CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

# *Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?*

*Heather M. Anderson and Farshid Vahid*

CEA@Cass Working Paper Series

WP–CEA–11-2006

# Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?

Heather M. Anderson and Farshid Vahid*

School of Economics
Australian National University
Canberra, ACT 0200
Australia

January 2006

**Abstract**

This paper develops multivariate factor models for forecasting volatility in Australian stocks. We suggest estimation procedures for approximate factor models that are robust to jumps when the cross-sectional dimension is not very large, and we also work with volatility measures that have been constructed so that they contain no jump components. Out of sample forecast analysis shows that multivariate factor models of volatility outperform univariate models, but that there is little difference between simple and sophisticated factor models.

**Keywords:** Realized volatility, bi-power variation, factor models, jumps, model selection, forecasting.

**JEL Classification:** C32, C53.

*Corresponding author. E-mail: Farshid.Vahid@anu.edu.au.

## 1. INTRODUCTION

Multivariate modelling of conditional heteroskedasticity has been an important research problem ever since ARCH models were introduced by Engle (1982). Complete multivariate generalization of a GARCH model involves the specification of a system of dynamic equations for the elements of a conditional variance-covariance matrix subject to positive definiteness constraints, as discussed in Engle and Kroner (1995). Typically such models involve many parameters even when the number of variables that are modelled jointly is only moderately large, and the computational difficulties and uncertainty caused by estimating too many parameters often outweighs the benefits of multivariate modelling. This has led researchers to consider various restricted versions of the general model, such as the constant conditional correlation model of Bollerslev (1990) or the dynamic conditional correlation model of Engle (2002). Researchers have also considered various restrictions that relate the evolution of variances and covariances to a reduced number of underlying factors. Our paper fits into this second stream of the literature.

It is easy to motivate factor models in financial applications. Theoretical asset pricing models often relate the dynamics of the returns for different assets to a small number of underlying factors, and this has led to factor models of the volatility of returns. Diebold and Nerlove (1989) consider a factor ARCH model for exchange rates, while Engle, Ng and Rothschild (1990) consider a factor ARCH model for the term structure of interest rates. King, Sentana and Wadhwani (1994) work with a multifactor model for aggregate stock returns for 16 countries. These models specify the complete conditional distribution of all variables, delivering internally consistent forecasts of means, variances and covariances.

Some authors, including Harvey, Ruiz and Shephard (1994) and Engle and Marcucci (2005) have focussed directly on building factor models of volatility, leaving other aspects of the joint distribution (and in particular the covariances) unspecified. We adopt a similar approach here, working with a measure of volatility known as realized volatility. Realized volatility has a long history in finance, appearing in early work by Merton (1980), but it has been strongly promoted as a volatility measure in recent work by Andersen, Bollerslev, Diebold and Labys (2001, 2003), and Barndorff-Nielsen and Shephard (2002, 2004, 2005), because it can provide a consistent estimate of (latent) integrated volatility in a continuous time diffusion model of the logarithm of an asset's price. In this paper we use realized volatility to develop multivariate models of the returns volatility of twenty one highly traded Australian stocks. Our goal is to investigate whether factor models can capture information that is useful for forecasting multivariate series of realized volatility.

The first stage in the development of a factor model is to determine the number of common factors. Engle and Kozicki (1993) show that if common factors are implied by some common statistical feature, that is, if the common factors have a statistical feature that is absent from the idiosyncratic factors, then as long as the number of variables is not too large, one can design common feature tests to determine the number of common factors (assuming that the cross sectional dimension is fixed and the sample size goes to infinity). This is not very difficult when one is modelling the conditional mean of a multivariate data set (see Anderson and Vahid 1998, for an example), but it becomes complicated in the case of conditional variances, as shown in Doz and Renault (2004). One advantage of working with pure measures of variance is that it facilitates the development of common features tests. Engle and Marcucci (2005) point out that non-normality and heteroskedasticity will usually imply that the standard canonical correlation based test statistics will not be useful in these circumstances. In such cases one has to turn to more robust tests for common features, such as those in Candelon, Hecq and Verschoor (2005).

We base our factor analysis on the approximate factor model developed by Chamberlain and Rothschild (1983), using the associated model selection criteria suggested by Bai and Ng (2002) to select the number of common factors. This approach relies on the consistency of principal components as estimators of common factors as the cross sectional and time series dimensions go to infinity. However, since our cross sectional dimension is quite small and our data contains substantial jumps, we argue that inference based on principal component estimator of the common factors may be severely distorted. We propose an instrumental variable estimator of common factors that is more robust to jumps, and we illustrate its use.

We also use the procedures discussed in Barndorff-Nielsen and Shephard (2005) to remove the jumps from our data. This allows us to focus on bi-power variation (i.e. realized volatility minus the jumps), which provides a consistent estimate of integrated volatility in a continuous time model of the (logarithm) of the stock price in the presence of jumps. We then build factor models of bi-power variation to capture the dynamics of this multivariate set of volatility measures. Since our decomposition of realized volatility into bi-power variation and jumps reveals that the jumps in our data set are not serially correlated, we are able to use our factor models to capture the dynamic behavior of both realized volatility and bi-power variation. We undertake an extensive forecast analysis as a means of evaluating our factor models, and find that they can do better than simple univariate models when forecasting the volatility of Australian stocks.

The structure of the rest of this paper is as follows. Section 2 provides a preliminary description of our data. Section 3 briefly explains approximate factor models and the determination of the number of factors in these models. Section 4 contains a discussion on how jumps can affect inference in approximate factor models, and it then suggests a procedure for choosing the number of factors that is robust to the presence of jumps and shows that this procedure works. This section also directs attention towards realized "bi-power variation" as the appropriate volatility measure to study, and then explores the properties of bi-power variation of the returns of Australian stocks. Section 5 develops univariate and multivariate models for the log-volatilities in our data set and then compares their out of sample performance with respect to forecasting both bi-power variation and realized volatility. Section 6 concludes.

## 2. DATA

We base our analysis on price data from the Securities Industry Research Centre of Asia and the Pacific (SIRCA) for stocks traded on the Australian Stock Exchange (ASX). Institutional details relating to trading on the ASX may be found on their web site (www.asx.com.au). Trading is on-line and is conducted through the Stock Exchange Automated Trading System (SEATS), which continuously matches bids and offers during normal trading hours from 10.00am to 4.00pm (EST) on Monday to Friday (public holidays excluded). Opening times for individual stocks are staggered but all stocks are trading by 10.10, and at the end of the day additional trading at volume weighted prices may continue until 4.20pm. Our data records the last price observed during every five minute interval within each working day for six years starting on January 1st 1996, but since there are too many five minute intervals in which there are no trades and hence no recorded price, we work with fifteen minute returns and restrict our attention to just twenty one frequently traded stocks. The names of the companies, their stock codes and their GICS (Global Industry Classification Standard) industry group are provided in Table 1.

[Table 1 about here]

Realized variance is calculated as the sum of all squared $\Delta$-period returns between time $t$ and $t+1$. That is, given the discretely sampled $\Delta$-period returns defined by $r_{t,\Delta} = p(t) - p(t-\Delta)$ where $p(t)$ is the natural logarithm of the price and $\Delta$ is small, realized variance is

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r^2_{t+j\Delta,\Delta}. \tag{1}$$

Given that the ASX is open for six hours in a normal working day, there are usually 120 fifteen minute time intervals in a five day week so that most of our weekly measures of realized variance are based on 120 raw data points, and $\Delta =$ 0.00825 (1/120). Some of our returns relate to shorter weeks that include public holidays (Easter Friday, Christmas, New Year, etc.), or trading halts that the ASX calls when firms are about to release price sensitive information (these halts can last anywhere between ten minutes to two days). In all cases involving less than 120 intra-week observations, we scale the measures of variance computed on the basis of the available fifteen minute returns up, so as to make them compatible with those measures computed from a full week of data.

[Table 2 about here]

We report summary statistics for weekly stock returns in Table 2. The most interesting aspect of this summary is that there is no evidence of ARCH in the weekly returns for most (14 out of 21) companies. The first column of Table 3 shows p-values for LM tests of the null hypothesis that there is no serial correlation in realized variance. Again, there is mixed evidence of predictability in volatility, with no evidence of predictability being found in 7 out of the 21 cases. These initial results suggest a very limited scope for pooling this data set to improve the forecastability of conditional variances, but after contrasting this evidence with the forecastability of filtered realized variance in Section 4, our interpretation is that significant idiosyncratic jumps in the prices of Australian stocks are responsible for giving the impression that conditional variances are unforecastable and dissimilar across different stocks. The jumps are large and are therefore very influential when one is estimating parameters, but they are also uncorrelated with the past history, and can therefore generate the impression that volatilities are unforecastable.

[Table 3 about here]

## 3. FACTOR MODELS OF REALIZED VOLATILITY

Raw intuition and formal theories in finance suggest that underlying market factors drive the movement of all asset returns, and there is a large empirical literature that summarizes the information in large sets of asset returns in terms of a few common reference factors. See Chapter 6 in Campbell, Lo and MacKinlay (1997), and Chapter 9 in Cochrane (2001) for discussion and a survey of relevant theoretical and empirical literatures. It is natural to think of principal component analysis as a technique that might be used for finding market factors, and this motivates the approximate factor literature in finance, originating in the work of Chamberlain and Rothschild (1983). This model is given by

$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r)}{A}\ \underset{(r\times 1)}{F_t} + \underset{(N\times 1)}{u_t}, \qquad (2)$$

where the $Y_t$ are assumed to have mean zero for simplicity, the vector $F_t$ contains $r$ common factors, and $u_t$ contains $N$ idiosyncratic factors that are independent of $F_t$. In the Chamberlain and Rothschild (1983) paper, the $N$ components in $Y_t$ are returns from $N$ assets. Here, we use the approximate factor model for modelling volatilities, so that the $Y_t$ in our setting will be a vector of volatilities. Andersen, Bollerslev, Diebold and Ebens (2001) provide a brief discussion that relates a factor structure in a multivariate set of realized volatilities to continuous time latent factor volatility models, and Lo and Wang (2000) develop and estimate a closely related factor model of weekly trading volumes.

Chamberlain and Rothschild (1983) show that the $r$ largest eigenvalues of $\frac{1}{T}\sum_{t=1}^{T} Y_t Y_t'$ will go to infinity as $N$ and $T$ go to infinity, while the $(r+1)^{th}$ eigenvalue remains bounded. Intuitively, the result holds because each additional cross sectional unit provides additional information about the common factors, but only local information about an idiosyncratic factor. Therefore, as $N \to \infty$, the information in the data about the common factors will be of order $N$, while the information about idiosyncratic factors will remain finite.

Bai and Ng (2002) use these results to develop four consistent model selection criteria for choosing the number of factors in approximate factor models. These are

$$PC_1(r) = \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \quad (3)$$

$$PC_2(r) = \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln(\min\{N,T\}), \quad (4)$$

$$IC_1(r) = \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \quad \text{and} \quad (5)$$

$$IC_2(r) = \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln(\min\{N,T\}), \quad (6)$$

where $ESS(r) = \sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it} - \hat{a}_i'\hat{F}_t)^2$, $\hat{F}_t$ is the vector of $r$ estimated common factors, $\hat{a}_i$ is the least squares estimate of factor loadings for asset $i$, and $r^{\max}$ is the largest possible $r$ considered by the researcher. The first two criteria compare the improvement (i.e., decrease in the error sum of squares) relative to a benchmark unrestricted model as $r$ increases, while the last two criteria work on the basis of the percentage improvement in the error sum of squares as $r$ increases. Bai and Ng (2002) use the principal component estimator of factors and factor loadings, which minimizes the sum of squared errors.

We apply these criteria to the square root of our realized variance measures for various values of $r^{\max}$ working with the square root measures so as to conform with the finance literature, where volatility usually refers to standard deviation. Results based on our realized variance series are qualitatively the same.

The results for $r^{\max} = 5$ are

| $r$ | $PC_1(r)$ | $PC_2(r)$ | $IC_1(r)$ | $IC_2(r)$ |
|---|---|---|---|---|
| 0 | 2.043 | 2.043 | -7.887 | -7.887 |
| 1 | 1.361 | 1.363 | -8.197 | -8.193 |
| 2 | 1.015 | 1.019 | -8.443 | -8.435 |
| 3 | 0.944 | 0.950 | -8.477 | -8.465 |
| 4 | 0.881 | 0.889 | -8.539 | -8.523 |
| 5 | 0.874 | 0.884 | -8.539 | -8.519 |

.

For our data set the first two criteria always choose $r^{\max}$ number of common factors. The last two criteria choose five and four common factors respectively. Bai and Ng (2002) observed (via simulation) that these model selection criteria select a large number of common factors relative to $N$ when $N$ is small, and Engle and Marcucci (2005) made a similar observation in their empirical study. Here, we argue that the relatively large number of common factors chosen in real data sets can be caused by large idiosyncratic jumps in asset prices.

Figure 1 illustrates how outliers can affect principal components. Based on a visual method suggested by Forni, Hallin, Lippi and Reichlin (2000), who worked with a more general factor model, the plots show the five largest eigenvalues of $\frac{1}{T}\sum_{t=1}^{T} Y_t Y_t'$ as $N$ is increased from 5 to 21, where $Y_t$ is the demeaned square root of realized variances. Diamonds, plus-signs, triangles, squares and circles respectively represent the first to fifth largest eigenvalues. The feature of interest is that the largest eigenvalue seems to be due to the variance of a single asset (LLC), because as soon as this asset is added to the set of $N$ variance measures, the diamond plot jumps to a value of around 2, and then stays at that value. This is clearly a symptom of small $N$, and the eigenvalue that is influenced by a single firm would eventually cease to dominate the analysis once $N \to \infty$. However, we show below that the realized variance process for most stocks are clearly dominated by large outliers, and we argue that one can get better estimates for the common factors by using alternative estimators that are more robust to jumps, or by purging jumps from the data set.

[Insert Figure 1 about here]

## 4. JUMPS

Many researchers have noted that models of asset returns that incorporate jumps fit the data better than models that don't allow for jumps (see Andersen, Bollerslev and Diebold 2005, and the references therein). Standard jump models are based on the assumption that the logarithm of an asset price follows a continuous time jump diffusion process given by

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t),$$

where $\mu(t)$ is a continuous function, $\sigma(t)$ is a strictly positive volatility process, and $\kappa(t)dq(t)$ is a jump process that allows for rare discrete jumps of size $\kappa(t)$ whenever $dq(t)$ equals 1. Under this assumption the realized variance defined in equation (1) converges to

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r^2_{t+j\Delta,\Delta} \xrightarrow{p} \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s<t+1} \kappa^2(s)$$

as $\Delta \to 0$. That is, the realized variance includes the squares of all jumps that have occurred between $t$ and $t+1$.

The dynamic properties of realized variance depend on the properties of both terms on the right hand side of the above equation, and both terms can be predictable. Andersen et al. (2001) provide a technical discussion on this issue. However, the degree of predictability in each of these terms can be quite different, and empirical evidence suggests that this is the case. Andersen et al. (2005) find that jumps are "distinctly less persistent than the continuous sample path variation process" and that "almost all of the predictability in daily, weekly and monthly return volatilities comes from the non-jump component". These authors cite several other papers that reach similar conclusions, and then they argue that this justifies the separate modelling and forecasting of continuous and jump components. We infer from this that jumps are largely unpredictable from the past history of stock prices, and that unless we explicitly account for them, it might be quite difficult to forecast realized variance directly, even if $\sigma(t)$ is predictable.

Below, we provide evidence that the presence of jumps in the price processes of multiple assets can mask the true factor structure of the stochastic volatility process when $N$ is relatively small. This means that we need to account for jumps when undertaking our factor analysis. In what follows we explore two ways of removing the influence of jumps on our analysis. Our first approach treats the jumps as a kind of measurement error, and uses instrumental variable methods to alleviate their effects on factor analysis. Our second approach uses a consistent estimate of $\int_t^{t+1} \sigma^2(s)ds$ which we take to be the predictable component of realized variance, and then it develops forecasting models for this component.

### 4.1 An instrumental variable estimator of common factors

Consider the model of the $N$ mean subtracted "volatilities" in $Y_t$

$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r)}{\Lambda} \underset{(r\times 1)}{C_t} + \underset{(N\times 1)}{J_t} + \underset{(N\times 1)}{u_t} \tag{7}$$

where $C_t$ are the "continuous" common factors, $J_t$ are mean-subtracted idiosyncratic jumps that are assumed to be unpredictable from the past, and $u_t$ are the idiosyncratic factors. We assume that the common factors $C_t$ are predictable from the past and that all regularity conditions on factor loadings $\Lambda$ and the cross sectional and time series dependence of common and idiosyncratic factors required for consistency of the principal component estimator of common factors as $\min\{N, T\} \to \infty$ are satisfied. See Bai and Ng (2002) for these regularity conditions. Further, we assume that the jumps $J_t$ have finite but large variances and are independent of each other and also independent of all other components. The addition of jumps to the usual approximate factor set-up does not change the asymptotic properties of the principal component estimator (because $J_t + u_t$ still satisfies the required conditions of idiosyncratic components). However, as Figure 1 and the above table of model selection criteria suggest, large idiosyncratic jumps might cause the principal component procedure to fit the noise rather than the signal when $N$ is small, so that the procedure identifies those variables with the largest jumps as the common factors, which then leads to incorrect inference about the number of common factors.

In order to mitigate the adverse effect of jumps in small samples, we introduce an instrumental variable estimator of the common factors. Consistent with our empirical example, we assume that $N < T$, but we will briefly discuss the case where $N \geq T$ in the remarks that follow our proposition. We start with a lemma that shows that the principal component estimator can be thought of as the least squares estimator of a reduced rank regression.

**Lemma 1** *The principal component estimator of common factors $\hat{A}'Y_t$, which by definition uses $\hat{A}$, the orthonormal eigenvectors corresponding to the largest $r$ eigenvalues of tr $(\mathbf{YY}')$, can be derived from fitting the best regression of rank $r$ of $Y_t$ on itself, i.e., it finds the $N \times r$ matrix $A$ and the $r \times N$ matrix $B$ in* *the reduced rank regression*

$$\mathbf{Y} = AB\mathbf{Y} + \mathbf{U} \tag{8}$$

*that minimizes tr* $\left(\mathbf{U}\mathbf{U}'\right)$ *subject to the normalization* $A'A = I_r$ *(where* $I_r$ *is the identity matrix of dimension* $r$*). Here* $\mathbf{Y}$ *is the* $N \times T$ *matrix of stacked* $Y_t$, $\mathbf{U}$ *is the* $N \times T$ *matrix of regression errors, and tr* $(.)$ *is the trace operator. Moreover,* $\hat{A}' = \hat{B}$.

**Proof.** This is a special case of Proposition A.5 in Lütkepohl (1991), which is a proposition about the generalized least squares estimator of parameters of a reduced rank regression of $\mathbf{Y}$ on $\mathbf{X}$. If one plugs in $\mathbf{Y}$ for $\mathbf{X}$ and an identity for the error variance matrix in the formula for the GLS estimator of reduced rank parameter matrices, then the results of this lemma follows. ■

Comparing equations (7) and (8), it becomes apparent that the principal component estimator of common factors is based on the idea of using $\hat{A}'Y_t$ as a proxy for $C_t$ and estimating equation (7) with least squares. The least squares estimator is a consistent estimator of common factors, despite the fact that the covariance of $\hat{A}'Y_t$ and $J_t + u_t$ is not zero in finite samples, because the regularity conditions guarantee that the variance of $\hat{A}'Y_t$ is of a larger order of magnitude than the variance of $J_t + u_t$ as $N \to \infty$, so that the "endogeneity" problem vanishes asymptotically. However, it is clear that this estimator might not be very good in finite samples, when there are large jumps. We therefore suggest the following estimator.

**Proposition 1** *Under the assumption that* $E\left(C_t[Y'_{t-1}, ..., Y'_{t-p}]\right)$ *has rank* $r$, *a consistent estimator of common factors as* $N, T \to \infty$ *with* $N < T$ *is* $\hat{A}'_{IV}Y_t$, *where* $\hat{A}_{IV}$ *consists of the eigenvectors corresponding to the* $r$ *largest eigenvalues of* $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ *and* $\hat{\mathbf{Y}}$ *is the orthogonal projection of* $\mathbf{Y}$ *on* $\mathbf{Y}_{-1}$. *Here,* $\mathbf{Y} = (Y_{p+1}, ..., Y_T)$ *is* $N \times (T-p)$ *and* $\mathbf{Y}_{-1}$ *is the* $Np \times (T-p)$ *matrix of lagged values, i.e.,*

$$\mathbf{Y}_{-1} = \begin{pmatrix} Y_p & & Y_{T-1} \\ \vdots & , ..., & \vdots \\ Y_1 & & Y_{T-p} \end{pmatrix}$$

*for some* $p > 0$. *Subject to the normalization that* $A'A = I_r$, *this estimator is also the ordinary least squares reduced rank regression estimator of* $A$ *in*

$$\mathbf{Y} = AB\mathbf{Y}_{-1} + \mathbf{U} \tag{9}$$

*that minimizes tr* $\left(\mathbf{U}\mathbf{U}'\right)$.

**Proof.** Consider the linear projection of $Y_t$ on $Y'_{t-1}, ..., Y'_{t-p}$ and denote all linear projections on this space by the subscript '$\mid t-1$'. We have

$$Y_{t|t-1} = \Lambda C_{t|t-1} + u_{t|t-1}, \tag{10}$$

where $J_{t|t-1} = 0$ because of the assumption that jumps are not predictable from the past. The assumption on the rank of covariance matrix of $C_t$ and lagged observations guarantees that the variance of $C_{t|t-1}$ is of full rank. The factor loadings in equation (10) are the same as those in equation (7), which by the assumption of approximate factor models guarantees that $C_{t|t-1}$ contributes to each and every element of $Y_t$ for all $N$. Moreover, the assumption of boundedness of the eigenvalues of $E\left(u_t u_t'\right)$ guarantees that the eigenvalues of $E(u_{t|t-1}u_{t|t-1}')$ are also bounded. Hence the regularity conditions of the approximate factor models are sufficient to ensure that the $r$ largest eigenvalues of $E(Y_{t|t-1}Y_{t|t-1}')$ diverge as $N \to \infty$, while the $(r+1)$-th largest eigenvalue remains bounded. Using the sample counterparts of this population entities, the eigenvectors of the $r$ largest eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$, where $\hat{\mathbf{Y}} = \mathbf{YP}$ with $\mathbf{P} = \mathbf{Y}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}$, provide consistent estimators of $r$ linear combinations of $Y_t$ whose orthogonal projections on the space of lagged $Y_t$ have largest possible variance diverging to infinity as $N$ increases. Since variance $\hat{A}_{IV}'\hat{Y}_t$ is always less than or equal to the variance of $\hat{A}_{IV}'Y_t$, it follows that $\hat{A}_{IV}'$ is also a consistent estimator of the eigenvectors corresponding to divergent eigenvalues of $E\left(Y_t Y_t'\right)$. For the second part of this proposition, note that $\mathbf{P}$ is symmetric and idempotent, and therefore $\hat{\mathbf{Y}}\hat{\mathbf{Y}}' = \mathbf{YPY}' = \mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$. The fact that the eigenvectors corresponding to the largest $r$ eigenvalues of this matrix are the estimates of $A$ in the reduced rank regression of $\mathbf{Y}$ on $\mathbf{Y}_{-1}$, is a well-known result in multiple regression theory (see Lütkepohl, 1991, Proposition A.5). ■

There are several remarks about the above proposition. Firstly, there is no requirement on the number of lags in $\mathbf{Y}_{-1}$, and in particular it is not necessary to add enough lags to explain all of the dependence of $Y_t$ on the past. The only requirement is that the covariance matrix of $C_t$ and the lags of $Y_t$ be of full rank. Basically, the consistency of the usual principal component estimator and our instrumental variable estimator hinges on different orders of magnitude of the variances of common and the idiosyncratic components in $Y_t$ as $N \to \infty$. Hence, incomplete dynamics or asymptotically negligible dependence between different components do not affect the consistency of this estimator. In practice, it is perhaps most appropriate to use a single lag of $Y_t$.

A second point is that we require that all common factors be predictable, and therefore we do not consider the possibility of common jumps. If some of common factors are common jumps, the instrumental variable approach only provides consistent estimators of the common forecastable factors.

A third remark is that if one is willing to add assumptions about the memory of common and idiosyncratic factors, then the problem can be cast as a common feature problem (see Engle and Kozicki 1993), and this would immediately suggest alternative solutions that do not require $N \to \infty$. For example, if one is willing to assume that common factors are autoregressive, while idiosyncratic factors are at most $m$-dependent, then this would imply that the $N$ elements of $Y_t$ are codependent (Gourieroux and Peaucelle 1988; Vahid and Engle 1997). In such a case, the projection on $Y_{t-m-1}$ will only include the common factors, i.e.,

$$Y_{t|t-m-1} = \Lambda C_{t|t-m-1},$$

and we could use a GMM test for codependence to determine the rank of $\Lambda$ (Vahid and Engle 1997). However, one may not want to make such an assumption in financial applications.

Our fourth remark is that we have assumed that $N < T$ because it applies to our empirical question. Conceptually, there is no substantive difference when $N \geqslant T$ because there is no part of the above argument that requires lags of all $N$ variables be used as instruments. The assumptions of factor loadings ensure that any subset of $N$ variables can be used as instruments.

Finally, consistent estimators of factor loadings $\Lambda$ can be estimated from a second stage least squares regression of $Y_t$ on $\hat{A}'_{IV} Y_t$. As in the case of the principal component estimator, even though $\hat{A}'_{IV} Y_t$ is a proxy for $C_t$ measured with error, the correlation between $\hat{A}'_{IV} Y_t$ and the errors becomes negligible as as $N \to \infty$, and hence least squares delivers consistent estimates of the factor loadings. Also, because the rate of convergence of $\hat{A}_{IV}$ to the eigenvectors of the divergent eigenvalues of $E(Y_t Y_t')$ is the same as that of the principal component estimator, we invoke Theorem 2 in Bai and Ng (2002) to construct model selection criteria for determining the number of common factors similar to those in equations (3), (4), (5) and (6) but with

$$ESS_{IV}\,(r) = \sum_{t=1}^{T} \sum_{i=1}^{N} (Y_{it} - \hat{\Lambda}_i' \hat{A}'_{IV} Y_t)^2$$

in place of $ESS\,(r)$. We evaluate the performance of these criteria in a simulation exercise reported in the next subsection. To indicate how projection on the space of just one lag can reduce the effect of outliers, we have plotted the largest 5 eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ for realized volatilities $\sqrt{RV}$ in Figure 2. Comparing Figure 2 with Figure 1, it is clear that the projection on one lag has eliminated the effect of idiosyncratic jumps.

[Insert Figure 2 about here]

Although the linear projection of $C_t$ on one lag can separate jumps and determine the number of forecastable factors, its implied forecast of $C_t$ is not the best possible forecast given the entire past information set. If it is reasonable to assume a parsimonious autoregressive dynamic model for $C_t$, then one can include the relevant lags in $\mathbf{Y}_{-1}$ and use the reduced rank regression (9) to deliver a "leading indicator" for forecasting $Y_t$. One can then make individual forecasting equations for each asset, using the leading index as an explanatory variable. This is one of the forecasting procedures that we pursue in this paper.

We close this section by comparing our estimator with two other estimators of common factors in the literature. Firstly, Forni et al. (2000) consider a more general dynamic factor model, in which each series may be affected by a different lag of the common factor. They transform their structure to the frequency domain and determine the number of common factors from the eigenvalues of spectral density matrices at different frequencies. Then they use the eigenvectors corresponding to their estimated eigenvalues to combine the spectral coordinates of the $N$ variables and hence obtain an estimate of the spectral density of the common factors. Then, through the inverse transform to the time domain, they find the weights of the filters that deliver the common factors. Since these filters are two-sided filters, they are not useful for forecasting. To overcome this problem, Forni, Hallin, Lippi and Reichlin (2003) find the projection of the factors on the past history to determine a one-sided estimate that can be used for forecasting. Our method can be viewed as a direct attempt to estimate the reduced form that is compatible with the factor structure. The advantage of the Forni et al. (2003) methodology is that their method delivers estimates of the covariances of the common and idiosyncratic factors, and using these, one can derive the parameters of an $h$-period ahead leading index for the common factors for any $h$. Our method would need to first specify $h$, and then choose the lags in $\mathbf{Y}_{-1}$ so as to deliver an $h$-step ahead leading indicator ( the most recent information in $\mathbf{Y}_{-1}$ would be $Y_{t-h}$).

Secondly, our method is closely related to methods of dimension determination in linear systems theory, such as those in Akaike (1976), Havenner and Aoki (1988) and Aoki and Havenner (1991). Recent work by Kapetanios and Marcellino (2004) has extended the last of these to the case with large $N$. Their approach determines the dimension of the state space that links the past to the current and future by examining the singular values of the covariance matrix between the past and the future. If "the future" is left out and one looks at the relationship between current and the past, then this method will be the same as examining the eigenvalues and eigenvectors of $\left(\mathbf{YY}'\right)^{-1}\mathbf{YY}'_{-1}\left(\mathbf{Y}_{-1}\mathbf{Y}'_{-1}\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$. Kapetanios and Marcellino (2004) replace $\left(\mathbf{YY}'\right)^{-1}$ with an identity matrix stating that under the assumptions of their model, in which the number of lags in $\mathbf{Y}_{-1}$ also increases to infinity, this substitution does not affect the consistency of their estimator for common factors. Here, we cannot get that result.

### 4.2 A Simulated Example

We have undertaken a small simulation exercise to illustrate (i) how large jumps can distort principal components based inference when $N$ is small; and (ii) how well our instrumental variables approach corrects for this. To do this, we generated 100 samples of 250 observations of 50 variables which shared a single common factor. Each variable was the sum of the common factor and an idiosyncratic factor. All idiosyncratic factors were independent of each other and independent of the common factor. The common factor and all idiosyncratic factors were autoregressive processes of order one with an autoregressive parameter of 0.5 and standard Normal innovations. In each replication, we generated 450 observations of each factor, and threw away the first 200 observations to minimize the effect of initial conditions.

We began our study by using information criteria to determine the number of common factors in 20 out of the 50 variables generated in each sample, and then we repeated the same exercise adding one extra variable at a time. In each case, we allowed for up to $N/2$ common factors, and estimated the common factors in two different ways: once using the usual principal component (PC) estimator of the common factors, and then again using the instrumental variable (IV) approach. Specifically, the PC estimators of $r$ common factors were eigenvectors corresponding to the $r$ largest eigenvalues of $\mathbf{YY}'$, whereas the IV estimator of the common factors were the eigenvectors corresponding to $r$ largest eigenvalues of $\mathbf{YY}'_{-1}\left(\mathbf{Y}_{-1}\mathbf{Y}'_{-1}\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$. For each $N$ (for $20 \leq N \leq 50$) and each $r \leq N/2$, the PC (IV) factor loadings were estimated using least squares regressions of each variable on the estimated PC (IV) common factors, and the two resulting sums of $N$ squared errors over 250 observations ($ESS_{PC}(r)$ and $ESS_{IV}(r)$) were used to minimize the information criteria

$$IC_J(r) = \ln\left(\frac{ESS_J(r)}{NT}\right) + r \times \frac{N+T}{NT}\ln\left(\frac{NT}{N+T}\right)$$

over $r$ for $J = \{PC, IV\}$. We use the notation $r_{PC}$ and $r_{IV}$ to denote the chosen value of $r$ in each case.

The top panel of Table 4 shows the results of this exercise for $N = 20$ to 50 with increments of 5. Both model selection criteria work perfectly, even with $N = 20$. This is not surprising as the design is quite simple and ideal for the model selection criterion, because all factors are independent of each other and the common to idiosyncratic variance ratio is $1:1$ for all variables.

[Table 4 about here]

Next, we repeated this exercise, but introduced "jumps" (i.e. additive outliers) into the generated variables. For each variable we set the probability of having an extra additive term at each time period equal to 0.005, which translates to an expected 1.25 jumps in 250 observations. Each of these additive jumps were generated independently from a uniform distribution on $[0, x]$. We performed the exercise with $x$ equal to 10, 20, 30, 40 and 50 times the standard deviation of the common factor and report the results for $x = 30$ and $x = 50$ in Table 4. Before discussing the results we point out that (i) $x$ determines the maximum magnitude of the jump and not the magnitude of each and every jump; and (ii) the average ratio of the largest jump to the standard deviation of the filtered volatility over the 21 stocks in our data set is about 30.

The bottom two panels of Table 4 show that these large outliers cause $IC_{PC}$ to over-estimate the number of common factors, particularly when $N$ is small. For example, when the maximum magnitude of jumps is 30 standard deviations, $IC_{PC}$ chooses an $r_{PC} > 1$, 68% of the time. In contrast, $IC_{IV}$ chooses $r_{IV} > 1$, only 15% of the time. As the theory predicts, the performance of both criteria improves as $N$ increases, but for every $N$, $IC_{IV}$ is more accurate than $IC_{PC}$. With larger jumps, this pattern is even more clear. With jumps that have a maximum magnitude of 50 standard deviations, $IC_{PC}$ almost always over-estimates $r$ when $N$ is small, and even when $N = 50$, it chooses the correct $r$, (i.e. $r = 1$) only 34% of the time and chooses $r \geq 4$ in 40% of the replications. In contrast, when $N = 50$, $IC_{IV}$ correctly chooses $r = 1$, 79% of the time, and it never chooses $r \geq 4$. Again, for each $N$, the performance of $IC_{IV}$ is significantly more accurate than that of $IC_{PC}$.

Although we have reported the results for only one of the four criteria designed by Bai and Ng (2002), the results for all other criteria are qualitatively the same. We conclude that the problem is not related to the penalty term in the information criterion. Rather, it appears that when $N$ is small, large outliers cause the principal component estimator to fit the noise rather than the signal, and it then chooses variables with large outliers as "common factors". As long as outliers are large enough to cause such distortions, altering the penalty term does not fix the problem.

To conclude this section, we emphasize that this simple simulation exercise was for demonstrative purposes only, and we do not claim that it is a comprehensive Monte Carlo study. The effect of large outliers on the PC estimator of common factors is a small sample problem, which will disappear as $\min\{N, T\} \to \infty$. To fully investigate the nature of this small sample phenomenon, one needs a comprehensive Monte Carlo study that changes every parameter in the design, including the number of common factors, the parameters of the generating processes for common and idiosyncratic factors, the ratio of the variance of the common and idiosyncratic factors, the frequency and the intensity of the jumps, and so on. Such a study is outside the scope of our paper. Our purpose with this example is simply to confirm that large outliers can influence the PC estimator of common factors, and that looking at the projection of the variables on a space that is orthogonal to jumps but not orthogonal to common factors can provide a workable solution to this problem.

### 4.3 Forecastable Component of Realized Variance

Another way to attenuate the effects of jumps is to remove them from the data. Barndorff-Nielsen and Shephard (2004) show that a properly normalized sum of the absolute value of adjacent $\Delta$-period returns converges to the integral of quadratic variation excluding the contribution of jumps. That is, as $\Delta \to 0$,

$$BV_{t+1}(\Delta) \equiv \frac{\pi}{2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta,\Delta}| \left|r_{t+(j-1)\Delta,\Delta}\right| \xrightarrow{p} \int_{t}^{t+1} \sigma^2(s)ds. \tag{11}$$

The $BV$ term in equation (11) is called the realized bi-power variation, and this equation shows how a consistent estimator of the quadratic variation of the continuous path process of 1-period returns can be calculated from $\Delta$-period returns. Equation (11) also implies that a consistent estimator for the jumps between times $t$ and $t+1$ is given by

$$RV_{t+1}(\Delta) - BV_{t+1}(\Delta) \xrightarrow{p} \sum_{t<s<t+1} \kappa^2(s).$$

Since the difference between realized variance and bi-power variation is not always positive, a preferred estimator for jumps is given by

$$J_{t+1}(\Delta) = \max\{RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0\}. \tag{12}$$

Barndorff-Nielsen and Shephard (2004) establish that for a process that has no jumps, $RV_{t+1}$ is a slightly better estimator for quadratic variation than

$BV_{t+1}$. However, since jumps are commonly believed to be present in asset volatilities, it is reasonable to expect that unless they are purged from the data, they will have a distortionary effect on the specification of any time series models developed to forecast realized volatility. Also, since jumps are not typically considered to be forecastable, it seems sensible to concentrate on building forecasting models for $BV$, since $BV$ will be the only forecastable component of $RV$. Andersen et al. (2005) show that realized jumps in exchange rate data that are computed using equation (12) have no forecasting power for realized volatility, whereas bi-power variation in (11) has considerable forecasting power.

We calculate weekly bi-power variation and jumps for our 21 Australian stocks using the fifteen minute returns as before. There are large estimated jumps in almost all of the 21 stocks, and Figure 3 shows the calculated time series of realized variance, bi-power variation and jumps for the AMCOR corporation (AMC), the Commonwealth Bank (CBA) and the Lend Lease Corporation (LLC). The realized variance and the jump time series are plotted with the same scale, while the bi-power variation plots have a different scale for a better visualization.

The plots show that some but not all of the large outliers in realized volatilities have been identified as jumps. Tests for the statistical significance of jumps have been developed in Barndorff-Nielsen (2004), but we do not use them here because our primary purpose is simply to use bi-power variation to forecast volatility. A user familiar with Australian company histories would know that some of jumps identified by equation (12) have been caused by buy-backs, bonus share issues or other forms of capital restructuring that cause a one-time jump in the share price without having any bearing on its volatility. In principle, when the information is available, one should remove such jumps from the price data prior to undertaking volatility analysis, but in practice this information might not be available for each and every stock. Indeed an advantage of using equation (12) is that it removes price movements that carry no news about volatility, including those that might not have come to the attention of the researcher prior to starting his/her analysis.

[Insert Figure 3 about here]

Table 3 reports the serial correlation properties of realized variance, bi-power variation and jumps for the 21 weekly return series. The entries in columns 2 to 5 of this table are the p-values of LM test for no serial correlation against the alternative of fourth order serial correlation in the realized variance, jumps, bi-power variation and the logarithm of the square root of the realized bi-power variation. We work with the last measure since the preferred measure of "volatility" in finance is the standard deviation rather than variance. The first point to note is that hardly any of the estimated jump components are serially correlated. The only clear exception is the jump component of NAB, with MIM and RIO being borderline at the 5% level of significance. The second point to note is that there is significant evidence of serial correlation in nearly all of the realized bi-power variations, and in all of the logarithms of the square root of the bi-power variations. We can then note that some of the unfiltered realized variances show no significant sign of serial correlation. This supports our earlier conjecture that large jumps in realized variance might be diluting the evidence of forecastability in variance. The ARMA(1,1) model fits all of the $\ln\left(\sqrt{BV}\right)$ series quite well. The estimated autoregressive parameters in these ARMA models are all large, and the MA polynomials have roots slightly smaller than the AR polynomials. Such ARMA models imply autocorrelations that are small but persistent. This pattern in the volatility of financial assets has been found in the past (regardless of how volatility has been measured), and this has sometimes led researchers to model volatilities as fractionally integrated processes (see e.g. Baillie, Bollerslev and Mikkelson, 1996).

Similarity between the univariate ARMA models that describe the log-volatility series suggests that the multivariate modelling of these series may be fruitful. For example, a VARMA model for a group of time series implies univariate representations that all have the same autoregressive polynomial. However, the proper identification of a 21-dimensional VARMA model is obviously quite difficult, and it is therefore natural to consider simple parsimonious factor structures.

## 5. EMPIRICAL RESULTS

The ability to compute an estimate of volatility from high frequency data provides a readily available time series for this unobserved variable, and opens up the possibility of modelling and forecasting it using standard methods that are usually used for the conditional mean. Since the results of the previous section indicate that bi-power variation is the only broadly forecastable component of realized variance, we investigate the multivariate modelling of $\ln\left(\sqrt{BV}\right)$, and call these series "log volatilities". We focus on finding good forecasting models for the log volatility of Australian stock returns, in order to determine whether the incorporation of common factors will improve forecasting performance.

We develop seven forecasting models, which include three univariate models and four factor models (described below). All models are developed using data from the first week of 1996 to the last week of 2000 (260 observations), and they are used to provide one-step ahead forecasts of log volatility for the 52 weeks of 2001. Absolutely no information from the forecast period is used in the development of any of the models. We produce two sets of forecasts for comparing models. The first set is based on a fixed estimation window. The second set is based on the same set of model specifications as the first, but the forecasting models are re-estimated to produce each one step ahead forecast, using a rolling window that maintains the original sample size (260 observations). Our evaluations are based on root mean squared forecast errors (RMSE), forecast encompassing regressions and Giacomini and White (2004) tests of predictive ability, and these are all based on forecasts of log volatility.

However, we also use our forecasts of log volatility to calculate forecasts of $BV$ and $RV$, to show that our forecasting models for log volatility are also useful for predicting these latter measures of volatility.

**5.1 Univariate Models**

We estimate univariate ARMA models, single exponential smoothing models and a pooled model for all 21 log-volatility series. The ARMA models are chosen following the Hannan-Rissanen (1982) methodology, which finds $p^{\max}$, the order of the best fitting AR model chosen by AIC, and then considers all ARMA models whose sum of the AR and the MA orders is less than or equal to $p^{\max}$ chosen by the Schwarz criterion. All but six of the chosen models are ARMA(1,1). Parameter values for the ARMA(1,1) models are reported in Table 3. As noted before, the fitted ARMA models have the common characteristic that the AR parameter is large, and the roots of the AR and MA polynomials are similar, reflecting autocorrelations that are small but persistent.

Single exponential smoothing models are often promoted as the most time effective method for forecasting a large number of time series. They are local level models parameterized in terms of a "smoothing parameter" rather than in terms of the signal to noise variance ratios (see Harvey 1991, p. 175). Estimated values of the smoothing parameters are reported in Table 5. They are relatively small, again reflecting the persistence in log volatility.

The pooled model estimates all univariate equations by jointly restricting all parameters (other than the mean) to be the same. If these restrictions are correct, then the cross sectional variation leads to more precise estimates of the parameters. In this approach, we allow the dynamics of each series to be given by a long autoregression. The pooled data shows evidence for seven lags, and in the pooled AR model the parameters of lags three to seven are small and close to each other, which is typical of processes with small but persistent autocorrelations.

[Table 5 about here]

**5.2 Factor Models**

We find that when we apply the Bai and Ng (2002) criteria to realized variances, they decline as the number of factors increase. However, once we purge the jumps from individual variance series and analyze series of $Y_t = \ln\left(\sqrt{BV_t}\right)$, then model selection criteria based on the principal components choose just one or two common factors. When we then use the criteria that use $Y_{t-1}$ as an instrument for $Y_t$, all select a one factor model. If we assume that the dynamics of this system can be well specified by a finite VAR and use the model selection criteria used in Vahid and Issler (2002) to choose the lag and rank of the VAR simultaneously, then we choose a lag of one and rank of two. This prompts our development of the four factor models described below.

Our first factor model is very simple, and takes the unweighted average of the 21 log volatility series as the estimate of the common factor. This average is plotted in Figure 4, and if there is only one common factor, then this series provides a consistent estimate of the common factor. To develop forecasting models based on this factor, we use lags of this variable as regressors and allow for ARMA errors. Most final models resemble the univariate models with the market variable included as a regressor. This model is denoted by EqW (equally weighted) model.

[Insert Figure 4 about here]

The second model is one that has two factors estimated by principal components. We chose two factors because there was a conflict among model selection criteria on whether one or two factors were appropriate. We made separate time series models for these factors and after regressing each log volatility on these factors to obtain factor loadings and the factor components, we made 21 ARMA models for the remainders (i.e. the idiosyncratic components). We then forecasted the factor and idiosyncratic components for each log volatility and added these to obtain forecasts of each log volatility. The two factors are plotted in Figure 5. The first factor is very similar to the average factor plotted in Figure 4, while the second factor looks like a slow moving underlying level. The time series model fitted to the first factor is an ARMA(2,1), while the fitted model for the second factor is an AR(4) model in which all AR parameters are constrained to be equal. This type of autoregressive model has been used quite recently, as an alternative for modelling persistent financial time series (see Andersen et al. (2005) and the references therein). The models for the idiosyncratic components are all low order ARMA models.

[Insert Figure 5 about here]

The third model uses the principal component analysis of the linear projection of $Y_t = \ln \sqrt{BV_t}$ on $Y_{t-1}$. All model selection criteria choose only one factor. As described in the previous section, this analysis also provides a leading indicator for the common factor, which is a linear combination of $Y_{t-1}$. Rather than making a separate ARMA model for the factor, we take this leading indicator and use it as a regressor in the equation for each log-volatility. We call the resulting models IVLI (instrumental variable-leading indicator) models. The leading index is plotted in Figure 6. As can be seen from this figure, the leading indicator looks like a good indicator for the market factor plotted in Figure 4.

[Insert Figure 6 about here]

The final model assumes that the log volatilities can be adequately modelled by a VAR, and uses model selection criteria to choose number of lags and rank of this VAR. This procedure chooses one lag and rank of two. Of course in a VAR with 21 variables it is unlikely that each equation will have white noise errors, but their serial correlation is too weak to warrant the addition of another lag (i.e., 441 parameters) to the VAR. We therefore check the errors of each equation and allow for serially correlated errors when this is needed. These models are denoted by CC (canonical correlation) models.

**5.3 Forecast Comparisons**

Table 5 reports the root mean squared forecast errors of the univariate models based on a fixed estimation sample. We can see that with only two exceptions (FGL and WOW), ARMA models produce better forecasts than the exponential smoothing models. Also, there are only six out of twenty one cases where the forecasts based on the pooled time series equations have smaller RMSE than the ARMA forecasts. This is perhaps not surprising because pooling (and also exponential smoothing) usually help when the time series dimension is too small to allow the precise estimation of a univariate model. Here, we have 260 observations for each variable in the estimation sample and pooling simply imposes a blanket restriction that may not be true. Of course we may have pooled too much. It may be that if we had only pooled the log-volatility data of stocks of the same industry group, or if we had used a data driven procedure for pooling such as that suggested in Vahid (1999), then we may have obtained better results. We do not pursue these issues here, and leave them for future research.

Table 6 reports the root mean squared forecast errors of the factor models based on a fixed estimation sample. It is evident that they outperform the univariate models in almost every case. When comparing multivariate models with each other, the only remarkable result is how well the simple average factor model performs. This model, under the heading of "EqW" performs best for 13 out of the 21 series, and performs second best in another 4. Of course an equally weighted estimate is a consistent estimator of the common factor when there is only one common factor in the model. Its strong performance in out of sample forecasting suggests that there is only one common factor in the volatilities of Australian stocks. It also shows that our attempts to get better estimates of this factor by using statistical procedures do not really pay off.

[Insert Table 6 about here]

Tables 5 and 6 show that the RMSEs associated with the factor models are usually smaller than those for the univariate models, but these differences are small in magnitude and may not be statistically significant. We have therefore run forecast encompassing regressions to examine statistical significance. The forecast encompassing regressions are regressions of the actual log-volatility on different pairs of forecasts for the out-of-sample period, and if the parameters for each forecast are insignificant then the forecasts are equally good, meaning that given one set of forecasts there is no significant information in the other. If both are significant, then neither encompasses the other and there is scope for combining them. If one is significant and the other is not, then the forecast with a significant coefficient encompasses the other forecast. Detailed regression results are not provided here, but the information revealed by these regressions is similar to the information drawn by comparing RMSEs. First, and perhaps not surprisingly, these regressions tell us that the multivariate models often encompass the univariate models. The strongest evidence relates to the model with equally weighted (EqW) factor estimates. When comparing EqW models with ARMA models, the former encompasses the latter at the 5% level of significance in 17 out of 21 cases, while the ARMA model never encompasses the EqW model. When comparing factor models, at the 5% level of significance, the EqW model encompasses the instrumental variable leading indicator (IVLI) model in forecasting four of the log-volatilities and is never itself encompassed. The EqW forecasts only encompass the canonical correlation forecasts twice and are themselves encompassed only once. The (IVLI) and the (CC) forecasts appear to be equivalent in all twenty one cases.

Summary statistics relating to our out of sample forecasts derived from using rolling windows for estimation are provided in Tables 7 and 8. Here, we compare forecasts derived from the ARMA models, the EqW model and the PC model. We present the RMSEs for these three models in Table 7, and use stars to indicate those cases when forecasts from the EqW model (or PC model) outperform those from the ARMA model. Here we use the tests of unconditional predictive ability proposed by Giacomini and White (2004) to assess predictive ability. These tests are based on tests of the null hypothesis that $E(e_{1t}^2) = E(e_{2t}^2)$, based on squared forecast errors $\widehat{e}_{1t}^2$ and $\widehat{e}_{2t}^2$. This test is just a t-test on the sample mean of $(\widehat{e}_{1t}^2 - \widehat{e}_{2t}^2)$, with a heteroskedasticity and autocorrelation consistent (HAC) standard error. We see that factor models outperform ARMA models in five out of six cases for material stocks and three out of five cases for banking stocks, suggesting that factors can be helpful when forecasting a stock that belongs to an industry group that is well represented in the sample. We provide standard deviations of errors measures for $\ln\sqrt{BV_t}$ and $\ln\sqrt{RV_t}$ in the last two columns of Table 7 as a benchmark. For each stock, the RMSEs from the factor models are smaller than the standard errors of the corresponding out of sample $\ln\sqrt{BV_t}$ observations (there are only three exceptions) and also smaller than the standard errors for out of sample $\ln\sqrt{RV_t}$ observations (with only one exception).

[Insert Tables 7 and 8 about here]

We explore the ability of our models to predict other volatility measures by converting our rolling window forecasts of $\ln\sqrt{BV_t}$ to forecasts of $BV_t$ and $RV_t$, and then measuring the correlation between our converted forecasts and our observed out of sample data. Our conversion of $\ln\sqrt{BV_t}^f$ to $BV_t^f$ is based on a simple exponential transformation (that makes no adjustment that assumes a normal distribution), and our conversion of $BV_t^f$ to $RV_t^f$ is based on coefficients obtained by regressing the in-sample observations of $RV_t$ on in-sample predictions of $BV_t$ (the exponential transformations of in-sample predictions of $\ln\sqrt{BV_t}$ from each model). Despite the crude conversion techniques that we have used, the correlations that we find when using the forecasts derived from factor models are significantly positive in the vast majority of cases, showing that the forecasts are able to track the observed data very well.

## 6. CONCLUSION

In this paper we argue that the principal component procedures that are typically used for factor analysis in approximate factor models can be misled by large outliers (be it measurement errors or jumps). These concerns are particularly relevant when forecasting the volatilities of asset returns because the price processes includes jumps, and volatilities can only be measured with error. As a solution, we propose a procedure that is based on principal component analysis of the linear projection of variables on their past. We then note that the usual principal component procedure, the canonical correlation procedure and our suggested procedure can be seen as different methods of estimating a reduced rank regression, and we give our procedure an instrumental variable interpretation in this context.

We use these procedures to determine the number of forecastable factors in the log-volatilities in the returns of 21 Australian stocks. Volatilities of weekly returns are estimated from fifteen minute returns, and jumps are isolated and removed by using the non-parametric method developed by Barndorff-Nielsen and Shephard (2004). Once jumps have been removed, the model selection criteria provide very similar estimates of the number of common factors.

We then ask whether these factors help in forecasting log-volatilities. The answer is yes. More interestingly, our results show that an equally weighted average of all log-volatilities can improve forecasts of log-volatility more than principal component or canonical correlation estimates of common factors. There are similar results about the superiority of equally weighted indices over indices with estimated weights elsewhere in the forecasting literature. For example, the business cycle coincident indicator of the Conference Board in the US is a simple average of four standardized variables, and it has performed remarkably well in post war history. Closer to this study is recent work by Deistler and Hamann (2005), who have found that simple average "factors" work better than standard factor models when forecasting stock returns.

There are two caveats that must be noted in relation to our results. First, our out-of-sample comparisons are based on only one year of weekly data (52 observation), so that the ordering of our models should be interpreted with caution. Second, we have only considered how well different models forecast the logarithm of volatility. The mapping from the forecast of log-volatilities to volatilities involves conditional moments other than just the conditional mean of the log-volatility process, and it is possible that the ordering of different models might change after this transformation.

It is possible that forecast averaging over our various factor models might further improve our forecasts (see Timmermann 2005), but we leave this for future research. Another avenue for future research is to evaluate our forecasts using empirically relevant loss functions. Here, we have compared forecasts based on squared error loss. We think that future work based on determining whether our simple factor models of volatility lead to improved derivative pricing would be an interesting avenue to explore.

*Acknowledgments*: We thank the Securities Industry Research Centre of Asia and the Pacific (SIRCA) for providing the data, and thank Nelufa Begum and Jing Tian for research assistance. We have benefited from comments of the Associate Editor, three anonymous referees, Peter Howard, and seminar participants at Monash University, the ANU and the "Common Features in London Conference". Both authors acknowledge the financial assistance of Australian Research Council Discovery Grants DP0449995 and DP0343811.

# References

Akaike, H. (1976), "Canonical correlation analysis of time series and the use of an information criterion," in *System Identification: Advances and Case Studies*, eds. R. K. Mehra and D. G. Lainiotis, Academic Press, pp. 27-96.

Andersen, T.G., Bollerslev, T., and Diebold, F.X. (2005), "Roughing it up: including jump components in measuring, modeling and forecasting return volatility", Working paper, Duke University.

Andersen, T.G., Bollerslev, T., Diebold, F.X., and Ebens, H. (2001), "The distribution of realized stock volatility", *Journal of Financial Economics*, 61 43-76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2003), "Modeling and forecasting realized volatility", *Econometrica*, 71, 579-625.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The distribution of realized exchange volatility", *Journal of the American Statistical Association*, 96, 42-55.

Anderson, H.M., and Vahid, F. (1998), "Detecting common nonlinear components using canonical correlations," *Journal of Econometrics*, 84, 1-36.

Aoki, M., and Havenner, A.M. (1991), "State space modelling of multiple time series", *Econometric Reviews*, 10, 1-59.

Bai, J., and Ng, S. (2002), "Determining the number of factors in approximate factor models", *Econometrica*, 70, 191-222.

Baillie, R.T., Bollerslev, T., and Mikkelsen, H.O. (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 74, 3-30.

Barndorff-Nielsen O.E., and Shephard N. (2005), "Econometrics of testing for jumps in financial economics using bi-power variation", forthcoming in the *Journal of Financial Econometrics*.

Barndorff-Nielsen O.E., and Shephard N. (2004), "Power and bi-power variation with stochastic volatility and jumps", *Journal of Financial Econometrics*, 2, 1-37.

Barndorff-Nielsen O.E., and Shephard N. (2002), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models", *Journal of the Royal Statistical Society, Series B,* 64, 253 - 280.

Bollerslev, T. (1990), "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model", *Review of Economics and Statistics*, 72, 498-505.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Candelon, B., Hecq, A., and Verschoor, W.F.C. (2005), "Measuring common cyclical features during financial turmoil: Evidence of interdependence not contagion", *Journal of International Money and Finance*, 24, 1317-1334.

Chamberlain, G., and Rothschild M. (1983), "Arbitrage, factor structure and mean-variance analysis in large asset markets", *Econometrica*, 51, 1305-1324.

Cochrane, J. H. (2001), *Asset Pricing*, Princeton University Press.

Deistler, M., and Hamann, E. (2005), "Identification of Factor Models for Forecasting Returns" *Journal of Financial Econometrics*, 3, 256-281.

Diebold, F.X., and Nerlove, M. (1989), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor Arch Model" *Journal of Applied Econometrics*, 4, 1-21.

Doz, C., and Renault, E. (2004), "Conditionally heteroskedastic factor models: Identification and instrumental variable estimation", Working paper, University of Montreal.

Engle, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation", *Econometrica*, 50, 987-1008.

Engle R. F., (2002), "Dynamic conditional correlation - A simple class of multivariate GARCH models", *Journal of Business and Economic Statistics*, 20, 339-350.

Engle R. F., and Kozicki, S. (1993), "Testing for common features", *Journal of Business and Economic Statistics*, 11, 369-380.

Engle, R.F., and Kroner, K.F. (1995), "Multivariate simultaneous generalized ARCH", *Econometric Theory*, 11, 122-150.

Engle, R.F., and, Marcucci, J. (2005), "A long-run pure variance common feature model for the common volatilities of the Dow Jones", forthcoming in the *Journal of Econometrics*.

Engle, R.F., Ng, V.K., and Rothschild, M. (1990), "Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills", *Journal of Econometrics*, 45, 213-238.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The generalized factor model: identification and estimation". *The Review of Economics and Statistics* 82, 540-554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003), "The generalized factor model: one-sided estimation and forecasting". LEM working paper series 2003/13.

Giacomini, R. and White, H. (2004), "Tests of conditional predictive ability", Working paper, University of California, Los Angeles and San Diego.

Gourieroux, C., and Peaucelle, I. (1992), "Series codependantes: application a l'hypothese de parite du pouvior d'achat", *Revue d'Analyse Economique* 68, 283-304.

Hannan, E.J., and Rissanen, J. (1982), "Recursive estimation of mixed autoregressive moving average order," *Biometrika*, 61, 81-94.

Harvey, A.C. (1991), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A.C., Ruiz, E., and Shephard, N. (1994), "Multivariate stochastic variance models", *Review of Economic Studies*, 61, 247-264.

Havenner, A.M., and Aoki, M. (1988), "An instrumental variable interpretation of linear systems theory", *Journal of Economic Dynamics and Control*, 12, 49-54.

Kapetanios, G., and Marcellino, M. (2004), "A parametric estimation method for dynamic factor models of large dimensions", Working paper, Queen Mary, University of London.

King, M., Sentana, E., and Wadhwani, S. (1994), "Volatility and links between national stock markets", *Econometrica*, 62, 901-934.

Lo, A. W., and Wang, J. (2000), "Trading Volume: definitions, data analysis and implications of portfolio theory", *The Review of Financial Studies*, 13, 257 - 300.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Heidelberg: Springer-Verlag.

Merton, R. C. (1980), "On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics*, 8, 323-361.

Timmermann, A. (2005),“Forecast Combinations”, forthcoming in the *Handbook of Economic Forecasting*, eds. G. Elliott, C.W.J. Granger and A. Timmermann, Amsterdam: North Holland.

Vahid, F. (1999), “Partial Pooling: A Possible Answer to ‘To Pool Or Not To Pool’ ” , in *Cointegration, Causality and Forecasting: Festschrift in honor of Clive W. J. Granger,* eds. R. F. Engle and H. White, Oxford University Press, pp. 410-428.

Vahid, F., and Engle, R.F. (1997), “Codependent cycles”, *Journal of Econometrics*, 80, 199-221.

Vahid, F., and Issler, J.V. (2002), “The importance of common cyclical features in VAR analysis: a Monte-Carlo study”, *Journal of Econometrics*, 109, 341-363.

**Table 1: The 21 most frequently traded stocks on the Australian Stock Exchange**( 1996 to 2001)

| Stock Code | Name of the Company | GICS Industry Group |
|---|---|---|
| AMC | AMCOR LIMITED | Materials |
| BHP | BHP BILLITON LIMITED | Materials |
| CSR | CSR LIMITED | Materials |
| MIM | MIM HOLDINGS LIMITED | Materials |
| RIO | RIO TINTO LIMITED | Materials |
| WMC | WMC RESOURCES LIMITED | Materials |
| STO | SANTOS LIMITED | Energy |
| WPL | WOODSIDE PETROLEUM LIMITED | Energy |
| ANZ | AUSTRALIA AND NEW ZEALAND BANKING GROUP LIMITED | Banks |
| CBA | COMMONWEALTH BANK OF AUSTRALIA | Banks |
| NAB | NATIONAL AUSTRALIA BANK LIMITED | Banks |
| SGB | ST GEORGE BANK LIMITED | Banks |
| WBC | WESTPAC BANKING CORPORATION | Banks |
| FGL | FOSTER'S GROUP LIMITED | Food, Beverage and Tobacco |
| SRP | SOUTHCORP LIMITED | Food, Beverage and Tobacco |
| BIL | BRAMBLES INDUSTRIES LIMITED | Commercial Services and Supplies |
| LLC | LEND LEASE CORPORATION LIMITED | Real Estate |
| MAY | MAYNE GROUP LIMITED | Health Care Equipment and Services |
| NCP | NEWS CORPORATION LIMITED | Media |
| QAN | QANTAS AIRWAYS LIMITED | Transportation |
| WOW | WOOLWORTHS LIMITED | Food and Staples Retailing |

Note: "GICS" stands for Global Industry Classification Standard. MIM is no longer traded on the Australian Stock Exchange and WMC is now known as WMR.

**Table 2: Summary statistics of weekly stock returns**
(First week of 1996 to last week of 2000)

| Stock | Mean | St.Dev. | Skewness | Kurtosis | ARCH | $\alpha+\beta$ | $\beta$ |
|---|---|---|---|---|---|---|---|
| Materials | | | | | | | |
| AMC | -.0023 | .0348 | -.5674 | 6.0796 | .0150 | .8581 | .7115 |
| BHP | .0000 | .0351 | .6086 | 4.8252 | .0011 | .9560 | .8741 |
| CSR | .0002 | .0358 | .1862 | 3.1434 | .7322 | | |
| MIM | -.0018 | .0543 | .0395 | 4.1165 | .0211 | .9102 | .8460 |
| RIO | .0015 | .0386 | .3014 | 3.3472 | .0107 | .9858 | .9579 |
| WMC | -.0005 | .0439 | -.4216 | 5.9406 | .1993 | | |
| Energy | | | | | | | |
| STO | .0017 | .0350 | .2989 | 3.5078 | .9981 | | |
| WPL | .0028 | .0364 | .2438 | 3.5483 | .9467 | | |
| Banks | | | | | | | |
| ANZ | .0031 | .0332 | -.2895 | 3.9104 | .6044 | | |
| CBA | .0040 | .0280 | -.4134 | 3.6174 | .6988 | | |
| NAB | .0032 | .0306 | -.3692 | 4.0922 | .9220 | | |
| SGB | .0023 | .0029 | -.1383 | 3.9486 | .0011 | .7400 | .6434 |
| WBC | .0031 | .0316 | -.4216 | 3.9681 | .0059 | .2293 | .1155 |
| Food & Bev | | | | | | | |
| FGL | .0028 | .0294 | .3047 | 3.5074 | .2508 | | |
| SRP | .0018 | .0393 | -.4515 | 6.2919 | .0637 | | |
| Other | | | | | | | |
| BIL | .0039 | .0337 | .2522 | 4.3419 | .8063 | | |
| LLC | -.0007 | .0567 | -6.7069 | 76.010 | .9998 | | |
| MAY | -.0001 | .0433 | -.1888 | 6.1116 | .6038 | | |
| NCP | .0025 | .0542 | .3309 | 5.6748 | .0001 | .9921 | .9437 |
| QAN | .0017 | .0407 | -.0648 | 4.0001 | .4054 | | |
| WOW | .0036 | .0299 | .2084 | 3.3449 | .6091 | | |

Notes: Entries in the 'ARCH' column are p-values of the LM test of the null hypothesis of no conditional heteroskedasticity against an ARCH(4) alternative. Results for ARCH(p) tests (for $1 \leq p \leq 12$) are largely robust to the choice of $p$. Estimates of GARCH(1,1) parameters are provided if there is significant evidence of conditional heteroskedasticity. A GARCH(1,1) specification implies ARMA(1,1) dynamics for the squared returns, i.e., $r_t^2 = \omega + (\alpha+\beta)\, r_{t-1}^2 - \beta v_{t-1} + v_t$, where $v_t$ is the expectation error, that is $v_t = r_t^2 - E\left(r_t^2 \mid \mathcal{I}_{t-1}\right)$.

**Table 3: Autocorrelation properties of weekly realized volatilities**
(First week of 1996 to last week of 2000)

| Stock | P-values of LM test for fourth order serial correlation | | | | ARMA(1,1) coefficients | |
|---|---|---|---|---|---|---|
| | $RV$ | $Jump$ | $BV$ | $\ln\left(\sqrt{BV}\right)$ | $\phi$ | $\theta$ |
| Materials | | | | | | |
| AMC | .9721 | .9999 | <.0001 | <.0001 | .9491 | .6875 |
| BHP | <.0001 | .1305 | <.0001 | <.0001 | .9092 | .5955 |
| CSR | .0004 | .5050 | <.0001 | <.0001 | .8977 | .5405 |
| MIM | <.0001 | .0427 | <.0001 | <.0001 | .9770 | .7714 |
| RIO | .0011 | .0552 | .0017 | <.0001 | .9652 | .8042 |
| WMC | <.0001 | .0705 | <.0001 | <.0001 | .8982 | .5632 |
| Energy | | | | | | |
| STO | .0399 | .7001 | .0002 | <.0001 | .8701 | .6266 |
| WPL | <.0001 | .2676 | <.0001 | <.0001 | .9588 | .7006 |
| Banks | | | | | | |
| ANZ | .0119 | .4183 | .1184 | <.0001 | .7149 | .3470 |
| CBA | .0334 | .9848 | .0033 | <.0001 | .6962 | .4038 |
| NAB | <.0001 | .0003 | <.0001 | <.0001 | .8454 | .5113 |
| SGB | .0872 | .8453 | .0011 | <.0001 | .7240 | .4642 |
| WBC | <.0001 | .3286 | <.0001 | <.0001 | .8655 | .6044 |
| Food & Bev. | | | | | | |
| FGL | .0951 | .9541 | <.0001 | <.0001 | .7518 | .4478 |
| SRP | .9977 | .9722 | .9925 | .0037 | .7589 | .5996 |
| Other | | | | | | |
| BIL | .0068 | .9994 | <.0001 | <.0001 | .8822 | .6685 |
| LLC | .9999 | .9999 | .0086 | <.0001 | .9302 | .7336 |
| MAY | .7525 | .9989 | .0107 | <.0001 | .9639 | .7970 |
| NCP | .0140 | .5652 | <.0001 | <.0001 | .9743 | .8063 |
| QAN | .3494 | .5106 | .1332 | <.0001 | .8516 | .6441 |
| WOW | .0048 | .4436 | <.0001 | <.0001 | .8494 | .6477 |

Notes: The LM tests are performed on realized variance ($RV$), the jump component ($Jump$), the realized bi-power variation ($BV$) and the logarithm of the square root of bi-power variation ($\ln\left(\sqrt{BV}\right)$). $\phi$ and $\theta$ are the estimated autoregressive and moving average parameters of the ARMA(1,1) model given by $\ln\left(\sqrt{BV_t}\right) = c + \phi \ln\left(\sqrt{BV_{t-1}}\right) - \theta\epsilon_{t-1} + \epsilon_t$.

**Table 4: Performance of the factor selection procedures**

| | Number of variables ($N$) in the model | | | | | | |
|---|---|---|---|---|---|---|---|
| | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
| | *No outliers* | | | | | | |
| $r_{PC}$ | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\geq 4$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{IV}$ | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\geq$4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | *Additive outliers: maximum 30 standard deviations* | | | | | | |
| $r_{PC}$ | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 32 | 51 | 70 | 74 | 77 | 78 | 81 |
| 2 | 15 | 17 | 18 | 17 | 18 | 17 | 15 |
| 3 | 10 | 10 | 6 | 7 | 4 | 4 | 4 |
| $\geq$4 | 43 | 22 | 6 | 2 | 1 | 1 | 0 |
| $r_{IV}$ | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 85 | 90 | 90 | 90 | 91 | 89 | 92 |
| 2 | 10 | 8 | 8 | 9 | 7 | 10 | 8 |
| 3 | 4 | 1 | 0 | 0 | 2 | 1 | 0 |
| $\geq$4 | 1 | 1 | 2 | 1 | 0 | 0 | 0 |
| | *Additive outliers: maximum 50 standard deviations* | | | | | | |
| $r_{PC}$ | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 3 | 9 | 8 | 16 | 27 | 34 |
| 2 | 0 | 0 | 3 | 4 | 9 | 13 | 20 |
| 3 | 1 | 0 | 2 | 4 | 2 | 6 | 6 |
| $\geq$4 | 97 | 97 | 86 | 84 | 73 | 54 | 40 |
| $r_{IV}$ | | | | | | | |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 39 | 53 | 63 | 75 | 72 | 78 | 79 |
| 2 | 10 | 8 | 12 | 12 | 17 | 17 | 16 |
| 3 | 12 | 8 | 5 | 4 | 5 | 1 | 5 |
| $\geq$4 | 37 | 31 | 20 | 9 | 6 | 4 | 0 |

Notes: $r_{PC}$ and $r_{IV}$ are the number of factors chosen by the principal components and IV criteria. The true number of factors ($r$) is 1 and $T = 250$.

**Table 5: Out of sample forecasting performance of univariate models** (forecasts of $\ln\left(\sqrt{BV}\right)$ based on a fixed estimation sample)

| Stock | ARMA | | SES | | Pooled |
|---|---|---|---|---|---|
| | $p,q$ | $RMSE$ | $\alpha$ | $RMSE$ | $RMSE$ |
| Materials | | | | | |
| AMC | 1,1 | **.2216** | .270 | .2265 | .2225 |
| BHP | 1,1 | .2844 | .320 | .2887 | **.2824** |
| CSR | 1,1 | **.2919** | .330 | .3002 | .2963 |
| MIM | 2,2 | **.1673** | .224 | .1738 | .1683 |
| RIO | 1,1 | .2352 | .168 | .2383 | **.2329** |
| WMC | 1,1 | **.2672** | .322 | .2736 | .2685 |
| Energy | | | | | |
| STO | 1,1 | **.2652** | .112 | .2720 | .2670 |
| WPL | 1,1 | .3886 | .262 | .4024 | **.3864** |
| Banks | | | | | |
| ANZ | 1,1 | **.2747** | .282 | .3057 | .2906 |
| CBA | 1,0 | **.2529** | .176 | .2832 | .2531 |
| NAB | 1,1 | **.2550** | .344 | .2724 | .2633 |
| SGB | 3,0 | **.2679** | .232 | .2740 | .2724 |
| WBC | 1,1 | **.2694** | .300 | .2839 | .2731 |
| Food & Bev. | | | | | |
| FGL | 1,1 | .2468 | .140 | .2467 | **.2450** |
| SRP | 1,0 | **.2765** | .030 | .2930 | .2903 |
| Other | | | | | |
| BIL | 1,1 | **.4611** | .206 | .4668 | .4681 |
| LLC | 1,1 | .2932 | .198 | .3005 | **.2843** |
| MAY | 2,0 | **.3352** | .150 | .3614 | .3454 |
| NCP | 1,1 | .2685 | .174 | .2738 | **.2610** |
| QAN | 1,0 | .4000 | .226 | .4062 | **.3969** |
| WOW | 1,0 | .2686 | .192 | .2659 | **.2610** |

Notes: $p$ and $q$ are the autoregressive and moving average order for the ARMA models and ($\alpha$) is the smoothing parameter for the single exponential smoothing model. In the "pooled" model, all autoregressive parameters are restricted to be equal across equations.

**Table 6: Out of sample forecasting performance of the factor models**
(forecasts of $\ln\left(\sqrt{BV}\right)$ based on a fixed estimation sample)

| Stock | RMSE | | | |
|---|---|---|---|---|
| | EqW | PC | IVLI | CC |
| Materials | | | | |
| AMC | **.1956** | .1976 | .2025 | .2010 |
| BHP | .2724 | **.2716** | .2660 | .2622 |
| CSR | **.2792** | .2800 | .2811 | .2814 |
| MIM | .1647 | .1659 | **.1637** | .1653 |
| RIO | **.2268** | .2291 | .2353 | .2340 |
| WMC | **.2548** | .2605 | .2615 | .2626 |
| Energy | | | | |
| STO | **.2533** | .2717 | .2691 | .2711 |
| WPL | **.3838** | .4172 | .3856 | .3898 |
| Banks | | | | |
| ANZ | **.2645** | .2700 | .2725 | .2731 |
| CBA | **.2498** | .2609 | .2539 | .2532 |
| NAB | **.2480** | **.2480** | .2561 | .2560 |
| SGB | .2650 | **.2580** | .2767 | .2666 |
| WBC | .2622 | **.2560** | .2609 | **.2560** |
| Food & Bev. | | | | |
| FGL | **.2367** | .2496 | .2395 | .2443 |
| SRP | **.2875** | .2916 | .2919 | .2905 |
| Other | | | | |
| BIL | .4477 | .4433 | **.4419** | .4438 |
| LLC | .3292 | **.3091** | .3233 | .3331 |
| MAY | **.3196** | .3261 | .3296 | .3265 |
| NCP | **.2556** | .2607 | .2599 | .2608 |
| QAN | .3866 | .3879 | **.3835** | .3898 |
| WOW | .2526 | **.2544** | .2506 | .2500 |

Notes: EqW models incorporate an equally weighted index of all log-volatilities, PC models incorporate the first two principal components, IVLI models incorporate a leading index and CC models incorporate two leading indices corresponding to the two largest canonical correlations.

**Table 7: Out-of-sample forecasting performance**

(forecasts of $\ln\left(\sqrt{BV}\right)$, based on rolling window estimation samples)

| Stock | RMSE for | | | Out of Sample $\sigma$ of | |
|---|---|---|---|---|---|
| | ARMA | EqW | PC | $\ln\left(\sqrt{BV}\right)$ | $\ln\left(\sqrt{RV}\right)$ |
| Materials | | | | | |
| AMC | .2211 | .1958** | **.1935**** | .2192 | .2589 |
| BHP | .2832 | **.2636**** | .2687** | .2772 | .5241 |
| CSR | .2924 | .2796** | **.2795**** | .2960 | .2805 |
| MIM | .1676 | .1648 | **.1640** | .1938 | .1823 |
| RIO | .2362 | .2306* | **.2289** | .2428 | .3004 |
| WMC | .2694 | **.2564*** | .2597 | .3252 | .4075 |
| Energy | | | | | |
| STO | .2651 | **.2516** | .2603 | .2656 | .2649 |
| WPL | .3985 | **.3870**** | .3873 | .4834 | .5099 |
| Banks | | | | | |
| ANZ | .2718 | **.2600**** | .2611 | .3103 | .3122 |
| CBA | .2467 | **.2451** | .2515 | .3014 | .3083 |
| NAB | .2508 | .2472 | **.2430*** | .3185 | .3751 |
| SGB | .2706 | .2653 | **.2574** | .2799 | .3236 |
| WBC | .2713 | .2643** | **.2546** | .3103 | .3242 |
| Food & Bev. | | | | | |
| FGL | .2459 | **.2394** | .2435 | .2415 | .2608 |
| SRP | **.2790** | .2866 | .2896 | .2869 | .2894 |
| Other | | | | | |
| BIL | .4584 | **.4534** | .4594 | .4619 | .6692 |
| LLC | **.2826** | .3036 | .2894 | .3022 | .3218 |
| MAY | .3335 | **.3192** | .3272 | .3572 | .3843 |
| NCP | .2670 | **.2544** | .2575 | .2773 | .3139 |
| QAN | .3972 | .3903 | **.3889** | .4181 | .4853 |
| WOW | .2693 | .2568 | **.2497**** | .2727 | .2534 |

Notes: One star (*) or two stars (**) indicate that these forecasts are significantly better than forecasts from the ARMA model at the 10% or 5% level of significance respectively.

**Table 8: Correlations of forecasts with actuals**
(all forecasts are based on rolling window estimation samples)

| Stock | ARMA | | EqW | | PC | |
|---|---|---|---|---|---|---|
| | with BV | with RV | with BV | with RV | with BV | with RV |
| Materials | | | | | | |
| AMC | 0.22 | 0.30 | 0.53 | 0.55 | 0.56 | 0.56 |
| BHP | -0.03 | -0.14 | 0.16 | -0.04 | 0.14 | -0.06 |
| CSR | 0.29 | 0.34 | 0.40 | 0.47 | 0.43 | 0.49 |
| MIM | 0.45 | 0.48 | 0.48 | 0.53 | 0.48 | 0.52 |
| RIO | 0.25 | 0.21 | 0.32 | 0.36 | 0.32 | 0.37 |
| WMC | 0.50 | 0.49 | 0.61 | 0.52 | 0.61 | 0.49 |
| Energy | | | | | | |
| STO | 0.28 | 0.25 | 0.73 | 0.71 | 0.62 | 0.58 |
| WPL | 0.26 | 0.02 | 0.35 | 0.07 | 0.43 | 0.12 |
| Banks | | | | | | |
| ANZ | 0.60 | 0.57 | 0.67 | 0.64 | 0.69 | 0.67 |
| CBA | 0.61 | 0.62 | 0.61 | 0.62 | 0.60 | 0.60 |
| NAB | 0.56 | 0.36 | 0.54 | 0.38 | 0.57 | 0.40 |
| SGB | 0.25 | 0.11 | 0.39 | 0.22 | 0.38 | 0.20 |
| WBC | 0.35 | 0.28 | 0.41 | 0.35 | 0.49 | 0.45 |
| Food & Bev. | | | | | | |
| FGL | 0.25 | 0.23 | 0.23 | 0.23 | 0.27 | 0.23 |
| SRP | 0.10 | 0.11 | 0.07 | 0.05 | 0.05 | 0.04 |
| Other | | | | | | |
| BIL | -0.17 | -0.20 | -0.02 | -0.07 | 0.00 | -0.06 |
| LLC | 0.37 | 0.22 | 0.15 | 0.48 | 0.58 | 0.40 |
| MAY | 0.29 | 0.26 | 0.43 | 0.27 | 0.40 | 0.34 |
| NCP | 0.17 | 0.21 | 0.32 | 0.23 | 0.31 | 0.25 |
| QAN | 0.14 | 0.07 | 0.18 | 0.13 | 0.21 | 0.16 |
| WOW | 0.23 | 0.18 | 0.36 | 0.30 | 0.41 | 0.34 |

Note: Correlations greater than .23 are statistically greater than zero at the 5% level of significance.

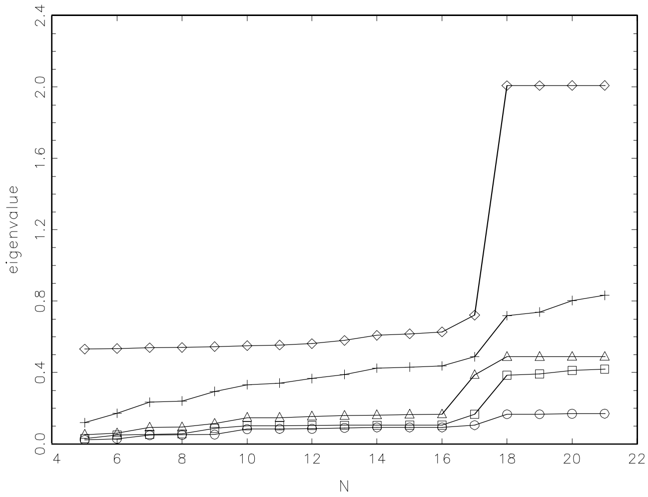


**Figure 1: The largest five eigenvalues of the variance matrix as $N$ increases**

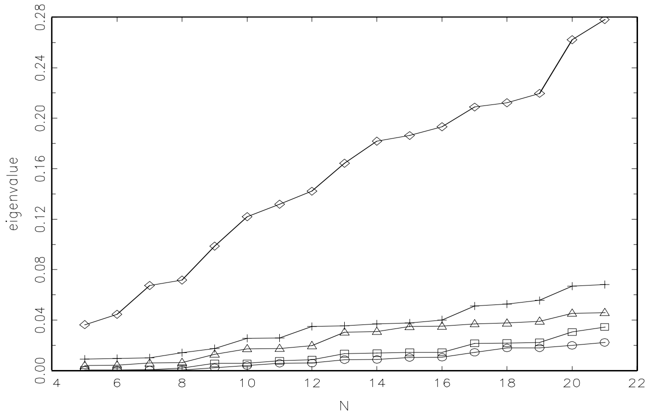


**Figure 2: The largest five eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ for $\sqrt{RV}$**

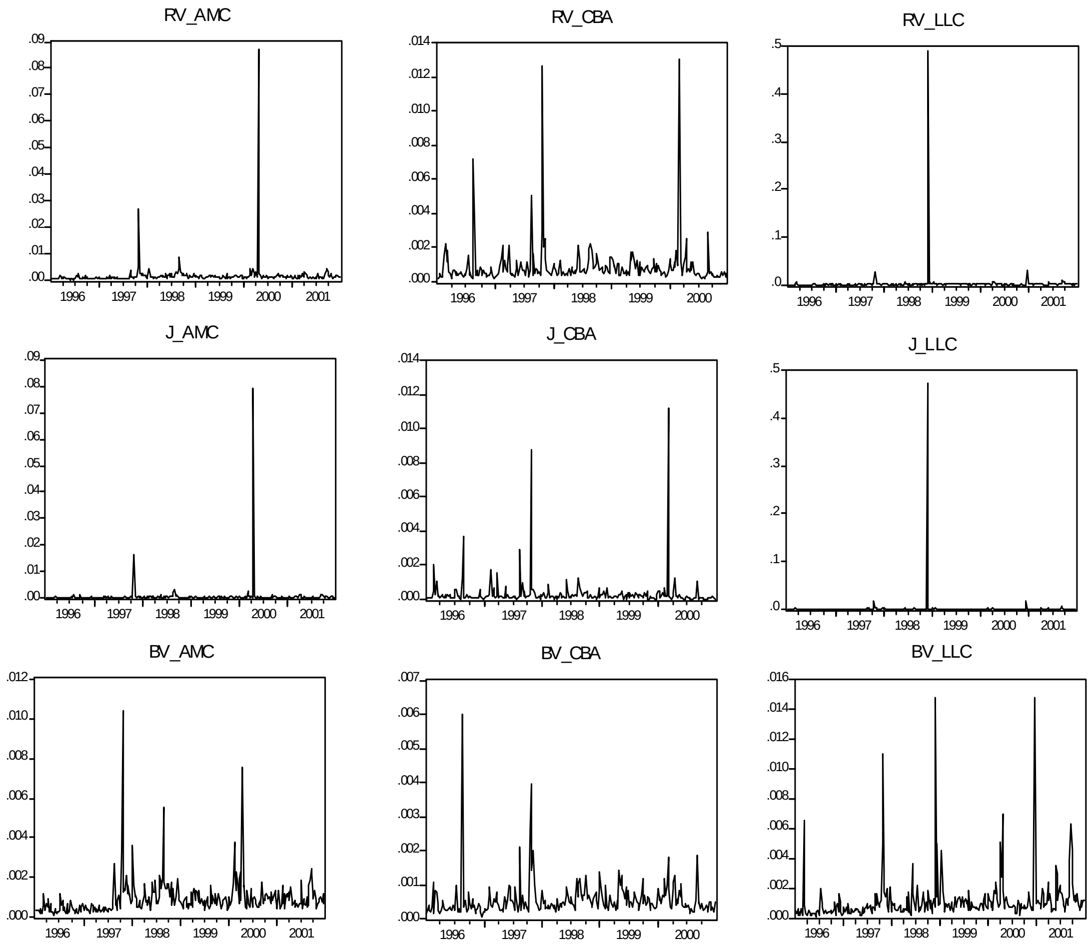


**Figure 3: Realized variance (RV_), jumps (J_) and bi-power variation (BV_) for a materials (AMC), a banking (CBA) and a real estate (LLC) company**

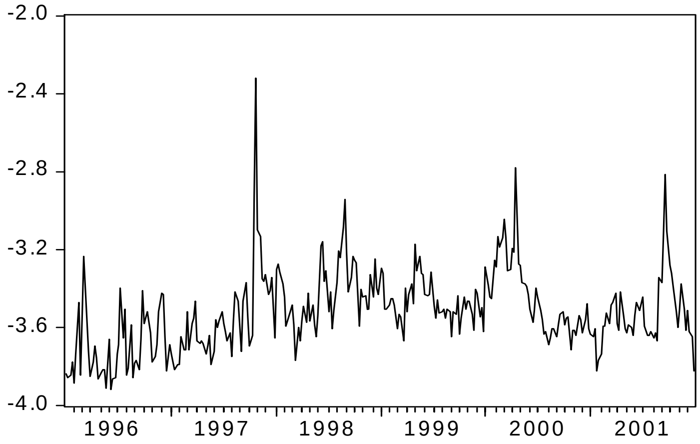


**Figure 4: The sample average of all 21 log volatilities**

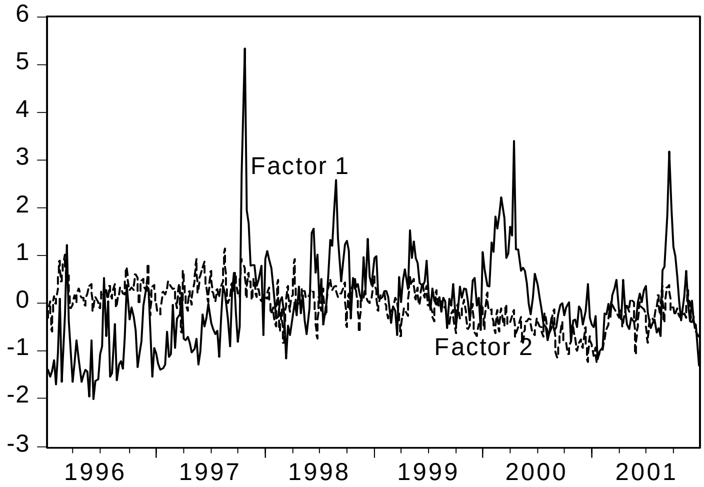


**Figure 5: The first two principal components of log-volatilities**

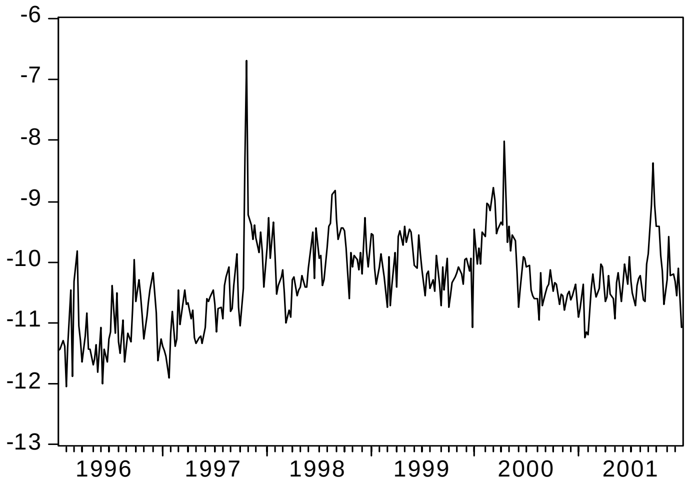


**Figure 6: The market leading indicator**