

# **MEASURING TRADED MARKET RISK: VALUE-AT-RISK AND BACKTESTING TECHNIQUES**

Colleen Cassidy and Marianne Gizycki

Research Discussion Paper  
9708

November 1997

Bank Supervision Department

Reserve Bank of Australia

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Reserve Bank of Australia. A number of people (both within the Reserve Bank and from other banks) provided useful comments. We are particularly grateful to Phil Lowe, Brian Gray and the bank that provided the data for testing.

## **Abstract**

The proposed market-risk capital-adequacy framework, to be implemented at the end of 1997, requires Australian banks to hold capital against market risk. A fundamental component of this framework is the opportunity for banks to use their value-at-risk (VaR) models as the basis of the market-risk capital charge. Value-at-risk measures the potential loss on a portfolio for a specified level of confidence if adverse movements in market prices were to occur. This paper examines the VaR measure and some of the techniques available for assessing the performance of a VaR model.

The first section of the paper uses a simple portfolio of two spot foreign exchange positions to illustrate three of the approaches used in the calculation of a VaR measure: variance-covariance, historical simulation and Monte-Carlo simulation. It is concluded that, although VaR is a very useful tool, it is not without its shortcomings and so should be supplemented with other risk-management techniques.

The second section of the paper focuses on the use of backtesting – the comparison of model-generated VaR numbers with actual profits and losses  $z$ – for assessing the accuracy of a VaR model. Several statistical tests are demonstrated by testing daily VaR and profit and loss data obtained from an Australian bank. The paper concludes that, although the tests are not sufficiently precise to form the basis of regulatory treatment of banks' VaR results, the tests do provide useful diagnostic information for evaluating model performance.

JEL Classification Number G21

## Table of Contents

1.	Introduction	1
2.	Value-at-risk	2
2.1	Defining Value-at-risk	3
2.2	An Example Portfolio	3
2.3	The Variance-covariance Approach	6
2.4	The Historical-simulation Approach	8
2.5	Monte-Carlo Simulation	10
2.6	A Comparison of the Three Methods	12
2.7	Advantages and Shortcomings of VaR	13
3.	Backtesting	14
3.1	Shortcomings of Backtesting	16
3.2	The Sample Portfolios	18
3.3	The Regulatory Backtest	21
3.4	Exceptions Testing	23
	3.4.1 Time between failures test	24
	3.4.2 Proportion of failures test	26
3.5	Estimates of Variance	27
3.6	Tests of Normality	28
3.7	Risk Tracking	32
4.	Conclusion	34
	References	36

# **MEASURING TRADED MARKET RISK: VALUE-AT-RISK AND BACKTESTING TECHNIQUES**

**Colleen Cassidy and Marianne Gizycki**

## **1. Introduction**

At the beginning of 1998 the capital-adequacy standards applying to Australian banks will be amended and banks will be required to hold capital against market as well as credit risk. Market risk is the risk that changes in the market prices of financial assets will adversely affect the value of a bank's portfolios. An important component of the proposed capital-adequacy arrangements is the opportunity for banks to use their internal value-at-risk (VaR) models, as opposed to standard regulatory formulae, as a basis for the capital calculation.

VaR is a measure of potential loss, where the potential loss is linked directly to the probability of occurrence of large adverse movements in market prices. The first part of this paper outlines the VaR measure and three different methods that are used in calculating it: the variance-covariance, historical-simulation and Monte-Carlo simulation methods.

The practical implementation of VaR models differs widely across banks. If banks are permitted to use their internally developed methodologies as a basis for regulatory capital requirements, regulators need to be satisfied as to the level of risk coverage and accuracy of those models. Hence, the testing of VaR model performance is a fundamental part of the proposed capital standards.

The second part of the paper discusses a number of tests that can be used to evaluate the performance of a VaR model. Since these tests focus on the past performance of a VaR model such testing is commonly referred to as backtesting. Backtesting assesses the relationship between the estimates of potential loss provided by a VaR model and the actual profits and losses realised by a bank's traders. The backtests are illustrated by applying the tests to VaR and profit and loss data collected from an Australian bank.

## 2. Value-at-risk

Internationally the use of VaR techniques has spread rapidly. This section defines the VaR measure and discusses its use within the Australian banking industry. We then discuss the three methods most commonly used to calculate a VaR estimate – the variance-covariance approach, the historical-simulation approach and the Monte-Carlo simulation approach.

In addition to its inclusion in the Basle Committee on Banking Supervision's guidelines for market risk capital adequacy,<sup>1</sup> a number of other official bodies and industry groups have recognised VaR as an important market-risk measurement tool. For example, the 'Fisher Report' (Euro-currency Standing Committee 1994) issued by the Bank for International Settlements in September 1994 made recommendations concerning the disclosure of market risk by financial intermediaries and advocated the disclosure of VaR numbers in financial institutions' published annual reports. Moreover, private-sector organisations such as the Group of Thirty (1993) (an international group of bankers and other derivatives market participants) have recommended the use of VaR methodologies when setting out best-practice risk-management standards for financial institutions.

VaR is one of the most widely used market risk-measurement techniques by banks, other financial institutions and, increasingly, corporates. Of the banks visited by the Reserve Bank's Market-Risk On-Site review team as at end October 1997, more than half had some form of VaR calculation in place.

The application of VaR techniques is usually limited to assessing the risks being run in banks' treasury or trading operations (such as securities, foreign exchange and equities trading). It is rarely only applied to the measurement of the exposure to the interest-rate and foreign-exchange risks that arise from more traditional non-traded banking business (for example, lending and deposit taking). Use of VaR by Australian banks ranges from a fully integrated approach (where VaR is central to the measurement and internal reporting of traded market risk and VaR-based limits are set for each individual trader) to an approach where reliance is placed on other techniques and VaR is calculated only for the information of senior management or annual reporting requirements. At this stage, it is most common to see VaR used to

---

<sup>1</sup> Basle Committee on Banking Supervision (1996a, 1996b).

aggregate exposures arising from different areas of a bank's trading activities while individual traders manage risk based on simpler, market-specific risk measures.

## **2.1 Defining Value-at-risk**

Value-at-risk aims to measure the potential loss on a portfolio that would result if relatively large adverse price movements were to occur.<sup>2</sup> Hence, at its simplest, VaR requires the revaluation of a portfolio using a set of given price shifts. Statistical techniques are used to select the size of those price shifts.

To quantify potential loss (and the severity of the adverse price move to be used) two underlying parameters must be specified – the holding period under consideration and the desired statistical confidence interval. The holding period refers to the time frame over which changes in portfolio value are measured – is the bank concerned with the potential to lose money over, say, one day, one week or one year. For example, a VaR measure based on a one-day holding period reflects the impact of daily price movements on a portfolio. It is assumed that the portfolio is held constant over the holding period. The Basle Committee's standards require that banks use a ten-day holding period – thus requiring banks to apply ten-day price movements to their portfolios. The confidence level defines the proportion of trading losses that are covered by the VaR amount. For example, if a bank calculates its VaR assuming a one-day holding period and a 99 per cent confidence interval then it is to be expected that, on average, trading losses will exceed the VaR figure on one occasion in one hundred trading days.

Thus, VaR is the dollar amount that portfolio losses are not expected to exceed, with a specified degree of statistical confidence, over a pre-specified period of time.

## **2.2 An Example Portfolio**

A simple portfolio of two spot foreign-exchange positions can be used to illustrate and compare three of the most common approaches to the calculation of VaR. In the following examples, a one-day holding period is assumed and VaR is defined in terms of both 95<sup>th</sup> and 99<sup>th</sup> percentile confidence levels.

---

<sup>2</sup> Value-at-risk may also be termed earnings-at-risk or a potential loss amount.

The example portfolio consists of a spot long position in Japanese Yen and a spot short position in US dollars (Table 1). Thus the value of the portfolio will be affected by movements in the JPY/AUD and USD/AUD exchange rates.

---

**Table 1: Portfolio of Two Positions**

---

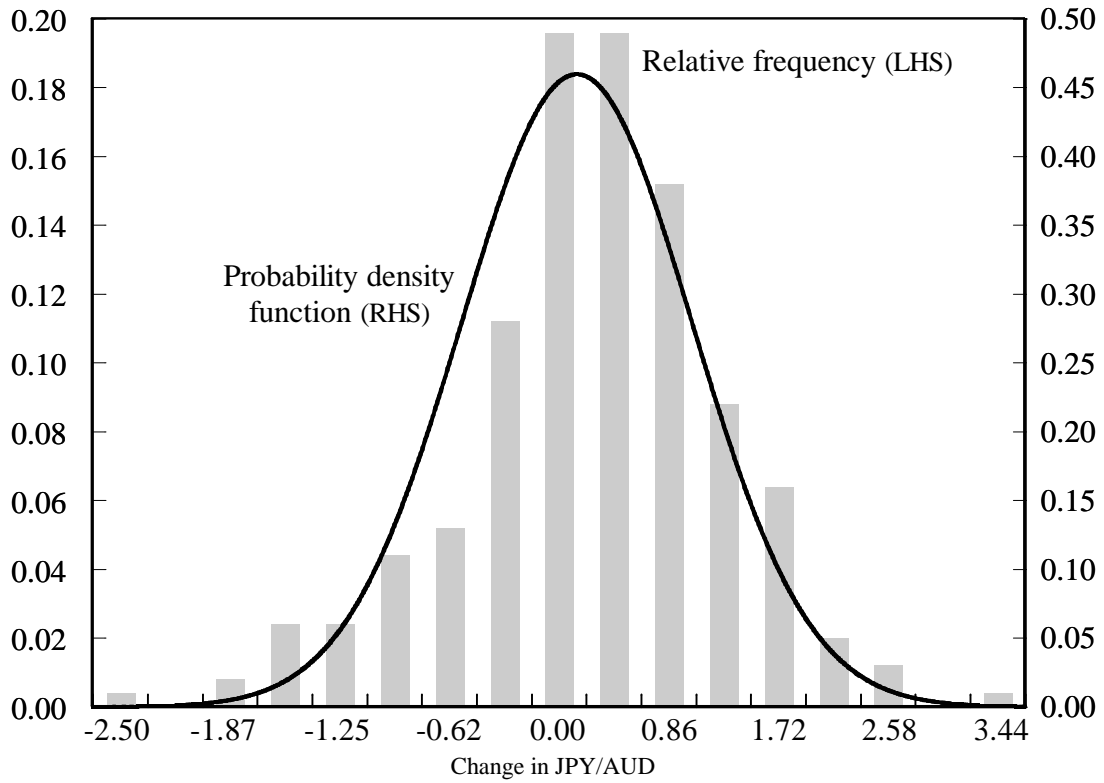
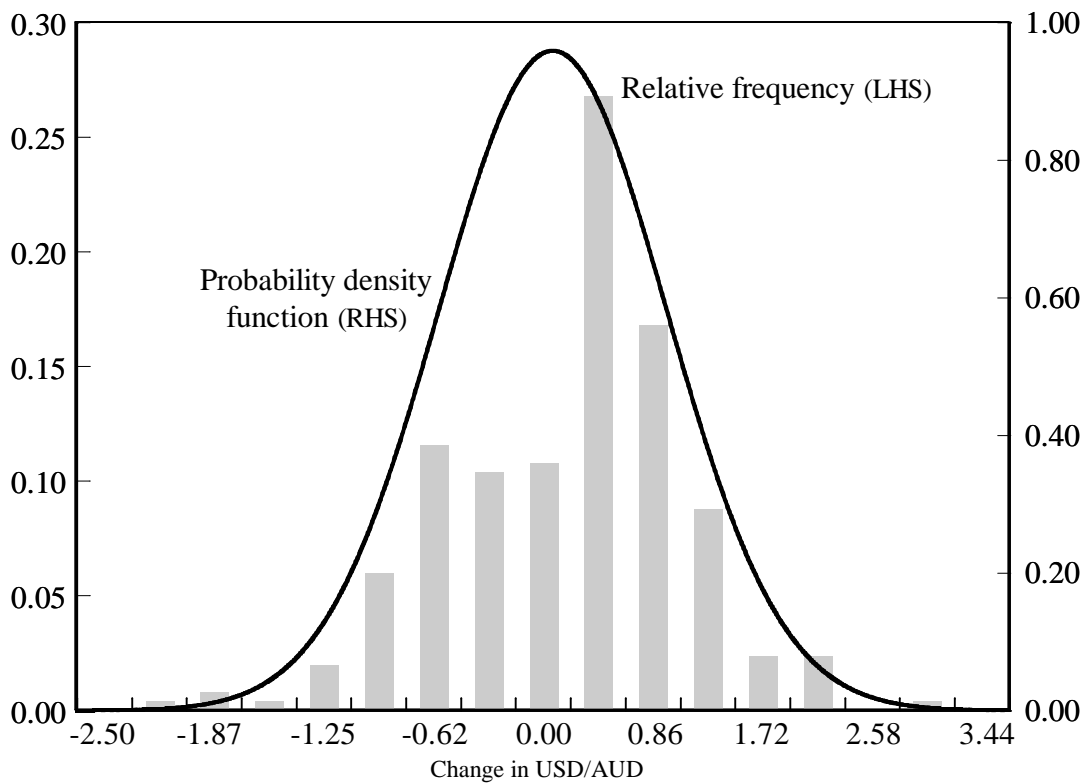
Position 1	100 000	JPY long
Position 2	-10 000	USD short

---

Estimation of a VaR figure is based on the historical behaviour of those market prices that affect the value of the portfolio. In line with the Basle Committee's requirements we use 250 days of historical data, from 9 June 1995 to 5 June 1996, to perform the VaR calculations below.

Figures 1 and 2 are histograms of the daily returns for the JPY/AUD and USD/AUD exchange rates. The smooth line in each chart represents a normal distribution with the same mean and standard deviation as the data. In both the upper and lower tails of each series, the actual frequency of returns is greater than that which would be expected if returns were normally distributed (that is, the observed distributions of daily returns have 'fatter tails' than implied by the normal distribution). Thus both series of daily returns appear more likely to be samples drawn from some distribution other than a normal distribution (such as a t-distribution). The implications of this result for the calculation of a VaR number will be considered later in this paper.

The starting point of all three VaR approaches is to revalue the portfolio at current market prices. Table 2 shows the revalued portfolio given the foreign exchange rates on 5 June 1996.

**Figure 1: Distribution of Daily Returns in JPY/AUD Exchange Rate****Figure 2: Distribution of Daily Returns in USD/AUD Exchange Rate**



**Table 2: Portfolio Value as at 5 June 1996**

	Spot FX rate	Position value	AUD equivalent
Position 1	JPY/AUD 86.46	100 000 JPY	1 156.60 (100 000/86.46)
Position 2	USD/AUD 0.7943	-10 000 USD	-12 589.70 (-10 000/0.7943)

### 2.3 The Variance-covariance Approach

In terms of the computation required, the variance-covariance method is the simplest of the VaR approaches. For this reason, it is often used by globally active banks which need to aggregate data from a large number of trading sites. Variance-covariance VaR was the first of the VaR approaches to be offered in off-the-shelf computer packages and hence, is also widely used by banks with comparatively low levels of trading activity.

The variance-covariance approach is based on the assumption that financial-asset returns and hence, portfolio profits and losses are normally distributed. The consequence of these two assumptions is that VaR can be expressed as a function of:

- the variance-covariance matrix for market-price returns; and
- the sensitivity of the portfolio to price shifts.

The **first** stage of the variance-covariance approach requires the calculation of a variance-covariance matrix using the 250 days of historical data for the two series of daily exchange rate returns. The variance-covariance matrix for this example is expressed as:

$$\begin{aligned}
 M &= \begin{bmatrix} \mathbf{S}_{JPY}^2 & \mathbf{S}_{JPY.USD} \\ \mathbf{S}_{JPY.USD} & \mathbf{S}_{USD}^2 \end{bmatrix} \\
 &= \begin{bmatrix} 0.753 & 0.228 \\ 0.228 & 0.173 \end{bmatrix}
 \end{aligned}$$

where  $\mathbf{s}_{JPY}^2$  is the variance of the series of daily returns for JPY/AUD,  $\mathbf{s}_{USD}^2$  is the variance of the series of daily returns for USD/AUD and  $\mathbf{s}_{JPY.USD}$  is the covariance between the two series.

The **second** step in this approach is to calculate the market price sensitivities or *deltas* of the portfolio; that is, the amounts by which the portfolio's value will change if each of the underlying market prices change by some pre-specified amount. To do this, movements in each of the market prices which affect the value of the portfolio are examined separately. Table 3 shows the change in the portfolio given a 1 per cent move in each of the spot FX rates.

<b>Table 3: Methodology for Calculating the Delta of Each Position in the Portfolio</b>			
	Current	Revalued (assuming a 1% increase in AUD)	
<b>FX rates</b>			
JPY/AUD	86.46	87.32	(1.01 x 86.46)
USD/AUD	0.7943	0.8022	(1.01 x 0.7943)
<b>Portfolio value (AUD)</b>			
Position 1	1 156.61	1 145.15	(100 000 / 87.32)
Position 2	-12 589.70	-12 465.05	(-10 000 / 0.8022)
<b>Change in portfolio value or delta (AUD)</b>			
Position 1	-11.45		
Position 2	124.65		

The **third** step in this approach is to calculate the standard deviation or volatility of total changes in portfolio value. Since total portfolio changes are assumed to be normally distributed, the volatility of portfolio changes can be expressed as a function of the deltas, the standard deviations of the two market-factor returns and the covariance between them. Let  $\mathbf{c}$  be the vector of market-price sensitivities or deltas. If the standard deviation of portfolio changes is  $v$  and the variance-covariance matrix of the market prices is  $\mathbf{M}$  then  $v$  is expressed as:

$$v = \sqrt{\mathbf{cM}\mathbf{c}}$$

In this example  $v$  is given by:

$$v = \sqrt{[d_{JPY} \quad d_{USD}] \begin{bmatrix} s_{JPY}^2 & s_{JPY.USD} \\ s_{JPY.USD} & s_{USD}^2 \end{bmatrix} \begin{bmatrix} d_{JPY} \\ d_{USD} \end{bmatrix}} \\ = 46.22$$

The standard deviation of changes in the portfolio's total value is 46 AUD.

To establish the VaR number of the portfolio for a given level of confidence the standard deviation must be multiplied by the relevant scaling factor, which is derived from the standard normal distribution. For example, if a 99 per cent level of confidence is desired the appropriate scaling factor is 2.33 since the probability of occurrence of a number less than -2.33 is 1 per cent. Scaling the standard deviation of the portfolio by this amount yields a VaR number which should only be exceeded 1 per cent of the time. Note that the choice of a 99 per cent confidence level and associated scaling factor of 2.33 assumes a one-tailed test in line with the Basle market risk requirements (that is, only large losses are considered, not large profits).

Table 4 shows the VaR amounts, given 95 and 99 per cent levels of confidence, for the example portfolio. Clearly the higher the level of confidence, the larger the VaR number will be: given the various assumptions there is a 5 per cent probability that the loss on the portfolio will exceed 76 AUD and only a 1 per cent probability that the loss on the portfolio will be larger than 108 AUD.

---

**Table 4: Value-at-risk using the Variance-covariance Approach**

---

Confidence level	Scaling factor	Value-at-risk number
95 per cent	1.645	76.02 AUD (46.21x1.645)
99 per cent	2.330	107.67 AUD (46.21x2.33)

---

## 2.4 The Historical-simulation Approach

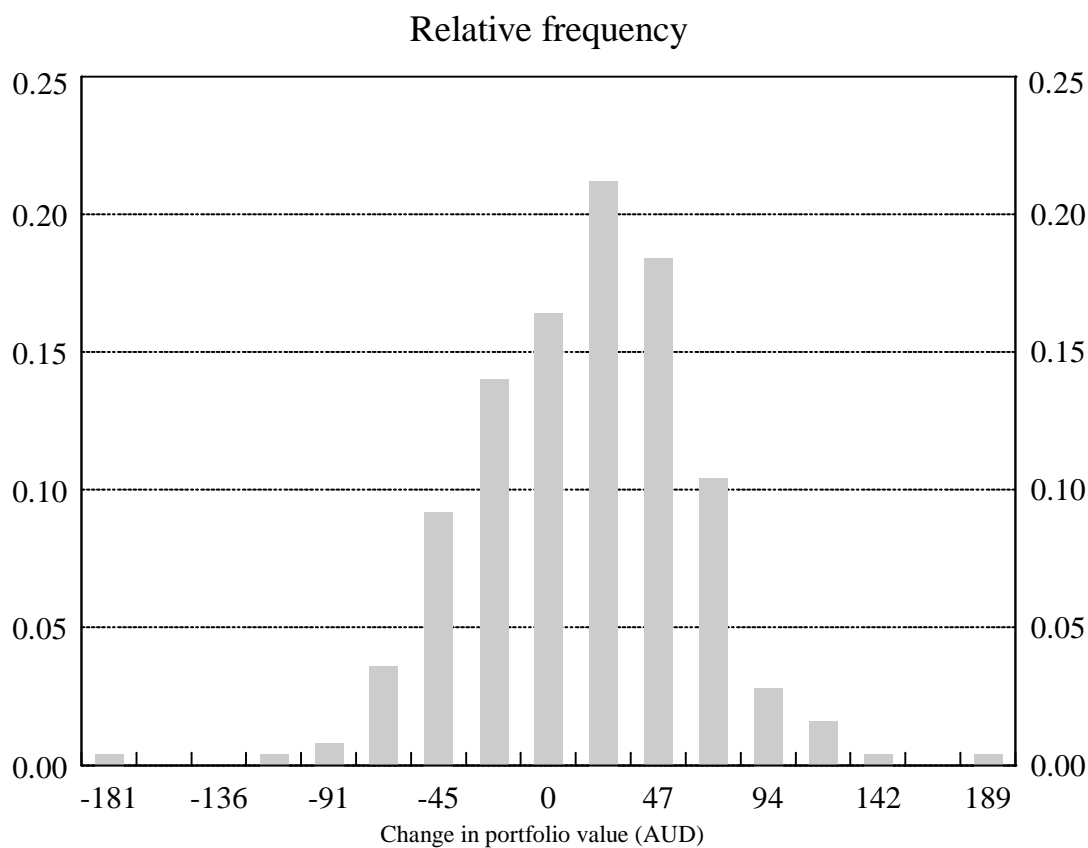
The historical-simulation method is more computationally intensive than the variance-covariance approach and its use emerged within the Australian banking industry a little later. While only three banks have been using historical simulation for some time, the development of historical databases of market prices, together

with more powerful (and less expensive) computer technology, has led several other banks to move towards the use of this approach.

The historical-simulation approach also uses historical data on daily returns to establish a VaR number, however, it makes no assumptions about the statistical distribution of these returns. The **first** step in this approach is to apply each of the past 250 pairs of daily exchange rate movements to the portfolio to determine the series of daily changes in portfolio value that would have been realised had the current portfolio been held unchanged throughout those 250 trading days.

To determine the revalued portfolio value two approaches can be used. The simpler approach requires the previously calculated delta amount for each position to be multiplied by each of the past changes in the relevant exchange rate. Recall that delta measures how much the position value will change if the exchange rate changes by 1 per cent. If the past actual change in the exchange rate is, say, 0.16 per cent then the portfolio value will change by  $0.16 \times \text{delta}$ . The second, more arduous approach is to revalue each position in the portfolio at each of the past exchange rates. For linear positions (that is, positions the values of which change linearly with changes in the underlying market prices) the two approaches will yield the same result. However, for non-linear positions, such as positions in complex options, the first approach may substantially under or overestimate the change in the value of the position and thus may not generate an accurate measure of market risk exposure.

The **second** step is to sort the 250 changes in portfolio value in ascending order to arrive at an observed distribution of changes in portfolio value. The histogram of these changes is shown in Figure 3. The VaR number will be equal to that percentile associated with the specified level of confidence. For a 95 per cent level of confidence, the VaR number is 68.17 AUD and equals the 5<sup>th</sup> percentile of the distribution of changes in portfolio value. The k<sup>th</sup> percentile means that the lowest k per cent of the sample of changes in portfolio value will exceed the VaR measure. Since there are 250 observations, essentially this means that 12.5 losses (or 5 per cent of the sample) will be larger than the VaR measure (the VaR measure is essentially the 13.5 lowest observation). Similarly, for a 99 per cent level of confidence the VaR number is 102.11 and equals the first percentile. These results are summarised in Table 5.

**Figure 3: Histogram of Changes in Portfolio Value****Table 5: Value-at-risk Using the Historical-simulation Approach**

Confidence level	Value-at-risk number
95 per cent	68.17 AUD
99 per cent	102.11 AUD

## 2.5 Monte-Carlo Simulation

This method is not widely used by Australian banks. Monte-Carlo techniques are extremely computer intensive and the additional information that these techniques provide is of most use for the analysis of complex options portfolios. To date, use of Monte-Carlo simulation has been limited principally to the most sophisticated banks and securities houses operating in the US. The Monte-Carlo method is based on the generation, or simulation, of a large number of possible future price changes that could affect the value of the portfolio. The resulting changes in portfolio value are then analysed to arrive at a single VaR number.

Briefly, the method requires the following steps.

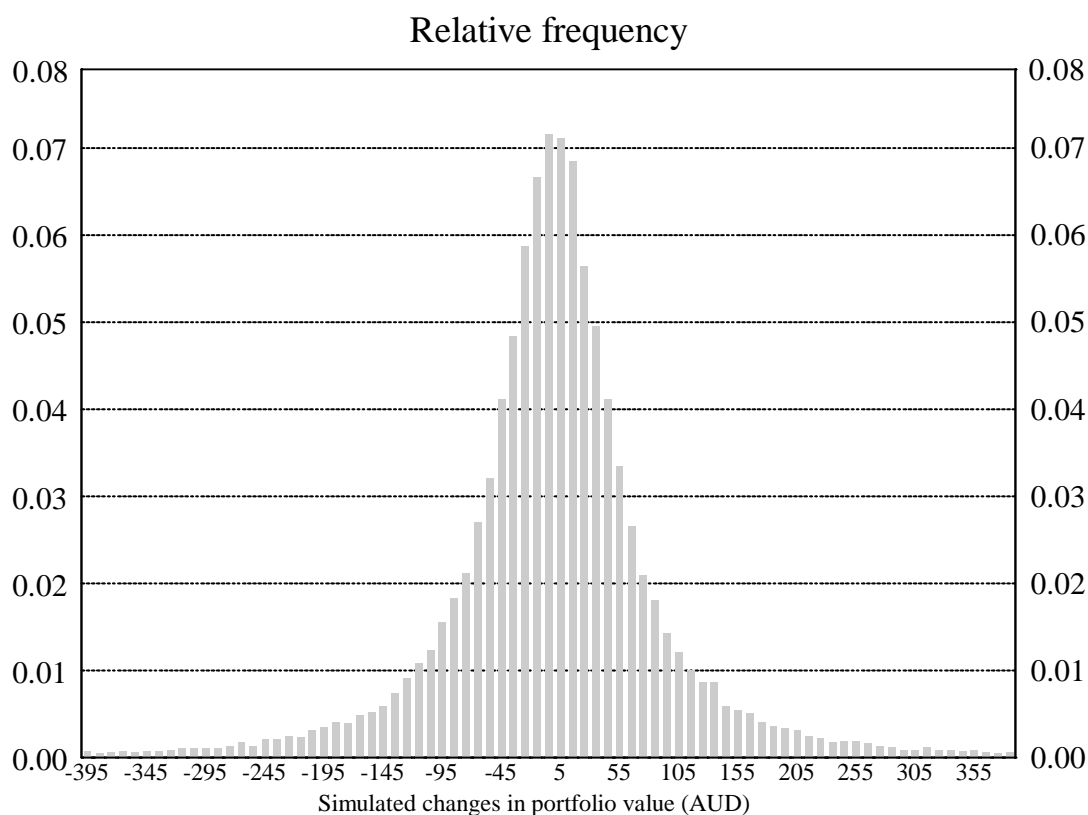
1. A statistical model of the market factor returns must be selected and the parameters of that model need to be estimated. For the purposes of our example, it is assumed that the two exchange-rate returns are drawn from a bivariate t-distribution with 5 degrees of freedom and a correlation of 0.63.<sup>3</sup> A t-distribution was chosen as it is able to capture the fat-tails characteristic observed in the data.
2. A large number of random draws from the estimated statistical model are simulated. This is done using a sampling methodology called Monte-Carlo simulation in which a mathematical formula is used to generate series of 'pseudo-random' numbers to simulate the market factors. In this example, the two exchange rates are simulated 50 000 times.
3. The portfolio is revalued for each pair of simulated exchange rates and the changes in portfolio value between the current value and these revalued amounts are then determined. Figure 4 shows the histogram of these changes.

In the same way as in the historical simulation approach, these changes in portfolio value are sorted in ascending order and the VaR number at a k per cent level of confidence is determined as the (100-k) percentile of these sorted changes. The resulting VaR measures are shown in Table 6.

---

<sup>3</sup> Maximum likelihood estimation of the degrees of freedom for a univariate t-distribution for each of the exchange rate returns series yielded an estimate of 5 degrees of freedom for both the USD/AUD and the JPY/AUD rates.

**Figure 4: Distributing of the Changes in Portfolio Value Given the Simulated Exchange Rate Levels**



**Table 6: Value-at-risk using the Monte-Carlo Simulation Approach**

Confidence level	Value-at-risk number
95 per cent	157.96 AUD
99 per cent	356.10 AUD

The Monte-Carlo process permits analysis of the impact of events that were not in fact observed over the historical period but that are just as likely to occur as events that were observed. It is this capacity to evaluate likely events that have not occurred that is one of the main attractions of this approach.

## 2.6 A Comparison of the Three Methods

The VaR numbers derived from the three approaches produce a wide range of results (Table 7). In this example, the historical simulation method which takes into account the actual shape of the observed distribution of profits and losses (shown in

Figure 3) yields the lowest risk estimates. The variance-covariance method's assumption of symmetry around a zero mean gives equal weight to both profits and losses, resulting in VaR estimates which are slightly higher than those of the historical simulation approach. The simulation of a bivariate t-distribution results in VaR estimates which are much larger than the estimates given by the other two methods. A t-distribution with the same mean and variance as a normal distribution will have a greater proportion of its probability mass in the tails of the distribution (in fact, in this case, the t-distribution also has longer tails than the empirical distribution). The prime focus of a VaR model is the probability of tail events, hence, the long tails of the t-distribution have a disproportionate effect on the VaR estimate. It can be seen that this effect becomes more marked the higher the confidence level. It should be noted that this ranking of results from the three methods is dependent on the data and also the statistical distribution used within the Monte-Carlo simulation technique. Other price series exhibiting different mean, skew and tail characteristics may result in the relative sizes of the three methods' VaR estimates being quite different.

---

**Table 7: Summary of VaR Results**

---

	95 per cent	99 per cent
Variance-covariance	76.02 AUD	107.67 AUD
Historical-simulation	68.17 AUD	102.11 AUD
Monte-Carlo simulation	157.96 AUD	356.10 AUD

---

## **2.7 Advantages and Shortcomings of VaR**

While VaR is used by numerous financial institutions it is not without its shortcomings. First, the VaR estimate is based solely on historical data. To the extent that the past may not be a good predictor of the future, the VaR measure may under or overestimate risk. There is a continuing debate within the financial community as to whether the correlations between different financial prices are sufficiently stable to be relied upon when quantifying risk. There is also debate as to how best to model the behaviour of volatility in market prices. Nevertheless, if an institution wishes to avoid relying on subjective judgments regarding likely future financial market volatility, reliance on history is necessary.



Second, a VaR figure provides no indication of the magnitude of losses that may result if prices move by an amount which is more adverse than that amount dictated by the chosen confidence level. For example, the dollar VaR provides no insight into what would happen to a bank if a 1 in 10 000 chance event occurred. To address the risks associated with such large price shifts, banks are developing, and bank supervisors are requiring, more subjective approaches such as stress testing to be adopted in addition to the statistically based VaR approach. Stress testing involves the specification of stress scenarios (for example, the suspension of the European exchange rate mechanism) and analysis of how banks' portfolios would behave under such scenarios.

Third, the comparative simplicity of a VaR calculation where exposures in a wide array of instruments and markets can be condensed into a single figure is both a strength and a weakness. This simplicity has been the key to the popularity of VaR, particularly as a means of providing summary information to a bank's senior management. The difficulty with this though, is that such a highly aggregate figure may mask imbalances in risk exposure across markets or individual traders.

One of the chief advantages of the VaR approach is that it assesses exposure to different markets (interest rates, foreign exchange, etc) in terms of a common base – losses relative to a standard unit of likelihood. Hence, risks across different instruments, traders and markets can be readily compared and aggregated. In addition, a dollar-value VaR can be directly compared to actual trading profit and loss results – both as a means of testing the adequacy of the VaR model and to assess risk-adjusted performance. Risk-adjusted returns can be quantified simply by looking at the ratio of realised profits/losses to VaR exposure. Such calculations provide a basis for a bank to develop sophisticated capital-allocation models and to remunerate individual traders not just for the volume of trading done, but to reflect the riskiness of each trader's activities.

### **3. Backtesting**

As the previous discussion has demonstrated, there is a range of methods in use for calculating VaR estimates. In practice, even where banks use the same broad methods to calculate VaR, there is considerable variation in the application of those VaR methodologies – different models may be used to measure the sensitivities of

particular instruments to price movements (particularly for options and other more complex products; see, for example, Cooper and Weston (1995)); different methods may be used to aggregate exposures across instruments; and different techniques for estimating price volatilities may be used.

Since there is such a divergence in the VaR methodologies and their application across banks, and given the debate about the veracity of the statistical assumptions underlying VaR calculations, it is useful to test the performance of VaR models. Such testing is often referred to as 'backtesting'.

Many banks that use VaR models routinely perform simple comparisons of daily profits and losses with model-generated risk measures to gauge the accuracy of their risk measurement systems. However, banks themselves are only just beginning to develop more sophisticated backtesting techniques and there are considerable differences in the types of tests performed.

In this paper we consider the following tests of VaR model performance:

- the regulatory backtest required as part of the capital-adequacy framework;
- exceptions testing which examines the frequency with which losses greater than the VaR estimate are observed;
- variance testing which compares the estimate of profit and loss variance implicit in a VaR estimate with the realised variability of profits and losses over time;
- tests to assess whether the profit and loss data are normally distributed; and
- a risk-tracking test evaluating the correlation between VaR estimates and the magnitude of daily profit and loss results.

To illustrate these tests we apply each of the tests to VaR and profit and loss data for a number of individual trading portfolios obtained from an Australian bank.

### 3.1 Shortcomings of Backtesting

Before presenting the tests themselves we note that there are a number of difficulties with the general approach to backtesting which uses realised profit and loss results. The most fundamental of these arises from the fact that such backtesting attempts to compare static portfolio risk with a more dynamic revenue flow. VaR is measured as the potential change in value of a static portfolio, at a specific point in time (typically end-of-day). Hence, the VaR calculation assumes that there is no change in the portfolio during the holding period; the portfolio can be viewed as representing a stock of risk at a given point in time. In practice, banks' portfolios are rarely static, but change frequently. Profits and losses are flows accruing over time as a bank takes on and closes out positions reflecting changes in portfolio composition during the holding period.

The difficulties that dynamic portfolios create can be illustrated most starkly by considering a trading desk that is not permitted to hold open positions overnight. During the day the desk may take positions and as a result experience large swings in profit and loss, but at the end of each day all positions must be closed out. Hence, an end-of-day VaR will always report a zero risk estimate, implying zero profit and loss volatility, regardless of the positions taken on during the day. More generally, where open positions remain at the end of the trading day, intra-day trading will tend to increase the volatility of trading outcomes, and may result in VaR figures underestimating the true risk embedded in any given portfolio.

To overcome this problem of dynamic portfolios, a backtest could be based on a comparison of VaR (using a one-day holding period) against the hypothetical changes in portfolio value that would occur if end-of-day positions were to remain unchanged. That is, instead of looking at the current day's actual profit or loss, the profit or loss obtained from applying the day's price movements to the previous day's end-of-day portfolio is calculated (this is often referred to as 'close-to-close' profit and loss). This hypothetical profit or loss result could then be compared to the VaR based on the same, static, end-of-day portfolio. In such a case, the risk estimate and the profit and loss would directly correspond. At this stage, several Australian banks do perform analysis on this basis.

The distortion in backtesting comparisons arising from changes in portfolio composition can be minimised by selecting a shorter holding period. Clearly

movements in portfolio composition will be greater the longer the chosen holding period. It is for this reason that the Basle Committee recommends that backtesting be conducted based on a one-day holding period even though the capital that a bank is required to hold against its market risk is based on a VaR with a ten-day holding period (Basle Committee on Banking Supervision 1996a).

Further difficulties in conducting backtests arise because the realised profit and loss figures produced by banks typically include fee income and other income not attributable to position taking. While identification of fee income is relatively straightforward, isolating profits generated solely from position taking may be more difficult. Acting as a market maker allows banks to earn profits by setting different bid and offer rates; even transactions conducted with a view to profiting from market movements may profit from the bid/offer spread. A more sophisticated approach to measuring profitability would involve a detailed attribution of income by source, including fees, spreads, market movements and intra-day trading results. In such a case, the VaR results should be compared with the income generated by market movements alone. While most Australian banks do separate fee income from trading profit and loss, more refined attribution of income focusing on isolating revenue derived from position taking is generally only done in limited cases (for example, where proprietary trading is conducted by traders separate from those involved in other trading activity).

It may be argued that fee income and other market-making income are an inherent part of a bank's business and hence, their variability should be taken into account when assessing the riskiness of the bank's trading operations and when evaluating the performance of risk-measurement techniques. However, the VaR models in use at most banks are designed to measure outright position risk rather than risk arising from volatility in fee income or from movements in bid/offer spreads which may require the use of other modelling techniques. Thus the objective of backtesting should be to compare measured position taking risk with pure position taking revenue.

It should be kept in mind that shortcomings in the construction and practical implementation of a VaR model may not be the only reason why models fail backtests. As discussed in Section 2.7, VaR models do have some fundamental shortcomings. A VaR model is reliant on historical data and cannot capture major regime shifts in markets. Large swings in intra-day trading or an unusual event in

trading income other than from position changes can lead to poor outcomes not reflective of the quality of the VaR model's construction.

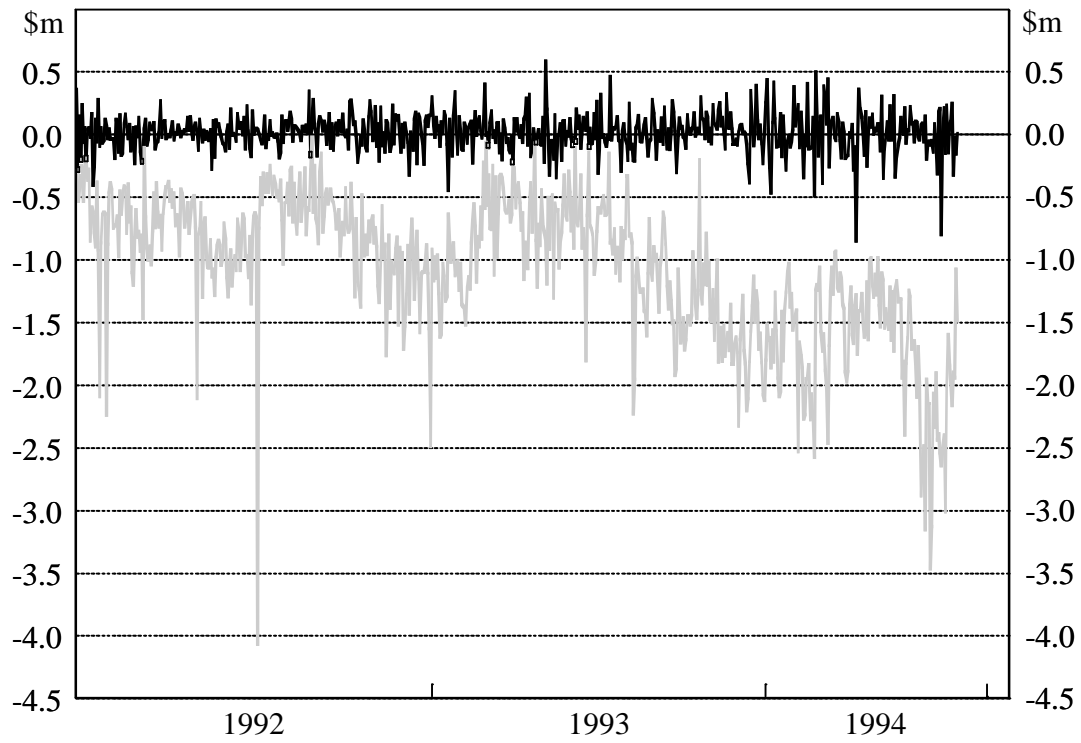
### **3.2 The Sample Portfolios**

The following sections step away from the data issues discussed above and look at some of the tests currently in use in banks in Australia and overseas. Those tests are applied to daily VaR and realised profit and loss data obtained from an Australian bank; the data cover the period from January 1992 to February 1995. The bank uses the variance-covariance approach to calculate a one-day holding period VaR.

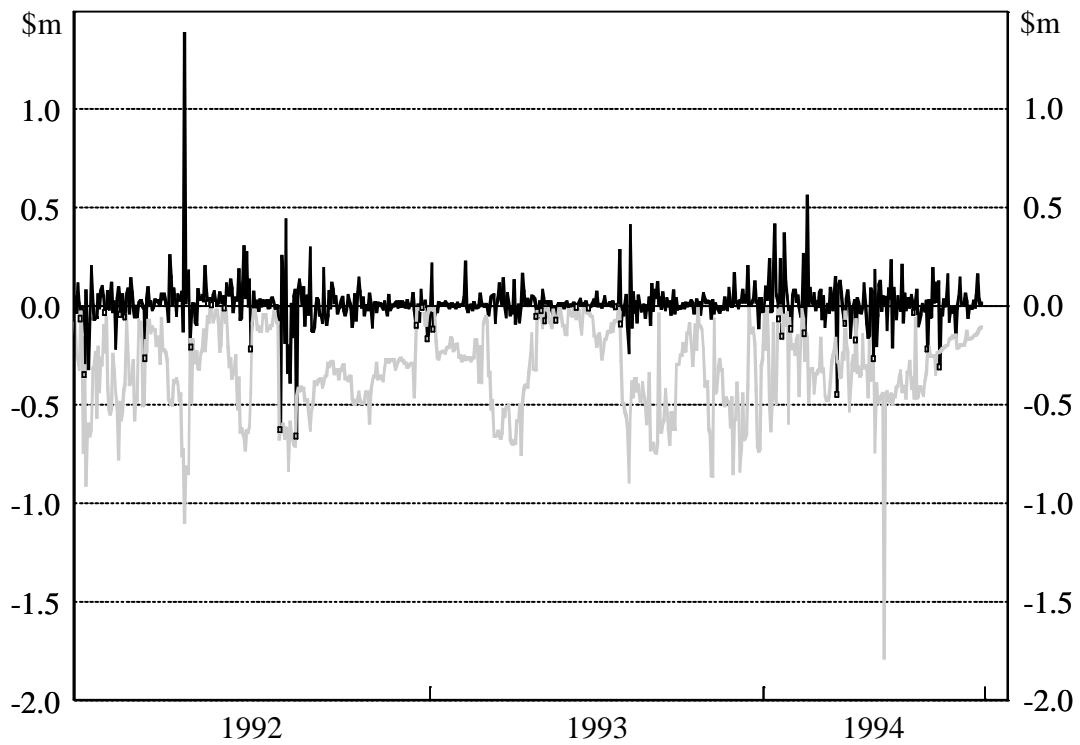
The daily VaR figures were obtained from the daily market risk reports produced by the bank. The reports detail both the bank's total VaR amount and the VaR for individual portfolios (spot foreign exchange (portfolio A), government securities (portfolio B), money market instruments (portfolio C), interest-rate swaps (portfolio D) and interest-rate options (portfolio E)). The testing here assesses VaR model performance at the individual portfolio level. Similarly, profit and loss results were obtained from the bank's internal management report detailing the profit performance of each portfolio. Actual, rather than hypothetical, profit and loss figures have been used. Hence, the profit and loss numbers include some fee income and other income not due to position taking. The daily VaR and profit and loss results for each portfolio are shown in Figure 5. The black lines depict daily profit and loss while the grey lines depict VaR results. Days where the actual loss exceeded the VaR estimate are highlighted.

**Figure 5: Individual Portfolio Profit and Loss and VaR**

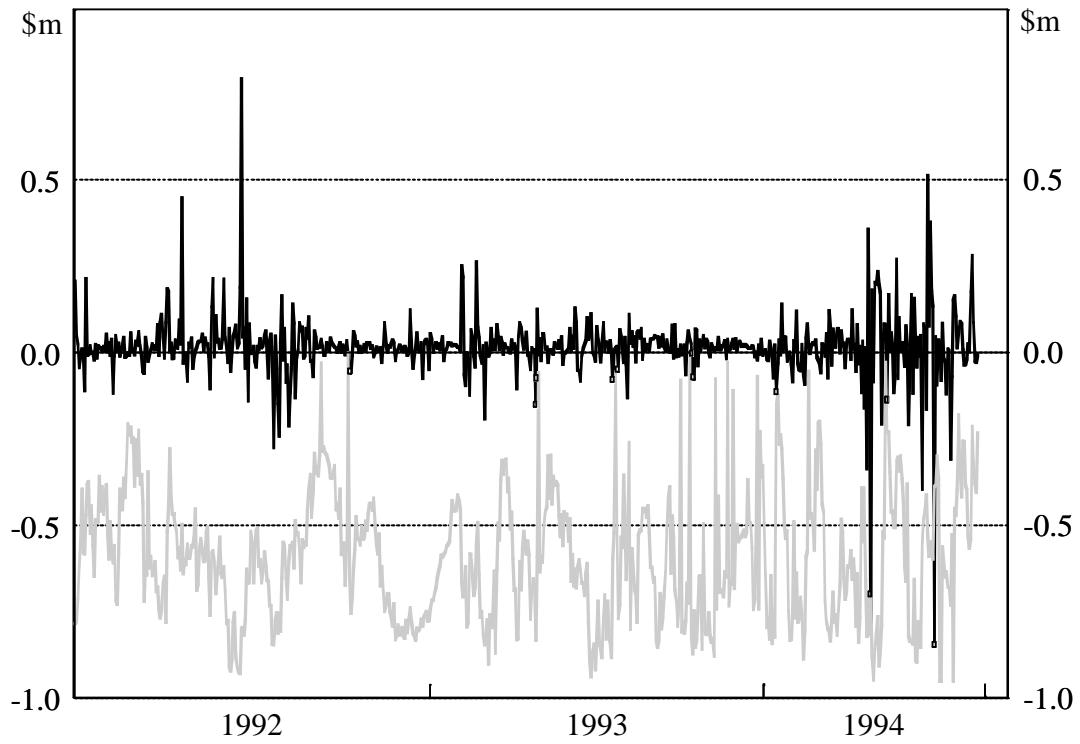
Portfolio A



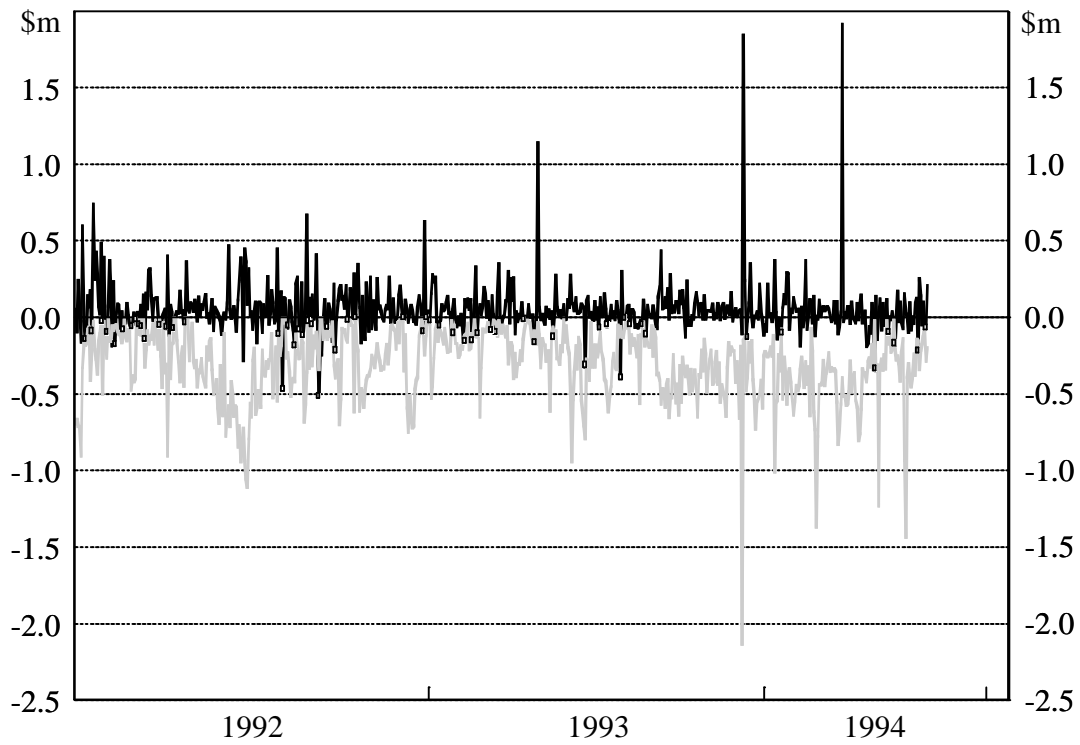
Portfolio B



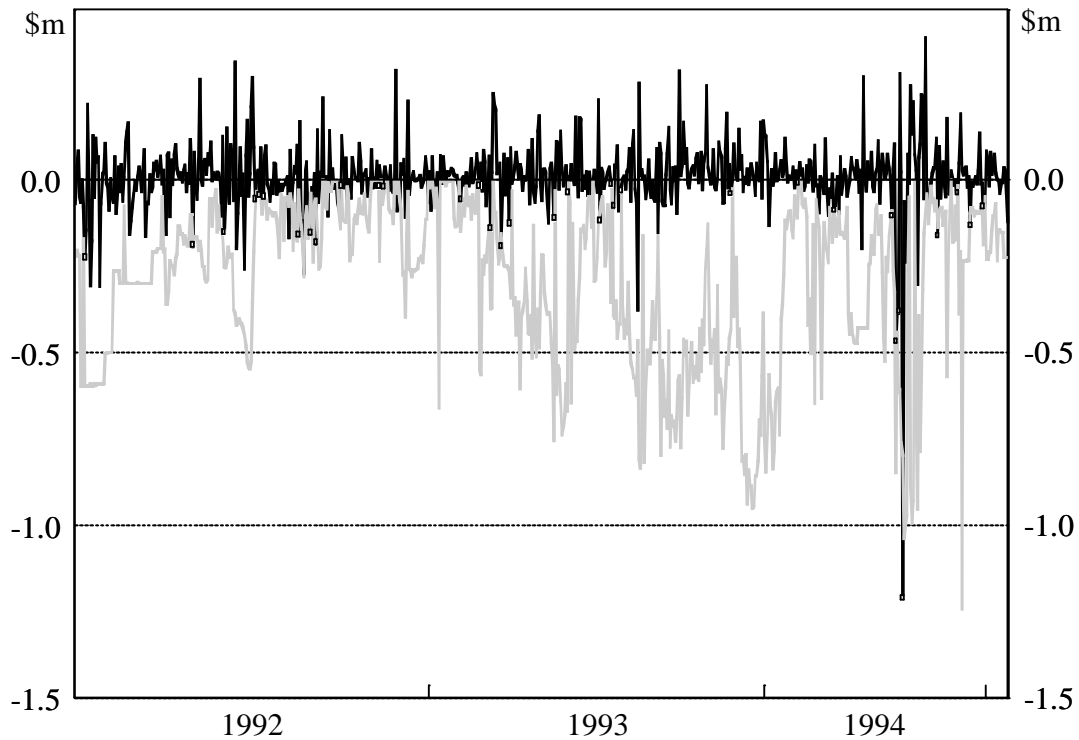
Portfolio C



Portfolio D



## Portfolio E



### 3.3 The Regulatory Backtest

Backtesting is a fundamental part of the market risk capital standards currently being put in place by supervisors around the world. Under the capital adequacy arrangements proposed by the Basle Committee, each bank must meet a capital requirement expressed as the higher of: (i) an average of the daily VaR measures on each of the preceding sixty trading days, adjusted by a multiplication factor; and (ii) the bank's previous day's VaR number. The *multiplication factor* is to be set within a range of 3 to 4 depending on the supervisor's assessment of the bank's risk management practices and on the results of a simple backtest (Basle Committee on Banking Supervision 1996a).<sup>4</sup>

The multiplication factor is determined by the number of times losses exceed the day's VaR figure (termed 'exceptions') as set out in Table 8 (Basle Committee on Banking Supervision 1996b). The minimum multiplication factor of 3 is in place to

---

<sup>4</sup> The market risk capital adequacy arrangements to apply to Australian banks (Prudential Statement No. C3) were released by the Reserve Bank in January 1997 and are broadly consistent with the Basle Committee's international standards.



compensate for a number of errors that arise in model implementation: simplifying assumptions, analytical approximations, small sample biases and numerical errors will tend to reduce the true risk coverage of the model (Stahl 1997). The increase in the multiplication factor is then designed to scale up the confidence level implied by the observed number of exceptions to the 99 per cent confidence level desired by regulators. In calculating the number of exceptions, banks will be required to calculate VaR numbers using a one-day holding period, and to compare those VaR numbers with realised profit and loss figures for the previous 250 trading days.

---

**Table 8: The Basle Committee's Three Zones**

---

	Number of exceptions (in 250 days)	Multiplication factor
Green zone	4 or less	3.00
	5	3.40
	6	3.50
Yellow zone	7	3.65
	8	3.75
	9	3.85
Red zone	10 or more	4.00

---

A simple approach to exceptions-based backtesting would be to assume that the selected data period provides a perfect indication of the long-run performance of the model. For example, if a VaR model was supposed to produce 99<sup>th</sup> percentile risk estimates, observed exceptions on any more than 1 per cent of days could indicate problems with the model. This is not realistic since with a finite number of daily observations it is quite probable that the actual number of times that losses exceed VaR estimates will differ from the percentage implied by the model's confidence interval, even when the model is in fact accurate. Hence, the Basle approach is to allocate banks into three zones based on the number of exceptions observed over 250 trading days. A model which truly covers a 99 per cent confidence interval has only a 5 per cent chance of producing more than four exceptions (yellow zone), and only a 0.01 per cent chance of producing more than ten exceptions (red zone).

Table 9 shows the number of regulatory exceptions and the scaling factor to be applied to each portfolio based on the last 250 days' data in the sample.<sup>5</sup> The regulatory backtest places two portfolios (D and E) in the yellow zone. Results in this range are plausible for both accurate and inaccurate models, although it is more likely that the model is inaccurate. Portfolio B is placed in the red zone which will, under the capital-adequacy framework, lead to an automatic presumption that a problem exists within the VaR model.

---

**Table 9: Regulatory Testing**

---

Portfolio	Basle exceptions	Scaling factor
A	0	3.00
B	10	4.00
C	2	3.00
D	6	3.50
E	8	3.75

---

### 3.4 Exceptions Testing

Kupiec (1995) presents a more sophisticated approach to the analysis of exceptions based on the observation that a comparison between daily profit or loss outcomes and the corresponding VaR measures gives rise to a binomial experiment. If the actual trading loss exceeds the VaR estimate the result is recorded as a failure (or 'exception'); conversely, if the actual loss is less than the expected loss (or if the actual trading outcome is positive) the result is recorded as a success.

If it can be assumed that a bank's daily VaR measures are independent, the binomial outcomes represent a sequence of independent Bernoulli trials each with a probability of failure equal to 1 minus the model's specified level of confidence; for example, if the level of confidence is 95 per cent the probability of failure on each trial will be 5 per cent.<sup>6</sup> Hence, testing the accuracy of the model is equivalent to a

---

<sup>5</sup> For internal risk management purposes the bank does not use a 99 per cent confidence interval; the Basle exceptions were calculated using VaR figures that had been rescaled to 99 per cent confidence equivalent amounts. All subsequent testing is based on the bank's own internal confidence level.

<sup>6</sup> In fact this is not a good assumption. The Ljung-Box Q-test for serial correlation found significant autocorrelation in the VaR series for all five portfolios.

test of the null hypothesis that the probability of failure on each trial equals the model's specified probability. Kupiec uses two tests to examine this hypothesis – the time between failures test and the proportion of failures test.

### 3.4.1 Time between failures test

The first test is based on the number of trading days between failures and is applied each time a failure is observed. This test is most useful in the case where a risk manager is monitoring the performance of a VaR model on a daily basis and focusing on the new information provided by the model. For example, a bank's risk manager could consider reviewing the model if a number of exceptions occur in succession. The test is less well suited to an analysis of long runs of *ex post* data on model performance.

To explain the test in more detail we use the following notation:

$v$  = the observed time (in days) between failures;

$p$  = true probability covered by the VaR model;

$p^*$  = the probability specified by the VaR model being tested,

(100 - confidence interval)%; and

$\tilde{p}$  = the maximum likelihood estimator of  $p$ , given by  $1/v$  in the test.

It can be shown that a likelihood ratio (LR) test is the most powerful for testing the null hypothesis that  $p = p^*$ . The LR statistic is given by:

$$LR = -2 \ln \left( \frac{p^* \cdot (1 - p^*)^{v-1}}{\tilde{p} \cdot (1 - \tilde{p})^{v-1}} \right)$$

The LR statistic is distributed as a chi square distribution with 1 degree of freedom.

This test is subject to a number of shortcomings. Firstly, the test has extremely poor power. This means that the test may not reject a VaR model although it continually underestimates risk. The problem is compounded by the fact that the power of the time between failures test is lower, the lower is the probability value under the null hypothesis. That is, for banks with higher stated confidence levels, the likelihood of

not rejecting the VaR model, when it is in fact underestimating risk, increases. Moreover, as the VaR model's confidence level increases, the extent by which potential loss is underestimated increases at an accelerating rate. For example, a VaR model purporting to cover a probability of 0.005 but which is in fact generating potential loss estimates consistent with a probability level of 0.025 will underestimate risk by a greater amount than a model with a specified probability level of 0.03 but which yields true coverage consistent with a 0.05 probability.

A second difficulty relates to the values of  $\nu$  (the number of days between successive failures) associated with the critical value of the LR statistic. The critical value of the distribution, at a 5 per cent level of significance, is 3.841. For a model which specifies a probability level of, say, 0.025 the LR statistic will exceed the critical value (and so the VaR model will be rejected) when  $\nu$  is less than 12 or greater than 878. For values of  $\nu$  less than 12, the test concludes that the model is underestimating risk. Conversely, for values of  $\nu$  greater than 878 the model is likely to be overestimating risk. From a supervisory perspective, the concern is on the former of these alternatives. The problem arises when the specified probability level is 0.05 or larger. In this case, the resulting LR statistic will be less than the critical value for values of  $\nu$  of two or larger and will be undefined for  $\nu$  equal to one (when two failures are observed on successive trading days). Hence, for these models there exists no critical value for  $\nu$  at which the model will be rejected.

Table 10 summarises, for each portfolio, the results of the time between failures test. The second column of the table reports the trading day on which the first failure is observed and column three reveals whether or not the test would reject the model after observing this first failure. Column four lists, for each portfolio, the number of failures that were observed before the VaR model is rejected. Columns five and six report the total number of failures observed over the sample period and the total number of times the null hypothesis was rejected at a 5 per cent level of significance.

**Table 10: Time between Failures Test**

Portfolio	First failure	Test result	Failures before first rejection	Failures	Rejections
A	4	reject	1	5	3
B	5	reject	1	27	17
C	203	do not reject	–	3	0
D	28	do not reject	2	31	17
E	88	do not reject	4	36	24

The results in Table 10 imply that the model performs well for portfolios A and C while portfolios B, D and E are identified as portfolios where the model does not seem to be adequately capturing the potential for losses to be observed. There are a number of occasions on which these three portfolios experience a run of rejections on successive trading days or trading days two or three days apart. This result suggests that the VaR model may not be able to cope with large movements in market prices; it could be that the price volatilities within the VaR are not updated with sufficient frequency to capture the time-varying nature of the profit and loss variability (that is, autoregressive conditional heteroscedasticity). It is noteworthy that if this LR test was applied by a risk manager to performance data on a daily basis this problem with the model would be detected quite early on in the sample period.

#### 3.4.2 Proportion of failures test

The second test is based on the proportion of failures observed over the entire sample period. A test of the null hypothesis that the VaR model's stated probability level is equal to the realised probability level covered by the VaR model ( $p = p^*$ ) is again best achieved using a LR test. The LR test statistic is given by:

$$LR = -2 \ln \left( \frac{p^{*x} \cdot (1 - p^*)^{n-x}}{\tilde{p}^x \cdot (1 - \tilde{p})^{n-x}} \right)$$

where  $n$  denotes the total number of outcomes in the sample period,  $x$  denotes a Bernoulli random variable representing the total number of observed failures and  $\tilde{p}$  is the maximum likelihood estimator, given by  $x/n$  ( $x \geq 1$ ).

While this test is better suited to an analysis of *ex post* data than is the time between failures test, it still has some problems. As was the case with the preceding test, the proportion of failures test has poor power characteristics which become worse as the confidence interval being tested increases. Further, although the power of the test improves as the sample period increases, a substantial sample size is required for the test to have significant power. Nonetheless, with the exception of the case of a single failure for which the test is equivalent to the time between failures test, the proportion of failures test has more power than the preceding test owing to its making use of more information. Moreover, the LR statistic is defined for all combinations of  $x$  and  $n$  (except  $x = n = 1$ ) and so can be applied for all stated probabilities.

Table 11 summarises the results of the proportion of failures test when applied to each of the portfolios. As with the time between failures test, it can be concluded that those portfolios whose models do not give rise to any failures over the period are not rejected. For all portfolios the results of this test are consistent with the time between failures test.

**Table 11: Proportion of Failures Test**

Portfolio	$x$	$n$	Proportion of failures (%)	Chi square	Significance
A	5	653	0.8	0.8	0.372
B	27	673	4.0	66.0	0.000
C	3	669	0.4	0.0	0.847
D	31	631	4.9	87.2	0.000
E	36	692	5.2	105.1	0.000

### 3.5 Estimates of Variance

A VaR figure can be regarded simply as a rescaled variance. This is most transparent in the variance-covariance calculation of VaR. Hence, it is possible to undo a VaR calculation to obtain the profit and loss volatility underlying the VaR calculation and to compare this against the observed variance of profits and losses. The test below compares the volatility derived from the average VaR over time with the variance of the actual profit and loss distribution over the same period. Assuming that profits and losses are normally distributed (the validity of this assumption is tested in the next section), an F test can be used to test whether the

two variance estimates are significantly different. Under the null hypothesis that the variances are equal the ratio of the estimates follows an F distribution:

$$\frac{\text{var } P\&L}{(\overline{VaR} / k)^2} = F_{n-1, n-1}$$

where  $\overline{VaR}$  denotes the average VaR over time,  $k$  denotes the number of standard deviations required for the specified confidence interval (for example, a 97.5 per cent confidence interval is equivalent to 1.96 standard deviations) and  $n$  denotes the number of daily observations.

Table 12 which follows shows the observed standard deviation of profit and loss,  $\overline{VaR} / k$  and the results of the F test for each portfolio. For all portfolios, except portfolio D, the VaR variance is not significantly less than the profit and loss variance suggesting that overall the VaR adequately captures true volatility.

**Table 12: Variance Testing**

Portfolio	$\overline{VaR} / k$ (\$ 000)	P&L Std Dev (\$ 000)	F	Significance
A	417.3	152.9	0.134	1.000
B	123.9	110.0	0.788	0.999
C	229.6	92.2	0.161	1.000
D	118.1	172.4	2.129	0.000
E	113.1	101.2	0.800	0.998

### 3.6 Tests of Normality

Where it is the case that a bank uses a VaR methodology that assumes that financial returns are normally distributed (such as the variance-covariance method), testing whether the observed profits and losses follow a normal distribution can serve as useful backtesting.

There is a wide variety of tests that may be used to test for normality. Two of the simplest tests focus on the skewness and kurtosis of the observed distribution.

Skewness measures the imbalance in a distribution, that is, whether observations occur more frequently above or below the mean. The skewness of each profit and loss series and the result of tests to determine whether the skewness is significantly non-zero are presented in Table 13. Tests for the significance of the skewness statistic are set out in Kendall and Stuart (1963).

In the opening section of this paper it was shown that large exchange rate movements are observed more often than is consistent with the data being normally distributed; that is, the empirical distributions for exchange rates have fatter tails than the normal distribution. In fact, it has been observed that a large number of financial series exhibit fat tails (see, for example, Bollerslev *et al.* (1994)). It is possible to quantify the divergence between the tails of the observed and normal distributions by measuring the kurtosis of the observed distribution as:

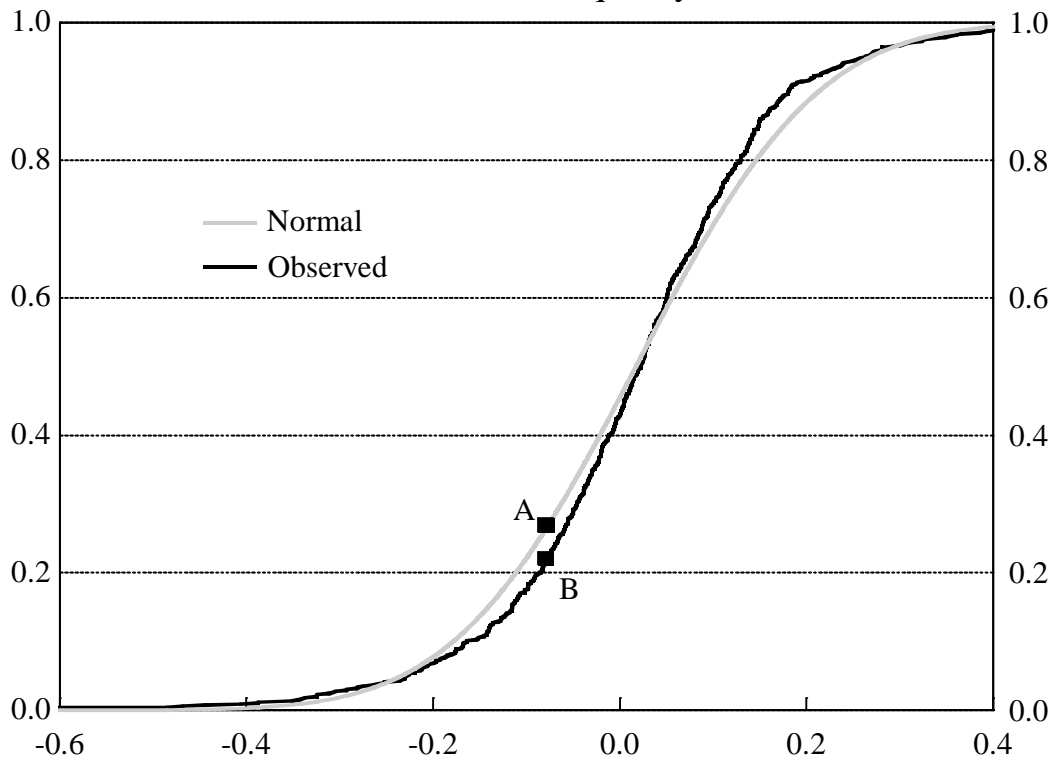
$$n^4 = \frac{1}{(n-2)(n-3)} \left[ \frac{n(n+1)}{(n-1)s^4} \sum (X_i - \bar{x})^4 - 3(n-1)^2 \right]$$

If  $v^4 = 0$  the tails on the distribution follow a normal distribution, if  $v^4$  is negative the distribution has thinner tails than a normal distribution, while if  $v^4$  is positive the distribution has fatter tails (known as leptokurtosis). Again, tests for the significance of the kurtosis statistic are given in Kendall and Stuart (1963).

Tests based on skewness and kurtosis each focus on one particular aspect of the observed distribution. It is also possible to devise tests that take into account the entire distribution by comparing the cumulative distribution functions of the observed data and the normal distribution. As an example, Figure 6 shows the observed cumulative distribution function for the profits and losses on portfolio A compared with the standard normal cumulative distribution function. There are a number of ways to quantify the difference between the two functions. One widely used test is the Kolmogorov-Smirnov test which takes as its test statistic the maximum value of the absolute difference between the two functions (the distance between points A and B in Figure 6).



**Figure 6: Empirical versus Normal Cumulative Distributions**  
Cumulative frequency



One weakness with this test is that it places greater weight on observations around the median. An alternative test which gives equal weight to all percentiles across the distribution is the Kuiper test (Crnkovic and Drachman 1996). This test is based on the statistic:

$$V = \max_{-\infty < x < \infty} [S(x) - N(x)] + \max_{-\infty < x < \infty} [N(x) - S(x)]$$

where  $S(\bullet)$  is the observed cumulative distribution function,  $N(\bullet)$  is the Normal cumulative distribution function and  $x$  denotes observations of profit and loss results.<sup>7</sup> That is, instead of looking at the largest distance between the functions, the Kuiper statistic takes into account the sum of the maximum distance of  $S(x)$  above  $N(x)$  and below  $N(x)$ . Press *et al.* (1992) describe the procedures for testing the significance of the Kolmogorov-Smirnov and Kuiper statistics.

---

<sup>7</sup> An unbiased estimator of the cumulative distribution function of the probability distribution from which the profit and loss figures are drawn,  $S(x)$ , is given by the fraction (out of the total number of observations) of data points that are less than  $x$ .

The skewness and kurtosis statistics, along with tests of their significance, and the results of Kolmogorov-Smirnov (KS) and Kuiper tests are presented in Table 13. The results of the four tests are in agreement; the tests yield a consistent conclusion for all five portfolios. All portfolios are found to be strongly non-normal. Since a variance-covariance VaR model is used by the bank, the finding of non-normality implies that the VaR measure will not correctly estimate the true risk exposure. In particular, since the profit and loss distributions are quite fat-tailed, the VaR model would be expected to significantly underestimate risk.

**Table 13: Tests Of Normality**

Portfolio	Skew	Significance	Kurtosis	Significance	KS	Significance	Kuiper	Significance
A	-0.462	0.000	3.292	0.000	1.307	0.066	2.627	0.000
B	2.347	0.000	41.477	0.000	4.087	0.000	8.212	0.000
C	-1.029	0.000	30.396	0.000	4.710	0.000	9.243	0.000
D	4.472	0.000	42.118	0.000	3.939	0.000	6.868	0.000
E	-2.598	0.000	32.455	0.000	3.042	0.000	5.976	0.000
Total bank	0.983	0.000	7.939	0.000	1.705	0.006	3.142	0.000

Given that the bank's profit and loss data does include fee income and other income derived from the bid/offer spread that a market maker is able to obtain, some negative skewness in the profit and loss distribution would be expected. In fact the test results are quite mixed, with two of the five portfolios exhibiting positive skewness. To the extent that there is positive skewness in the profit and loss distribution, the variance-covariance VaR calculation will overestimate true risk, offsetting any underestimation resulting from the failure to capture the leptokurtosis (fat tails) of the distribution.

Given that individual portfolios may be quite non-normal, this could result in VaR models that do assume normality significantly underestimating risk. It is at the aggregate level (that is, the sum of individual portfolio's profit and loss), however, that regulatory tests will apply in the first instance. It is expected that non-normality should be less evident at the aggregate level than for individual portfolios. This is based on the central limit theorem which states that taking the sum of increasing numbers of individual statistical processes will result in the distribution of the sum tending towards normality regardless of the distribution of the individual

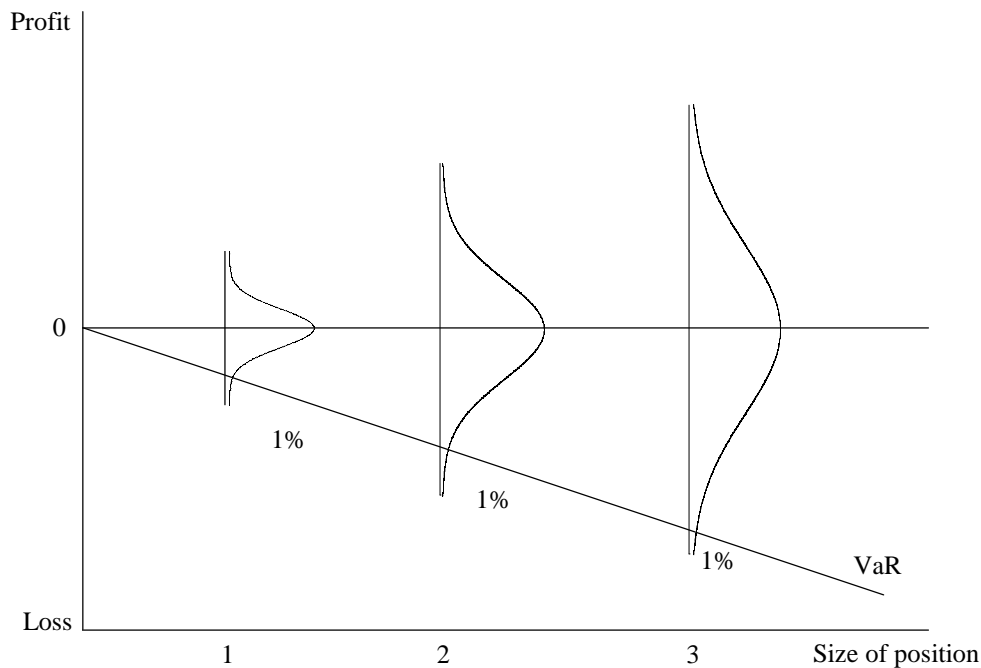
components. In fact, as shown in Table 13, this effect does not seem to be very strong. The total profit and loss for the bank remains strongly non-normal.

### **3.7 Risk Tracking**

There are two dimensions to evaluating the performance of any risk measure – the level of coverage provided and the efficiency of the measure. The tests in the previous sections have focused on coverage – what level of losses would a VaR-based capital charge insure against. A good risk measure needs to be more than just appropriately conservative, it should also be strongly correlated with the portfolio's true risk exposure. A conservative but inefficient risk measure would tend to overestimate risk in periods of low risk.

The very simplest test that can be performed is to assess the relationship between the VaR estimates and the absolute value of each day's profit or loss. In a very simple sense, large VaR figures should be accompanied by large profits or losses while small VaR estimates should be associated with small profit and loss results. Consider a portfolio containing positions in a single asset: as the exposure to that asset grows the variance of the profit and loss distribution will increase and the VaR should increase to reflect the greater exposure (Figure 7). As the exposure increases, larger profit and loss observations become more likely. To the extent that the VaR is tracking the increased variance of the underlying profit and loss distribution, there should be some correlation between the size of profits and losses and the VaR. This effect may not be strong, however, as the bulk of the profit and loss observations will be drawn from the centre of the distribution.

In the broadest terms, both VaR and profit and loss are a function of two processes – changes in market prices and the composition of the bank's portfolio. The bank considered here uses a relatively long run of data to calculate historical price volatilities and updates the price volatility history quarterly. The bank does not use a time weighting scheme such as exponential weighting but rather assumes that the covariance matrix is fixed independent of time. As a result, the bulk of the time variation in the VaR figures comes from changes in the composition of the bank's portfolios.

**Figure 7: Var and the Distribution of Profit and Loss**

If the bank's portfolios were constant (and assuming that the statistical process driving market prices are unchanging through time), then the profit and loss distributions would be constant through time. A perfectly efficient VaR, in such a case, would also be constant. Hence, profit and loss magnitudes and VaR estimates would be quite uncorrelated despite the fact that the VaR is perfectly efficient. However, banks' portfolios are not constant; as can be seen from Figure 5, the bank considered here is an active trader whose portfolio composition and profit and loss distribution change markedly from day to day.

Given that the assumption of normality of both the absolute value of profit and loss and VaR is problematical, it is useful to consider tests which do not require any assumption about the distribution of the two series. One such test is based on the rank correlation coefficient. Let  $X$  denote the series of profit and loss magnitudes and  $Y$  the VaR series. Ranks  $1, 2, \dots, n$  are then assigned to the  $X$ 's and  $Y$ 's. Let  $U_i = \text{rank of } X_i$  and  $V_i = \text{rank of } Y_i$  then the rank correlation coefficient,  $R$ , is defined as the correlation between  $U$  and  $V$ . For sufficiently large samples

$R = \frac{d}{n-1} N\left(0, \frac{1}{n-1}\right)$ . Tests of the significance of the relation between the two series can be based on this.

The rank correlations for each portfolio are presented in Table 14. Overall, the tests indicate that the VaR numbers are not strongly correlated with the magnitudes of the profit and loss outcomes. While four of the five rank correlations are significantly positive, the magnitude of the estimated correlations are small.

---

**Table 14: Correlation between VaR and the Absolute Value of Profit and Loss**

---

Portfolio	Rank correlation	Significance
A	0.142	0.000
B	0.240	0.000
C	0.129	0.001
D	0.017	0.672
E	0.182	0.000

---

#### 4. Conclusion

Despite their weaknesses, the tests employed in this paper provide some useful diagnostics for evaluating VaR model performance. The exceptions-based testing identifies portfolios B, D and E as areas where the VaR model is not adequately covering realised losses. In the case of portfolio D, this assessment is supported by the results of the variance-based test. These weaknesses in the VaR model can, to a large extent, be attributed to the observed non-normality of profit and loss. Given that the risk-tracking performance of the model is not strong, there does appear to be considerable scope for the model to be further refined.

The use of VaR models has spread remarkably rapidly across the financial industry over the past six or so years. As the techniques have become more widespread, the range of methods for calculating VaR estimates has broadened. That said, the use of VaR models are at comparatively early stages of development for many Australian banks. The need for further refinement of newly developed models on the part of banks, and the need for regulators to assess the performance of such models, suggests that backtesting should play an important role in furthering the evolution of VaR techniques.

Given that our sample bank was just starting the process of implementing and refining its VaR model at the time when the data were collected, it would be expected that the model might not perform well. Over the past year, many banks that use VaR measures as part of their risk-management process (including the sample bank discussed here) have significantly upgraded their VaR model; for example, by improving the quality of the historical price data underlying the calculation, more accurately modelling the price sensitivities of individual instruments or by moving from a variance-covariance approach to a Monte-Carlo simulation approach. Hence, the test results shown here should not be taken as describing the performance of Australian banks' models more broadly.

The tests that have been applied in this paper are not sufficiently precise to form the basis of regulatory treatment of banks' VaR estimates in the way that the Basle Committee defines its three zones. Notwithstanding their shortcomings, these tests do appear to provide some useful information to both risk managers within banks and bank supervisors. The tests may aid in identifying those portfolios where VaR performs poorly, and may provide some insight into why a VaR model may not capture the risks of particular portfolios well, thus guiding the process of improving the analysis and modelling of trading risks.

## References

**Basle Committee on Banking Supervision (1996a)**, ‘Amendment to the Capital Accord to Incorporate Market Risks’, January.

**Basle Committee on Banking Supervision (1996b)**, ‘Supervisory Framework for the use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements’, January.

**Bollerslev, T., R. Engle and D. Nelson (1994)**, ‘ARCH Models’ in R. Engle and D. McFadden (eds), *Handbook of Econometrics*, 4, Elsevier Science, Amsterdam, Chapter 49, pp. 2961–3038.

**Cooper, S. and S. Weston (1995)**, ‘The Pricing of Over-the-counter Options’, *Bank of England Quarterly Bulletin*, 35(4), pp. 375–381.

**Crnkovic, C. and J. Drachman (1996)**, ‘Quality Control’, *Risk*, September, pp. 138–143.

**Euro-currency Standing Committee (1994)**, ‘A discussion paper on Public Disclosure of Market and Credit Risks by Financial Intermediaries’, prepared by a Working Group of the Euro-currency Standing Committee of the Central Banks of the Group of Ten Countries, Basle, September.

**Group of Thirty (1993)**, ‘Derivatives: Practices and Principles’, Global Derivatives Study Group, July.

**Kendall, M.G. and A. Stuart (1963)**, *The Advanced Theory of Statistics*, 1, Charles Griffin & Co, London.

**Kupiec, P. (1995)**, ‘Techniques for Verifying the Accuracy of Risk Measurement Models’, Finance and Economics Discussion Series, Federal Reserve Board, Washington DC, pp. 95–24.

**Press, W.H., B.P. Flannery, S.A. Teukolsky and W. Vetterling (1992)**, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

**Stahl, G. (1997), 'Three Cheers', *Risk*, May, pp. 67–69.**