Filtering Historical Simulation. Backtest Analysis

Article · May 2000

Source: CiteSeer

3 authors:

Giovanni Barone-Adesi
University of Lugano

Kostas Giannopoulos
Neapolis University

Les Vosper
AOL

Filtering Historical Simulation. Backtest Analysis[1]

By Giovanni Barone-Adesi, Kostas Giannopoulos and Les Vosper

March 2000

A new generation of VaR models, based on historical simulation (bootstrapping), is being increasingly used in the risk management industry. It consists of generating scenarios, based on historical price changes, for all the variables in the portfolio. Since the estimated VaR is based on the empirical distribution of asset returns it reflects a more realistic picture of the portfolio's risk. Unfortunately this methodology has a number of disadvantages. To overcome some of them Barone-Adesi, Bourgoin and Giannopoulos(1998) and Barone-Adesi, Giannopoulos and Vosper (1999) introduce filtered historical simulation (FHS hereafter). They take into account the changes in past and current volatilities of historical returns and make the least number of assumptions about the statistical properties of future price changes.

In this paper we backtest the FHS VaR model on three types of portfolios invested over a period of two years. The first set of backtests consists of LIFFE financial futures and options contracts traded on LIFFE. In the second set of backtests we examine the suitability of the FHS model on interest rate swaps. Finally, we backtest a set of mixed portfolios consisting of LIFFE interest rate futures and options as well as plain vanilla swaps.

We go beyond the strict criteria of the BIS recommendations by evaluating daily risk at four different confidence levels and five different trading horizons for a large number of realistic portfolios[2] of derivative securities.

In the first section we describe the backtesting methodology and report the results for the LIFFE portfolios. To enable us to appraise the different components of risk measurements on each of the three types of portfolios we run three sets of backtests, relaxing the following assumptions in each test: in the first backtest we keep constant implied volatilities and FX rates. Our analysis focuses on how well FHS predicts losses due to futures and options market price changes. In our second backtest we simulate implied volatilities while in the third backtest we also take into account the portfolios' FX exposure. Our results show that fixed implied volatility performs better at short VaR horizons, while at longer ones (5 to 10 days) our stochastic implied volatility performs better.

[1] Universita' della Svizzera Italiana and City University Business School, Westminster Business School and London Clearing House. We are grateful to The London Clearing House for providing us the data and financing the project.

[2] For backtest 1, a total of 75,835 daily portfolios; backtests 2 and 3 have a total of 75,985 portfolios.

In the second series of backtests we investigate the performance of FHS on books of interest rate swaps. We compare each book's daily values with the FHS lower forecasted value. For each book we produce two types of forecast; an aggregate market value risk expressed in GBP and a set of currency components (plain vanilla swaps in USD, JPY, DEM, and GBP are used). In the swap portfolios we find that our methodology is too conservative at longer horizons.

In the final set of backtests we investigate the performance of the FHS model on diversified portfolios across plain vanilla swaps and futures and options traded on LIFFE. This study shares the same data with the separate LIFFE and Swaps backtests but restricts the number of portfolios to 20 among the largest members on LIFFE. By adding to each LIFFE portfolio one of the four swap books[3] used in the swaps backtest we form 20 combined (combo) portfolios.

Our analysis is based on two criteria: statistical and economic. The former examine the frequency and the pattern of losses exceeding the VaR predicted by FHS (breaks); the latter examine the implications of these breaks in economic terms, with reference to the total VaR allocated. Overall our findings sustain the validity of FHS as a risk measurement model. Furthermore, we find that diversification reduces risk effectively across the markets we study.

[3] Four SWAP books consisting of 500 Swaps each were formed for the SWAP backtest. Details of the portfolios are available from the authors.

1 Overview of VaR models.

VaR models play a core role in the risk management of today's financial institutions. A number of VaR models are in use. All of them have the same aim, to measure the size of possible future losses at a predetermined probability. There are a variety of approaches used by VaR models to estimate the potential losses. Models differ in fact in the way they calculate the density function of future profits and losses of current positions, as well as the assumptions they rely on. Although VaR analysis has been used since early 1980's by some departments of few large financial institutions it wasn't until the middle 1990's that it became widely accepted by banks and also imposed by the regulators. The cornerstone behind this wide acceptance was a **linear** VaR model, based on the **variance-covariance** of past security returns, introduced by JP Morgan, RiskMetrics (1993). The variance-covariance approach to calculate risk can be traced back to the early days of Markowitz's (1959) Modern Portfolio Theory, which is now common knowledge among today's risk managers.

Linear VaR models, however, impose strong assumptions about the underlying data. For example, the density function of daily returns follows a theoretical distribution (usually normal) and has constant mean and variance[4]. The empirical evidence about the distributional properties of speculative price changes provides evidence against these assumptions[5], e.g. Kendall (1953) and Mandelbrot (1963). Risk managers have also seen their daily portfolio's profits and losses to be much larger than those predicted by the normal distribution. The RiskMetrics VaR method has two additional major limitations. It linearises derivative positions and it does not take into account expiring contracts. These shortcomings may result in large biases, particularly for longer VaR horizons and for portfolios weighed with short out-of-the money options.

To overcome problems of linearising derivative positions and to account for expiring contracts, risk managers have begun to look at **simulation** techniques. Pathways are simulated for scenarios for linear positions, interest rate factors and currency exchange rate and are then used to value all positions for each scenario. The VaR is estimated from the distribution (e.g. 1st percentile) of the simulated portfolio values. Monte-Carlo simulation is widely used by financial institutions around the globe. Nevertheless, this method can attract severe criticisms. First, the generation of the scenarios is based on random numbers drawn from a theoretical distribution, often normal, which not only does not conform to the empirical distribution of most asset returns, but also limits the losses to around three or four standard deviations when a very large number of simulation runs is carried out. Second, to maintain the multivariate properties of the risk factors when generating scenarios, historical correlations are used; during market crises, when most correlations tend to increase rapidly, a Monte Carlo system is likely to underestimate the possible losses. Third, historical simulation tends to be slow, because a large number of scenarios which has to be generated[6].

Barone-Adesi and Giannopoulos (1996) argue that the covariances (and correlations) are unnecessary in calculating portfolio risk[7]. They suggested the creation of a synthetic security by multiplying current portfolio weights by the historical returns of all assets in the portfolio. They fit a conditional volatility model on these historical returns to estimate the last trading day's volatility and then calculate the portfolio's VaR as in RiskMetrics (1993). Barone-Adesi, Bourgoin and Giannopoulos (1988)

[4] The RiskMetrics VaR approach recognises the fact that variances and covariances are changing over time and use a simple method (exponential smoothing) to capture these changes. They contradict themselves, however, and use a constant volatility assumption for the multi-period VaR (i.e. the last trading day's volatility is scaled with the time).
[5] However the extent to which these assumptions are violated depend on the frequency of the data. Daily data, which are of interest in risk management, tend to deviate to a great extend from normality.
[6] Jamshidian and Zhu (1997) proposed a method that limits the number of portfolio valuations.
[7] Markowitz introduced the variance-covariance matrix in his portfolio risk approach but had a different objective to risk managers. Markowitz aim was to find optimal risk/return portfolio weights rather than trying to measure risks of a (known) portfolio.

have suggested to draw random standardised returns[8] from the portfolio's historical sample and after rescaling these standardasied historical returns with the current volatility, to use them as innovations in a conditional variance equation for generating scenarios for both future portfolio variance and price (level)[9]. This method not only generates scenarios that conform with to the past history of the current portfolio's profits and losses, but also overcomes an additional limitation of the variance-covariance model. It allows both past and future volatility to vary over time. The creation of a synthetic security simplifies the computational effort to a large extent. However, it suffers from the above criticisms when handling non-linear positions (it uses the options delta to linearise them).

Recognising the fact that most asset returns cannot be described by a theoretical distribution, an increasing number of financial institutions are using **historical simulation**. Here, each historical observation forms a possible scenario, see Butler and Schachter (1998). A number of scenarios is generated and in each of them all current positions are priced. The resulting portfolio distribution is more realistic since it is based on the empirical distribution of risk factors. This method has still some serious drawbacks. Historical simulation ignores the fact that asset risks are changing all the time. The historical returns that are used as if they are random numbers are not i.i.d., and so, are unsuitable for any simulation. Thus the VaR value will be biased. During high volatile market conditions the historical simulation will underestimate risk. Furthermore, historical simulation uses constant implied volatility to price the options under each scenario. Some positions which may be appear well-hedged under the constant implied volatility hypothesis, may become very risky under a more realistic scenario. It would be difficult to determine the extent of this problem as sensitivity analysis is difficult with historical simulation.

1.1 Filtered historical Simulation

To overcome the shortcomings of historical simulation it is necessary to **filter** historical returns; that is to adjust them to reflect current information about security risk. Our complete filtering methodology is discussed in Barone-Adesi, Giannopoulos and Vosper (1999). A brief synopsis is presented below.

1.1.1 Simulating a Single Pathway

Many pathways of prices are simulated for each contract (asset or interest rate) in our dataset over several holding periods. In our backtests we use 5000 simulations results over 10 days. The algorithm is described by starting with the simulation of a single pathway out of the 5000, for a single contract. From this we can generalise to the simulation of many pathways for many contracts, and their aggregation into portfolio pathways. The set of portfolio pathways for each day in the holding period defines 10 empirical distributions i.e. over holding periods from 1 to 10 days.

Our methodology is non-parametric in the sense that simulations do not rely on any theoretical distribution on the data as we start from the historical distribution of the return series. We use two years of earlier data to calibrate our GARCH models, (Bollerslev 1986), for asset returns and to build the data bases necessary to our simulation. By calibrating GARCH models to the historical data we form residual returns from the returns series. Residual returns are then filtered to become identically and independently distributed, removing serial correlation and volatility clusters. As the computation of the i.i.d. residual returns involves the calibration of the appropriate GARCH model,[10] the overall approach can be described as semi-parametric. For example, assuming a GARCH(1,1) process with both

[8] The portfolio's standardised returns must be i.i.d. If this is not the case then a mean equation that yields i.i.d. residuals is fitted.
[9] See section 1.1.1
[10] For example one asset may use GARCH (1,1) with no AR or MA terms, another may employ AGARCH with an MA term; we examine a variety of processes to attempt to fit the appropriate model to the data series of each asset.

moving average (θ) and autoregressive (m) terms, our estimates of the residuals ε_t and the variance h_t are :

$$r_t = m\, r_{t-1} + \theta e_{t-1} + e_t \qquad e_t \sim N(0, h_t) \tag{1}$$

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

To bring residuals close to a stationary i.i.d. distribution, so that they are suitable for historical simulation, we divide the residual e_t by the corresponding daily volatility estimate:

$$e_t = \frac{e_t}{\sqrt{h_t}} \tag{3}$$

We now randomly draw standardised residual returns from the dataset and use them to form a pathway of variances to be used in our backtest. To do this the first-drawn standardised residual is scaled by the deterministic volatility forecast one day ahead:

$$z_{t+1} = e_1 \cdot \sqrt{h_{t+1}} \tag{4}$$

This forecast is used to form the one-day ahead forecast of the asset price:

$$p_{t+1} = p_t + p_t(m\, r_t + \theta\, z_t + z_{t+1}) \tag{5}$$

Forecasts of volatility for subsequent days ahead are simulated by the recursive substitution of scaled residuals into the variance equation (2). Thus the first-drawn standardised residual from (3) with which we form the price forecast one-day ahead in (5) also allows for the simulation of the volatility forecast two days ahead. This volatility depends on the return simulated on the first day. Therefore it is stochastic. We then scale the second-drawn standardised residual, which is used to simulate the price two days ahead.

Similarly the volatility three days ahead is formed from the previous second-drawn scaled residual and allows for the scaling of the third-drawn residual and so on up to ten days ahead. Generally we have for the volatility pathway:

$$\sqrt{h_{t+i}} = \sqrt{w + a(z_{t+i-1})^2 + b h_{t+i-1}} \qquad i \geq 2 \tag{6}$$

Our procedure enables successive scaling of drawn residuals and for the price pathway to be constructed for the holding period. Repetition of the approach enables the formation of many pathways of the asset prices.

1.1.2 Simulating Multiple Pathways

When we have many assets, pathways are simulated for each asset by application of the above approach. The correlation between different assets is modelled implicitly by the random drawing of strips of residuals i.e. for a given pathway and day in the holding period (“node x,y”), each separate asset’s simulated price and volatility is produced from standardised residuals at the same date. Hence the empirical price co-movements between assets are represented across corresponding nodes in volatility and price pathways[11]. The price pathways for assets in a portfolio can therefore be aggregated, to produce portfolio price distributions, without resorting to the correlation matrix, or assuming any particular distribution for the data. The procedure is non-parametric apart from assumptions used in the estimation of residuals in the GARCH process (see equation 1).[12]

For assets in different currencies, the methodology can produce pathways of simulated currency exchange rates, so that all values are expressed in a common currency. The historical residuals derived from changes in exchange rates are included in the dataset so that they are a part of the random residuals strips used during simulation, so that FX moves are produced simultaneously with asset (or interest rate) moves.

In this way, assets' pathways, FX pathways and interest rate pathways[13] are constructed from historical returns modified through GARCH filters. We go beyond historical simulation by scaling the random set of returns to produce i.i.d. standardised residuals; these standardised residuals are then scaled again to reflect current and forecast volatilities[14]. For certain scenarios, the successive prices in a pathway will have been constructed from one large return following another, so that extreme scenarios are generated, beyond those generally used in Value at Risk estimations.

If a portfolio contains non-linear derivatives e.g. options, their pathways are produced by using the appropriate options model with the pathways of underlying assets as inputs. Implied volatility can either be assumed to be constant over the holding period, or pathways of implied volatilities can be produced that relate to the simulated variance pathways given suitable assumptions. In our backtesting, we consider both possibilities.

Section 1: Backtesting interest rate futures and options

This section describes the first set of backtests, applied to LIFFE member portfolios. For each trading date over a period of nearly 2 years (4 January 1996 to 12 November 1997), we use the filtered historical simulation (therefore FHS) to measure the market risk on LCH members with positions in LIFFE financial products. We compare the daily profits and losses for each portfolio with FHS-generated risk measures. We use only information available at the end of each trading date (positions and closing prices) to measure the market risk of each portfolio for horizons of up to 10 working days. Consequently, we compare the actual trading results of these portfolios[15] with the risk values predicted by FHS. This process is known as "backtesting", it is recommended by the Basle Committee (1996) and has been adopted by many financial institutions to gauge the quality and accuracy of their risk measurement models.

2 The Backtesting process

For each business day from 4 January 1996 until 12 November 1997 we use the FHS methodology to calculate the market risk of all LIFFE members with positions in interest rate contracts (German

[11] For a detailed description of the algorithm see Barone-Adesi, Giannopoulos and Vosper (1999).
[12] GARCH constants have been maximum likelihood-estimated by assuming residuals in (1) are normally distributed.
[13] These constitute points on zero coupon curves allowing us to produce simulations of entire curves, with co-movements implicitly linked to all other prices and rates.
[14] Volatilities themselves are simulated over the holding period for each scenario or pathway.
[15] Portfolio weights i.e. net positions, are kept constant for 1 to ten days.

Bund, BTP, Long Gilt, Short Sterling, Euromark, 3 month Swiss, Eurolira respectively). In this section we run three sets of backtests. During the first we keep constant the FX and the implied volatility (i.e. neither is allowed to change during the VaR horizon). In the second set the implied volatility on each option is modelled as a function of the stochastic volatility of the underlying asset; this enables the generation of more realistic scenarios to be generated. In the final set, we generate pathways for FX-to-sterling rate. This enables us to incorporate the FX risk of the pathways of futures and options prices. In most VaR systems FX risk increases the complexity of a VaR system by a large factor. In our case, however, there is no additional computational complexity; the size of the problem increases linearly with the number of currencies in the portfolio.

In each of our three backtests we stored the risk measures of five different VaR horizons (1, 2, 3, 5 and 10 days) and four different probability levels (0.95, 0.98, 0.99 and 0.995). We estimate daily risk measures for about 158 portfolios for a period of 480 days. These values are subsequently compared to the actual ones and the number of breaks is recorded.

2.1 Calculation of Value Losses

Value losses are calculated from price pathways. In a futures or option pathway the contract's close of business price is subtracted from each of the ten pathway prices. This procedure forms a pathway of ten value changes for the contract which may be positive, negative or zero. For example, we have a pathway of futures prices for a given scenario:

Closing Price	day1	Day2	day 3	day 4	Day 5	Day 6	day 7	day 8	day 9	day 10
96.00	97.53	97.94	96.32	97.10	96.02	95.00	95.50	96.48	95.99	96.00

The pathway of value changes measures deviations from the closing price and is:

day1	day2	day 3	Day 4	day 5	day 6	day 7	day 8	day 9	day 10
1.53	1.94	0.32	1.10	0.02	-1.0	-0.50	0.48	-0.01	0.00

Conversion of value changes to a common currency is then performed, for contracts in different currencies, using the pathway of currency exchange rates for the corresponding simulation scenario. The position-weighted value changes for all the contracts are then aggregated to form the portfolio value changes for each of the 5000 scenarios.

The Value at Risk is determined directly at the desired percentile e.g. 99th percentile from the distribution of 5000 portfolio value changes[16] for a given number of days ahead in the holding period.

[16] Value changes were introduced to simplify the calculations when using stochastic currency exchange rates. However, calculating value losses as deviations from closing prices allows certain portfolios e.g. long straddles to have a positive VaR. This will be observed occasionally in backtest 1, which uses constant implied volatility pathways. The defence against this is to estimate stochastic implied volatility pathways: a long straddle position will make a loss if option implied volatilities are reduced. However as the current algorithm for producing implied volatility paths is not stochastic for one day ahead in the holding period, occasional positive VaR's can still be observed, mainly at lower time horizons. While this is not statistically significant, a fully safe approach, which will always give a negative VaR, would be to calculate VaR from the simulated median portfolio value instead of the closing price value. This would not however, address any miss-specification in the implied volatility model. A further alternative would be to alter the algorithm that produces implied volatility pathways.

2.2 *Calculation of Breaks*

2.2.1 Re-marking Portfolios to Market

For a given close of business date, portfolios are re-marked to market for each of the subsequent ten days, to correspond to the days ahead in the holding period. This is done in an analogous fashion to forming value losses in FHS: for the first day ahead in the holding period, the actual closing price for this date has the closing price at the close of business subtracted from it. This is repeated for each of the dates corresponding to the holding period i.e. the close of business price is subtracted from the closing prices on these dates.

The actual portfolio gain or loss from closing price values is therefore computed in the same way that FHS calculates portfolio value changes, using the actual corresponding currency exchange rates.

2.2.2 The Breaks

As VaR's and portfolio actual gains or losses are calculated consistently, they can be compared directly to each other, for the corresponding number of days ahead in the holding period. The objective in FHS is to exceed the actual portfolio losses a certain percentage of the time corresponding to the confidence level used. This means that some of the time the VaR will not be sufficient to cover an actual loss. For example for 99% confidence, "breaks" should occur 1% of the time.

To compute whether a break has occurred:

If (VaR > change in actual portfolio value) then a break has occurred.[17]

2.2.3 Expiring Positions

(i) Positions in delivery (i.e. the contract has ceased trading before the current close of business) are out of scope and are not included in portfolios.

(ii) Futures and options contracts that expire on the close of business[18] are not included in the portfolio (i.e. contracts with positions having zero days to expiry).

(iii) Futures and options contracts that expire the next day are not included in the portfolio (i.e. contracts with positions having one day to expiry).

(iv) Futures and options contracts which expire between 2 and 10 days inclusive during the holding period have previous day values in the pathways "frozen in": for each of the 5000 pathways we keep the simulated price constant at its value one day prior to the expiry day. For example, a contract expires on day 6 during the holding period. One pathway is illustrated:

day1	day2	day 3	day 4	day 5	day 6	day 7	day 8	day 9	day 10
98.55	97.94	96.32	97.10	97.52	*97.52*	*97.52*	*97.52*	*97.52*	*97.52*

The day-5 simulated price is carried forward (in italics) to the remaining days in the holding period.

A different scenario will have different simulated values in the pathway but the same treatment:

[17] Note that as VaR's can occasionally be positive (see footnote 16) then "false breaks" will be recorded if the actual portfolio shows a gain, which is nevertheless smaller in value than the VaR. It is incorrect to record this as a break but the frequency with which this happens during backtesting is not sufficient to significantly bias the results.

[18] We refer to the "close of business" date as the date for which we run the simulation.

day1	day2	day 3	day 4	day 5	day 6	day 7	day 8	day 9	day 10
99.52	98.44	97.99	97.65	96.48	*96.48*	*96.48*	*96.48*	*96.48*	*96.48*

2.3 Actual Portfolios

Actual portfolio values are calculated to reflect the above treatment i.e. a contract's actual closing price one day prior to its expiry date is "copied forward" to the expiry date and subsequent dates corresponding to the holding period. This is done to ensure that re-marking to the market is consistent with FHS.

3 Backtest I, part A (Constant FX & Implied Volatility)

During part A of the first backtest we hold implied volatility constant. Furthermore, to isolate the currency risk from market value risk we translate all returns to sterling at the close of business FX rates. Here are the summary results:

3.1 Overall frequency tests

In Table 1, we show the Number of breaks across all portfolios for the 2-year period (total of 75,835 daily portfolios). The number of breaks[19] recorded across all portfolios for the entire backtest period are reported in each column (D1, D2, D3, D5, and D10). D1, ... D10 are the 1, 2, 3, 5 and 10-day VaR horizons. We record the breaks at each of the four confidence levels (percentiles) used in the backtest. Below each number of breaks we report the corresponding percentage on the number of predictions. The expected number of those percentage breaks should be equal to one minus the corresponding confidence interval.

Table 1 *Breaks across All Portfolios*

c.i.	D1	D2	D3	D5	D10
95%	3583	3701	3864	3580	3438
	4.725%	4.880%	5.095%	4.721%	4.534%
98%	1906	1914	1846	1560	1390
	2.513%	2.524%	2.434%	2.057%	1.833%
99%	1296	1194	1093	927	746
	1.709%	1.574%	1.441%	1.222%	0.984%
99.5%	938	810	711	561	433
	1.237%	1.068%	0.938%	0.740%	0.571%

The percentages of breaks are within 0.5% of the expected ranges, except at high confidence levels and short horizons. However, some portfolios are more likely to report breaks than others are. We identified three such portfolios, "A", "B" and "C". Table 2 reports the aggregate number of breaks on the three portfolios. Below the number of breaks is the percentage of breaks on the number of days these portfolios have traded during the 2-year period. In the line below we show the percentage of the

[19] A break occurs when the portfolio trading loss is greater to the one predicted by BAGV for that VaR horizon.

breaks of these three portfolios on the overall number of breaks. As we can see, the three portfolios account for up to 21.5% on the overall number of breaks[20].

Table 2 *Cumulative Break Analysis for Portfolios A, B, and C*

c.i.	D1	D2	D3	D5	D10
95%	316	256	225	188	123
	22.316%	*18.079%*	*15.890%*	*13.277%*	*8.686%*
	8.819%	*6.917%*	*5.823%*	*5.251%*	*3.578%*
98%	256	197	175	139	74
	18.079%	*13.912%*	*12.359%*	*9.816%*	*5.226%*
	13.431%	*10.293%*	*9.480%*	*8.910%*	*5.324%*
99%	225	175	132	112	60
	15.890%	*12.359%*	*9.322%*	*7.910%*	*4.237%*
	17.361%	*14.657%*	*12.077%*	*12.082%*	*8.043%*
99.5%	201	146	119	91	46
	14.195%	*10.311%*	*8.404%*	*6.427%*	*3.249%*
	21.429%	*18.025%*	*16.737%*	*16.221%*	*10.624%*

Table 3 reports the same statistics as Table 1 excluding the above three portfolios. To gauge the statistical significance of these results, under the hypothesis of independence the above percentages are distributed normally around their expected values, with standard deviations ranging from 0.2% (at 95% level) to 0.06% (at 99.5% level). A two-standard deviation interval may be heuristically doubled to account for the dependence across portfolios, leading to a tolerance of 4 standard deviations. As we can see when the three portfolios are excluded the backtest results marked with an asterisk still show some significant differences from our success criteria.

Table 3 *All portfolios except A, B and C*

c.i.	D1	D2	D3	D5	D10
95%	3267	3445	3639	3392	3315
	**4.390%*	*4.629%*	*4.890%*	**4.558%*	**4.455%*
98%	1650	1717	1671	1421	1316
	**2.217%*	**2.307%*	**2.245%*	*1.909%*	**1.768%*
99%	1071	1019	961	815	686
	**1.439%*	**1.369%*	**1.291%*	*1.095%*	*0.922%*
99.5%	737	664	592	470	387
	**0.990%*	**0.892%*	**0.795%*	*0.632%*	*0.520%*

3.2 Individual Firm Tests

These tests determine whether breaks occur randomly in our sample or cluster for some firms for which risk may be miss-specified. Under the null hypothesis of randomness the number of breaks in the two halves of our backtesting period are independent. Therefore a cross-sectional regression of the breaks, which each firm reports in the first half on the number of breaks reported in the second half, should have zero slope. The regression analysis on the breaks for each sub-period shows a bias at each VaR horizon and each confidence level. However, when the three portfolios mentioned above are excluded from the sample there is no significant correlation between the breaks reported in two sub-periods. Therefore the evidence of miss-specification is limited to those three anomalous portfolios. Table 4 reports some typical results (the significant slope is denoted by asterisk). The values in brackets are the standard errors.

[20] i.e. on 1 day VaR at 99.5% probability.

Table 4 (3 Day VaR horizon at 99%)

All portfolios		Excluding three anomalous portfolios	
a	***b***	***a***	***b***
0.271	0.599*	1.869	0.029
(0.442)	(0.080)	(0.389)	(0.097)

3.3 Time clustering effect

Clustering tests assess whether days with large number of breaks across all the firms tend to be followed by other days with large numbers of breaks, pointing to a miss-specification of the time-series model of volatility. The evidence of that can be detected by autocorrelations in the aggregate number of breaks occurring each day. We applied the Ljung-Box (1978) test and we found no evidence of significant serial correlation's[21] (order 1 to 6) for any confidence level at the 1-day VaR horizon. The overlapping of measurement intervals makes this test not applicable at longer horizons, because portfolios cannot then be regarded as being independent.

Table 5 serial correlation test on 1 day VaR

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
95%	0.03	0.68	2.61	3.29	5.18	5.18
98%	0.92	1.51	1.72	1.78	1.89	2.59
99%	0.02	0.8	0.81	0.87	0.92	1.17
99.50%	0.04	1	1	1.06	1.06	1.17

3.4 Economic Criteria

Breaks are more uniformly distributed when longer time horizons (5 or 10 days) are used. At shorter horizons breaks cluster more (31 Dec 96, 28 Nov 96 and 7 March 96 are the days with the largest concentrations of breaks). The worst day is 31 Dec 96, with 46 breaks adding up to £139M on a VaR of £285M (at 95% level and 1 day horizon). The sum of breaks in the worst day decreases slightly when the horizon increases to 10 days, but it decreases dramatically when the confidence level increases (it drops to 40% of the above figures, down to £42 M at 99.5% level and 10 days).

The sum of VaR's (in Sterling) over the different days for all portfolios ranges within the following intervals:

- 300M to 600M at 95% level and 1 day horizon
- 750M to 1,800 M at 95% level and 10 day horizon
- 400M to 1,000 M at 99.5% level and 1 day horizon
- 1,200M to 3,000M at 99.5% level and 10 day horizon

The ratio of total breaks on the worst day over total VaR ranges from 48% to 2%, depending on the chosen horizon and confidence level. It should be recalled that breaks are computed by not allowing any portfolio change or accounting for cash settlement to market over the VaR horizon. Our aggregate statistics beyond the number of breaks are not significantly affected by the three portfolios identified above, because the sizes of their VaR and their breaks are minuscule (always less than £40,000, mostly close to zero).

[21] Under the null hypothesis of zero correlation our test statistic is a chi-square with six degrees of freedom. At the 95% level, the critical value for the Chi-square test is 12.6.

Figure 1 Sum of ten largest breaks on 10-Day VaR at 99%

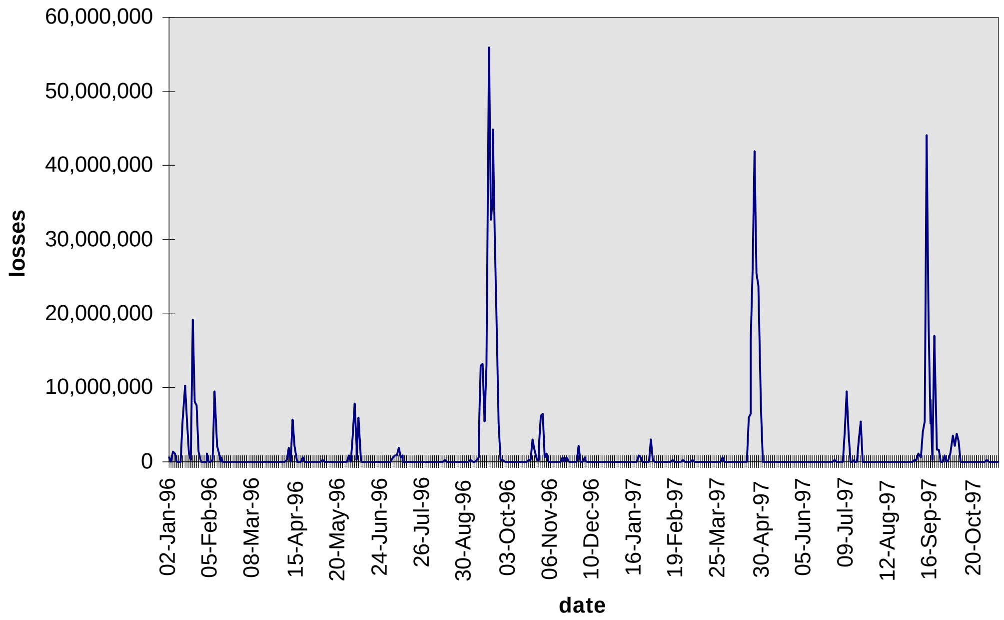


Figure 2 VaR across all portfolios for a 10-day horizon at 99%

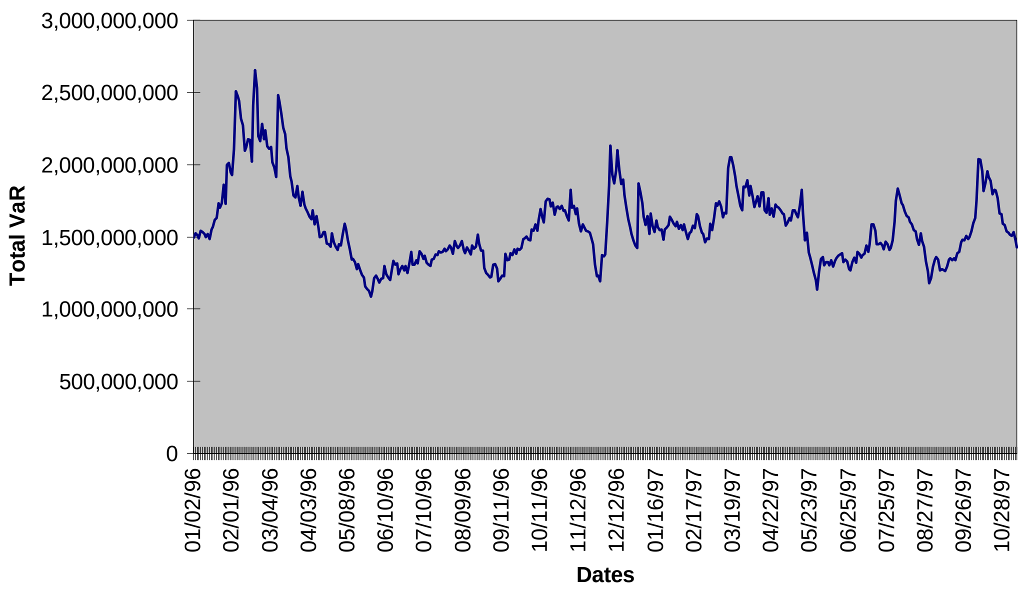
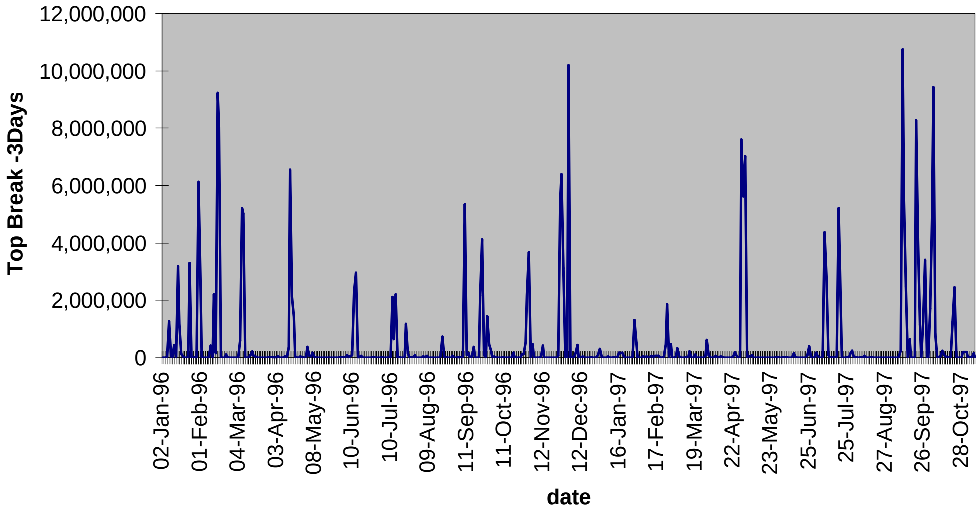
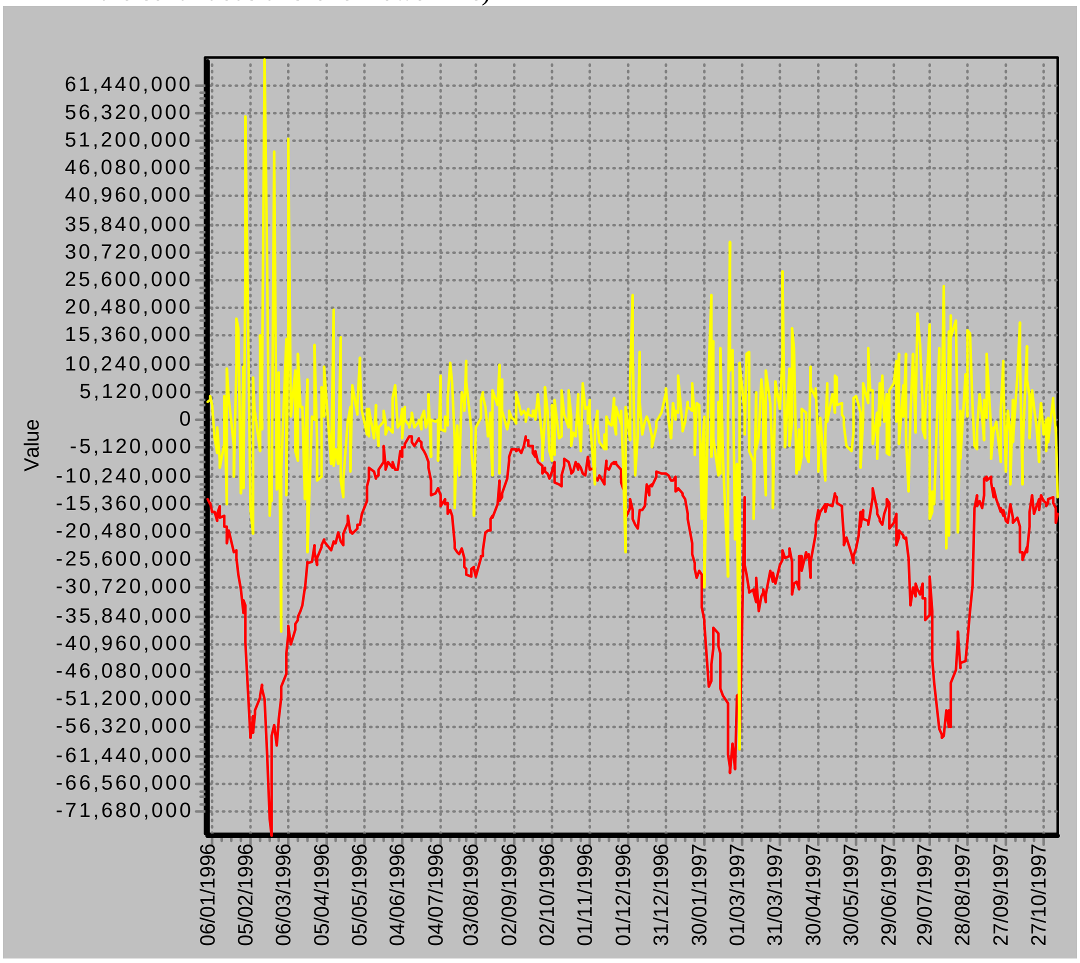


Figure 3 Largest Daily Break for a 3-Day VaR at 99%



Figures 1, 2 and 3 are obtained by aggregating individual member breaks and VaR numbers. A sample figure for one portfolio is reported below where breaks are given by the segments below the continuous VaR curve.

Figure 4 99% Confidence Level at 1 day horizon for a portfolio against profits and losses (VaR is the continuous unbroken lower line)



4 Backtest I, part B (constant **fx, stochastic iv)**

During part B of the first backtest we model implied volatility as stochastic. For each simulation run we create 10-day pathways for implied volatility. The implied volatility pathways are not created ad hoc but are conditional on the futures simulated prices and their, parallel, volatility pathways. The aim is to consider the effect of returns on future implied volatility in a fashion consistent with the evolution of return volatility. To achieve this we compute the integral of expected GARCH volatility from current date to option maturity. This integral equals the product of implied volatility times time to maturity plus a miss-specification error. The evolution of the GARCH volatility at each step in our simulation determines the evolution of implied volatility. The miss-specification error is assumed to be proportional to time left to maturity and is not affected by GARCH innovations.

4.1. Overall Frequency Tests

The main purpose of this backtest is to assess any improvements obtained by simulating future implied volatility. Our approach to simulating i.v. leads to VaR numbers > 0 (see footnote 16) in some instances (their frequency is <0.4% in our sample, mostly at the 1 or 2 day horizon in the presence of extreme volatility shocks). Although the impact of these occurrences on our aggregate results at longer horizons is negligible, safeguards (such as bounding VaR to 0, or taking the larger of the numbers obtained with constant or stochastic i.v., or as mentioned in footnote 16) could be included.

To trigger the implied volatility simulations it is necessary to introduce a time lagged return volatility[22]. This causes the backtest performance to deteriorate at short horizons. At longer horizons the number of breaks for the worst performing firms is reduced. However the number of breaks on the worst days increases because of the anomalous VaR numbers in the presence of volatility shocks.

Table 6 reports the number of breaks found in the second backtest. The performance deteriorates substantially at short horizons because of the necessary introduction of a lag. At longer horizons results become marginally more consistent than the ones in the first backtest. However, the number of breaks at the 95% level is still less than expected.

Table 6 *Breaks across all Portfolios*

c.i.	D1	D2	D3	D5	D10
95%	4981	3639	3640	3449	3444
	*6.56%	4.79%	4.79%	*4.54%	*4.53%
98%	2993	1805	1703	1459	1391
	*3.94%	*2.38%	*2.24%	1.92%	1.83%
99%	2186	1121	1012	833	758
	*2.88%	*1.48%	*1.33%	1.10%	1.00%
99.5%	1680	785	637	482	432
	*2.21%	*1.03%	*0.84%	0.63%	0.57%

Aggregated breaks for the three portfolios that were responsible for the majority of the breaks in part B, backtest 1 are reported in table 7. With the exception of the one-day horizon, the numbers of breaks with stochastic i.v. are smaller than the numbers found with constant i.v. This suggests that the stochastic i.v. model is more capable of modelling the risk of those firms.

[22] i.e. the two GARCH volatility forecasts made on the close of business and on the previous day are deterministic; this causes the first i.v. in the simulation i.v. pathway to be the same value over all 5000 simulations, rather than being stochastic. The effect of this diminishes as we progress along the pathway to the higher time horizons. In fact implied volatility values alter along the pathway to reflect the expected change in total variance up to the option maturity due to the simulated returns used in the futures pathway.

Table 7 *Aggregated Breaks for Portfolios: A, B and C*

	D1	D2	D3	D5	D10
95%	471	149	104	68	83
	33.26%	10.52%	7.35%	4.80%	5.86%
	0.62%	0.20%	0.14%	0.09%	0.11%
98%	418	121	76	39	41
	29.52%	8.55%	5.37%	2.75%	2.90%
	0.55%	0.16%	0.10%	0.05%	0.05%
99%	393	109	63	32	32
	27.75%	7.70%	4.45%	2.26%	2.26%
	0.52%	0.14%	0.08%	0.04%	0.04%
99.50%	374	100	58	25	20
	26.41%	7.06%	4.10%	1.77%	1.41%
	0.49%	0.13%	0.08%	0.03%	0.03%

In table 8 we report the aggregated results of the break analysis for all portfolios excluding A, B and C. The frequency of breaks is now higher for VaR horizons of one day; for longer horizons, however, there are no significant differences compared to the results obtained when implied volatility was fixed rather than stochastic.

Table 8 *Breaks for All Members excluding A, B and C*

	D1	D2	D3	D5	D10
95%	4510	3490	3536	3381	3361
	6.05%	4.68%	4.74%	4.53%	4.51%
98%	2575	1684	1627	1420	1350
	3.45%	2.26%	2.18%	1.90%	1.81%
99%	1793	1012	949	801	726
	2.40%	1.36%	1.27%	1.07%	0.97%
99.50%	1306	685	579	457	412
	1.75%	0.92%	0.78%	0.61%	0.55%

Table 9 reports the numbers of VaR estimates with the wrong sign found in each test (i.e. percentile, holding period). This problem arises because VaR numbers are not computed as differences from the median of portfolio values, but as changes from close of business values (see footnote 16). Volatility shocks may therefore induce the wrong (positive) sign on VaR for some portfolios. Our choice ensures congruence of the second and third backtests (the third backtest estimates stochastic fx rates), but it downgrades the performance of the second backtest at short horizons. Beneath the number of positive VaR's is the percentage of positive VaR's on the total number of predictions.

Table 9 Number of Positive VaR's for All Members

	D1	***D2***	***D3***	***D5***	***D10***
95%	1868	442	251	123	53
	2.46%	0.58%	0.33%	0.16%	0.07%
98%	1386	320	167	65	19
	1.82%	0.42%	0.22%	0.09%	0.03%
99%	1155	269	134	48	12
	1.52%	0.35%	0.18%	0.06%	0.02%
99.50%	977	238	112	34	3
	1.29%	0.31%	0.15%	0.05%	0.00%

4.2. Individual firm tests

The regressions of breaks in the first year over breaks in the second year point to i.v. model miss-specification up to 3-day horizon (denoted by asterisk). Results for 5 and 10-day horizons support the adequacy of the stochastic volatility model for longer horizons. Table 10 reports typical results:

Table 10 (3-Day VaR horizon at 99%)

All portfolios		Excluding three anomalous portfolios	
a	***b***	***a***	***b***
1.383	0.212*	1.927	0.002
(0.437)	(0.090)	(0.379)	(0.083)

4.3. Time clustering effect

The autocorrelation tests for a one-day VaR are reported in table 11. The high values of the test statistics (see footnote 21) point to a miss-specification of the time series properties of the implied volatility model. There is an increased likelihood of a break being closely followed by another break.

Table 11

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
95%	4.88	5.07	7.06	8.53	13.62*	14.59*
98%	11.81*	11.91*	12.17*	14.25*	17.7*	23.99*
99%	12.75*	13.29*	13.37*	15.38*	18.15*	25.02*
99.50%	14.23*	15.9*	16.13*	16.68*	19.68*	28.04*

4.4 Economic criteria

The aggregate VaR across firms is almost identical to the one in the first backtest: the graphs of results are very similar. Only short horizons exhibit higher variability with larger VaR's on high volatility days. At longer horizons, there is a reallocation of total VaR across portfolios, leading to a better coverage of risk on the same total VaR. The sum of breaks on the worst day decreases from £42 million to £35 million at 99.5% and 10-day horizon. The number of breaks on the worst day as well as their size decreases at longer horizons. This reduced clustering points to a better allocation of VaR across portfolios. The ratio of breaks to maximum VaR requirement goes as low as 1.2%. Overall, at longer horizons, i.e. 5 or more days, either the constant or stochastic implied volatility model is adequate. If shorter horizons are chosen some modification of the stochastic implied volatility model is advisable. This can be seen comparing figures 3 and 7.

Figure 5: Sum of 10 largest breaks on 10-Day VaR at 99%

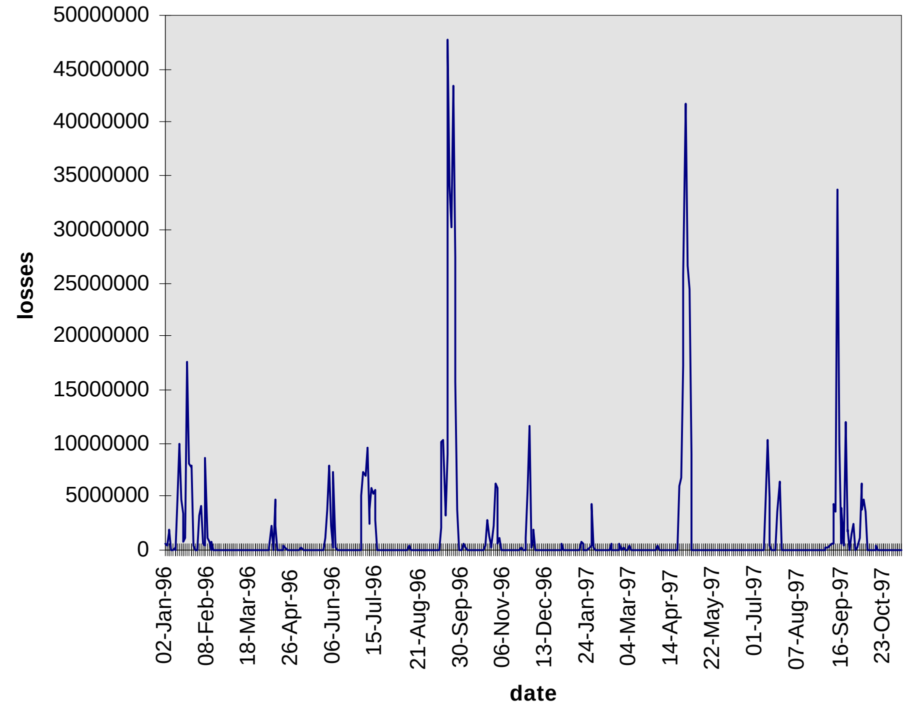


Figure 6 - Daily VaR: Total for All Portfolios // 10Day horizon at 99%

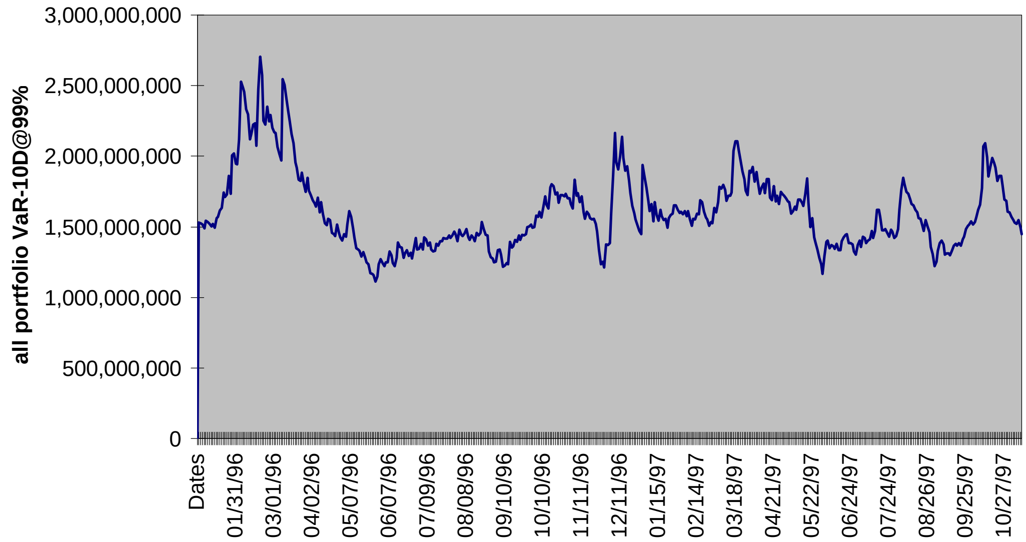
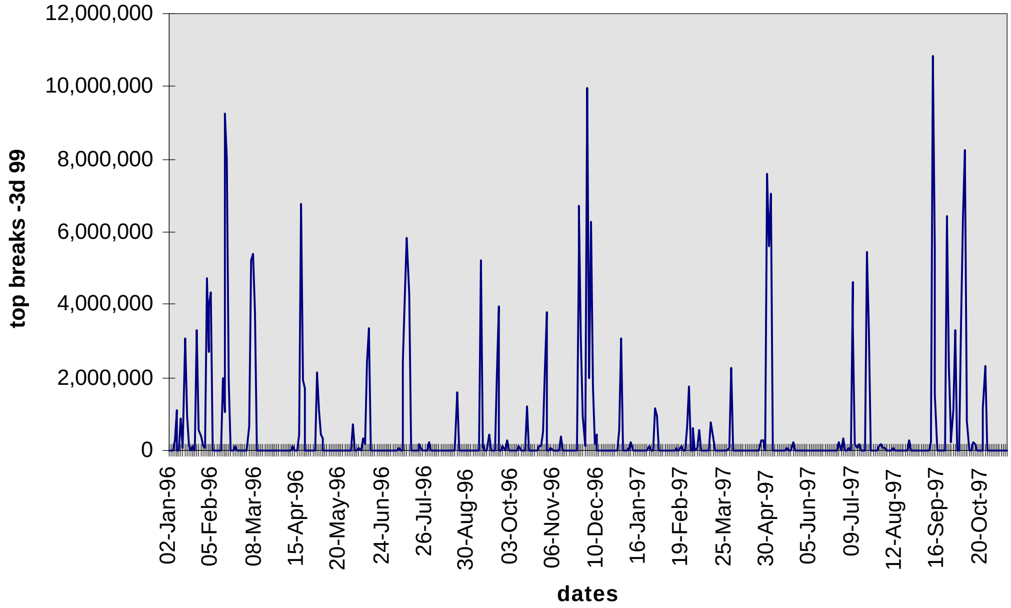


Figure 7 Highest daily breaks - on a 3-day VaR at 99%



5 Backtest I, part C (stochastic iv, fx)

This backtest introduces stochastic currency exchange rate pathways for contracts quoted in currencies other than British sterling. In addition the stochastic implied volatility model for option evaluation is used. For each simulation run we form 10-day pathways for the fx rates which convert non-Sterling prices to Sterling. The actual fx rates for the corresponding days ahead in the holding period are therefore used for portfolio gains/losses, instead of the close of business fx rates.

5.1. Overall Frequency Tests

Tables 12 and 13 report the number of breaks found in the third backtest. They are similar to previous backtests where FX rates were kept constant but the implied volatility was stochastic.

Table 12 Breaks across All Portfolios

c.i.	D1	D2	D3	D5	D10
95%	4968	3634	3628	3428	3410
	6.54%	4.78%	4.78%	4.51%	4.49%
98%	2986	1801	1690	1444	1380
	3.93%	2.37%	2.22%	1.90%	1.82%
99%	2184	1116	1010	820	751
	2.87%	1.47%	1.33%	1.08%	0.99%
99.50%	1679	789	641	484	419
	2.21%	1.04%	0.84%	0.64%	0.55%

Table 13 Breaks for All Portfolios Excluding A, B and C

	D1	D2	D3	D5	D10
95%	4500	3485	3524	3361	3327
	6.04%	4.67%	4.73%	4.51%	4.46%
98%	2568	1680	1614	1406	1339
	3.44%	2.25%	2.16%	1.89%	1.80%
99%	1792	1007	948	788	718
	2.40%	1.35%	1.27%	1.06%	0.96%
99.50%	1305	687	584	460	401
	1.75%	0.92%	0.78%	0.62%	0.54%

5.1.1 Individual portfolios test

Table 14 reports typical results. The regressions of breaks in the first year over breaks in the second year are explained by i.v. model miss-specification up to 3-day horizon (denoted by asterisk) i.e. results are similar to part B.

Table 14 (3Day VaR horizon at 99%)

All portfolios		Excluding three anomalous portfolios	
A	***b***	***a***	***b***
1.446	0.194*	1.942	-0.005
(0.4380	(0.091)	(0.376)	(0.082)

5.3. Time Clustering effect

The autocorrelation tests for a one-day VaR are reported in table 15. Results are similar to part B. The high values of the test statistics (see footnote 21) re-iterate a miss-specification of the time series properties of the implied volatility model.

Table 15 serial correlation (1 day VaR)

	Q1	Q2	Q3	Q4	Q5	Q6
95%	4.94	5.1	7.06	8.59	13.56	14.56
98%	12.16	12.27	12.55	14.54	18	24.34
99%	12.89	13.33	13.42	15.56	18.37	25.56
99.50%	14.44	15.91	16.21	16.85	20.28	28.62

5.4. Economic Criteria

The aggregate VaR is almost identical to the one of the second backtest. Figures 8 to 10 show that. For the portfolios we used, the inclusion of stochastic fx rates does not significantly alter the results achieved in the second backtest.

Figure 8 Sum of ten largest breaks on 10 day VaR at 99%

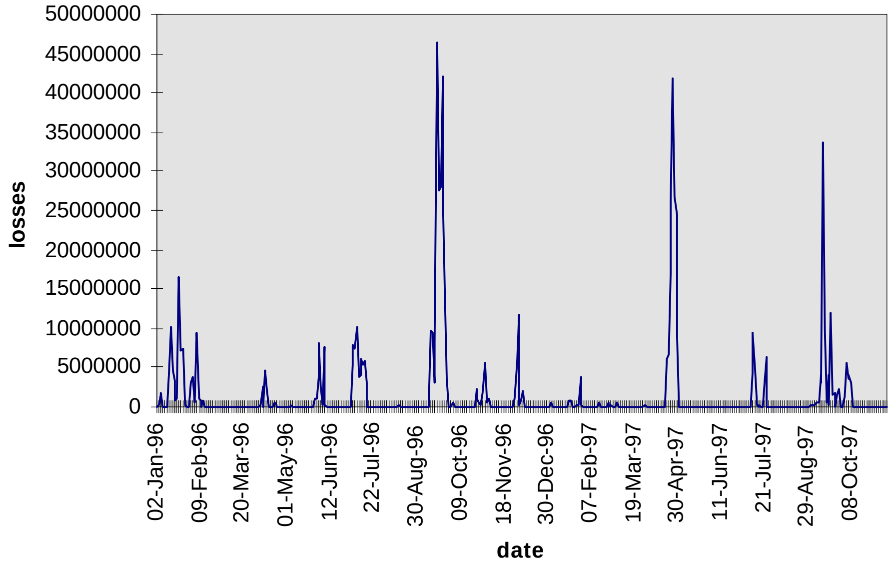


Figure 9 Total VaR on 10 Days at 99%

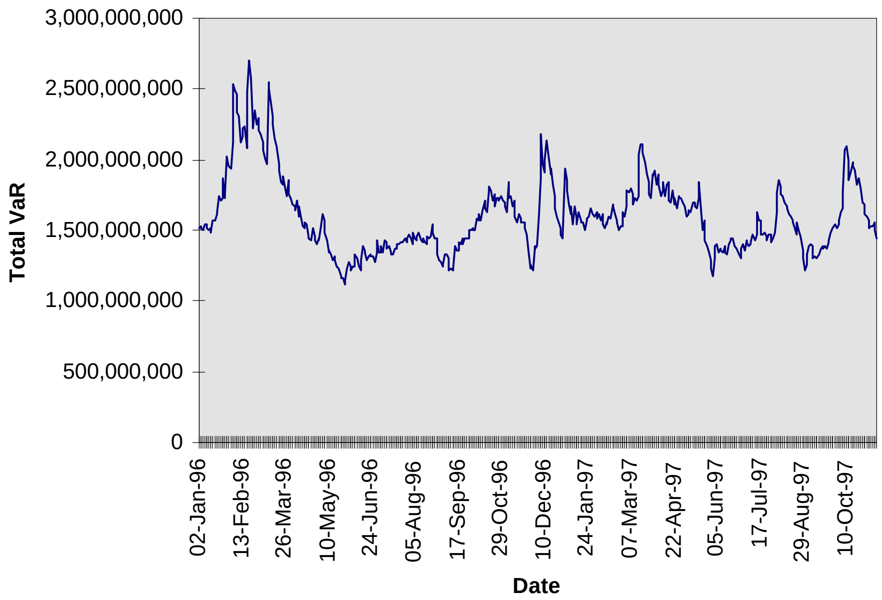
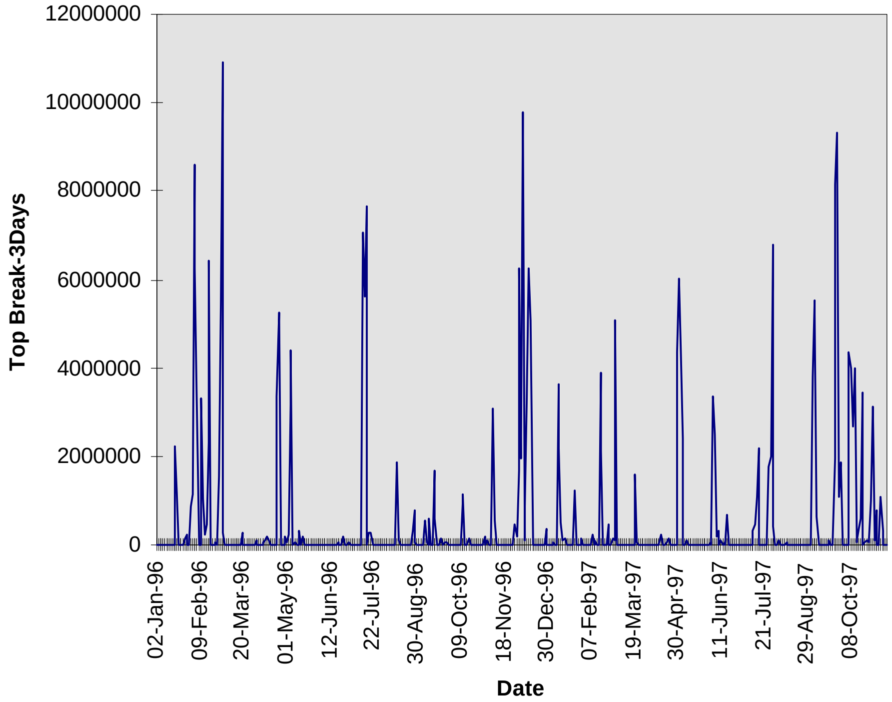


Figure 10 Largest Daily Break for a 3 Days VaR at 99%



Section 2: Backtesting on Swaps

6 The Swaps Backtesting methodology

Following the FHS methodology, for each trading day and each currency we simulate 5000 scenarios for 14 factors from the zero curve based on maturities from overnight (ON) to 10 years (10Y), (1W, 1M, 2M, 3M, 6M, 9M, 12M, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y). On each scenario we price all swaps. Therefore, for each factor[23] 5000 forecasts are generated using that factor's own volatility estimates and historical daily returns. One of the advantages of our approach is to allow the near end of the yield curve to be more volatile than the far end. On the other hand the (parallel) bootstrapping used in FHS provides the means to capture the interdependence between factors across the term structure and also across other currency rates and different financial instruments[24].

We use daily money and coupon rates (swap)[25] for four currencies for the period 2 January 1994 to 10 November 1997. The first two years' data set is used as in-sample database, and provides the feed for our historical simulation. Therefore our backtests start on the first trading date of 1996. At each business date and for each of the four swap portfolios and their currency components we generate 5000 simulated prices for days 1, 2, 3, 5, and 10. The risk measure is determined directly for the desired percentile (e.g. 99th percentile) from the distribution of 5000 portfolio (simulated) values for a given number of days ahead in the VaR period.

A *break* occurs when:

$$P_{a,t+i} < Min_{t+i}$$

where $P_{a,t+i}$ is the portfolio's actual value at the ith VaR day ahead in the holding period and Min_{t+i} the lower predicted value for that probability level.

In order for FHS to meet the statistical criteria the difference in the number of breaks from their expected value must be statistically insignificant (i.e. at 95% probability on average we will have 5 breaks for a portfolio over 100 days.

6.1 *Expiring Swaps and Coupons*

It is important that the portfolios remain identical over the VaR horizon for each business date during the backtest. To satisfy this, for each business day we run FHS we use the following rules:

[23] For a trading date we generated risk measures for 1 to 10 days ahead. Hence for each zero rate factor 5000 ten-day pathways were created.
[24] See Barone-Adesi, Giannopoulos and Vosper (1999) for a description of how BAGV takes into account the possible daily co-movements between different financial instruments.
[25] The data providers are Olsen Associates (money rates) and Dart (swap rates).

- SWAP contracts that expire on the close of business[26] are not included in the portfolio (i.e. contracts with positions having zero days to expiry).
- SWAP contracts that expire the next day are not included in the portfolio (i.e. contracts with positions having one day to expiry).
- Coupons paid during the holding period are ignored (added back) to ensure comparability of results.
- Swap contracts, which expire between 2 and 10 days inclusive during the holding period, have previous day values in the pathways "frozen in" as done in section 1.
- Actual portfolio values are calculated to reflect the above treatment e.g. a swap's actual closing price one day prior to its expiry date is "copied forward" to the expiry date and subsequent dates until the last day of the VaR horizon. This is done to ensure that marking to market is done consistently with FHS.

7 Backtest on 24 small SWAP books

Three sets of backtests are performed. For the first two the same set of 24 portfolios mixed across the four currencies are employed. In the third backtest we use four currency mixed portfolios consisting of 500 swaps each. The scope of the third backtest is to investigate the effects of diversification in larger swap portfolios and to analyse the portfolios that will provide the swap components for the final combo backtest. During the first backtest we translate all returns to Sterling at the close of business FX rates. The scope of the second backtest is to investigate the impact of currency risk. Therefore, during the second backtest we generate pathways for both currency-to-Sterling FX rates and zero coupon rates[27]. The third backtest is run with stochastic FX.

7.1.1 Overall frequency tests

A break occurs when the portfolio trading loss is greater than the one predicted by FHS for that VaR horizon. On Table 16, we show the number of breaks across the 24 portfolios for the 2-year period (total of 11,639 daily portfolios)[28]. The number of breaks recorded across all portfolios for the entire backtest period are reported in each column, D1, D2, D3, D5, D10, to denote the 1, 2, 3, 5 and 10 day VaR horizons.

We record the breaks for four confidence levels (percentiles), 95%, 98%, 99% and 99.5%. Below each number of breaks we report the corresponding percentage on the number of predictions. The expected number for those percentages should be equal to one minus the corresponding confidence level.

Table 16 *Breaks across 24 Portfolios (Constant FX)*

c.i.	D1	D2	D3	D5	D10
95%	495	515	511	496	411
	4.25%	4.42%	4.39%	4.26%	3.53%
98%	234	217	227	203	136
	2.01%	1.86%	1.95%	1.74%	1.17%
99%	133	131	134	109	67
	1.14%	1.13%	1.15%	0.94%	0.58%
99.5%	90	90	85	66	32
	0.77%	0.77%	0.73%	0.57%	0.27%

[26] We refer to the "close of business" date as the date for which we ran the simulation.
[27] Simulated pathways for the FX rates have been generated in the same way as those of the yield rates. The correlation between the former and latter was preserved with the parallel bootstrapping.
[28] For the last 311 days of the first two backtests, there was an additional portfolio.

In table 17 we report the number of breaks and the percentages breaks on the same portfolios. This time the FX rates as well as the yield factors are forecast.

Table 17 *Breaks across 24 Portfolios (Stochastic FX)*

c.i.	D1	D2	D3	D5	D10
95%	519	518	505	479	412
	4.46%	4.45%	4.34%	4.12%	3.54%
98%	233	219	235	214	125
	2.00%	1.88%	2.02%	1.84%	1.07%
99%	143	133	142	113	70
	1.23%	1.14%	1.22%	0.97%	0.60%
99.5%	97	91	85	66	39
	0.83%	0.78%	0.73%	0.57%	0.34%

Results in Tables 16 and 17 show that FHS tends to over-estimate risk slightly. This effect is more pronounced at long horizons. Clearly most table values at short horizons are not significantly different from theoretical values. At longer horizons FHS becomes too conservative. There is no significant difference, in the risk prediction, between the first (constant FX) and second (stochastic FX) backtest.

7.1.2 Individual Portfolio Tests

Individual portfolio tests try to determine whether breaks occur randomly in our sample. Clusters of breaks indicate portfolios for which risk may be miss-specified. Under the null hypothesis of randomness the number of breaks in the two halves of our backtesting period are independent. Therefore a cross-sectional regression of the breaks reported for each firm in the first half on the number of breaks reported in the second half should have zero slope (i.e. *b*=0).

The regression analysis on the breaks for each sub-period in Table 18 confirms that the slope, b, is not significant. There is therefore no evidence of FHS failing to capture the risk of any of our tested portfolios.

Table 18 (3-Day VaR horizon at 99%)

24 portfolios - constant FX		24 portfolios - stochastic FX	
A	***B***	***a***	***b***
3.176*	-0.219	2.659*	-0.023
(1.023)	(0.278)	(0.970)	(0.242)

(numbers in brackets are standard errors)

7.1.3 Time clustering effect

Clustering tests assess whether days with large number of breaks across all the firms tend to be followed by other days with large numbers of breaks, pointing to a miss-specification of the time-series model of volatility. The evidence of that can be detected by the autocorrelations in the aggregate number of breaks occurring each day. The results are displayed in table 19; there is no evidence of significant serial correlations[29] (order 1 to 6) for any confidence level at the 1 day VaR horizon[30].

[29] At the 95% level, the critical value for the Chi-square test is 12.6.

[30] The overlapping of measurement intervals makes this test not applicable at longer horizons, because portfolios can no longer be regarded as being independent.

Table 19 Serial Correlation Test on 1-Day VaR (constant FX)

	Q1	Q2	Q3	Q4	Q5	Q6
95%	1.12	1.15	1.64	2.27	3.62	5.1
98%	2.28	2.95	4.4	4.66	6.09	7.22
99%	0.38	0.41	0.59	1.18	6.2	6.99
99.50%	0.52	0.56	0.58	0.6	4.56	5.4

Similar results were obtained when FX rates were modelled as stochastic.

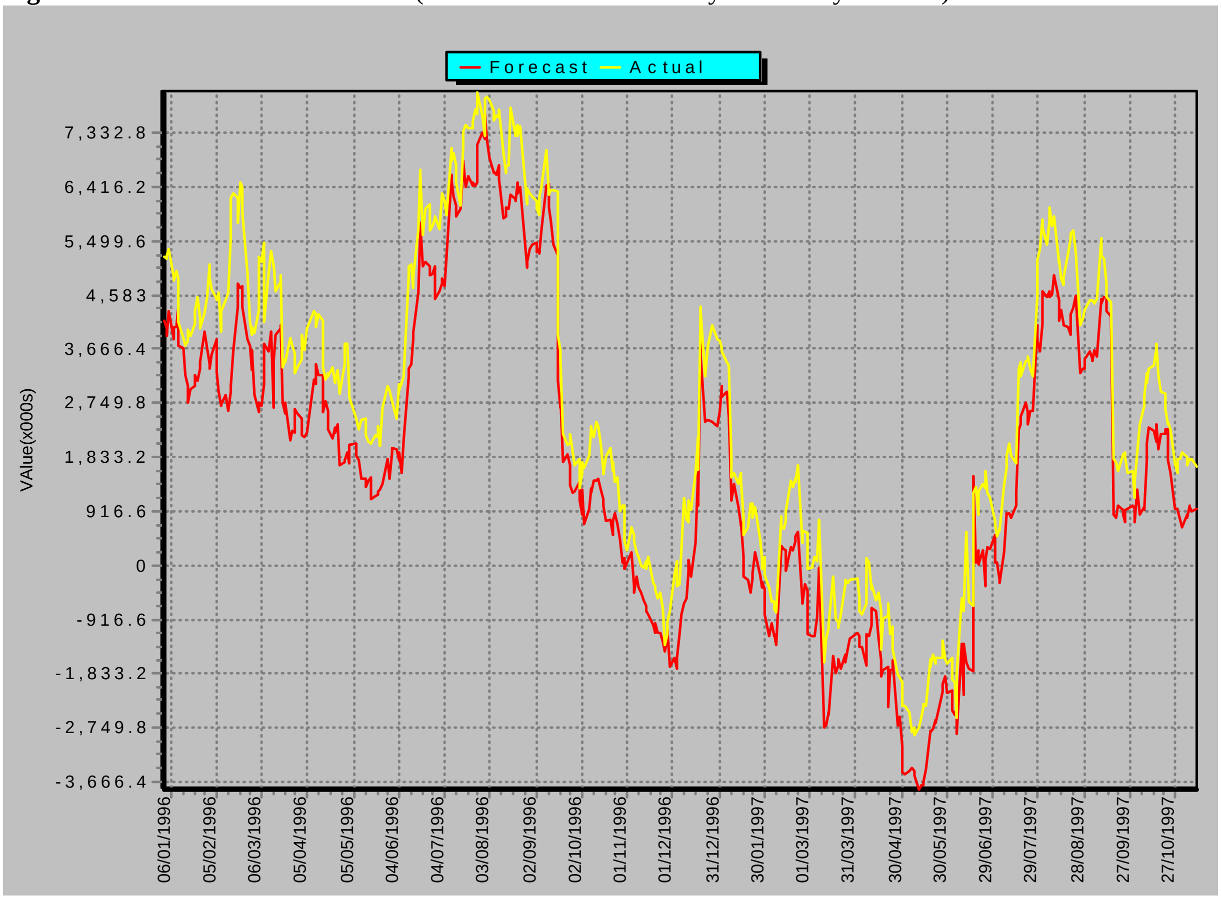
7.2 Economic significance

Table 20 reports the sum of breaks (i.e. losses greater to those predicted by FHS) on the worst day during our backtesting period at different confidence levels and VaR horizons. Note the decreases reported when moving from a one to a two-day horizon. The sums of breaks are a modest percentage of the sums of VaR over all 24 portfolios. Figure 11 reports the evolution of the portfolio values through time (upper line) against our 3-day forecast (lower line) at the 99% level.

Table 20 Cumulative Largest Loss (Break) in One Day (constant FX)

c.i.	D1	D2	D3	D5	D10
95%	3,629	2,628	3,961	4,084	4,433
98%	3,206	1,895	2,109	1,941	2,566
99%	2,871	1,562	1,668	1,409	2,205
99.50%	2,077	1,301	1,311	1,253	1,876

Figure 11 Forecast vs. Actual (Forecast at 99% Probability for a 3-day Horizon)



8 Backtest on four large swap portfolios.

Part C of backtest II is run on four large swap portfolios. Each of these portfolios consists of swaps in all four currencies. They differ in terms of maturity and type of position (pay or receive fixed/floating). The data sample used in the backtest is a sample of large swap books provided by anonymous banks. Table 21 describes these portfolios.

Table 21 Portfolio Description

Portfolio	**Duration**	**Position**
1	Any	Pay fixed
2	short	Pay fixed/float
3	medium	Pay fixed/float
4	any	Pay fixed/float

The analysis for the breaks across our four portfolios is shown in table 22. As in the first two back-tests, violations on the FHS predictions tend to be in line with the probability level for short horizons. For longer horizons, FHS tends to be conservative (over-estimates risk). However the limited number of portfolios leads to large fluctuations in the frequency of breaks, ranging up to 1%.

Table 22 Breaks across Four Portfolios

	D1	D2	D3	D5	D10
95%	72	81	79	65	51
	3.81%	4.29%	4.18%	3.44%	2.70%
98%	36	38	29	24	11
	1.91%	2.01%	1.54%	1.27%	0.58%
99%	25	20	21	9	4
	1.32%	1.06%	1.11%	0.48%	0.21%
99.50%	17	9	3	2	0
	0.90%	0.48%	0.16%	0.11%	0.00%

To investigate whether the results are uniform across the different currencies we break down the four portfolios into the four currency components. Tables 23-26 report the breaks and percentage breaks for each currency component.

Table 23 Breaks across all DEM Portfolios

	D1	D2	D3	D5	D10
95%	67	72	77	76	62
	3.55%	3.81%	4.08%	4.03%	3.28%
98%	30	32	42	25	23
	1.59%	1.70%	2.23%	1.32%	1.22%
99%	20	18	18	13	5
	1.06%	0.95%	0.95%	0.69%	0.27%
99.50%	13	12	7	7	2
	0.69%	0.64%	0.37%	0.37%	0.11%

Table 24 Breaks across all GBP Portfolios

	D1	D2	D3	D5	D10
95%	58	46	59	49	34
	3.07%	2.44%	3.13%	2.60%	1.80%
98%	17	21	20	19	6
	0.90%	1.11%	1.06%	1.01%	0.32%
99%	10	9	7	11	1
	0.53%	0.48%	0.37%	0.58%	0.05%
99.50%	6	3	3	5	0
	0.32%	0.16%	0.16%	0.27%	0.00%

Table 25 Breaks across all JPY Portfolios

	D1	D2	D3	D5	D10
95%	116	81	84	98	77
	6.14%	4.29%	4.45%	5.19%	4.08%
98%	40	36	38	44	15
	2.12%	1.91%	2.01%	2.33%	0.79%
99%	27	25	32	22	8
	1.43%	1.32%	1.70%	1.17%	0.42%
99.50%	22	19	23	10	2
	1.17%	1.01%	1.22%	0.53%	0.11%

Table 26 Breaks across all USD Portfolios

	D1	D2	D3	D5	D10
95%	81	73	78	54	43
	4.29%	3.87%	4.13%	2.86%	2.28%
98%	33	26	32	20	11
	1.75%	1.38%	1.70%	1.06%	0.58%
99%	20	21	17	12	3
	1.06%	1.11%	0.90%	0.64%	0.16%
99.50%	13	15	9	4	1
	0.69%	0.79%	0.48%	0.21%	0.05%

Again, our results confirm that the frequency of breaks on FHS risk predictions closely matches the corresponding confidence level across all currencies examined. There is however, a tendency to under-estimate risk on the JPY book for short horizons. This is not a surprise since there is a dramatic drop on Japanese rates during the second half of the backtest period. Nevertheless this is corrected after few days (on three or longer day horizon). On ten days FHS becomes conservative even for the JPY book.

Section 3: Backtesting on Combined Portfolios

9 Data and methodology

In this final section we analyse the results of backtests performed on 20 portfolios consisting of swaps and interest rate futures and options traded on LIFFE . We combine our four large swap portfolios with 20 portfolios used in the futures and options backtests (see section I). Each swap portfolio is therefore assigned to 5 members. To make the risks for swap and option positions comparable, all the notional amounts of swap contracts are scaled[31].

9.1 The Data

Our data consist of LIFFE futures and options prices and money and coupon rates (swap) and cover the period from 2 January 1994 to 10 November 1997[32]. The first two years of our data set is used as an in-sample database and provides the feed for our historical simulation.

We begin our backtest on the 2nd January 1996 and for each consecutive business date up to 10th November 1997, use FHS to make predictions (pathways), for 1 to 10 days ahead, for all the futures and options[33] contracts traded on that date. Parallel to these predictions we generate the pathways for money and coupon rates; these pathways are used to price each of the 2000 swaps in the database. We then apply the futures and options pathways to each of the futures and options positions[34]. The resulting values are added to the swap simulated prices to compute the pathways for the combo portfolio[35]. The risk measure is determined directly for the desired percentile (e.g. 99th percentile) from the distribution of 5000 portfolio (simulated) values for the desired VaR horizon.

A *break* occurs when:

$$P_{a,t+i} < Mint_{+i}$$

where $P_{a,t+i}$ is the portfolio's actual value at the i^{th} VaR day ahead in the holding period and Min_{t+i} the lower predicted value for that probability level.

[31] We found that by multiplying all swap notionals by a factor of 10, the risk of our 20 portfolios is approximately equally balanced between swaps and interest rate futures and options.
[32] For a full description see sections I and II.
[33] The approach we use to form the pathways for the options is the one followed in the 3rd backtest for LIFFE, stochastic implied volatility and stochastic FX.
[34] A full description on how we handled the expiring contracts is provided in the previous sections.
[35] We apply the FX risk to the market value of the swap portfolio. That assumes that the risk is measured in the domestic currency regardless of the swap currency.

10 Backtesting Results

One set of backtests is performed where the fx rates and implied volatilities are forecast in parallel with the futures, option and swap prices[36]. Since all twenty portfolios are currency-mixed, all forecast prices are translated to GBP using the stochastic FX rates. Similarly, to calculate the breaks we translate all actual contract and swap prices to GBP. The statistical results are reported in this section.

10.1 Overall frequency tests

A break occurs when the portfolio trading loss is greater than the one predicted by FHS for that VaR horizon. In Table 27, we show the number of breaks across the 20 mixed portfolios for the 2-year period (total of 9,440 daily portfolios). The number of breaks recorded across all portfolios for the entire backtest period are reported in each column, D1, D2, D3, D5, D10, to denote the 1, 2, 3, 5 and 10 day VaR horizons.

We record the breaks at four confidence levels (percentiles), 95%, 98%, 99% and 99.5%. Below each number of breaks we report the corresponding percentage on the number of predictions. The expected number of breaks at each confidence level should be equal to one minus the corresponding confidence level.

Table 27 *Breaks across All Portfolios (LIFFE+Swaps)*

	D1	*D2*	*D3*	*D5*	*D10*
95%	402	419	395	321	281
	4.26%	4.44%	4.18%	3.40%	2.98%
98%	203	178	149	117	79
	2.15%	1.89%	1.58%	1.24%	0.84%
99%	126	92	77	50	32
	1.34%	0.98%	0.82%	0.53%	0.34%
99.50%	82	41	32	19	10
	0.87%	0.43%	0.34%	0.20%	0.11%

Most table values at short horizons are not significantly different from theoretical values. However, at longer horizons FHS becomes too conservative.

10.2 Time clustering effect

The results of tests of clustering are displayed in table 33; there is no evidence of significant serial correlation (order 1 to 6) in the numbers of daily breaks for any confidence level at the 1 day VaR horizon.

Table 28 Serial Correlation Test on 1-Day VaR (constant FX)

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
95%	0.94	2.72	3.77	4.01	4.28	4.95
98%	0.72	1.54	1.76	2.02	2.72	3.33
99%	0.59	1.02	1.25	1.48	1.88	2.42
99.50%	0.30	0.60	0.70	0.86	2.15	2.42

[36] Stochastic FX and i.v.

11 Economic significance of backtest results

Table 29 reports the largest breaks (i.e. losses greater to those predicted by FHS) on the worst day during our backtesting period at different confidence levels and VaR horizons[37]. The largest single break on a five day horizon, at the 95%, 98% and 99% levels, occurs on predictions made on 25th November 1996 for a portfolio containing the third swap book. However the largest loss at 99.5% occurs for a portfolio containing the fourth swap book.

Table 29 Largest Loss (Break) in One Day at 5-Day Horizon (in '000's)

c.i.	LIFFE VaR	SWAP VaR	Combo VaR	SWAP minimum	Combo minimum	SWAP actual	Combo actual	Break combo
95%	19,539	48,236	56,604	-481,758	-491,566	-511,164	-539,150	-47,584
98%	24,107	62,496	73,117	-496,015	-508,079	-511,164	-539,150	-30,171
99%	27,253	71,319	83,638	-504,840	-518,600	-511,164	-539,150	-20,550
99.50%	4,617	29,681	34,198	124,726	129,870	132,312	116,219	-13,651

Although the breaks appear to be quite large this is because all swap notional principals have been magnified by a factor of ten[38]. The scaling has been necessary to make similar the size of the risks between the swap and LIFFE components of member portfolios.

Figure 13-14 shows the 1 day VaR analysis at 99% for the portfolios ABC and XYZ respectively over the backtest period. The line in red[39] shows the sum of the daily VaR's of the two main components, LIFFE and swaps. The line in yellow shows the combined VaR, i.e. the VaR when full offsetting is allowed across all LIFFE contracts and swaps in different currencies. We can that, because of diversification, there is a significant reduction in VaR in both portfolios during most of the backtesting period.

[37] The sum of the breaks is not meaningful here because the swap portfolios are not independent (i.e. there are only 4 swap portfolios allocated across 20 futures and options portfolios, a break in one of the swap portfolios will be reflected on 5 futures and options portfolios).
[38] By magnifying each notional by a factor of 10 we emulate the risk of a portfolio of up to 100 times larger.
[39] On the hard copy the red is shown with the upper and the yellow with the lower line.

Figure 12 VaR analysis for portfolio ABC (1-day var. at 99%)

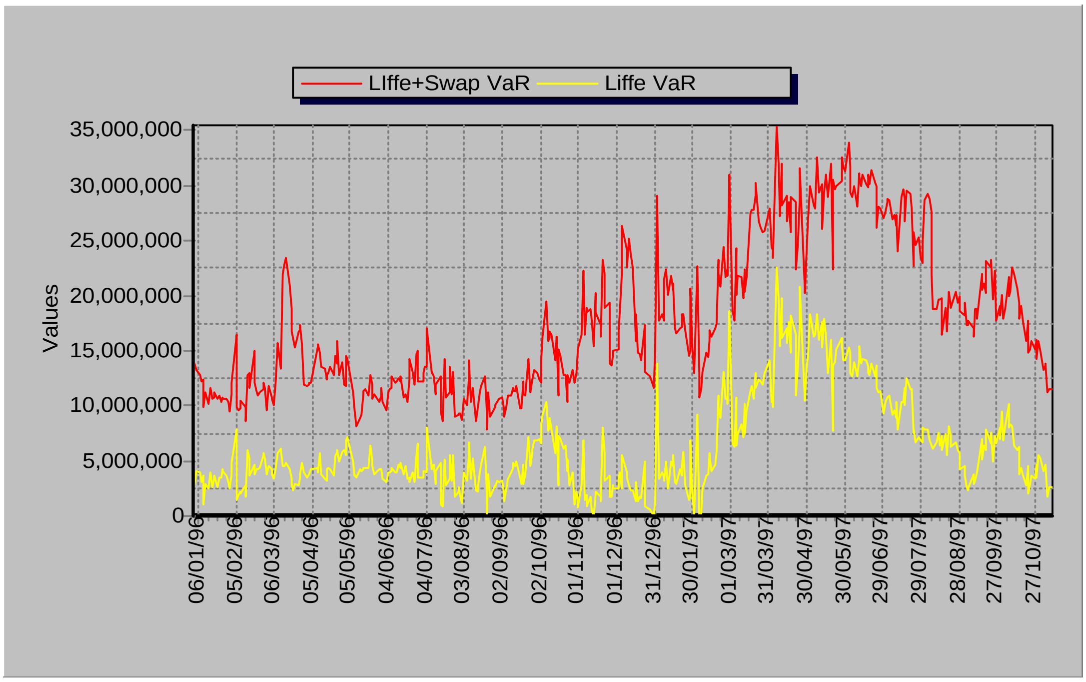
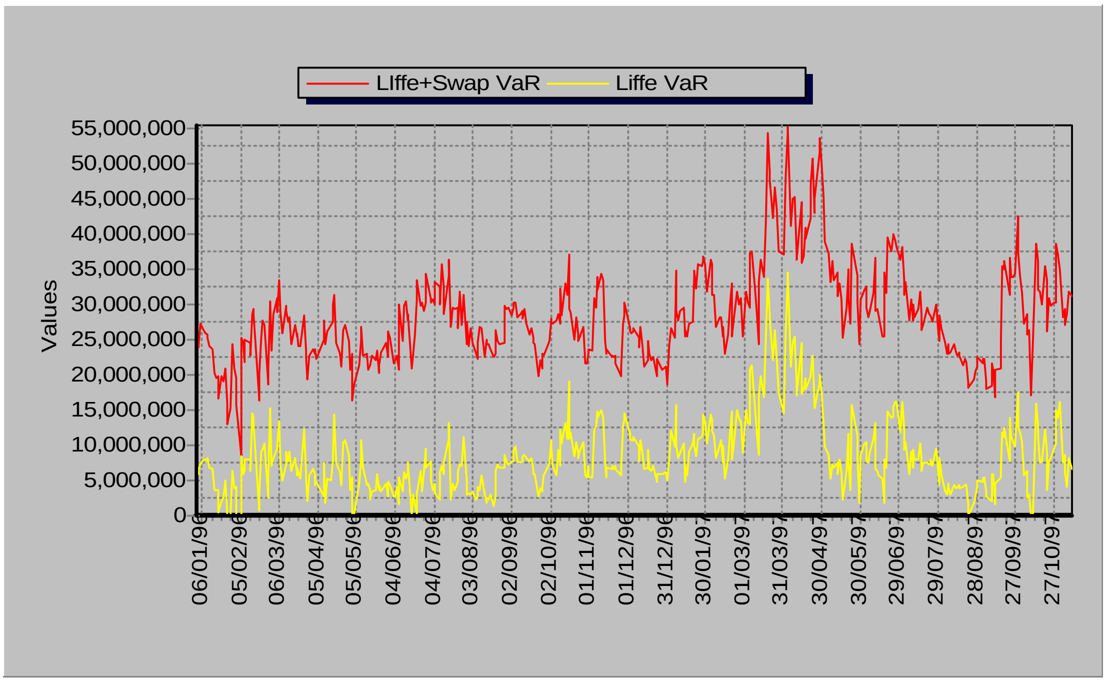


Figure 13 VaR analysis for portfolio XYZ (1-day var. at 99%)



12 Offsetting risk across contracts and currencies

One of the features of FHS is to account for the benefits of risk diversification across different contracts and currencies (swaps). The proportion of risk reduced because of diversification can be measured as follows:

$$\frac{VaR_{t,combo}}{VaR_{t,A} + VaR_{t,C} + VaR_{t,G} + VaR_{t,L} + VaR_{t,S} + VaR_{t,U} + VaR_{t,W} + VaR_{t,DEM} + VaR_{t,GBP} + VaR_{t,JPY} + VaR_{t,USD}}$$

where $VaR_{t,combo}$ is the VaR calculated for day t on the mixed portfolio as described in section 1. $VaR_{t,A}, \ldots VaR_{t,W}$ are the VaR of positions on contracts A to W (LIFFE) for that probability level and investment horizon. The analogous swap (currency) components in each portfolio are shown as $VaR_{t,DEM}, \ldots\ VaR_{t,USD}$. After multiplying each of the eleven VaR numbers by the corresponding GBP FX rate we add them together to create a new series of daily VaR values, as in the denominator of the above ratio. Each of these aggregated daily values reflects the risk associated with the relevant portfolio; we denote this as undiversified risk[40].

[40] This happens under the extreme hypothesis that worst losses (e.g. at 99%) for each LIFFE contract and swap book all occur simultaneously.

Figure 14 Risk Ratio (5-day var. at 99%) for portfolio ABC

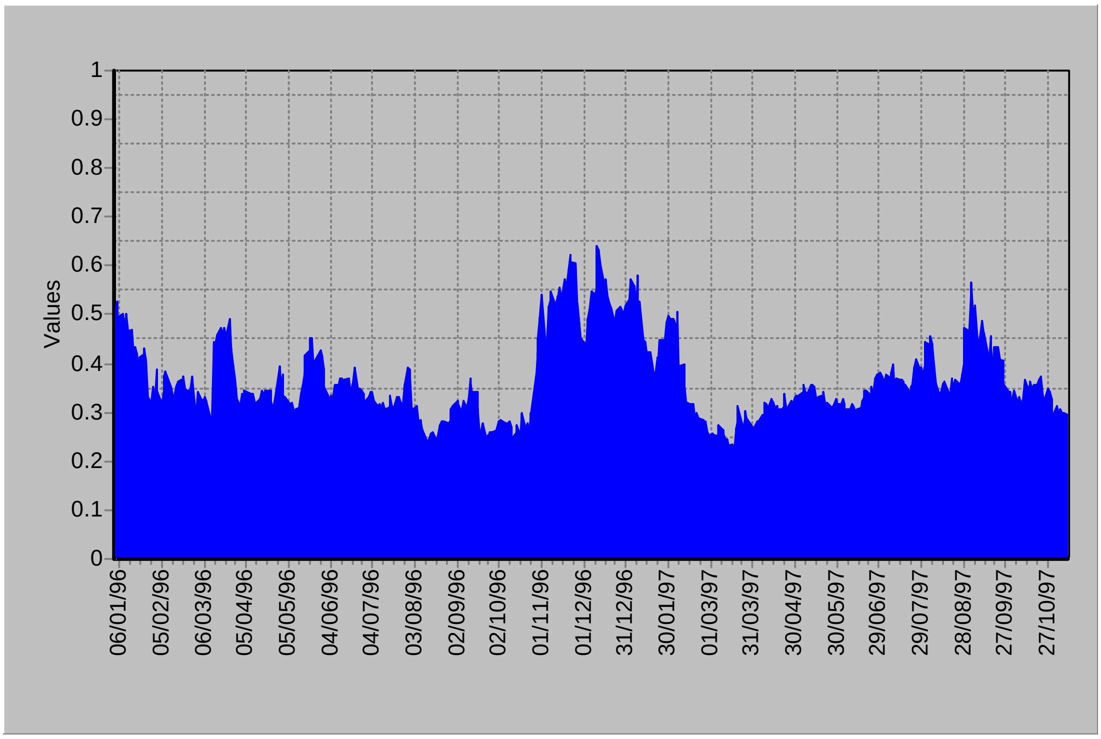
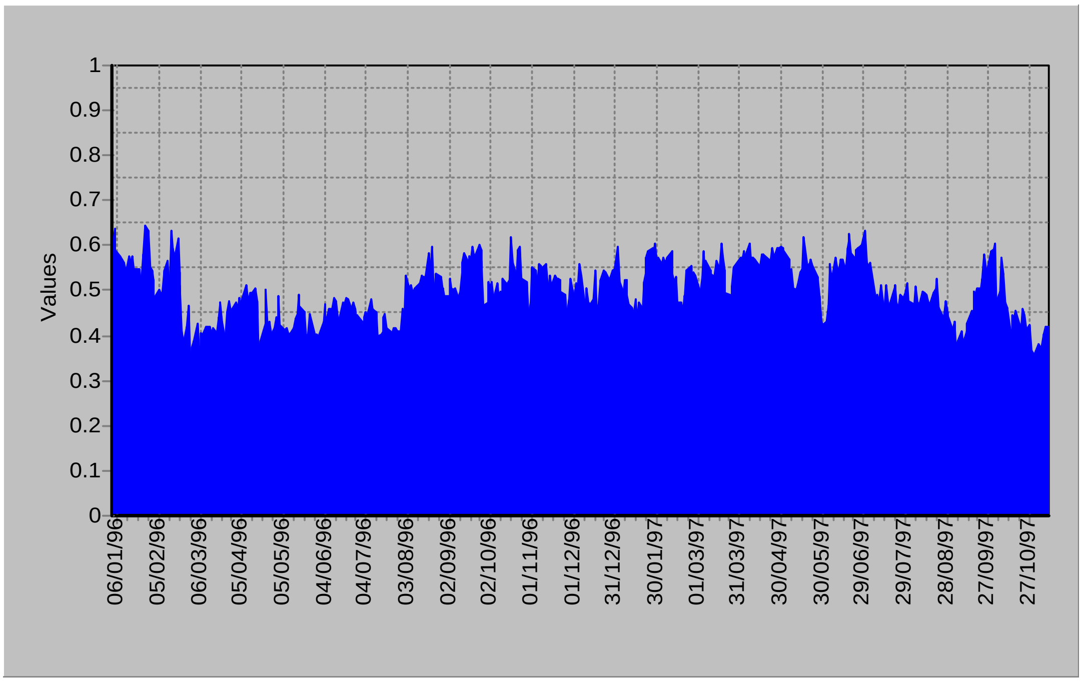


Figure 15 Risk Ratio[41] (5 day VaR @ 99%) for portfolio XYZ



Figures 14 and 15 show the proportion of the combo VaR over the aggregated VaR of individual components of all LIFFE contracts and a single currency swap book for a 5-day VaR at 99%. There is a

[41] This is the VaR of the combo portfolio over the sum of VaR components.

large reduction in risk. In a portfolio the combo VaR can be as low as 20% of the aggregated VaR's. The proportion of risk that is reduced because of diversification is given by 1 minus the above ratio.

13 Conclusions

The results of the three backtests on futures and options indicate that FHS provides an effective framework to measure risk of portfolios of interest rate futures and options. No version of FHS dominates alternative versions in every dimension. Therefore actual implementations should be based on careful consideration of management priorities.

There seems to be little room for improvement in the results of the first backtest, keeping implied volatilities constant. The results of the first backtest are adequate except for the three identified portfolios. However these portfolios are subject to risks that do not fit well in the constant volatility framework.

The second backtest suggests that the stochastic implied volatility model is adequate at long horizons (5 or 10 days). However there is room from improvement of the methodology of the second backtest to reduce clustering of breaks or to improve its performance at short horizons. It is unlikely that these potential improvements would lead to further reductions in break size in the worst days beyond the large ones already observed in our backtest.

The third backtest does not alter the above conclusions. This is to be expected with the futures and options contracts, where currency exchange risk is small.

It is important to remember that our backtest results are based on comparing the VaR value with portfolio profits (losses) on the same horizon (e.g. 3 day VaR vs 3 day portfolio profits /losses). No cash settlement-to-market is allowed for within the VaR horizon. In practice however, FHS VaR numbers at longer horizons could be used to monitor daily risk. Breaks would be very rare when compared to portfolio profits/losses under daily cash settlement to market.

Our swap backtest evaluate the risk of plain vanilla swaps on four currencies (DEM, GBP, JPY, and USD) and for (residual) maturities ranging from 2 days to 10 years. We ran daily simulations on mixed and single currency portfolios for a period of nearly 2 years. Our results indicate that errors in FHS predictions are agreeable with the pre-set VaR probability levels. We find a tendency in FHS to over-estimate risk at longer horizons. The results do not change between single and multi-currency portfolios. The analysis of the risk diversification reveals that a significant amount of risk (in some cases as much as 66% on selected days) is eliminated through diversification across currencies.

Finally, we backtest portfolios invested in interest rate futures and options and plain vanilla swaps. There was no major difficulty in simulating futures, options and swaps using consistent pathways. The overall results show that FHS is accurate at shorter horizons, but tends to become conservative at long horizons, as previously reported for swaps. The fact that VaR values are more stable for swaps than for futures and options may be partly due to our choice of portfolios, but the lack of secondary markets for swaps may contribute to that. Managers in fact may implement their active strategies more easily in markets that allow for secondary trading. In all cases it appears that significant risk diversification is achieved. In fact combining swaps and futures portfolios cuts up to 70-80% of risk.
Our extensive testing suggests that filtered historical simulation, introduced in FHS, is capable of modelling market risk accurately for portfolios of derivative contracts.

BIBLIOGRAPHY

Barone-Adesi G, F Bourgoin and K Giannopoulos (1998), “Don’t Look Back”, *Risk*, **11**, August, 100-104.

Barone-Adesi G and K Giannopoulos (1996), “A Simplified Approach to the Conditional Estimation of Value at Risk”, *Futures and Options World,* October, 68-72.

Barone-Adesi G, K Giannopoulos and L Vosper (1999), “VaR Without Correlations for non-linear Portfolios”, *Journal of Futures Markets*, **19**, August, 583-602.

Basle Committee on Banking Supervision (1996), *Supervisory Framework for the Use of Backtesting in conjunction with the Internal Models Approach to Market Risk Capital Requirements,* Basle.

Bollerslev T (1986), “Generalised Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, **31**, 307-327.

Butler J S and B Schachter (1998), “Estimating Value at Risk by Combining Kernel Estimation with Historical Simulation, *Review of Derivatives Research*, **1**, 371-390.

Kendall M (1953), "The Analysis of Economic time-Series", *Journal of the Royal Statistical Society*, **96,** 11-25.

Jamshidian F and Y Zhu (1997), “Scenario Simulation Model: Theory and Methodology”, *Finance and Stochastics,* **1,** 43-67.

Ljung G.M. and G.E.P. Box (1978), “On a Measure of Lack of Fit in Time Series Models”, *Biometrika*, **67**, 297-303.

Mandelbrot B (1963), “The Variation of Certain Speculative Prices”, *Journal of Business*, **36**, 394-419.

Markowitz H (1959), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley, NY.

RiskMetrics, (1993), 1st edition, JP Morgan publication.