Available online at www.sciencedirect.com
ScienceDirect
Journal of Banking & Finance 32 (2008) 2220–2236
Journal of BANKING & FINANCE
www.elsevier.com/locate/jbf

Developing a stress testing framework based on market risk models ☆

Carol Alexander, Elizabeth Sheedy *

ICMA Centre, University of Reading, P.O. Box 242, Reading RG6 6BA, UK
Macquarie Applied Finance Centre, Macquarie University, Sydney, Australia

Received 29 May 2007; accepted 31 December 2007
Available online 15 January 2008

Abstract

The Basel 2 Accord requires regulatory capital to cover stress tests, yet no coherent and objective framework for stress testing portfolios exists. We propose a new methodology for stress testing in the context of market risk models that can incorporate both volatility clustering and heavy tails. Empirical results compare the performance of eight risk models with four possible conditional and unconditional return distributions over different rolling estimation periods. When applied to major currency pairs using daily data spanning more than 20 years we find that stress test results should have little impact on current levels of foreign exchange regulatory capital.

JEL classification: G18; G19; G21

Keywords: Value-at-Risk models; Stress testing; Market risk; Exchange rates; GARCH

1. Introduction

A stress test is a risk management tool used to evaluate the potential impact on portfolio values of unlikely, although plausible, events or movements in a set of financial variables (Lopez, 2005). They are designed to explore the tails of the distribution of losses beyond the threshold (typically 99%) used in Value-at-Risk (VaR) analysis. Since the end of 1997 financial institutions using internal VaR models to assess capital adequacy have been required to implement stress testing (see Basel Committee on Banking Supervision, 1996). They provide an input to decisions concerning, amongst other things, hedging, limit setting, portfolio allocations and capital adequacy.

The Basel 2 Accord recommends a more direct link between stress tests and risk capital, i.e. 'A bank must ensure that it has sufficient capital to... *cover the results of its stress testing*' (Basel Committee on Banking Supervision, 2006, at paragraph 778 (iii), p. 218, emphasis added). As a result leading industry practitioners have called for a re-examination of stress testing methodologies, see Rowe (2005).

A recent survey of stress testing practice (Committee on the Global Financial System, 2005) shows that most stress tests are currently designed around a series of scenarios based either on historical events, hypothetical events, or some combination of the two. These methods have been criticised by Berkowitz (2000) and Greenspan (2000) for their lack of rigour. They are typically conducted without a risk model so the probability of each scenario is unknown, making its importance difficult to evaluate. There is also a distinct possibility that many extreme yet plausible scenarios are not even considered.

Stress tests conducted in the context of a risk model can provide a useful alternative or complement to the current ad hoc methods of stress testing. Several authors have attempted to build such a bridge between stress tests and risk models including Kupiec (1998) who examines cross-market effects resulting from a market shock and Aragones

☆ This paper was accepted by Prof. Giorgio Szego while he was the Managing Editor of The Journal of Banking and Finance, and by the past Editorial Board.
* Corresponding author. Tel.: +61 2 9850 7755; fax: +61 2 9850 7281.
E-mail addresses: c.alexander@icmacentre.rdg.ac.uk (C. Alexander), esheedy@mafc.mq.edu.au (E. Sheedy).

0378-4266/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2007.12.041

et al. (2001) who incorporate hypothetical stress events into an Extreme Value Theory (EVT) framework.

Our research begins by asking a more fundamental question: What is the most suitable risk model in which to conduct a stress test? If the model is misspecified, then our approach is vulnerable to a considerable degree of model risk. Hence a significant part of this paper performs extremely thorough backtests, which are designed to reduce the model risk in risk models that are used for stress testing.

The risk model used for stress testing need not necessarily have the same features as that used for daily VaR models; indeed it could be argued that there are advantages in using different models for cross-checking purposes. The empirical performance of VaR models used at commercial banks has been analysed by Berkowitz and O'Brien (2002) and Berkowitz et al. (2006). These studies suggest that banks' VaR models may be misspecified since they do not fluctuate much while actual volatility of trading revenues is clearly time-varying. This is almost certainly because many banks use simple unconditional models to estimate VaR. The possibility that bank VaR models are misspecified creates further incentive to ensure that an appropriate model is selected for stress testing purposes.

We therefore backtest eight risk models including both conditional and unconditional models and four possible return distributions. We include in our analysis risk models that are already popular in the industry and those that are relatively accessible to financial institutions. Our backtesting methodology is designed with stress testing applications in mind, that is, we assess the ability of risk models to forecast extreme percentiles (up to 99.9%) of returns distributions over relatively short multi-day horizons. We also use assessment criteria based on expected tail loss (ETL) as well as Value-at-Risk (VaR) to ensure that the model performs well for extreme outcomes beyond VaR. Our findings support the use of conditional risk models with non-normal innovations such as the empirical distribution and Student's *t* distribution.

We choose major exchange rates as the context for analysis due to their importance in the portfolios of financial institutions in many countries, and because of the availability of high quality data over a lengthy historical period. We investigate daily returns for three of the most traded currency pairs[1]: the USD/JPY, GBP/USD and the AUD/USD. The importance of these three currencies for the risk of financial institutions is highlighted in a recent paper on carry trade activity (see Galati et al., 2007).

Having identified our preferred risk models, we then develop and illustrate a model-based stress testing methodology. The methodology includes specification of an initial shock event and analyses the subsequent market response to that shock using simulation. Alternatively analysts can specify hypothetical shock events and use the risk models only to assess its after effects as volatility increases in response to the shock. The methodology can readily be extended to handle multiple assets/risk factors and to incorporate changing liquidity conditions. Market participants may also use this framework to assess their own response to a market crisis, i.e. immediate versus gradual hedging. The methodology is evaluated by comparing it to current stress testing techniques and regulatory requirements.

The outline of the paper is as follows: Section 2 explains the risk models that will be examined in this study, and the reasons for their selection. Section 3 presents an empirical analysis of these models in currency markets. Section 4 takes the best-performing models and shows how they may be adapted for stress testing purposes. Section 5 evaluates the stress tests and Section 6 summarises and concludes.

2. Risk models

What guidance can the market risk capital assessment literature provide regarding the selection of appropriate risk models for stress testing? Previous research, which typically explores the 99th percentile of outcomes at 1-day horizons, suggests that accurate risk models will capture two key characteristics: volatility clustering and heavy tails.[2] But stress testing requires exploration of more extreme outcomes and, since immediate hedging may not be practical in a market crisis, longer horizons. Danielsson and De Vries (2000) make the point that the most extreme market moves tend to exhibit reduced dependency between successive daily returns. This suggests that unconditional risk models may be sufficient for our application, provided they have sufficiently heavy tails. We therefore examine a range of unconditional risk modelling approaches, each with its conditional counterpart, to give a total of eight univariate risk models.

Berkowitz (2000) has argued that the distribution of an asset or risk factor during periods of market stress is very different from its usual distribution. Therefore, we have included a two-component normal mixture distribution as a candidate return distribution; here the two component distributions can be interpreted as corresponding to 'nor-

[1] According to Bank for International Settlements Bank for International (2005); Triennial Central Bank Survey: Foreign exchange and derivatives market activity in 2004 (Bank for International Settlements, Basel) these were the three most actively traded currency pairs after the Euro/USD in 2004, and the Euro/USD could not be included in the analysis due to lack of historical data.

[2] The RiskMetrics Group popularised the Exponentially Weighted Moving Average method for estimating volatility. The heavy tails apparent in financial return distributions led to the popularity of VaR methods based on historical simulation, which employs an empirical distribution (see Dowd and Rowe, 2004; Introduction to Value-at-Risk models. In: Carol Alexander, Elizabeth Sheedy (Eds.), The Professional Risk Manager's Handbook (PRMIA Publications)). More recently researchers have explored conditional non-normal innovation distributions for VaR modelling, finding them superior to their conditional normal counterparts. See discussion later in Section 2.

mal' and 'stress' market conditions. We compare the normal mixture distribution with two more standard and accessible approaches designed to capture extreme outcomes: the Student's t distribution and a smoothed empirical distribution.

An obvious variation on the latter would be an application of the EVT (Extreme Value Theory – see Longin, 2000; Aragones et al., 2001; Longin, 2005). We choose not to include EVT in the backtesting experiment that follows since excellent results are achieved using simpler, more accessible models. Previous research applying EVT to risk modelling has typically focused on a single period horizon and has assumed independence in returns. Applying EVT in a conditional, multi-day setting is far more demanding and likely to be inaccessible to many financial institutions (although we note that McNeil and Frey (2000) have successfully implemented this approach).

For each model we estimate Value-at-Risk (VaR) and expected tail loss (ETL), i.e. the expected loss conditional on exceeding VaR. We now describe each of the risk models, outlining their merits and explaining the calculation of VaR and ETL for each as a percentage of the current portfolio value.

2.1. Unconditional normal

This approach is included for the purpose of benchmarking the candidate risk models. Here we assume that $\varepsilon_t = y_t - \bar{y}$, where y_t, the daily log returns, are independent and normally distributed:

$$\varepsilon_t \sim N(0, \sigma^2). \tag{1}$$

Under this assumption the VaR and ETL for an horizon of h days at significance level α are:

$$\mathrm{VaR}_{h,\alpha} = \Phi^{-1}(\alpha)\sigma_h \tag{2}$$

and[3]:

$$\mathrm{ETL}_{h,\alpha} = \alpha^{-1}\varphi(\Phi^{-1}(\alpha))\sigma_h, \tag{3}$$

where Φ denotes the standard normal cumulative distribution function, φ denote the standard normal density function and $\sigma_h = \sigma\sqrt{h}$.

2.2. Unconditional empirical

This method is chosen by virtue of its popularity in the industry. Indeed, a recent survey found that of the 64.9% of firms that disclose their methodology 73% report the use of historical simulation (Perignon and Smith, 2006). It makes no assumption about the distribution of past returns, other than the assumption that returns are independent and identically distributed (*i.i.d*).[4] If the empirical return distribution has heavy tails then VaR measured at high confidence levels will be greater than under the assumption of normality. There is no assumption that the distribution of returns is symmetric so VaR and ETL will differ according to whether the portfolio is long or short the asset (in our case the commodity currency expressed in units of the terms currency). For the long asset case, the VaR_α estimate is the absolute value of the lower α percentile of the empirical return distribution and ETL_α is the absolute value of the average of all empirical return outcomes below VaR. Risk measures for the short asset case can be calculated using the same method, but first multiplying the return series by -1.

The failings of historical simulation have been well described by Pritsker (2006). Of particular importance to this research is the fact that VaR estimates are sample specific, particularly for high confidence levels. Following Sheather and Marron (1990), Butler and Schachter (1998) and Chen and Tang (2005) we therefore smooth the empirical density by fitting the Epanechnikov kernel to a sample of daily returns. Kernel estimation is designed to fit a smooth curve to a random sample such that it provides the best possible representation of the density of a random variable. This method is explained in Dowd (2002) and involves selecting a kernel (that is, a density function centred on the data point), and a parameter called the bandwidth that is analogous to the bin-width in a histogram. Silverman (1986) describes a range of possible kernels (e.g. Epanechnikov, Gaussian, triangular) but finds that the choice between them is typically not critical. The software used for analysis in this study is Matlab, where kernel estimation is a standard function in the Statistics toolbox. The software automatically selects the optimal bandwidth to minimise the squared difference between the empirical density and the fitted density.

The daily VaR estimate for a long portfolio is the absolute value of the return at the lower α percentile of the kernel density. To calculate ETL we take an average of tail VaRs as recommended by Dowd (2002, p. 45). To calculate VaR and the ETL over an h-day horizon we multiply the daily VaR or ETL by the square root of h, consistent with common industry practice. In addition we have performed an empirical analysis to justify the choice of square-root scaling, described in Appendix A.[5]

2.3. Unconditional Student's t

Huisman et al. (1998) were among the first to specifically accommodate heavy tails in VaR estimation with the Student's t distribution. We assume the mean-adjusted returns ε_t are distributed as follows:

$$\varepsilon_t\sigma(v/(v-2))^{1/2} \sim T_v, \tag{4}$$

[3] See McNeil et al., 2005. Quantitative Risk Management (Princeton University Press) at p. 45.

[4] A referee has pointed out that historical simulation is effective under mild dependence of returns (such as GARCH).

[5] Scaling with the square root of time rule is still only an approximation. A better, though very computationally intensive method, is to simulate paths of length h and determine h-day VaR directly.

where T_v is the standardised Student's t distribution with v degrees of freedom, zero mean and unit variance and σ is the standard deviation of the mean-adjusted returns. To estimate the α% h-day VaR we follow the method described by Dowd (2002):

$$\mathrm{VaR}_{h,\alpha} = T_v^{-1}(\alpha)((v-2)/v)^{1/2}\sigma_h, \tag{5}$$

where the parameter v is estimated using the method of moments. Appendix B derives the ETL:

$$\mathrm{ETL} = \alpha^{-1}(v-1)^{-1}(v-2+T_v^{-1}(\alpha)^2)t_v(T_v^{-1}(\alpha))\sigma_h,$$

where t_v is the standardised Student's t density function.

2.4. Unconditional normal mixture

A finite normal mixture distribution function is a probability weighted sum of a finite number of normal distribution functions. Here we use two normal distributions: the first reflects 'typical' market conditions while the second, having higher volatility and lower probability, reflects stressed market conditions. Depending on the means and variances of the individual normal densities the mixture distribution can capture the positive excess kurtosis and negative skew often observed in financial markets. These characteristics suggest that a normal mixture distribution is well-suited to stress testing. Moreover, it allows us to perform stress tests using only the distribution that governs stressed market conditions.

A mixture of two normal densities has distribution:

$$G(\varepsilon) = \pi F_1(\varepsilon) + (1-\pi)F_2(\varepsilon), \tag{6}$$

where $F_i(\varepsilon)$ is a normal distribution function with mean μ_i and standard deviation σ_i for $i = 1$ and 2, and $0 < \pi < 1$. Note that while we have defined the overall mixture distribution to have zero mean, the two component distributions can have different, non-zero means. It is this feature that creates the possibility of skewness in the mixture distribution and therefore risk measures can differ depending on whether one is long or short the currency.

We estimate the parameters from historical data using the EM algorithm. Then to estimate the 1-day VaR (expressed as a percentage of portfolio value) we apply a numerical algorithm to solve the following expression[6]:

$$\pi \mathrm{Prob}(Z < [\mathrm{VaR}_\alpha - \mu_1]/\sigma_1) + (1-\pi)\mathrm{Prob}(Z < [\mathrm{VaR}_\alpha - \mu_2]/\sigma_2) = \alpha,$$

where $Z \sim N(0,1)$. Appendix B derives the ETL as

$$\mathrm{ETL}_\alpha = \alpha^{-1}[\pi\varphi(\sigma_1^{-1}G_0^{-1}(\alpha))\sigma_1 + (1-\pi)\varphi(\sigma_2^{-1}G_0^{-1}(\alpha))\sigma_2] - [\pi\mu_1 + (1-\pi)\mu_2], \tag{8}$$

where φ is the standard normal density function and G_0 is the distribution function for a mixture of two zero-mean normal distributions with standard deviations σ_1 and σ_2. Both VaR and ETL are scaled using the square root of time rule for the case of horizons beyond 1 day.

2.5. Conditional normal

Here the mean adjusted returns are assumed to be conditionally normally distributed with conditional variance following the symmetric GARCH(1,1) process of Bollerslev (1986)[7]:

$$\sigma_t^2 = \gamma_1 + \gamma_2\varepsilon_{t-1}^2 + \gamma_3\sigma_{t-1}^2 \quad \gamma_1 \geqslant 0, \quad \gamma_2, \gamma_3 > 0, \quad \gamma_2 + \gamma_3 < 1. \tag{9}$$

To calculate VaR and ETL we simulate the process (9) over an h-day horizon with 30,000 paths.[8] For a long (short) position the α% h-day VaR is the absolute portfolio return calculated at the lower (upper) α percentile and the corresponding ETL is the absolute value of the mean of all simulated returns below (above) the VaR. This model, along with the other conditional models described below, will capture the volatility clustering that is potentially significant in a stressful market. That is, a large price disturbance will lead to higher volatility and hence to further large price changes.

2.6. Conditional empirical

This method is based on the filtered historical simulation model of Barone-Adesi et al. (1998). First we fit a normal GARCH process (9) to historical data then we standardise each return in the sample by dividing by the conditional volatility estimate corresponding to that return. Then we scale all the returns in the sample to the conditional volatility applying on the day that the VaR and ETL are estimated. That is, we multiply the standardised returns by the current conditional volatility estimate. We fit the Epanechnikov kernel to the scaled returns to provide a smoothed distribution of standardised returns and then simulate the GARCH model forward over the h-day risk horizon using innovations that are sampled from the kernel density. The α% h-day VaR estimate for a long (short) posi-

[6] For portfolios with a short asset position the same method can be used, but the portfolio returns should first be multiplied by −1.

[7] Use of an asymmetric GARCH model was considered but discarded for two reasons. First, the present study focuses on currency returns where symmetric GARCH models are typical (see Vilasuso, 2002. Forecasting exchange rate volatility. Economic Letters 76, 59–64 and So and Yu, 2006. Empirical analysis of GARCH models in Value-at-Risk estimation. International Financial Markets, Institutions & Money 16, 180–197). Second, we tested the data for asymmetry in conditional volatility and found no consistent evidence to support its presence. In other markets an asymmetric GARCH specification may be preferred.

[8] We cannot simply plug in the GARCH forecast of the h day variance into the VaR model because this approach assumes that all the innovations during the risk horizon are 'typical', i.e. that the square of each the innovation in the stress test is equal to its expected value. This assumes no major shocks which could precipitate an increase in volatility under the assumptions of the model. Only simulation can produce large price discontinuities and assess their effect on VaR. With many simulations a few large innovations will occur, even assuming conditional normality.

tion is the absolute portfolio return calculated at the lower (upper) α percentile and the corresponding ETL is the absolute mean of all simulated returns below (above) the VaR.

2.7. Conditional Student's t

Several heavy-tailed distributions have been combined with GARCH models in the VaR estimation literature. Venter and de Jongh (2002), for example, use the normal inverse Gaussian distribution. The conditional Student's t model has arguably been investigated by the greatest number of researchers and is a relatively standard option in statistical packages. Mittnik and Paolella (2000) estimate VaR for East-Asian currencies with an asymmetric generalised t distribution in combination with an APARCH model. Giot and Laurent (2003) apply a similar approach in equity markets and Angelidis et al. (2004) apply the simpler t-EGARCH model for estimating VaR in equity markets. In an analysis of major currencies, So and Yu (2006) find that the t-GARCH method performs well in VaR estimation for major currency returns.

The method used here is identical to that for Section 2.5, but now the innovations are drawn from a Student's t distribution with v degrees of freedom, in order to capture the conditional excess kurtosis present in empirical data.

2.8. Conditional normal mixture

We start by fitting a standard GARCH(1,1) model with normal innovations to historical returns, and then standardise the historical returns as in Section 2.6. Then using the EM algorithm we fit a normal mixture distribution to the standardised returns having two component densities. To calculate VaR at the h-day horizon we simulate (9) over an h-day horizon using innovations drawn from the fitted normal mixture distribution. The VaR for a long (or short) position is the absolute portfolio return calculated at the α (or $1-\alpha$) percentile and the corresponding ETL is the absolute mean of all simulated returns below (above) the VaR.

3. Model selection

In this section we backtest our risk models using the Kupiec (1995) test for coverage, the Christoffersen (1998) test for conditional coverage and a method for backtesting ETL due to McNeil and Frey (2000). Test procedures are explained in Appendix C. The empirical analysis here differs from previous empirical research in several respects, namely:

- Most previous research has analysed only VaR at the 99.0% confidence level. To better reflect the stress testing context of this research we consider confidence levels for VaR and ETL of 99.0%, 99.5% and 99.9%. The accuracy of our risk measures is assessed using over 15,000 out-of-sample returns.
- To account for market liquidity constraints we consider a 3-day risk horizon in addition to the more typical 1-day horizon.
- The choice of estimation window is an important source of model risk.[9] The Basel regulations insist on an estimation window of at least 1 year (equivalent to around 250 trading days). Thus we consider a range of possible estimation windows (250, 500, 1000 and 2000 days) for estimating parameters, volatility and percentiles.

Hence we conduct a rigorous set of backtests that are specifically designed to reduce the model risk in the risk models that are used for stress testing. This is an interesting area that warrants further empirical research. The present paper presents results for currency markets and forthcoming research by the authors examines equity index markets.

We evaluate the risk models in the context of daily returns on the Australian dollar in terms of US dollars (AUD/USD), the Great Britain Pound in terms of US dollars (GBP/USD) and the US dollar in terms of Japanese Yen (USD/JPY).[10] We have 5691 observations on AUD/GBP dating from the float of the AUD in December 1983 until June 2006 and 8154 observations from January 1974 until June 2006 for the USD/JPY and GBP/USD. For each series we repeatedly estimate both VaR and ETL for each of the eight risk models. The maximum sample size for estimation of parameters is 2000 days, so the first estimation occurs 2000 days (nearly 8 years) into the data sample. In the case of the USD/JPY and the GBP/USD this allows us to calculate over 6000 estimates of VaR and ETL having a 1-day horizon, or around 2000 non-overlapping estimates of VaR and ETL having a 3-day horizon. Each time VaR and ETL are estimated they are based on revised parameter estimates (or percentiles) using the most recent estimation window of 250, 500, 1000 or 2000 days. At the end of each risk horizon we calculate actual profit and loss for the trading portfolio. An exceedance occurs if the loss is greater than the estimated VaR for that risk horizon. These exceedances are the inputs to the tests of coverage, conditional coverage and ETL referred to above. Tables 1–4 present results for each of the eight risk models using a sample period of 2000 days.

Clearly the assumption of normality of returns cannot be justified, whether in the conditional or unconditional context. Column (i.c.) of Table 1 shows that in almost every case we can comfortably reject the hypothesis that the actual number of exceptions is equal to the expected number of exceptions. The unconditional normal model

[9] See the discussion of sample period in relation to historical simulation in Section 9.2 of Hull (2006). Risk Management and Financial Institutions (Pearson Prentice Hall). The issue of how many data points are needed to estimate the parameters of GARCH reliably is discussed in Straumann (2004) in Section 5.10.

[10] Data were obtained from (www.rba.gov.au/Statistics/HistoricalExchangeRates/index.html) and (www.federalreserve.gov/releases/h10/Hist/default.htm).

Table 1
Backtest with normal innovations (Estimation window = 2000 days[A])

Currency pair	$1-\alpha$	(i) Unconditional normal			(ii) Conditional normal		
(a)	(b)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)
1-day horizon							
Long AUD/Short USD (%)	99.0	0.00	0.02	0.00	0.00	0.02	0.00
	99.5	0.00	0.00	0.00	0.00	0.02	0.00
	99.9	0.00	0.02	0.00	0.00	0.02	0.00
Short AUD/Long USD (%)	99.0	1.24	4.24	0.00	2.74	8.30	0.50
	99.5	0.00	0.02	0.00	0.42	1.26	5.60
	99.9	0.00	0.00	0.00	5.23	14.95	6.00
Long GBP/Short USD (%)	99.0	0.00	0.00	0.00	0.00	0.02	0.00
	99.5	0.00	0.00	0.00	0.00	0.02	0.00
	99.9	0.00	0.00	0.50	0.00	0.02	2.20
Short GBP/Long USD (%)	99.0	9.52	6.30	0.00	7.37	20.16	0.10
	99.5	5.44	1.67	0.00	5.44	11.78	0.00
	99.9	0.00	0.00	0.10	0.10	0.43	0.00
Long USD/Short JPY (%)	99.0	0.00	0.00	0.00	0.00	0.00	0.00
	99.5	0.00	0.00	0.00	0.00	0.00	0.00
	99.9	0.00	0.00	0.00	0.00	0.00	0.00
Short USD/Long JPY (%)	99.0	12.15	17.81	0.00	23.63	48.65	0.00
	99.5	0.24	0.18	0.00	0.40	1.10	0.20
	99.9	0.00	0.00	0.00	0.00	91.77	0.10
3-day horizon							
Long AUD/Short USD (%)	99.0	0.06	0.08	0.30	2.35	5.34	2.00
Short AUD/Long USD (%)	99.0	63.36	75.96	4.50	9.84	24.52	98.10
Long GBP/Short USD (%)	99.0	0.00	0.00	0.00	0.00	0.02	17.90
Short GBP/Long USD (%)	99.0	74.38	46.75	0.60	16.96	27.15	0.90
Long USD/Short JPY (%)	99.0	0.00	0.00	0.00	0.00	0.00	0.00
Short USD/Long JPY (%)	99.0	90.95	81.58	0.70	90.95	81.59	5.60

(a) The trading portfolio used for the analysis. (b) The confidence level at which VaR and ETL are calculated. (c) Test of the null hypothesis that the actual number of violations is equal to the expected number of violations. (d) Test of the null hypothesis that violations are spread evenly over time (as opposed to being clustered) and (c) above. (e) Test of the null hypothesis that the ETL does not consistently understate the true potential for losses beyond the VaR.
[A] Estimation window refers to the number of observations used to estimate VaR/ETL. The sample is rolled over daily, keeping the sample size constant, to generate 15 years of out-of-sample VaR forecasts over both 1-day and 3-days horizon. The backtests are based on the 'violations', i.e. the returns that exceeded the prediction of VaR.

performs no better in a test of conditional coverage (column (i.d.)), and performance in measuring ETL is poor. Column (i.e.) shows that in every case we can reject the null hypothesis (with 95% confidence) that ETL does not consistently understate the true potential for losses beyond the VaR. Turning to the conditional normal risk model presented in panel (ii) of Table 1 we observe that while modelling heteroscedasticity improves performance, especially in terms of conditional coverage, the results do not generally support the use of this risk model.

The risk models based on the empirical distribution are much more suited to stress testing, especially in the conditional case. Results for the unconditional empirical model are presented in panel (i) of Table 2. In the majority of cases considered we cannot reject the hypotheses tested at 95% confidence. We note, however, that in some cases there is a tendency for exceptions to be clustered, and the hypothesis with regard to conditional coverage is rejected (at 95%) in 8 out of 24 cases. The conditional empirical risk model in panel (ii) of Table 2 eliminates this problem. Use of the Student's t distribution can also be justified for stress testing purposes, especially in the conditional t case. Table 3 presents the unconditional t results in panel (i), with reasonable results for most portfolios, the obvious exceptions being long AUD/USD and long USD/JPY. This can be explained by the negative skewness observed in both currency pairs which potentially could be addressed with the use of an asymmetric t distribution. Column (i.d.) also provides some evidence that this risk model is flawed by clustering of exceedances, even in the conditional case. However, the ETL test results are in fact the best of all the risk models considered in this study. In only 1 out of 24 cases can we reject the ETL measure at the 95% confidence level. Even at very low levels of α the measure of ETL under this risk measure conservatively estimates the potential for losses beyond the VaR. Indeed, it could be argued that the conditional t risk model is too conservative in some cases since no exceedances are recorded for two of the portfolios.

In order to conserve space we present abbreviated results for the normal mixture distribution in Table 4. As explained earlier, the normal mixture distribution is intuitively appealing for modelling extreme risk as it allows

Table 2
Backtest with empirical innovations (Estimation window = 2000 days[A])

Currency pair	$1-\alpha$	(i) Unconditional empirical			(ii) Conditional empirical		
(a)	(b)	Unconditional coverage VaR *p*-value (c)	Conditional coverage VaR *p*-value (d)	ETL *p*-value (e)	Unconditional coverage VaR *p*-value (c)	Conditional coverage VaR *p*-value (d)	ETL *p*-value (e)
1-day horizon							
Long AUD/Short USD (%)	99.0	40.90	56.93	14.50	98.68	68.73	93.00
	99.5	56.04	74.84	5.80	27.79	52.63	58.70
	99.9	27.03	53.94	7.40	9.60	25.02	0.00
Short AUD/Long USD (%)	99.0	50.51	62.43	23.70	51.12	59.83	52.10
	99.5	42.14	21.86	38.90	91.60	91.04	74.60
	99.9	5.23	0.62	70.30	71.02	93.11	62.20
Long GBP/Short USD (%)	99.0	95.20	8.83	96.00	9.60	16.88	72.20
	99.5	74.74	82.76	94.80	9.50	22.93	48.80
	99.9	63.06	88.72	12.60	95.06	99.23	35.60
Short GBP/Long USD (%)	99.0	1.80	0.76	73.10	55.67	49.37	86.80
	99.5	6.11	3.01	60.40	28.14	50.55	72.40
	99.9	15.75	36.77	3.90	63.06	88.72	89.70
Long USD/Short JPY (%)	99.0	57.14	0.26	15.80	84.39	54.32	10.40
	99.5	45.39	0.29	22.20	68.97	77.29	13.90
	99.9	15.51	1.43	16.70	73.82	93.82	2.10
Short USD/Long JPY (%)	99.0	20.97	9.98	54..00	94.58	89.35	68.00
	99.5	61.18	26.08	51.60	48.70	21.83	18.40
	99.9	95.06	99.23	28.10	47.66	76.81	21.50
3-day horizon							
Long AUD/Short USD (%)	99.0	63.36	75.96	91.20	84.25	85.32	94.50
Short AUD/Long USD (%)	99.0	9.84	24.52	71.50	9.84	24.52	99.00
Long GBP/Short USD (%)	99.0	4.88	3.03	82.20	45.07	42.44	61.20
Short GBP/Long USD (%)	99.0	4.06	11.45	4.60	58.78	66.51	35.60
Long USD/Short JPY (%)	99.0	33.54	0.72	19.40	24.21	36.13	0.20
Short USD/Long JPY (%)	99.0	12.54	28.07	38.30	73.44	79.04	13.50

(a) The trading portfolio used for the analysis. (b) The confidence level at which VaR and ETL are calculated. (c) Test of the null hypothesis that the actual number of violations is equal to the expected number of violations. (d) Joint test of the null hypothesis that violations are spread evenly over time (as opposed to being clustered). (e) Test of the null hypothesis that the ETL does not consistently understate the true potential for losses beyond the VaR.
[A] Estimation window refers to the number of observations used to estimate VaR/ETL. The sample is rolled over daily, keeping the sample size constant, to generate 15 years of out-of-sample VaR forecasts over both 1-day and 3-days horizon. The backtests are based on the 'violations', i.e. the returns that exceeded the prediction of VaR.

for a different distribution of returns in stressful market conditions. Yet our backtest results for the normal mixture model are disappointing. This may be specific to currency markets and regime specific behaviour may be more evident in equity markets. The analysis suggests that the distribution of currency returns during stressful periods (as well as during typical periods) can be adequately described by a single heavy-tailed distribution in combination with a GARCH model. In other words, occasional large shocks are observed from time to time, which are then followed by further large price movements consistent with volatility clustering.

Finally, our results confirm that large estimation windows (say 2000 days) are preferable to smaller estimation windows (say 250 days) for risk estimation. This is especially true for the four conditional models and also for the unconditional empirical model. For the remaining models the sample size appears less important. Space does not permit presentation of results for all estimation windows but we compare estimation windows of 250 and 2000 days for the conditional t in Table 5.

4. Stress testing methodology

Section 3 highlighted the importance of volatility clustering and heavy tails in modelling extreme risk. Our stress testing methodology therefore incorporates these elements: we first specify the initial shock event based on a heavy-tailed distribution (Section 4.1), and then specify the subsequent response to that shock over several days as volatility increases (Section 4.2). Alternatively, a hypothetical shock event could be selected and the model used only for analysing its consequences, but we do not consider such an approach here. Extension of the methodology to the multivariate case and to more formally address liquidity risk is discussed in Sections 4.3 and 4.4, respectively.

4.1. Initial shock event

We start by considering the initial shock event, i.e. a large gap or discontinuity in prices typically caused by important unanticipated information entering the market. The traditional approach described in Committee on the

Table 3
Backtest with Student's t innovations (Estimation window = 2000 days[A])

Currency pair	$1-\alpha$	(i) Unconditional t			(ii) Conditional t		
(a)	(b)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)
1-day horizon							
Long AUD/Short USD (%)	99.0	0.03	0.04	9.70	14.71	11.44	69.30
	99.5	1.35	2.41	6.20	30.66	51.33	92.20
	99.9	0.06	0.27	87.70	33.46	62.71	73.00
Short AUD/Long USD (%)	99.0	61.28	66.68	10.80	17.45	31.61	99.70
	99.5	21.58	16.37	32.00	0.60	2.25	88.40
	99.9	27.03	1.20	68.30	NA	NA	NA
Long GBP/Short USD (%)	99.0	15.34	1.58	95.10	74.41	53.55	96.10
	99.5	20.73	35.65	99.70	88.96	85.50	98.70
	99.9	35.33	64.83	87.10	15.75	36.77	86.60
Short GBP/Long USD (%)	99.0	5.21	2.38	24.40	7.14	13.50	97.30
	99.5	20.34	9.97	24.10	2.21	6.87	84.30
	99.9	63.06	88.72	37.30	5.08	14.84	76.20
Long USD/Short JPY (%)	99.0	0.10	0.00	0.10	7.37	4.59	23.30
	99.5	0.00	0.00	10.10	5.44	0.00	36.10
	99.9	0.10	0.05	56.00	7.87	20.90	54.70
Short USD/Long JPY (%)	99.0	7.14	3.32	56.60	0.51	1.10	100.00
	99.5	37.62	17.54	86.80	1.23	4.13	100.00
	99.9	35.33	64.83	75.50	NA	NA	NA
3-day horizon							
Long AUD/Short USD (%)	99.0	7.54	11.92	58.10	12.65	23.81	99.40
Short AUD/Long USD (%)	99.0	49.57	73.04	91.80	4.50	13.02	99.80
Long GBP/Short USD (%)	99.0	0.10	0.18	87.30	7.61	15.16	89.30
Short GBP/Long USD (%)	99.0	7.40	18.66	3.80	74.38	74.67	59.50
Long USD/Short JPY (%)	99.0	0.62	0.02	2.40	0.62	0.74	3.90
Short USD/Long JPY (%)	99.0	7.40	18.66	48.80	19.91	39.26	63.00

(a) The trading portfolio used for the analysis. (b) The confidence level at which VaR and ETL are calculated. (c) Test of the null hypothesis that the actual number of violations is equal to the expected number of violations. (d) Joint test of the null hypothesis that violations are spread evenly over time (as opposed to being clustered) and (c) above. (e) Test of the null hypothesis that the ETL does not consistently understate the true potential for losses beyond the VaR.

[A] Estimation window refers to the number of observations used to estimate VaR/ETL. The sample is rolled over daily, keeping the sample size constant, to generate 15 years of out-of-sample VaR forecasts over both 1-day and 3-days horizon. The backtests are based on the 'violations', i.e. the returns that exceeded the prediction of VaR. NA: No violations.

Table 4
Backtest with normal mixture innovations[A] (Estimation window = 2000 days)

Currency pair	$1-\alpha$	(i) Unconditional normal mixture			(ii) Conditional normal mixture		
(a)	(b)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)
1-day horizon							
Long AUD/Short USD (%)	99.0	2.74	3.85	0.00	0.03	0.10	0.00
	99.5	0.12	0.32	0.00	0.00	0.01	0.00
	99.9	0.00	0.02	2.10	0.00	0.02	0.00
Long GBP/Short USD (%)	99.0	4.27	3.25	1.40	0.00	0.02	0.00
	99.5	1.61	3.97	13.00	0.00	0.01	0.00
	99.9	15.50	35.79	3.30	0.00	0.02	0.40
Long USD/Short JPY (%)	99.0	2.37	0.00	0.00	0.00	0.00	0.00
	99.5	0.14	0.00	0.00	0.00	0.02	0.00
	99.9	0.00	0.00	0.30	0.00	0.02	0.00
3-day horizon							
Long AUD/Short USD (%)	99.0	20.29	35.02	5.30	2.35	5.34	5.50
Long GBP/Short USD (%)	99.0	0.05	0.10	3.70	0.01	0.03	12.70
Long USD/Short JPY (%)	99.0	3.04	0.04	0.00	0.05	0.10	0.00

[A] For reasons of space results from only three currency pairs are displayed here. All portfolios performed badly in the unconditional coverage and ETL tests.

Table 5
Impact of estimation window on backtest of conditional t

Currency pair	$1-\alpha$	(i) Estimation window = 250 days			(ii) Estimation window = 2000 days		
(a)	(b)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)	Unconditional coverage VaR p-value (c)	Conditional coverage VaR p-value (d)	ETL p-value (e)
1-day horizon							
Long AUD/Short USD (%)	99.0	0.21	0.88	0.50	14.71	11.44	69.30
	99.5	0.01	0.03	13.80	30.66	51.33	92.20
	99.9	1.97	6.45	1.60	33.46	62.71	73.00
Short AUD/Long USD (%)	99.0	98.68	68.73	95.70	17.45	31.61	99.70
	99.5	55.84	78.59	97.70	0.60	2.25	88.40
	99.9	9.60	25.02	0.00	NA	NA	NA
Long GBP/Short USD (%)	99.0	23.63	21.65	53.30	74.41	53.55	96.10
	99.5	45.39	61.83	31.90	88.96	85.50	98.70
	99.9	28.30	55.46	27.10	15.75	36.77	86.60
Short GBP/Long USD (%)	99.0	94.58	89.35	30.20	7.14	13.50	97.30
	99.5	68.97	77.29	14.80	2.21	6.87	84.30
	99.9	15.51	35.81	38.20	5.08	14.84	76.20
Long USD/Short JPY (%)	99.0	0.00	0.00	0.30	7.37	0.00	23.30
	99.5	0.00	0.00	1.10	5.44	0.00	36.10
	99.9	0.01	0.01	1.20	7.87	20.90	54.70
Short USD/Long JPY (%)	99.0	7.14	13.67	98.30	0.51	1.10	100.00
	99.5	6.11	16.11	93.20	1.23	4.13	100.00
	99.9	NA	NA	NA	NA	NA	NA
3-day horizon							
Long AUD/Short USD (%)	99.0	2.35	5.34	9.40	12.65	23.81	99.40
Short AUD/Long USD (%)	99.0	4.50	13.02	80.40	4.50	13.02	99.80
Long GBP/Short USD (%)	99.0	3.04	7.59	82.10	7.61	15.16	89.30
Short GBP/Long USD (%)	99.0	74.38	74.67	24.60	74.38	74.67	59.50
Long USD/Short JPY (%)	99.0	0.62	2.05	0.10	0.62	0.74	3.90
Short USD/Long JPY (%)	99.0	42.22	62.86	28.40	19.91	39.26	63.00

(a) The trading portfolio used for the analysis. (b) The confidence level at which VaR and ETL are calculated. (c) Test of the null hypothesis that the actual number of violations is equal to the expected number of violations. (d) Test of the null hypothesis that violations are spread evenly over time (as opposed to being clustered). (e) Test of the null hypothesis that the ETL does not consistently understate the true potential for losses beyond the VaR.

Global Financial System (2005) has been to base this on an historical or hypothetical event specified by analysts, management or regulators. An alternative or complementary approach is to consider an extreme outcome as defined under the risk model.

For a long position VaR_α is a loss that should be exceeded in only α% of cases. Consequently the model-based stress test equates α to the probability of a market shock occurring on a given day where the size of the shock is $-\mathrm{VaR}_{1,\alpha}$.[11] A typical value for α is 0.0002, corresponding to a loss that we are 99.98% confident will not be exceeded over one day. The value of α should be fixed at board level to reflect the risk profile of the organisation and, if relevant, its target credit rating.

Under the empirical approach the initial shock ε^* for the long asset position is simply the α percentile of the empirical distribution using a large sample of data. We recommend using 15 years or more of daily returns. But if we assume a specific return distribution then the initial shock is derived from that distribution. For example under the Student's t distribution:

$$\varepsilon_T^* = t_v^{-1}(\alpha)((v-2)/v)^{1/2}\bar{\sigma}_T,$$

where $\bar{\sigma}_T$ is the equally weighted sample standard deviation using all available data up to time T. The initial shock can also be defined using EVT. Gencay and Selcuk (2004) show that EVT-based VaR estimates are often more accurate for small values of α in emerging markets. Under the EVT the initial shock for a long position is

$$\varepsilon_T^* = -u - \beta\xi^{-1}[(n\alpha)^{-\xi}n_u^{\xi} - 1],$$

where u is the threshold, ξ and β are the tail index and scale parameters of the generalised Pareto distribution, n is the sample size and n_u is the number of observations exceeding the threshold.

Table 6 illustrates these alternative approaches for defining the initial shock event using the three long portfolios. We define the initial shock using all data up to June 2006. We note that the Student's t distribution always gives the smallest shocks in an absolute sense but the other two models provide similar results. Thus it is difficult to justify the use of the more complex EVT methodology in this case. We hypothesise that this may be due to the fact that the estimated tail index is relatively low for currency data com-

[11] For a short position the size of the initial shock to returns is $\mathrm{VaR}_{1,(1-\alpha)}$.

Table 6
Estimating the initial shock for stress tests[a]

Initial shock by risk model		Long AUD/ Short USD (%)	Long GBP/ Short USD (%)	Long USD/ Short JPY (%)
$\alpha = 0.0002$	EVT	−4.56	−3.73	−4.68
	Empirical	−4.41	−3.75	−5.12
	Student's *t*	−3.90	−3.61	−4.18
$\alpha = 0.0005$	EVT	−3.85	−3.22	−3.94
	Empirical	−4.06	−3.22	−3.84
	Student's *t*	−3.26	−3.02	−3.45
Tail index, $\hat{\xi}$		0.0957	0.0521	0.0924

[a] Initial shock is the large gap or discontinuity in prices assumed to occur on the first day of the stress test. Again, for reasons of space results are shown for only three of the six currency pairs.

pared with those observed in other markets by authors such as Gencay and Selcuk (2004).

4.2. Modelling the after-shock

The consequences of a shock event can include some or all of the following: further large moves in the same market (as volatility clustering predicts), large moves in other markets and higher correlations between markets, increased implied volatility in option markets and reduced market liquidity. In this section we explain a modelling strategy suited to the current application: single currency, linear positions in major currency markets. Bangia et al. (2002) show that market liquidity risk is relatively unimportant for major currencies so we adopt here only a crude adjustment for liquidity when modelling the after-shock. That is, the horizon of the stress test is extended beyond a single day to reflect possible delays in closing a position.

Portfolio returns are assessed for $h-1$ days after the initial shock which occurs at time T so that the total stress test, including the initial shock, has a horizon of h days. The 'stress loss' for a long asset is -1 times the lower ϱ percentile of the simulated h-day returns when ranked from highest to lowest returns.[12] We propose Monte Carlo simulation to evaluate possible asset price paths subsequent to the shock. Whilst simulation can be criticised for computational intensity we note that stress tests (unlike VaR analysis) need not be conducted on a daily basis. And simulation has many advantages for stress testing including the ability to compare the impact of gradual and immediate hedging strategies on portfolio returns. Note that the maximum loss need not occur at the end of the risk horizon; in fact it often occurs at an interim point. Simulation allows analysts to evaluate a large array of possible paths including those with extreme interim losses, assessing their implications for limits, margin calls and funding liquidity.

The analysis in Section 3 shows that of all the approaches considered for forecasting extreme risks the conditional risk models are most suitable. Hence we describe the stress testing methodology for conditional models only.[13] On the first day of the test (at time T) the standard deviation is set equal to its long-term value $\bar{\sigma}_T$ to reflect 'typical' market circumstances, but there is an extreme innovation corresponding to the initial shock event defined above. At time $T+1$ variance will increase significantly in response to the shock event. Adapting (9) the variance on day $T+1$ is defined as

$$\hat{\sigma}^2_{T+1} = \hat{\gamma}_1 + \hat{\gamma}_2\varepsilon^2_T + \hat{\gamma}_3\bar{\sigma}^2_T.$$

Hereafter the simulation proceeds in the usual way, i.e. for $i = 1,\ldots,h-1$:

$$\hat{\sigma}^2_{T+i+1} = \hat{\gamma}_1 + \hat{\gamma}_2\varepsilon^2_{T+i} + \hat{\gamma}_3\hat{\sigma}^2_{T+i}$$

with innovations drawn from the chosen distribution (e.g. normal, empirical or Student's t), scaled to the appropriate conditional variance, and applying (9) to determine the variance on subsequent days during the life of the stress test. We simulate 30,000 paths each with $h-1$ realizations and aggregate the daily returns along each path to give the simulated h-day return under the assumption that the portfolio is held constant for the duration of the stress test.

4.3. Stress testing in the multivariate case

The multivariate GARCH literature is surveyed by Bauwens et al. (2004). While a number of these approaches are potentially useful for stress testing applications, the dynamic conditional correlation method of Engle (2002) is arguably the most useful for portfolios with many assets. VaR based stress tests for multivariate distributions can account for transmission of volatility between assets and correlation clustering as well as volatility clustering. We select an asset/factor which will sustain an initial shock. The initial shock is defined as the α percentile of the marginal returns distribution, either derived from the empirical marginal or the marginal in a parametric risk model. Depending on the particular model specified, the shock can lead to an increase in volatility in all assets as well as an increase in correlation. The algorithm for modelling the aftershock with two returns is as follows:

1. At time $T+i$, for $i = 2, 3, \ldots, h$ take two independent random draws $z_{1,T+i}$ and $z_{3,T+i}$ from a standard normal *i.i.d.* process.
2. Set $z_{2,T+i} = \hat{\varrho}_{T+i}z_{1,T+i} + \sqrt{1-\hat{\varrho}^2_{T+i}}z_{3,T+i}$ where $\hat{\varrho}_{T+i}$ is the GARCH correlation forecast at time $T+i$.
3. Set $\hat{\varepsilon}_{1,T+i} = \hat{\sigma}_{1,T+i}z_{1,T+i}$ and $\hat{\varepsilon}_{2,T+i} = \hat{\sigma}_{2,t}z_{2,T+i}$.

[12] In the short asset case the same analysis is performed but path returns are first multiplied by −1.

[13] The methodology for unconditional stress tests is available from the authors on request.

4. Find $\hat{\sigma}_{1,T+i+1}, \hat{\sigma}_{2,T+i+1}$ and $\hat{\sigma}_{12,T+i+1}$ from $\hat{\sigma}_{1,T+i}, \hat{\sigma}_{2,T+i}, \hat{\sigma}_{12,T+i}$, $\hat{\varepsilon}_{1,T+i}$ and $\hat{\varepsilon}_{1,T+i}$ using the GARCH model.
5. Return to step 1 replacing $T+i$ by $T+i+1$.

More generally for n correlated returns the Cholesky matrix of the conditional correlation matrix is employed in step 2 above. The process can be repeated with initial shocks occurring in different asset classes.

Multivariate stress tests based on unconditional VaR models are more flexible than in the univariate case. The best fitting unconditional copula can be selected as described, for example, in Kole et al. (2007). It is even possible to combine copulas with GARCH, as described in Patton (2006). Stress testing with multivariate VaR models is a very broad area for future research.

4.4. Modelling market liquidity risk

The issue of market liquidity is of utmost importance to portfolio stress testing and its analysis is required by regulators. Traders hedging large positions after a shock event may find that they cannot transact efficiently, resulting in exposure to adverse market conditions for longer periods, wider spreads or adverse market impact. Thus liquidity has both an endogenous and an exogenous element as explained in Jorion (2007, Chapter 13). Liquidity varies between markets and within a given market liquidity can also vary over time, as illustrated by Borio (2000). In particular, this study notes the link between market liquidity and volatility that can be incorporated to our stress-testing methodology. Other researchers (see Bangia et al., 2002; Deuskar, 2006) have also noted this relationship.

Jarrow and Subramanian (1997) were among the first to estimate liquidity-adjusted VaR (LA-VaR), taking account of the expected execution lag in closing a position and the market impact of prices being adversely effected by a quantity discount that varies with the size of the trade. Bangia et al. (2002) propose similar measures of LA-VaR, focussing on the exogenous dimension of liquidity risk. Their methodology incorporates an 'add-on' based on the bid-ask spread which is assumed to be a random variable. LA-VaR is estimated analytically under the simplifying assumption that extreme market events occur concurrently with extreme liquidity events.

The methodology we propose borrows some of these ideas but applies them in the context of Monte Carlo simulation with time varying volatility. Since market liquidity is linked to volatility, liquidity risk can readily be incorporated into the framework described in Section 4.2. To achieve this it is necessary to make assumptions about the likely delays in execution and the necessity for gradual liquidation, and the relevant quantity discount that will apply following the market shock specified in Section 4.1. Ideally these assumptions will be based on an analysis of trading patterns in stressful market conditions but in the short-run could be based on the subjective judgement of experienced traders. The Monte Carlo simulation proceeds in a manner similar to that described in Section 4.2, but builds in the gradual liquidation strategies and quantity discounts.

In a forthcoming empirical study by the authors, we assume that a large position is divided into five equal parcels traded on successive days. The assumed price on each of these days is adjusted to allow for the quantity discount appropriate to the parcel size. We assume that the price adjusted for quantity, $P(q)$ is given by[14]:

$$P(q) = P_0(1 - \theta \ln q),$$

where P_0 is the price just before the trade occurs (determined by simulation) and q is the quantity traded. Note that $P(q)$ is assumed to be a mid-price, ignoring any bid-offer spread. To capture variation in the bid-ask spread we refer to the analysis of Plerou et al. (2005) and Angelidis and Benos (2006) and assume a log-linear relationship between spread and volatility. At each trade an adjustment is made to the quantity adjusted price above, using the spread that is determined from the simulated volatility. For a long (short) position, we subtract (add) half the spread from $P(q)$ to determine the net transaction price. The simulation continues until the position is completely liquidated. At the end of the simulation we simply calculate the average close-out price, over all simulations, and compare this to the price at the start of the stress test. The difference is the overall cost of liquidation, taking account of both market and liquidity risks.

5. Evaluating model based stress tests

This section explains the basic parameters used in our stress tests and compares stress tests results obtained from different risk models with those obtained from more traditional stress tests. We also focus on the implications of model-based stress testing for capital adequacy.

5.1. Stress testing parameters

We illustrate model-based stress testing using the Conditional Empirical, Conditional t, and Unconditional Empirical risk models, anticipating a superior result from the two conditional models. We present results using all available data from the sources described in Section 3, noting that the results are robust to different sample periods.[15] There are three basic parameters for these tests:

[14] A simple linear price-quantity function is assumed by Jorion (2007). *Value at Risk* (McGraw Hill) at p. 342. However, the empirical results of Plerou et al. (2005) indicate that a log-linear relationship may be more appropriate.

[15] Results for other sample periods are available from the authors on request.

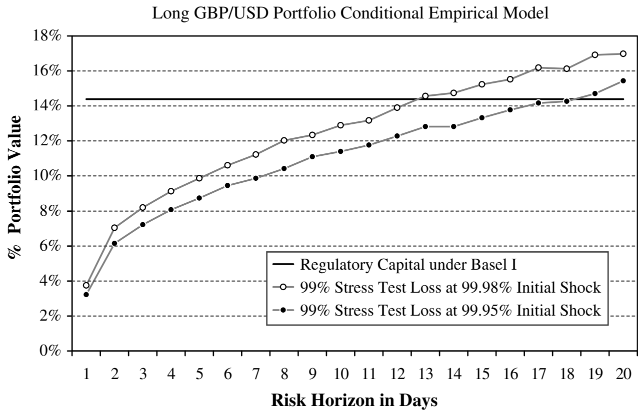

Fig. 1. Comparison of Basel I regulatory capital and stress test losses.

- The size of the initial shock: We set $\alpha = 0.0002$ (or 0.0005), equivalent to an initial shock that just breaches the 99.98% (or 99.95%) confidence level.
- The percentile for calculating the stress loss: We chose ϱ equal to 0.01, equivalent to a one-sided 99% confidence interval for stress test outcomes.
- The holding period: This is the number of days that the portfolio is held before it is fully hedged against further losses. We examined holding periods of 1, 2, ..., 20 days. Note that there is no regulatory prescription regarding the risk horizon of stress tests.

Fig. 1 presents the conditional empirical stress test results for a long GBP/USD currency portfolio. For comparison the VaR-based regulatory capital requirement is depicted by the horizontal line (see explanation in Section 5.5 below). The risk horizon for the stress tests is on the horizontal axis and the 99% stress loss corresponding to an initial shock at the 99.98% and the 99.95% percentiles are shown on the chart. The results indicate that VaR-based regulatory capital requirement exceeds our stress loss provided the horizon is not less than 13 or 19 days (depending on the size of the initial shock). That is, model-based stress tests would not necessitate any increase in regulatory capital provided the horizon for the stress test is less than 13 or 19 days. A similar conclusion is reached for AUD/USD and USD/JPY, although losses from a 10-day stress test would marginally exceed VaR-based regulatory capital for USD/JPY (see Table 7).

What risk horizon should be applied to model-based stress tests? A much shorter risk horizon (say, 3 days) may be sufficient to hedge typical positions in major currencies since even the least liquid of the currency pairs we are examining (the AUD/USD) has average daily turnover of USD90 billion.[16] In addition, the ability of institutions to gradually hedge positions (and thereby reduce the average risk horizon) will provide some relief almost immediately. However, the stress test horizon should also take account of reduced market liquidity in a crisis, the size of the position relative to the market and potential delays in managerial reaction to a shock event. Taking account of these factors, Deutsche Bank use a horizon of 10-days for stress-testing its very liquid currency positions such as GBP/USD.[17]

5.2. Comparing stress tests by risk model

Table 7 displays the stress test results for three of the six portfolios and for risk horizons of 3 and 10 days. Consider, for example, the 3-day stress test performed on the long AUD portfolio with $\alpha = 0.0002$, presented in panel 1 of Table 7. The assumed initial shock to currency returns varies between models from −4.41% (empirical) to −3.90% (conditional t), but a much larger discrepancy between the risk models is apparent when comparing the overall stress test outcomes over 3-day and 10-day horizons. The conditional models incorporate volatility clustering and therefore have a wider range of outcomes. The conditional empirical model suggests that when a stress event occurs we are 99% confident that losses over 3 days will be no greater than 9.88% of portfolio value. The best possible outcome (with 99% confidence) is a loss of −0.002%, i.e. a very small profit!

We also repeated the analysis with initial shocks calculated using EVT.[18] We estimated ξ and β from the same empirical sample and used these parameter estimates to calculate VaR, which in turn was used to determine the initial shock in the stress test. But calculating initial shocks in this way had negligible impact on the final stress test outcomes. Note that in every case the estimated value of the crucial tail index parameter was less than 0.1, indicating that the tails of currency returns distributions are not excessively heavy.

5.3. Model-based stress tests over time

Standard application of the GARCH approach is known to induce VaR estimates that vary significantly over time, a characteristic that can present practical difficulties for financial institutions using VaR as an input to the gearing decision.[19] GARCH models have been eschewed by some institutions since GARCH-based VaR estimates can increase suddenly and dramatically following a market shock and it is typically not possible to raise capital in a very short time-frame.

[16] Turnover figures were obtained from BIS (2005). In the USD/JPY average daily turnover is USD296 billion, while in GBP/USD average daily turnover is USD245 billion.

[17] We know this from personal correspondence with the Market Risk Analytics team at Deutsche Bank, London. Longer horizons are used for less liquid assets, with a typical stress testing horizon being 2 months.

[18] Results available from the authors on request.

[19] A better application of GARCH-based VaR models may be to the assessment of trading limits. In this context a VaR measure that is sensitive to current market conditions is potentially attractive.

Table 7
Stress test results

	Unconditional empirical	Conditional empirical	Conditional t
1. Long AUD/Short USD portfolio: maximum historical loss over 3 days = 7.27%, Over 10 days = 12.61%			
Long term volatility estimate	NA	10.36%	10.36%
VaR-based regulatory capital	17.84%	16.93%	16.05%
99% Stress loss:			
$h = 3,\ \alpha = 0.0002$	7.07%	9.88%	8.03%
$h = 3,\ \alpha = 0.0005$	6.72%	9.07%	7.04%
$h = 10,\ \alpha = 0.0002$	10.05%	15.42%	12.46%
$h = 10,\ \alpha = 0.0005$	9.70%	14.54%	11.12%
2. Long GBP/Short USD portfolio: maximum historical loss over 3 days = 7.45%, Over 10 days = 15.89%			
Long term volatility estimate	NA	9.60%	9.60%
VaR-based regulatory capital	15.94%	14.39%	13.97%
99% Stress loss:			
$h = 3,\ \alpha = 0.0002$	6.12%	8.20%	7.53%
$h = 3,\ \alpha = 0.0005$	5.60%	7.15%	6.37%
$h = 10,\ \alpha = 0.0002$	8.79%	12.84%	11.74%
$h = 10,\ \alpha = 0.0005$	8.26%	11.28%	10.26%
3. Long USD/Short JPY portfolio: maximum historical loss over 3 days = 7.05%, Over 10 days = 12.77%			
Long term volatility estimate	NA	10.44%	10.44%
VaR-based regulatory capital	18.16%	16.55%	15.12%
99% Stress loss:			
$h = 3,\ \alpha = 0.0002$	7.82%	10.93%	8.33%
$h = 3,\ \alpha = 0.0005$	6.55%	8.47%	7.04%
$h = 10,\ \alpha = 0.0002$	10.86%	17.06%	12.34%
$h = 10,\ \alpha = 0.0005$	9.58%	13.19%	10.72%

(a) *Long term volatility estimate* is sample standard deviation of all daily log returns, expressed in annual terms. (b) *Regulatory capital* expressed as a percentage of the portfolio amount, calculated as $-3 \times \mathrm{VaR}_{0.01,10\text{-day}}$ using the relevant risk model. VaR calculations use all available historical data for estimating parameters and percentiles and, where needed, the long-term volatility estimate. (c) *Stress Loss* refers to -1 times the stress test outcome at the ρ percentile over an h-day horizon, expressed as a percentage of initial portfolio value, assuming an initial shock.

We therefore assess the stability of the model-based stress tests over time, and hence their suitability for guiding decisions regarding capital adequacy. Fig. 2 depicts the results of model-based stress tests over time in the case of the Long GBP/Short USD portfolio using the conditional t risk model. The stress analysis is repeated every quarter, using the available data from the start of the sample to the estimation date. The minimum sample size is thus 20 years. We find that the stress loss estimates are in fact quite stable over time. The gradual downtrend in stress loss is linked to a decrease in both long term volatility and the reaction of currencies to market shocks.

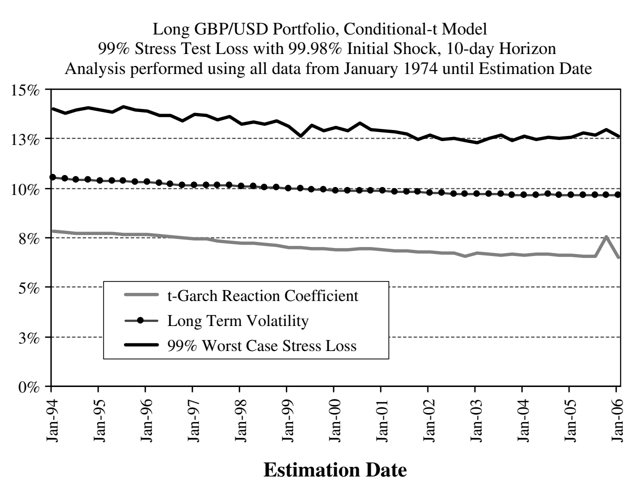

Fig. 2. Stress test analysis over time.

5.4. Comparing traditional stress tests with model based stress tests

According to the Committee on the Global Financial System (2005) the most commonly used stress tests are based on historical events that would be potentially disastrous for the portfolio being analysed. For the portfolios in this study we have identified the worst recorded currency movements from our data over horizons of 3 and 10 days, as shown in the headers of Table 7. Taking the Long AUD portfolio as an example, the worst recorded 3-day loss for this portfolio is 7.27% which is within the range predicted by both conditional stress tests with 99.98% initial shock. Note, however, that it is not within the range predicted by the unconditional stress test, providing further evidence that conditional risk models are preferable for stress testing purposes.

Are model based stress tests superior to traditional stress tests based on historical extremes? Consider the Long AUD position, where the conditional t risk model assumes an initial shock to returns of -3.90%, equivalent to a 99.98% percentile loss. There have been only four days since the float of the AUD in 1983 on which returns have been lower than -3.90%, so when the stress test is limited to historical data it relies on a very small sample. In contrast the model based stress test can simulate any number of exchange rate paths which are different from, but consistent with, those experienced historically. A larger sample allows the analyst to form a far more reliable conclusion concerning the likely impact of the stress event.

A further advantage of the model based approach is that it avoids subjective assumptions regarding the implications of the stress event for the portfolio. Using the traditional approach, analysts may over-estimate, underestimate or even ignore the potential for increased volatility and therefore further large returns subsequent to the initial shock. The model based approach plots the likely consequences of a shock based on established characteristics of markets that have been verified using historical data over more than 20 years.

The weakness of the model based approach is its vulnerability to model risk. This includes the risk that an inappropriate model is used to describe market dynamics and the uncertainty about the true model. Our backtesting experiment suggests that unconditional historical simulation, currently the most popular VaR methodology in the industry according to Perignon and Smith (2006), is likely to be misspecified and is therefore unsuited for stress testing purposes. We have attempted to mitigate model risks with an extensive, tailored backtesting procedure (Section 3) and by using long data sets for parameter estimation, but there remains the risk that market dynamics may change significantly in the near future. The methodology we have outlined relies on extensive data. While lengthy data sets are available in many markets they are not available in all. In cases where markets have recently been deregulated, or where deregulation is considered likely in the near future, historical data may be available but not relevant. In these cases any stress test based on historical data (whether model based or not) will be problematic.

An example of this type of model risk can be illustrated with reference to the breakdown of the European Monetary System (EMS) in 1992. In September 1992 the GBP came under speculative attack as market participants anticipated the withdrawal of the GBP from the EMS and its subsequent depreciation. The greatest historical depreciation in the GBP/USD occurred at that time, with a loss of 15.89% over a 10-day period. Note that this loss falls outside the bounds predicted by the model-based stress test in Table 7. The model predicted a 99% stress loss over 10 days of only 12.84%, based on an initial shock at the 99.98% percentile. However the sharp depreciation in the GBP had been widely anticipated at the time. In such cases it is prudent to apply a hypothetical initial shock to the stress test model. In this case a hypothetical return of −5.00%, which seems conservative in the circumstances, would have been sufficient to generate a 99% stress loss of 17.40% over 10 days. This example illustrates the need for risk managers to constantly monitor markets for potential structural breaks that even a sophisticated risk model cannot adequately predict. In these cases hypothetical shocks should be used as a complement to model-based shocks.

5.5. Implications of model based stress tests for risk capital

This section assesses the model based stress tests in the light of regulatory capital requirements. According to Basel Committee on Banking Supervision (2006), regulated financial institutions using internal models to assess capital are now required to hold sufficient capital to cover stress tests. This is in addition to the requirement that they have capital sufficient to meet VaR-based criteria. That is, market risk capital requirements are set at:

$$\mathrm{Max}[\text{Most recent estimate of} -\mathrm{VaR}_{0.01,10\text{-day}}, k \times \text{Average} -\mathrm{VaR}_{0.01,10\text{-day}} \text{for past 60 days}],$$

where k is a multiplier determined by local regulators having a minimum value of 3. Note that in the case of interest rate related instruments and equity securities the VaR-based criteria also requires capital to reflect the specific risk of the bank trading portfolio.

In Table 7 we calculate the VaR-based capital requirement as $-3\mathrm{VaR}_{0.01,10\text{-day}}$ using all available data for estimating risk model parameters/percentiles and, where necessary, using the long-term standard deviation. This measure, while slightly different from the equation above, is designed to represent a ‘typical’ capital requirement assuming relatively benign market conditions. Although the VaR-based capital requirement could exceed this value, we observe that our capital requirement estimates are always greater than the losses generated by model based stress tests except for very long stress test horizons. For instance, a position of £1 m for a USD investor has 99% stress loss over 3 days (following an initial shock at the 99.98% percentile) of £82,000. But the regulatory capital estimate is almost double this at £143,900.

6. Summary and conclusions

Stress testing is an established component of daily risk management for portfolios exposed to market risk. Indeed many financial institutions are required by regulators to implement stress tests and their importance is to be extended under the new capital accord. However traditional stress tests can be criticised for being conducted outside the context of a risk model, hence the probability of an extreme outcome is unknown and many extreme yet plausible possibilities are ignored. Many stress tests also fail to incorporate the characteristics that markets are known to exhibit in crisis periods, namely, increased probability of further large movements, increased co-movement between markets, greater implied volatility and reduced liquidity.

The first objective of this paper was to identify the most suitable risk models in which to conduct a stress test. We considered eight relatively well-known and parsimonious risk models, of which four incorporate volatility clustering.

We also explored the importance of heavy tails and skewness by comparing the normal, empirical and Student's t distributions. Analysis was performed using daily returns for the AUD/USD from December 1983 to June 2006 and for the GBP/USD and USD/JPY from January 1974 to June 2006.

Our model specification procedure was designed to focus on extreme market movements over short risk horizons, as this is most relevant to stress testing. Conditional and unconditional coverage tests were performed at very high percentiles for horizons up to 3 days on a 15-year daily out-of-sample returns series for each model. We also tested the accuracy of the expected tail loss estimates to ensure that our risk models adequately explain losses beyond VaR. We presented strong evidence that the most suitable risk model for stress testing purposes is the conditional empirical model, i.e. the model with volatility adjusted historical innovations. Our preferred risk models benefit from estimation windows in excess of 1000 days. Despite the relative simplicity of the risk models the rigour of the backtesting ensured that the preferred models achieved excellent results. We found that the use of more complex models such as EVT and mixture models may not be warranted to model returns in currency markets.

The second and major objective of this paper was to develop a methodology for conducting stress tests in the context of a risk model. We proposed a two-stage approach whereby an initial shock event is linked to the probability of its occurrence. The risk model is then used to model the consequences of that shock event in simulation. We implemented this stress testing procedure for three major currency pairs and found that results compared favourably with the traditional 'historical scenario' stress testing approach in all but one extraordinary case.

Finally, the implications for capital adequacy under the new Basel recommendations were explored. A detailed comparison indicated that current regulatory requirements for trading portfolios in liquid currency markets are already conservative enough to cover the results of stress tests. While there is legitimate debate concerning the appropriate risk horizon for a stress test we conclude that in liquid currency markets the typical risk horizons will not be long enough for stress tests to have an impact on capital requirements.

The advantages of our stress test methodology are: greater objectivity, the ability to link stress tests to a targeted probability, greater statistical reliability, ability to examine market response following a shock event (including volatility clustering) and the ability to evaluate alternative hedging strategies along with the impact of limits or margins on expected losses and funding liquidity. A disadvantage of the methodology is vulnerability to model risk. The problem is most acute in markets undergoing structural change (such as deregulation). Model risk can be mitigated by using hypothetical shock events when markets are changing. Long sample periods for parameter estimation and extensive backtesting to guide the choice of risk model are also very important.

Acknowledgements

Many thanks to two anonymous referees for extensive comments on earlier drafts, which helped to improve this paper considerably. The authors also wish to acknowledge the financial support of the Australian Prudential Regulatory Authority (APRA) and Macquarie University, and the excellent research assistance of Luke Liu. Many thanks also to Philippe D'Ornano of Ernst and Young, David Rowe of Sunguard and Richard Evans of Deutsche Bank, all of whom offered insightful comments and much useful advice.

Appendix A. Scaling empirical VaR and ETL

When a distribution is ξ-stable then the whole distribution, including the percentiles, scales as $h^{1/\xi}$. In a normal distribution $\xi = 2$ and its scale exponent is 1/2. For other stable distributions $0 < \xi < 2$ so scaling by the square root of time will lead to an underestimation of the percentile of h-day returns, as discussed by Danielsson and Zigrand (2004). We can estimate ξ for the distribution of a log return X as follows. Let $x_{h,\alpha}$ denote the lower α percentile of the h-day discounted log returns. We seek ξ such that $x_{h,\alpha} = h^{1/\xi} x_{1,\alpha}$. In other words:

$$\xi = \frac{\ln(h)}{\ln(x_{h,\alpha}) - \ln(x_{1,\alpha})}.$$

Hence ξ can be estimated as the slope of graph with $\ln(x_{h,\alpha}) - \ln(x_{1,\alpha})$ on the horizontal axis and $\ln(h)$ on the vertical axis. Fig. 3 shows the graphs for the GBP rate with $\alpha = 5\%$, based on daily data from January 1975 to September 2007. If the distribution is stable the graph is a straight line and ξ will not depend on the choice of α. Nor should it vary much when different samples are used if the sample contains sufficient data to estimate the percentiles accurately.

The scale exponent is the reciprocal of the slope of the graph. In this case $\hat{\xi}^{-1} = 0.5286$. A similar calculation for the AUD rate gives $\hat{\xi}^{-1} = 0.5220$. The estimated scale exponents decrease marginally as we move to more extreme percentiles. For instance in the GBP/USD rate with $\alpha = 1\%$, we obtain $\hat{\xi}^{-1} = 0.4957$, but the estimation error also increases as α decreases. With $\alpha = 5\%$ all data give

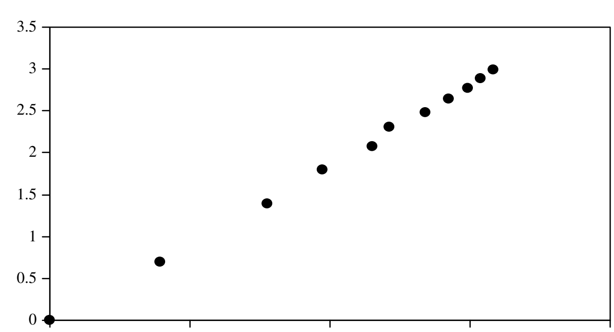

Fig. 3. Log–log plot for the power law scale exponent, GBP/USD.

log-log plots that are nearly straight lines with slope $\approx 1/2$ and we conclude that the square root of time rule is a suitable scaling law for the percentiles of currency returns distributions.

Appendix B. Derivation of analytic formulae for ETL

If X has distribution function $F(x)$ and density function $f(x)$ then:

$$\mathrm{ETL}_\alpha(X) = -\alpha^{-1}\int_{-\infty}^{F^{-1}(\alpha)} xf(x)\mathrm{d}x. \tag{10}$$

We prove the following:

(i) If X has a Student's t distribution with mean μ and standard deviation σ then

$$\mathrm{ETL}_\alpha(X) = \alpha^{-1}(v-1)^{-1}(v-2 + T_v^{-1}(\alpha)^2)t_v(T_v^{-1}(\alpha))\sigma - \mu, \tag{11}$$

where T_v is the standardised Student's t distribution with v degrees of freedom, zero mean and unit variance and t_v is its density function, i.e.

$$t_v(x) = ((v-2)\pi)^{-1/2}\Gamma\left(\frac{v}{2}\right)^{-1}\Gamma\left(\frac{v+1}{2}\right)(1+(v-2)^{-1}x^2)^{-(1+v)/2}.$$

(ii) If $X \sim \mathrm{NM}(\pi,\mu,\sigma^2)$ where $\pi = (\pi_1,\ldots,\pi_n)$, $\mu = (\mu_1,\ldots,\mu_n)$ and $\sigma^2 = (\sigma_1^2,\ldots,\sigma_n^2)$ then

$$\mathrm{ETL}_\alpha(X) = \alpha^{-1}[\pi\varphi(\sigma_1^{-1}G_0^{-1}(\alpha))\sigma_1 + (1-\pi)\varphi(\sigma_2^{-1}G_0^{-1}(\alpha))\sigma_2] - [\pi\mu_1 + (1-\pi)\mu_2], \tag{12}$$

where φ is the standard normal density function and G_0 is the distribution function for a mixture of two zero-mean normal distributions with standard deviations σ_1 and σ_2.

Proof. (i) The result follows on showing that the ETL in a standardised Student's t-distribution with v degrees of freedom, unit standard deviation and zero mean is given by

$$\alpha^{-1}(v-1)^{-1}(v-2+T_v^{-1}(\alpha)^2)t_v(T_v^{-1}(\alpha)). \tag{13}$$

Write $t_v(x) = A(1+ax^2)^b$ where

$$A = ((v-2)\pi)^{-1/2}\Gamma(v/2)^{-1}\Gamma((v+1)/2), a = (v-2)^{-1}, \quad \text{and } b = -(1+v)/2.$$

Then substituting $y = 1 + ax^2$:

$$\int_{-\infty}^{T_v^{-1}(\alpha)} xt_v(x)\mathrm{d}x = A\int_{-\infty}^{T_v^{-1}(\alpha)} x(1+ax^2)^b\mathrm{d}x = \frac{A}{2a}\int_{-\infty}^{Y} y^b\,\mathrm{d}y, \quad Y = 1 + (v-2)^{-1}T_v^{-1}(\alpha)^2.$$

But $\int_{-\infty}^{Y} y^b\,\mathrm{d}y = \frac{Y^{b+1}}{b+1} = \frac{2Y^{(1-v)/2}}{1-v}$ and $A = t_v(T_v^{-1}(\alpha))Y^{(1+v)/2}$ so:

$$\int_{-\infty}^{T_v^{-1}(\alpha)} xt_v(x)\mathrm{d}x = \frac{t_v(T_v^{-1}(\alpha))Y^{(1+v)/2}}{2(v-2)^{-1}}\cdot\frac{2Y^{(1-v)/2}}{1-v} = -(v-1)^{-1}(v-2)Yt_v(T_v^{-1}(\alpha))$$

and this proves (13).

(ii) Suppose X has normal mixture distribution G_0 with zero means in the components, and write its density function as $\sum_{i=1}^n \pi_i f_i(x)$ where each $f_i(x)$ is a zero mean normal density with standard deviation σ_i.

$$\mathrm{ETL}_\alpha(X) = -\alpha^{-1}\sum_{i=1}^n \pi_i \int_{-\infty}^{G_0^{-1}(\alpha)} xf_i(x)\mathrm{d}x.$$

But

$$\int_{-\infty}^{G_0^{-1}(\alpha)} xf_i(x)\mathrm{d}x = -\sigma_i\varphi(\sigma_i^{-1}G_0^{-1}(\alpha)).$$

Hence for the zero-mean normal mixture

$$\mathrm{ETL}_\alpha(X) = \alpha^{-1}\sum_{i=1}^n (\pi_i\sigma_i\varphi(\sigma_i^{-1}G_0^{-1}(\alpha)))$$

and the result follows on noting that $\sum_{i=1}^n \pi_i\mu_i$ is the mean of the normal mixture. $\square$

Appendix C. Description of backtests

The test for unconditional coverage is a likelihood ratio (LR) test of the form:

$$\mathrm{LR} = \frac{\pi_{\exp}^{n_1}(1-\pi_{\exp})^{n_0}}{\pi_{\mathrm{obs}}^{n_1}(1-\pi_{\mathrm{obs}})^{n_0}}; \quad -2\ln\mathrm{LR} \sim \chi_1^2,$$

where $\pi_{\exp}$ is the expected proportion and π_{obs} is the observed proportion of returns that lie in the prescribed interval of the distribution, n_1 is the number of returns that lie inside the interval (i.e., the number of violations) and n_0 is the number of returns that lie outside the interval (we can call these returns the 'good' returns).

The test for conditional coverage is

$$\mathrm{LR} = \frac{\pi_{\exp}^{n_1}(1-\pi_{\exp})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}; \quad -2\ln\mathrm{LR} \sim \chi_2^2$$

where n_{10} is the number of times a violation is followed by a good return; n_{11} is the number of times a violation is followed by another violation; n_{01} is the number of times a good return is followed by a violation; and n_{00} is the number of times a good return is followed by another good return.

Also let

$$\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}} \quad \text{and} \quad \pi_{11} = \frac{n_{11}}{n_{10}+n_{11}},$$

i.e. π_{01} is the proportion of exceedances, given that the last return was a 'good' return, and π_{11} is the proportion of exceedances, given that the last return was an exceedance.

The ETL test first computes the standardised exceedance residuals, $r = \hat{\sigma}_{t+h}^{-1}(\varepsilon_{t+h} - \mathrm{ETL}_{h,\alpha})$ if $\varepsilon_{t+h} < -\mathrm{VaR}_{h,\alpha}$, and $r = 0$ otherwise. The null hypothesis is that these have zero mean, against the alternative that their mean is positive. The distribution of the test statistic $t = \bar{r}/est.s.e.\bar{r}$ is found using the standard bootstrap simulation introduced by Efron and Tibshirani (1993, pp. 224–7).

References

Angelidis, T., Benos, A., Degiannakis, S., 2004. The use of GARCH models in VaR estimation. Statistical Methodology 1, 105–128.

Angelidis, T., Benos, A., 2006. Liquidity adjusted value-at-risk based on the components of the bid–ask spread. Applied Financial Economics 16 (11), 835–851.

Aragones, J., Blanco, C., Dowd, K., 2001. Incorporating stress tests into market risk modeling. Derivatives Quarterly. Spring 2001, 44–49.

Bangia, A., Diebold, F., Schuermann, T., Stroughair, J., 2002. Modeling liquidity risk with implications for traditional market risk measurement and management. In: Levich, R., Figlewski, S. (Eds.), Risk Management: State of the Art. Springer.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don't look back. Risk 11, 100–103.

Basel Committee on Banking Supervision, 1996. Amendment to the Capital Accord to Incorporate Market Risks.

Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive Version (Bank for International Settlements, Basel).

Bauwens, L., Laurent, S., Rombouts, J., 2004. Multivariate GARCH models: A Survey, CORE Discussion Paper (Louvain).

Berkowitz, J., 2000. A coherent framework for stress-testing. Journal of Risk 2, 1–11.

Berkowitz, J., Christoffersen, P., Pelletier, D., 2006. Evaluating VaR Models with Desk-level Data. McGill University.

Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? Journal of Finance, 1093–1111.

Bollerslev, T., 1986. Generalised autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 309–328.

Borio, C., 2000. Market liquidity and stress: Selected issues and policy implications. BIS Quarterly Review, 38–51.

Butler, J.S., Schachter, B., 1998. Estimating VaR with a precision measure by combining kernel estimation with historical simulation. Review of Derivatives Research 1, 371–390.

Chen, S., Tang, C., 2005. Nonparametric inference of value-at-risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Committee on the Global Financial System, 2005. A survey of stress tests and current practice at major financial institutions.

Danielsson, J., De Vries, C., 2000. Value-at-risk and extreme returns. In: Embrechts, Paul (Ed.), Extremes and Integrated Risk Management. Risk Books, London.

Danielsson, J., Zigrand, J.-P., 2004. On time-scaling of risk and the square-root-of-time rule, EFA 2004 Maastricht Meetings.

Deuskar, P., 2006. Extrapolative Expectations: Implications for Volatility and Liquidity. NYU Stern School of Business, New York.

Dowd, K., 2002. Measuring Market Risk. Wiley Finance.

Dowd, K., Rowe, D., 2004. Introduction to value-at-risk models. In: Alexander, C., Sheedy, E. (Eds.), The Professional Risk Manager's Handbook. PRMIA Publications.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalised autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20, 339–350.

Galati, G., Heath, A., McGuire, P., 2007. Evidence of carry trade activity. BIS Quarterly Review, 27–41.

Gencay, R., Selcuk, F., 2004. Extreme value theory and VaR: Relative performance in emerging markets. International Journal of Forecasting 20, 287–303.

Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 18, 641–664.

Greenspan, A., 2000. Greenspan's plea for stress testing. Risk 5.

Huisman, R., Koedijk, K., Pownall, R., 1998. VaR-x: Fat tails in financial risk management. Journal of Risk 1, 47–61.

Hull, J., 2006. Risk Management and Financial Institutions. Pearson Prentice Hall.

Jarrow, R., Subramanian, A., 1997. Mopping up liquidity. Risk 10, 170–173.

Jorion, P., 2007. Value at Risk. McGraw Hill.

Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. Journal of Banking and Finance 31, 2405–2423.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives, 73–84.

Kupiec, P., 1998. Stress testing in a Value-at-Risk framework. Journal of Derivatives 6, 7–24.

Longin, F., 2000. From Value-at-Risk to stress testing: The extreme value approach. Journal of Banking and Finance 24, 1097–1130.

Longin, F., 2005. The choice of the distribution of asset returns: How extreme value theory can help? Journal of Banking and Finance 29, 1017–1035.

Lopez, J., 2005. Stress tests: Useful complements to financial risk models. FRBSF Economic Letter, 119–124.

McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271–300.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press.

Mittnik, S., Paolella, M., 2000. Conditional density and Value-at-Risk prediction of asian currency exchange rates. Journal of Forecasting 19, 313–333.

Patton, A., 2006. Copula-Based Models for Financial Time Series. London School of Economics.

Perignon, C., Smith, D., 2006. The Level and Quality of Value-at-Risk Disclosure by Commercial Banks. Simon Fraser University, Vancouver.

Plerou, V., Gopikrishnan, P., Stanley, H.E., 2005. Quantifying fluctuations in market liquidity: Analysis of the bid–ask spread. Physical Review E 71, 046131.

Pritsker, M., 2006. The hidden dangers of historical simulation. Journal of Banking and Finance 30, 561–582.

Rowe, D., 2005. The new market risk challenge. Risk 18 (9), 103.

Settlements, Bank for International, 2005. Triennial Central Bank Survey: Foreign Exchange and Derivatives Market Activity in 2004 (Bank for International Settlements, Basel).

Sheather, S., Marron, J.S., 1990. Kernel quantile estimators. Journal of the American Statistical Association 85, 410–416.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

So, M., Yu, P., 2006. Empirical analysis of GARCH models in value at risk estimation. International Financial Markets, Institutions and Money 16, 180–197.

Straumann, D., 2004. Estimation in Conditionally Heteroscedastic Time Series Models. Lecture Notes in Statistics. Springer-Verlag, Berlin.

Venter, J., de Jongh, P., 2002. Risk estimation using the normal inverse Gaussian distribution. Journal of Risk 4, 1–5.

Vilasuso, J., 2002. Forecasting exchange rate volatility. Economics Letters 76, 59–64.