A, 3B and 3C here]

(1) For alternatives A, B and C in Table 3B, the BB and BD tests clearly show that the DCC values, assuming the Student-t distribution, first order dominate Gaussian distributions for all three GARCH models. As these results imply a higher likelihood of higher DCC under the Student-t distribution, they are preferred to the Gaussian distribution. Figure 7 shows the Cumulative Distribution Function (CDF) and the integrated CDF (ICDF) for alternative A. Regarding CDF in the left panel, dashed line, DCC produced by GARCH for the Student-t distribution lies ahead of the solid line, DCC produced by GARCH assuming the Gaussian distribution. The right panel in Figure 7 shows that DCC produced by GARCH assuming the Student-t distribution second order dominates DCC produced by GARCH assuming the Gaussian distribution, as the dashed line is always below the solid one. This means that GARCH assuming the Gaussian distribution would be preferred by a financial risk manager.

(2) Assuming the Gaussian distribution, alternatives D, E and F, the three tests conclude that neither EGARCH nor GJR FSD GARCH, and GJR does not FSD EGARCH. Figures 8-10 show the CDFs and ICDFs for the alternatives for one of the 500 daily simulations for DCC. In each case, Y CDF (dashed line) crosses X CDF (solid line), therefore we do not find FSD. Not having found FSD, we check for SSD, but again, the hypothesis is rejected in all three cases. The rejection rates for SSD are given in Table 3C. Even though we cannot establish SD of any order, it is worth examining Figure 8 to shed light on the richness of the information provided by the DCC probability distributions. For low values of DCC, GARCH provides a higher likelihood of higher DCCs (the solid line lies ahead of the dashed line), but this is reversed for high values of DCC. The right panel shows the dashed line (EGARCH) lies above the blue line (GARCH), and the difference between them (dashed minus solid) is always positive. This means that DCC produced by GARCH SSD DCC produced by EGARCH. GARCH would be preferred to EGARCH for forecasting DCC as the higher expected DCC of GARCH for low DCC values can be compensated for a lower degree of uncertainty. Similar intuition can be drawn from Figure 10 for alternative F comparing DCC probability distributions produced by EGARCH (solid) and GJR (dashed). Even though these tests cannot establish either first or second order stochastic dominance for this particular
simulation, the CDFs nevertheless cross. The left panel shows that the dashed line generally lies ahead of the solid line. However, in the right panel it seems that the higher expected DCC produced by GJR is compensated by lower uncertainty (dashed line, beneath solid line in the right panel), making this strategy more desirable. Therefore, based on Figures 8 and 10, GARCH and GJR would be preferred to EGARCH for forecasting DCC (for the Gaussian case).

(3) Assuming the Student-t distribution, EGARCH FSD GARCH (alternative G) and EGARCH SSD GJR (alternative O). Neither GJR FSD GARCH (alternative H) nor GARCH FSD GJR. Figure 11 for alternative G shows that the CDF of DCC produced by EGARCH lies ahead of the solid line, CDF of DCC produced by GARCH. This implies a higher likelihood of high DCC values under EGARCH, in which case GARCH would be preferred. The left panel in Figure 12 shows the solid CDF (GJR) underneath the blue CDF (GARCH) at low quantiles, then crosses the dashed one, stays above during some quantiles, and then returns below the dashed CDF. Accordingly, we do not find either first or second order SD. Figure 13 shows the last case, EGARCH versus GJR using the Student-t distribution (alternatives I and O). Even though the tests in Table 3B show that EGARCH FSD GJR (alternative O), the selected case\(^2\) for Figure 13 does not provide clear evidence in favour of the hypothesis.

In summary, the Gaussian distribution is preferred to Student-t for forecasting DCC. EGARCH seems to provide a higher likelihood of higher DCC when compared with GARCH and GJR, when using the Student-t distribution. Using the Gaussian distribution for forecasting DCC does not lead to either first or second order stochastic dominance. However, on the basis of the CDF and integrated CDF, it seems that the higher expected DCC values of GJR or GARCH may be compensated by lower uncertainty than for EGARCH. These results lend support to the empirical findings in Table 5 of Chang et al. (2011), where it is shown that EGARCH provides the highest average DCC for all periods in comparison with GARCH and GJR. Moreover, Chang et al. (2011) also show that GJR always provides higher DCC values than GARCH. These results notwithstanding, the SD criterion does not seem to show any dominance between these two models.

\(^2\) Individual simulation number 250 of 500 simulations was chosen for illustrative purposes.
7. Conclusions

In the spectrum of financial assets, VIX futures prices are a relatively new product. As with any financial asset, VIX futures are subject to risk. In this paper we analyzed the performance of a variety of strategies for managing the risk, through forecasting VaR, of VIX futures under the Basel II Accord.

The alternative strategies for forecasting VaR of VIX futures, and for managing financial risk under the Basel II Accord, are several univariate conditional volatility models, specifically GARCH, EGARCH and GJR, with each based on either the Gaussian and Student t distributions. The main criterion for choosing among the alternative strategies was minimizing average daily capital charges. In the paper we used a methodology based on stochastic dominance that permits partial ordering of strategies by accommodating the entire distribution of DCC values. This methodology provides a search for uniformly higher ranked volatility models, based on large classes of evaluation functions and the entire DCC distribution.

The main empirical findings of the paper are as follows:

1. The Gaussian models are generally preferred to their Student-t counterparts.
2. SD relations between DCC values produced by Gaussian models are generally not uniformly ranked. An analysis of CDFs and ICDF seems to show, however, that EGARCH provides DCC distributions with greater uncertainty, so the other models would be preferred. A lack of uniform rankings by SD also indicates that there exist special utility/evaluation functions that may provide complete, albeit subjective, rankings.
3. Within the class of Student-t distributions, EGARCH SD both GARCH and GJR, implying that EGARCH would be uniformly preferred to GARCH and GJR by a financial risk manager.
4. In general, a stochastic dominance criterion can be used to rank different models of VIX futures and distributions, as illustrated in the previous empirical results. Even in cases of no FSD and SSD, the tests provide additional information about the entire distribution over specific ranges.
5. The graphs of the CDFs of each pair of models allow a comparison globally for the whole distribution, and also locally for a given range of DCC values and probabilities. This allows
more specific comparisons than previously afforded based on the mean and other moments of the relevant distributions.

In this paper we have not found an optimal model in the sense that it outperforms the other models during the whole sample period. On the other hand, we have restricted attention to a set of widely used, though not necessarily exhaustive set of forecasting models and distributions. The paper does not pay explicit attention to the number of violations, as defined by the Basel Accords, except in the computation of the DCC. The analysis provided in the paper highlights the need for specific evaluation functions when an optimal model is to be identified.

The results of the paper suggest that further work is needed to compare, not only univariate models, but also combinations of models, such as based on the mean or median. This framework presented above should also be applied to a portfolio of assets to determine the usefulness of the stochastic dominance approach. This paper performed pairwise comparisons for a variety of models. The extension to comparisons among multivariate models is a topic for future research, as is a detailed analysis of the useful information that is contained in the CDF and ICDF.
References


Donald, S. and Y. Hsu (2013), Improving the power of tests of stochastic dominance. mimeo. [details]


Ling, S. and M. McAleer (2002b), Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models, Econometric Theory, 18, 722-729.


Figure 1
VIX and 30-day Maturity VIX Futures Closing Prices
26 March 2004 - 29 November 2013
Figure 2
30-day Maturity VIX Futures Returns
26 March 2004 - 29 November 2013
Figure 3
Histogram, Normal and Student-t Distributions
30-day Maturity VIX Futures Returns
26 March 2006 - 29 November 2013
Figure 4
Volatility of 30-day Maturity VIX Futures Returns
26 March 2004 - 29 November 2013

Figure 5. Probability density functions and CDFs of DCC Models

Note: In the left panel, the solid line is for model 1, a Gaussian density with mean 4.5 and standard deviation 1. The dashed line, for model 2, depicts the density of a Chi-square with 4 degrees of freedom. The right panel depicts the CDFs for models 1 (Gaussian) and 2 (Chi-square).
Figure 6.

DCC and ICDF for models 1 and 2 and their difference

Note: In the left panel, the solid line represents the integrated cumulative distribution function (ICDF) for a Gaussian variable with mean 4.5 and standard deviation 1. The dashed line for model 2 depicts the ICDF of a 4-degrees of freedom Chi-square variable. The right panel depicts the difference between the ICDFs given in the left panel.

Figure 7

Alternative A.

The solid line is GARCH-t and the dashed line is GARCH-Gaussian

Note: In the left panel, the solid line is the CDF of DCC produced by a GARCH model assuming a Student-t distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by a GARCH model assuming Gaussian distribution of VIX futures returns. In the right panel, solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown on the left panel.
Figure 8
Alternative D

The solid line is GARCH-Gaussian and the dashed line is EGARCH-Gaussian

Note: In the left panel, the solid line is the CDF of DCC produced by a GARCH model assuming Gaussian distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by an EGARCH model assuming Gaussian distribution of VIX futures returns. In the right panel, the two lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown on the left panel.
Figure 9
Alternative E
The solid line is GARCH-Gaussian and the dashed line is GJR-Gaussian

CDF
Integrated CDF

Note: In the left panel, the solid line is the CDF of DCC produced by a GARCH model assuming a Gaussian distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by a GJR model assuming a Gaussian distribution of VIX futures returns. In the right panel, the solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown in the left panel.

Figure 10
Alternative F
The solid line is EGARCH-Gaussian and the dashed line is GJR-Gaussian

CDF
Integrated CDF

Note: In the left panel, the solid line is the CDF of DCC produced by an EGARCH model assuming a Gaussian distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by a GJR model assuming a Gaussian distribution of VIX futures returns. In the right panel, solid and dashed lines depict the integrated cumulative distribution functions (ICDFs) of the CDFs shown in the left panel.
Figure 11
Alternative G
The solid line is GARCH-t and the dashed line is EGARCH-t

CDF
Integrated CDF

Note: In the left panel, the solid line is the CDF of DCC produced by a GARCH model assuming a Student-t distribution of VIX futures returns. The dashed line is the CDF of DCC produced by an EGARCH model assuming a Student-t distribution of VIX futures returns. In the right panel, solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown in the left panel.

Figure 12
Alternative H
The solid line is GARCH-t and the dashed line is GJR-t

CDF
Integrated CDF

Note: In the left panel, the solid line is the CDF of DCC produced by a GARCH model assuming a Student-t distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by a GJR model assuming a Student-t distribution of VIX futures returns. In the right panel, solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown in the left panel.
The Blue line is EGARCH-t and the green line is GJR-t

CDF

Integrated CDF

Note: In the left panel, the solid line is the CDF of a DCC produced by an EGARCH model assuming a Student-t distribution of VIX futures returns. The dashed line is the CDF of a DCC produced by a GJR model assuming a Student-t distribution of VIX futures returns. In the right panel, solid and dashed lines depict the integrated cumulative distribution functions (ICDFs) of the CDFs shown in the left panel.

<table>
<thead>
<tr>
<th>Zone</th>
<th>Number of Violations</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>0 to 4</td>
<td>0.00</td>
</tr>
<tr>
<td>Yellow</td>
<td>5</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.85</td>
</tr>
<tr>
<td>Red</td>
<td>10+</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: The number of violations is given for 250 business days. The penalty structure under the Basel II Accord is specified for the number of violations and not their magnitude, either individually or cumulatively.
Table 2

30-day Maturity VIX Futures Returns
Histogram and Descriptive Statistics
26 March 2004 - 29 November 2013

<table>
<thead>
<tr>
<th>Series: RETURNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 26/03/2004 29/11/2013</td>
</tr>
<tr>
<td>Observations 2525</td>
</tr>
<tr>
<td>Mean    -0.014896</td>
</tr>
<tr>
<td>Median   -0.273598</td>
</tr>
<tr>
<td>Maximum  25.81866</td>
</tr>
<tr>
<td>Minimum  -23.32617</td>
</tr>
<tr>
<td>Std. Dev. 3.800655</td>
</tr>
<tr>
<td>Skewness 0.697327</td>
</tr>
<tr>
<td>Kurtosis 7.350578</td>
</tr>
<tr>
<td>Jarque-Bera 2195.970</td>
</tr>
<tr>
<td>Probability 0.000000</td>
</tr>
</tbody>
</table>
Table 3A
Definition of pairs of alternative models to be tested for stochastic dominance

<table>
<thead>
<tr>
<th>Alternative</th>
<th>GARCH Gaussian</th>
<th>GARCH Student-t</th>
<th>EGARCH Gaussian</th>
<th>EGARCH Student-t</th>
<th>GJR Gaussian</th>
<th>GJR Student-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>X</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>X</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>

Note 1: The first column denotes 14 alternatives out of 30 permutations of the models considered: GARCH, EGARCH and GJR and two different distributions, Gaussian and Student-t, of VIX futures returns. FSD and SSD are tested for each alternative using three different tests: BB, BD, and LMW for the null hypothesis: \( H_0: Y \) SD X.

Note 2. For example, in alternative A, X represents the empirical distribution of DCC produced using a GARCH model assuming a Gaussian distribution of the VIX futures returns. Alternative Y represents the empirical distribution of DCC produced using a GARCH model assuming a Student-t distribution of the VIX futures returns. Failing to reject \( H_0 \) would imply that DCC produced by option Y SD option X, therefore under option Y there are higher chances of producing a higher DCC than under option Y.
### Table 3B

Rejection Rates for First-order SD Tests

<table>
<thead>
<tr>
<th>Design</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB(12)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
<td>0.50</td>
<td>0.47</td>
<td>0.02</td>
<td>0.49</td>
<td>0.68</td>
<td>0.50</td>
<td>0.51</td>
<td>0.55</td>
<td>0.87</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>BD(12)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.45</td>
<td>0.43</td>
<td>0.42</td>
<td>0.01</td>
<td>0.45</td>
<td>0.67</td>
<td>0.42</td>
<td>0.46</td>
<td>0.47</td>
<td>0.87</td>
<td>0.25</td>
<td>0.11</td>
</tr>
<tr>
<td>BB(24)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
<td>0.30</td>
<td>0.32</td>
<td>0.01</td>
<td>0.27</td>
<td>0.51</td>
<td>0.34</td>
<td>0.34</td>
<td>0.37</td>
<td>0.62</td>
<td>0.15</td>
<td>0.07</td>
</tr>
<tr>
<td>BD(24)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.28</td>
<td>0.29</td>
<td>0.00</td>
<td>0.25</td>
<td>0.50</td>
<td>0.30</td>
<td>0.31</td>
<td>0.34</td>
<td>0.62</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>LMW</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.21</td>
<td>0.26</td>
<td>0.02</td>
<td>0.25</td>
<td>0.52</td>
<td>0.28</td>
<td>0.27</td>
<td>0.28</td>
<td>0.18</td>
<td>0.13</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### Table 3C

Rejection Rates for Second-order SD Tests

<table>
<thead>
<tr>
<th>Design</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB(12)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.44</td>
<td>0.38</td>
<td>0.37</td>
<td>0.00</td>
<td>0.40</td>
<td>0.67</td>
<td>0.40</td>
<td>0.46</td>
<td>0.48</td>
<td>1.00</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>BD(12)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.41</td>
<td>0.36</td>
<td>0.34</td>
<td>0.00</td>
<td>0.39</td>
<td>0.67</td>
<td>0.37</td>
<td>0.43</td>
<td>0.45</td>
<td>1.00</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>BB(24)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
<td>0.34</td>
<td>0.31</td>
<td>0.01</td>
<td>0.30</td>
<td>0.55</td>
<td>0.33</td>
<td>0.39</td>
<td>0.39</td>
<td>0.97</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>BD(24)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
<td>0.33</td>
<td>0.29</td>
<td>0.00</td>
<td>0.30</td>
<td>0.44</td>
<td>0.32</td>
<td>0.39</td>
<td>0.38</td>
<td>0.97</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>LMW</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.29</td>
<td>0.26</td>
<td>0.30</td>
<td>0.06</td>
<td>0.24</td>
<td>0.44</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.94</td>
<td>0.09</td>
<td>0.02</td>
</tr>
</tbody>
</table>