THE AUSTRALIAN NATIONAL UNIVERSITY

WORKING PAPERS IN ECONOMICS AND ECONOMETRICS

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

Working Paper No. 451

March, 2005

ISBN: 086831 451 X

Abstract: This paper develops univariate and multivariate forecasting models for realized volatility in Australian stocks. We consider multivariate models with common features or common factors, and we suggest estimation procedures for approximate factor models that are robust to jumps when the cross-sectional dimension is not very large. Our forecast analysis shows that multivariate models outperform univariate models, but that there is little difference between simple and sophisticated factor models.

Keywords: Realized volatility, factor models, jumps, model selection, forecasting.

JEL Classification: C32, C53.
* Corresponding author. E-mail: Farshid.Vahid@anu.edu.au.
Acknowledgments: We thank Nelufa Begum and Jing Tian for research assistance. We have benefited from comments of the participants of the "Common Features in London" Conference.

1 Introduction

Multivariate modelling of conditional heteroskedasticity has been an important research problem ever since ARCH models were introduced by Engle (1982). Full 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 (see Engle and Kroner 1995). Fully general models model 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 prices of different assets to a small number of underlying factors. Engle, Ng and Rothschild (1990) propose a factor ARCH model for the term structure of interest rates. King, Sentana and Wadhwani (1994) consider a multifactor model for aggregate stock returns for 16 countries. These models specify the complete conditional distribution of all variables, so that they deliver internally consistent forecasts of means, variances and covariances. On the other hand, authors such as Harvey, Ruiz and Shephard (1994) develop explicit multivariate models of just the logarithms of squared returns, and then they consider the possibility of common random factors in the variances of their return series. Engle and Marcucci (2004) adopt a similar approach. Motivated by applications in which only the forecasts of variances are of interest, these authors specify a long-run pure variance model of thirty Dow Jones stocks in which just the conditional variances of returns are modelled jointly, leaving other aspects of the joint distribution (and in particular the covariances) unspecified. This is achieved by assuming that squared returns, or certain transformations of squared returns have a common long-run feature. In their model, the volatility of each asset depends on a small number of common factors and an idiosyncratic factor, and their analysis compares and contrasts the performance of canonical correlation and principal component based estimators of these common factors.

Recently with the increased availability of high frequency data and improved data storage and computing capabilities, and also with advances in theory and empirics of continuous time finance, there has been renewed interest in using high frequency returns between period $t-1$ and t to obtain a consistent estimator of volatility for time t. This measure of volatility, known as "realized volatility", was first used by French, Schwert and Stambaugh (1987) and has been strongly promoted in recent work by Andersen, Bollerslev, Diebold and Labys (2003). Here, we use realized volatility to develop a pure variance model for the returns of twenty one highly traded Australian stocks. Our goal is to investigate whether a parsimonious multivariate model can do better than simple univariate models with respect to forecasting the realized volatility of Australian stocks.

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 (under the assumption that N is fixed and T goes to infinity). This is not very difficult when one is modelling the conditional mean of a multivariate data set, (see, e.g., Anderson and Vahid, 1998), but it becomes quite complicated in the case of conditional variances (see Doz and Renault, 2004). One advantage of pure variance models is that they are more convenient for developing common features tests for determining the number of common factors. Non-normality and heteroskedasticity will usually imply that the usual canonical correlation based test statistics will not be useful in these circumstances (see Engle and Marcucci, 2004), so that one has to turn to more robust tests for common features (see Candelon, Hecq and Verschoor, 2004). However, even under ideal conditions (i.i.d. normal errors), the performance of such tests in finite samples will deteriorate as the number of variables N becomes large.

When common and idiosyncratic factors have similar statistical properties, one can determine the number of factors by comparing the fit of estimated models that each use a different number of common factors. This is somewhat cumbersome, because it requires the complete specification and estimation of each model. Alternatively, one can make constructive use of the cross sectional dimension, as in the approximate factor models of Chamberlain and Rothschild (1983). Connor and Korajczyk (1993) provide a test for the number of factors and Bai and Ng (2002) suggest several model selection criteria that produce consistent estimates of the number of common factors when N and T go to infinity.

In this paper we argue that the presence of jumps in time series of realized volatilities can distort inference relating to common factors, and then we outline modifications to model selection criteria that are likely to be more robust to jumps. We also argue that since jumps are unpredictable, there is little to be gained by including them in forecasting models. We therefore remove jumps from our data by using the procedures discussed in Barndorff-Neilsen and Shephard (2004), and then we build factor models for the forecastable component of volatility.

The structure of the rest of this paper is as follows. Section 2 provides a 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. This section also suggests using realized "bipower variation" (i.e. realized volatility minus the jumps) instead of realized volatility for developing forecasting models, and explores the properties of bi-power variation of the returns of Australian stocks. Section 5 develops univariate and multivariate models for forecasting the log-volatilities in our data set and compares their out of sample performance. Section 6 concludes.

2 Data

We base our analysis on price data for stocks traded on the Australian Stock Exchange (ASX).[1] 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.

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

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

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.

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

[1] The data is provided by the Securities Industry Research Centre of Asia and the Pacific (SIRCA).

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 volatilities of stock prices of Australian companies are responsible for giving the impression that conditional variances are constant or very dissimilar across different stocks. The jumps are large and are therefore very influential when one is estimating parameters, but they are also quite unpredictable and hence generate the impression that volatilities are unpredictable.

3 Factor models of realized volatility

Raw intuition and more formal theories in finance suggest that underlying market factors drive the movement of all asset returns. This prompts the use of information on all asset returns to extract a few common reference factors, and it is natural to think of principal component analysis as a technique that might be used for this purpose. This motivates the approximate factor literature in finance, originating in the work of Chamberlain and Rothschild (1983). This model is given by

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

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. 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 $(r+1)^{\text{th}}$ 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

$$\begin{aligned}
PC_1(r) &= \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \\
PC_2(r) &= \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln\left(\min\{N,T\}\right), \\
IC_1(r) &= \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \quad \text{and} \\
IC_2(r) &= \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln\left(\min\{N,T\}\right),
\end{aligned}$$

where $ESS(r) = \sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it} - \hat{a}_i'\hat{F}_t)^2$, $\hat{F}_t$ are the r estimated common factors 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[2] for various values of $r^{\max}$. 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. The observation that these model selection criteria select a large number of common factors relative to N when N is small has been noted in the simulation study of Bai and Ng (2002) and in the empirical study of Engle and Marcucci (2004). Here, we would like to 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 principle components.[3] 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 variance process for LLC is clearly dominated by jumps, and we argue that by purging these jumps and other jumps, and by also considering alternative estimators that are more robust to jumps, one can get better estimators for common factors.

[2] We take the square root of our realized variance series to conform with the finance literature, where "volatility" usually refers to the standard deviation. Results based on our realized variance series are qualitatively the same.

[3] This visual method is suggested by Forni et al (2000), who work with a more general factor model.

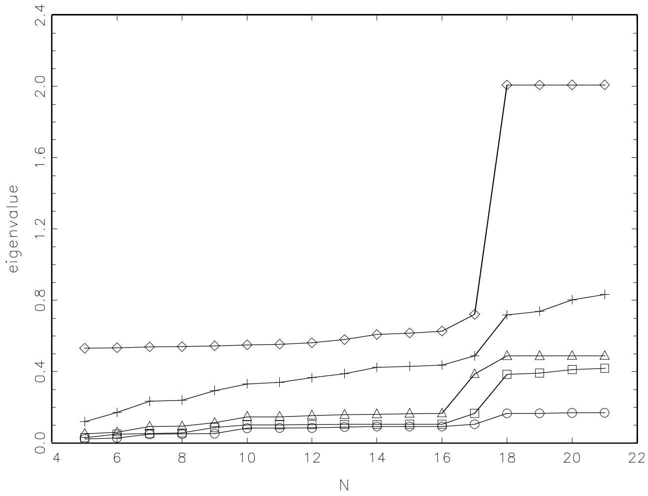

Figure 1: The largest five eigenvalues of the variance matrix as N increases

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 et al, 2003, and the reference 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

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

as $\Delta \to 0$. That is, the realized variance includes all jumps that have occurred between t and $t+1$. Since jumps are unpredictable, realized variance may seem to be unpredictable even if $\sigma(t)$ is predictable. Further, when analyzing the realized variances of multiple assets, large jumps can potentially mask the fact that the $\sigma(t)$ of different assets depend on a small number of common factors.

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. Our second approach uses a consistent estimate of $\int_t^{t+1} \sigma^2(s)ds$, which is the predictable component of realized variance, and then develops forecasting models for these components.

4.1 Instrumental variable estimators of common factors

Consider the model of the N mean subtracted "volatilities":

$$\underset{(N\times 1)}{Y_t} = \underset{(N\times r_1)}{A_1} \underset{(r_1\times 1)}{C_t} + \underset{(N\times r_2)}{A^J} \underset{(r_2\times 1)}{J_t} + \underset{(N\times 1)}{j_t} + \underset{(N\times 1)}{u_t} \tag{3}$$

where C_t are the "continuous" common factors and J_t are the common jump factors. The idiosyncratic components are $j_t + u_t$, where j_t are idiosyncratic jumps and u_t are idiosyncratic dynamic factors. If factor loadings A_1 and A^J are not the same, then it is likely that the factors corresponding to largest principal components will be those identified by the common jumps. While these factors explain a large proportion of contemporaneous variation of volatilities, they will not be useful for forecasting since the jumps are unpredictable. Moreover, in practice, idiosyncratic jumps can also be quite large, and in finite samples the principal component procedure may identify those variables with the largest jumps as the common factors.

Consider the linear projection of Y_t on Y_{t-1} (note that we are using Y_{t-1}, and not the entire history of information), and denote all linear projections on this space by the subscript '$\mid t-1$'. We have

$$Y_{t|t-1} = A_1 C_{t|t-1} + u_{t|t-1}, \tag{4}$$

since jumps are not predictable from the past, and this implies that $J_{t|t-1} = j_{t|t-1} = 0$. If the idiosyncratic factors u_t are serially correlated, then the number of common factors cannot be identified by a serial correlation common feature (SSCF, Engle and Kozicki, 1993) test, because $u_{t|t-1}$ will be a non-trivial function of Y_{t-1}. With fixed N, one solution within the common feature framework is to assume that common factors are autoregressive, while idiosyncratic factors are m-dependent, i.e. they have finite memory, as in Gourieroux and Peaucelle (1988) or Vahid and Engle (1997). In such a case, the projection on Y_{t-m-1} will only include the common factors, i.e.,

$$Y_{t|t-m-1} = A_1 C_{t|t-m-1},$$

and we can use a GMM test for codependence to determine the rank of A_1 (Vahid and Engle, 1997). However, one may not want to make such an assumption in financial applications.

As in Bai and Ng (2002), we can make constructive use of the large cross sectional dimension to develop model selection criteria for the determination of the common factor rank, without imposing m-dependence on the idiosyncratic factors. We need to choose a penalty function such that the addition to the fit generated by modelling the idiosyncratic components (i.e., increasing the rank of the parameter matrix that links Y_{t-1} to Y_t beyond r_1) becomes negligible as

$N \to \infty$. In a way, such a set-up can be viewed as instrumental variable estimation, with Y_{t-1} taken as instruments for C_t. Ordinary principle components can be viewed as least squares estimates of Y_t on linear indices made from Y_t. These indices are proxies for C_t. The combination of large jumps and small N, makes these proxies unreliable. Treating the jumps as being similar to measurement errors, one can use Y_{t-1} as instruments to obtain better proxies for C_t. While each of the idiosyncratic components is correlated with one of the instruments, the average (over N) correlation of each idiosyncratic component with all instruments goes to zero, while the average correlation between the instruments and the r common factors does not go to zero.

Proposition 1 *A consistent estimator of the forecastable common factors under the assumption that $N, T \longrightarrow \infty$ with $N < T$ is $\hat{A}_1' Y_t$, where $\hat{A}_1$ consists of the eigenvectors corresponding to the r_1 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 any $p > 0$. Subject to the usual normalization that $A_1' A_1 = I_{r_1}$, this estimator of $\hat{A}_1$ is also the ordinary least squares reduced rank regression estimator of A_1 in

$$\mathbf{Y} = A_1 B_1 \mathbf{Y}_{-1} + \mathbf{U} \tag{5}$$

that minimizes tr $(\mathbf{UU}')$. This is also the ordinary least squares reduced rank estimator of A_1 in

$$\hat{\mathbf{Y}} = A_1 B_1 \mathbf{Y}_{-1} + \mathbf{U}^*. \tag{6}$$

Proof: The approximate factor structure is preserved after linear projection on lagged Y_t, but with only r_1 forecastable factors, as can be seen from equation (4). This does not depend on the number of lags in the projection matrix, and the factor loadings in A_1 also stay the same if the lag structure in the projection matrix is changed. Therefore, under appropriate regularity conditions, the principal component of the variance covariance matrix of $Y_{t|t-1,...,t-p}$ produces consistent estimates of $C_{t|t-1,...,t-p}$ and A_1, as $N \longrightarrow \infty$. Since $\hat{Y}_t$ is a consistent estimator of $Y_{t|t-1,...,t-p}$, the eigenvectors of the first r_1 eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ are consistent estimators of A_1. Note that $\hat{\mathbf{Y}} = \mathbf{YP}$ where $\mathbf{P} = \mathbf{Y}_{-1}' \left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1} \mathbf{Y}_{-1}$ is the orthogonal projection matrix on the space of lagged Y. Since $\mathbf{P}$ is symmetric and idempotent, it follows that $\hat{\mathbf{Y}}\hat{\mathbf{Y}}' = \mathbf{YPY}' = \mathbf{YY}_{-1}' \left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1} \mathbf{Y}_{-1}\mathbf{Y}'$. The fact that the eigenvectors corresponding to the largest r_1 eigenvalues of this matrix are the estimates of A_1 in the reduced rank regression of $\mathbf{Y}$ on $\mathbf{Y}_{-1}$, is another well known result in multiple regression theory (see Lütkepohl, 1991, Proposition A.5). Finally, the last part follows because $\mathbf{YY}_{-1}' \left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1} \mathbf{Y}_{-1}\mathbf{Y}' = \hat{\mathbf{Y}}\mathbf{Y}_{-1}' \left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1} \mathbf{Y}_{-1}\hat{\mathbf{Y}}'$.

Corollary 2 *We can determine the number of forecastable factors* r_1*, by designing model selection criteria that are analogous to those in Bai and Ng (2002). In these new criteria,* $ESS\,(r_1)$ *will be the error sum of squares of the estimated reduced rank regression (5), which will be equal to tr* $\left(\mathbf{YY}'\right)$ *minus sum of the* r_1 *largest eigenvalues of* $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$*. Also* $ESS\,(r_1^{\max})$ *will be the error sum of squares of the full rank regression, which is equal to tr* $\left(\mathbf{YY}'\right) - tr(\hat{\mathbf{Y}}\hat{\mathbf{Y}}')$*.*

Figures 2 plots the largest 5 eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ for realized volatilities $\sqrt{RV}$. Comparing Figure 2 with Figure 1, it is clear that the projection on one lag has eliminated the effect of idiosyncratic jumps.

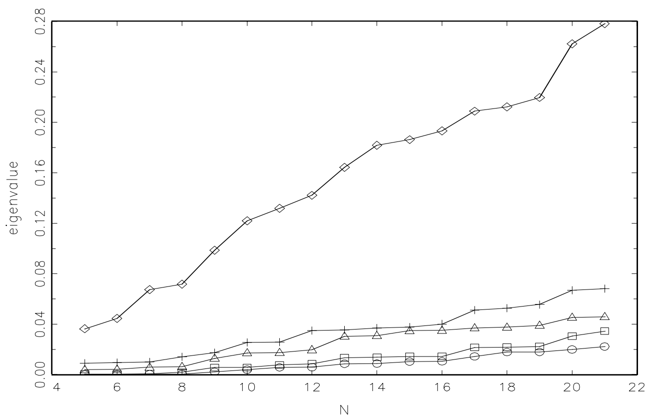

Figure 2: The largest five eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ for $\sqrt{RV}$

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 (5) 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.

There is possible serial correlation in the residuals of equation (5) (where u_t is the t-th column of $\mathbf{U}$) and the variances of u_{it} and u_{jt} can be different. Also, since these remainders can contain common jumps, they can be contemporaneously correlated. However, the factor structure implies that Eu_tu_{t-s}' is diagonal for $s > 0$, or that it will be block diagonal if we allow the idiosyncratic components of assets in the same industry group to be correlated. Hence, an appropriate GLS correction would take account of all of these considerations and would involve the complete specification of the entire structure. In that case, one may as well estimate the entire system jointly. A partial GLS correction that ignores the serial correlation in the errors and only corrects for cross sectional heteroskedasticity and contemporaneous correlation among the errors, leads to the canonical covariate estimators of the common components. If the idiosyncratic components are not serially correlated and the errors are normally distributed, then this estimator would be the maximum likelihood estimator of the common factor and its best prediction given the information in $\mathbf{Y}_{-1}$. These results are stated in the next proposition.

Proposition 3 *If the idiosyncratic components are not serially correlated, then the variables in Y_t have r_1 common serial correlation features. Then, (i) the GLS estimator of the reduced rank regression is MLE under the assumption that the errors are normal; and (ii) this estimator minimizes $\left|\mathbf{UU}'\right|$. Under the normalization that $A_1'A_1 = I_{r_1}$, the columns of A_1 consist of the eigenvectors corresponding to the r_1 largest eigenvalues of $\left(\mathbf{YY}'\right)^{-\frac{1}{2}}\mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'\left(\mathbf{YY}'\right)^{-\frac{1}{2}}$, which correspond with the eigenvalues of $\left(\mathbf{YY}'\right)^{-1}\mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$. These eigenvalues are the squared sample canonical correlations between Y_t and $Y_{t-1}, Y_{t-2}, ..., Y_{t-p}$.*

Proposition 4 *If the factor structure is correct but the idiosyncratic components are serially correlated, then the canonical variates corresponding to the r_1 largest eigenvalues of $\left(\mathbf{YY}'\right)^{-1}\mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$ need not provide a consistent estimator of A_1.*

All parts of Proposition 3 are well-known results in multivariate statistics (see, e.g, Anderson, 1984, and Vahid and Engle, 1993), but Proposition 4 needs some explanation. The eigenvalues of $\left(\mathbf{YY}'\right)^{-\frac{1}{2}}\mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'\left(\mathbf{YY}'\right)^{-\frac{1}{2}}$ are, by construction between zero and one, and they show the highest, the second highest, ... squared correlation between every possible linear combination of Y_t and the past. Under the assumed factor structure, the r_1 largest eigenvalues of $\frac{1}{T}\mathbf{YY}_{-1}'\left(\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\right)^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$ increase to infinity at rate N, and at least the r_1 largest eigenvalues of $\frac{1}{T}\mathbf{YY}'$ increase to infinity at the same rate. This implies that the r_1 smallest eigenvalues of $\left(\frac{1}{T}\mathbf{YY}'\right)^{-1}$ go to zero at rate N^{-1}. However, this does not tell us if the largest eigenvalue of the product of these matrices still reflect any information about the common factor. Canonical correlations procedures will always reveal if the dimension of the dynamic system is smaller than N, but a common factor structure with predictable idiosyncratic components need not have reduced rank dynamics. This does not mean that the canonical correlation procedure "does not work". It only tells us that the canonical correlation procedure cannot be used to identify the common and idiosyncratic components, when both types of components have the same feature.

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 et al (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 Marcelino (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 $(\mathbf{YY}')^{-1}\mathbf{YY}'_{-1}(\mathbf{Y}_{-1}\mathbf{Y}'_{-1})^{-1}\mathbf{Y}_{-1}\mathbf{Y}'$. Kapetanios and Marcelino (2004) replace $(\mathbf{YY}')^{-1}$ 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 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 Δ-period returns converges to the integral of quadratic variation excluding the contribution of jumps. That is, as $\Delta \to 0$,

$$BV_{t+1}(\Delta) \equiv \frac{\pi}{2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta,\Delta}| \left|r_{t+(j-1)\Delta,\Delta}\right| \xrightarrow{p} \int_{t}^{t+1} \sigma^2(s)ds. \tag{7}$$

The BV term in equation (7) 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 Δ-period returns. Equation (7) 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

$$J_{t+1}(\Delta) = \max\left\{RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0\right\}. \tag{8}$$

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 (2003) show that realized jumps in exchange rate data that are computed using equation (8) have no forecasting power for realized volatility, whereas bi-power variation in (7) 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.[4] A user familiar with Australian company histories would know that some of these large price changes were 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. For example, Lend Lease Corporation issued a well publicized one for one bonus share issue on 12 December 1998, which brought their share price down from around 38 to around 19 Australian dollars. However, it is not always obvious if a price movement that follows some capital restructuring is purely a jump. Comparison of the realized variance with the realized bi-power variation provides a simple way to isolate pure jumps.

[4] Barndorff-Neilson and Shephard (2004) develop tests of the significance of jumps, but we do not apply them here, given that our primary purpose is simply to use bipower variation as a basis for forecasting realized volatility.

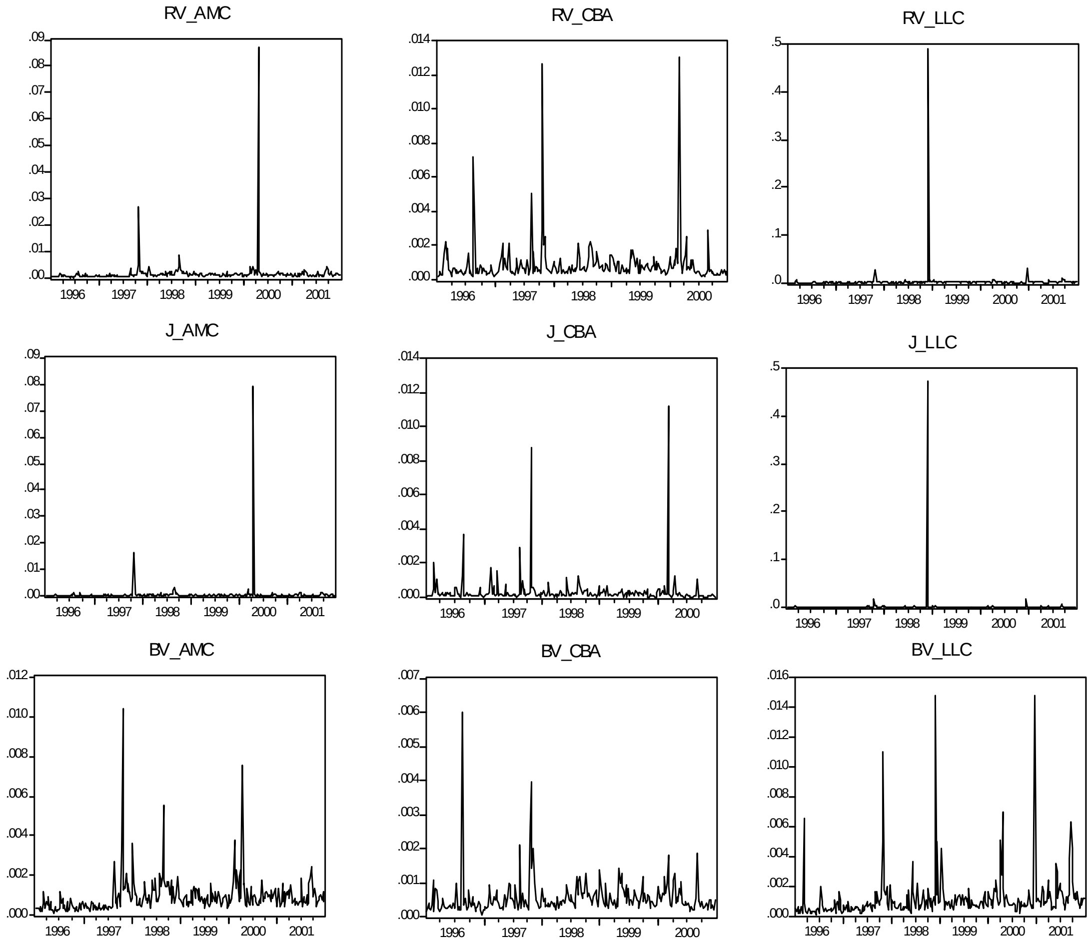

Figure 3: Realized variance (RV_), jumps (J_) and bi-power variation (BV_) for a mining (AMC), a banking (CBA) and a real estate (LLC) company

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[5]. The first point to note is that hardly any of the estimated jump com-

[5] Since $\ln\left(\sqrt{BV}\right) = 0.5\ln(BV)$, taking the square root is of no real consequence. However, since the standard deviation rather than variance is the preferred measure of "volatility" in finance, we work with $\ln\left(\sqrt{BV}\right)$.

ponents 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 et al, 1996). We report our estimates of the degree of integration estimated for each of the $\ln\left(\sqrt{BV}\right)$ series in the last column of Table 3[6]. While most are statistically significant at the 5% level, none are numerically large, and hence we prefer to model this persistence using parsimonious ARMA models, rather than taking explicit account of fractional integration (and possible fractional cointegration).

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 forecastable component of realized variance, we investigate the multivariate modelling of the $\ln\left(\sqrt{BV}\right)$ series, and use the term "log volatility" to denote $\ln\left(\sqrt{BV}\right)$. We now focus on finding good forecasting models for the log-volatility of Australian stock returns, and to determine whether the incorporation of common factors will improve the forecasting performance of the model. All models are developed using data from the first week of 1996 to the last week of 2000, and they are used to provide

[6] See Geweke and Porter-Hudak (1983) for computational details.

one-step ahead forecasts of $\ln\left(\sqrt{BV}\right)$ for the 52 weeks of 2001. Absolutely no information from the forecast period is used in the development of any of the models. We report the square root of the mean squared forecast error (RMSE) as our measure of forecast accuracy, and we also report some results of forecast encompassing regressions for the out of sample period.

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). As noted before, the fitted ARMA models have the common characteristic that their AR parameter is large, and the roots of the AR and MA polynomials are close. This is a parsimonious way of modelling a variable with 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, page 175). 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 4 reports the root mean squared forecast error of these models for each of the log-volatility series for the out of sample period. We can see that with only two exceptions (FGL and WOW), ARMA models almost always 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.

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 bi-power variation, 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.

We report the forecasting performance of several multivariate models. The first one is a one factor model that takes the simple average of the 21 $\ln\sqrt{BV}$ series as the estimate of common factor. This is plotted in Figure 4. If there is only one common factor, this provides a consistent estimate of the common factor. We add 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.

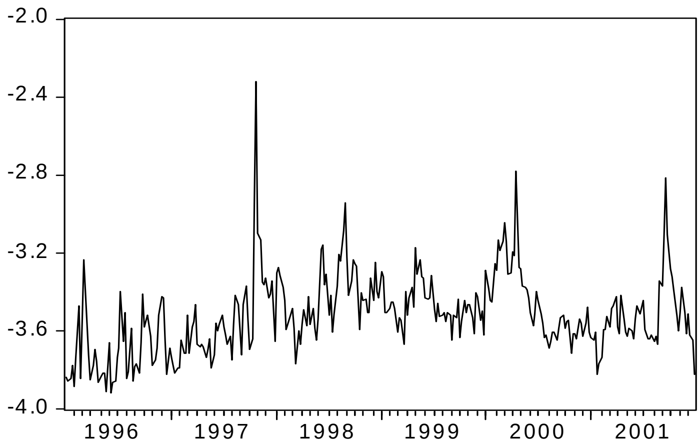

Figure 4: The sample average of all 21 log-volatilities

The second model is one that has two factors estimated by the principal component procedure. We chose two factors because there was a conflict among model selection criteria on whether there was one or two factors. We made separate time series models for these factors and 21 time series models for the remainders (i.e. the idiosyncratic components). We then forecasted each component, and obtained forecasts of each log volatility by adding up the forecasted idiosyncratic component and the estimated factor loadings applied to the forecasted factors. 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 trend. 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 financial time series (see Andersen et al, 2003 and the references therein). The models for the idiosyncratic components are all low order ARMA models.

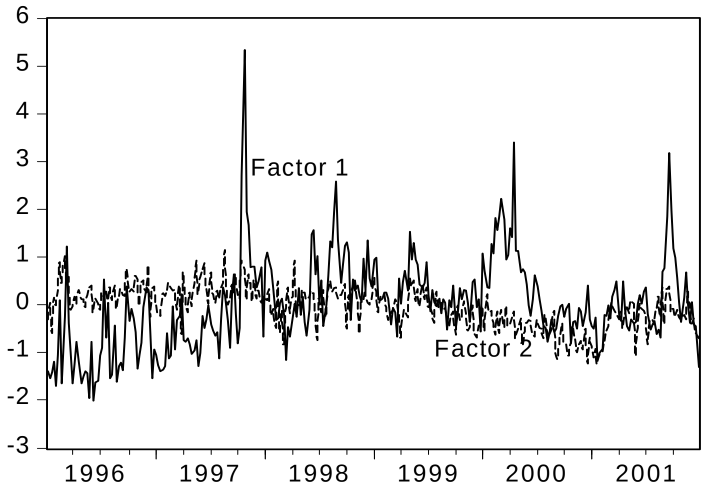

Figure 5: The first two principal components of log-volatilities

The third model uses the principal component analysis of the linear projection of Y_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.

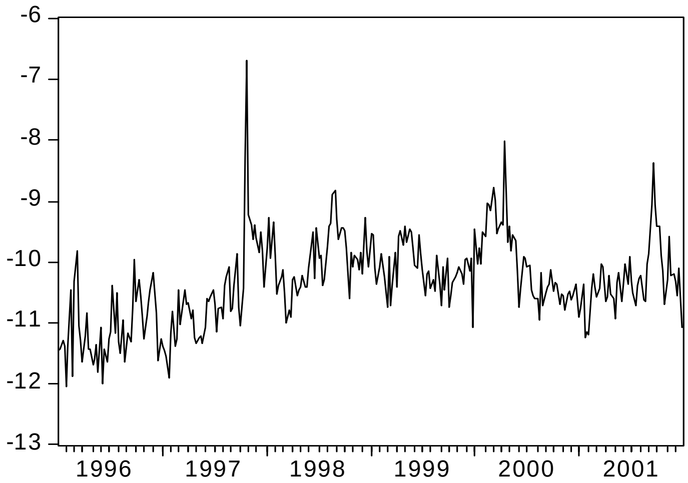

Figure 6: The market leading indicator

The final model is the model assumes that variables can be adequately modelled by a VAR, and uses model selection criteria to choose number of lags and rank of the 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 when this is needed. These models are denoted by CC (canonical correlation) models.

Table 5 reports the out of sample performance of the multivariate models. 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 Australian stocks. It also shows that our attempts to get better estimates of this factor by using statistical procedures do not really pay off.

As a final comparison of different forecasts, we have run forecast encompassing regressions (regressions of the actual log-volatility on different pairs of forecasts for the out-of-sample period). If the parameters for each forecast are insignificant, then the forecasts are equally good, meaning that given one, 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 conclusions drawn by comparing RMSEs. First, and perhaps not surprisingly, these regressions tell us that the multivariate models encompass the univariate models. The strongest evidence is for the model with equally weighted (EqW) factor estimates. 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 encompassed by it. 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.

6 Conclusion

In this paper we argue that the principle 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 methods may also deliver factors that are non-forecastable. These concerns are particularly relevant when forecasting the volatilities of asset returns, because the process 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-Neilsen 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-volatilties. 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 averages over averages 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. Also, it is often found that simple averages of forecasts that do not encompass each other often provide better forecasts than do combinations that are based on estimated weights.

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, especially since the RMSE figures in Table 5 are often the same until three digits after the decimal. 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.

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, 27-96.

Andersen, T.G., T. Bollerslev, F.X. Diebold (2003), "Some like it smooth, some like it rough: untangling continuous and jump components in measuring, modeling and forecasting asset return volatility", Working paper, Duke University.

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

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

Anderson, T.W. (1984), An Introduction to Multivariate Statistical Analysis. New York: Wiley.

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

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

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

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

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.

Candelon, B., A. Hecq and W.F.C. Verschoor (2004), "Measuring common cyclical features during financial turmoil: Evidence of interdependence not contagion", Journal of International Money and Finance, forthcoming.

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

Connor G. and R.A. Korajczyk (1993), “A test for the number of factors in an approximate factor model”, Journal of Finance, 48, 1263-1291.

Doz C. and E. Renault (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 S. Kozicki (1993), “Testing for common features”, Journal of Business and Economic Statistics, 11, 369-380.

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

Engle, R.F., and J. Marcucci (2004), “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., V.K. Ng and M. Rothschild (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., 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., Reichlin, L., (2003), “The generalized factor model: one-sided estimation and forecasting”. LEM working paper series 2003/13.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), “Expected stock returns and volatility”, Journal of Financial Economics, 19, 3-29.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models", Journal of Time Series Analysis, 4, 221 - 238.

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

Hannan, E.J. and J. Rissanen (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., E. Ruiz and N. Shephard (1994), "Multivariate stochastic variance models", Review of Economic Studies, 61, 247-264.

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

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

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

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

Vahid, F. (1999), "Partial Pooling: A Possible Answer to 'To Pool Or Not To Pool' " , in R. F. Engle and H. White (Eds) Cointegration, Causality and Forecasting: Festschrift in honor of Clive W. J. Granger, Oxford University Press.

Vahid, F. and R.F. Engle (1993), "Common trends and common cycles", Journal of Applied Econometrics, 8, 341-360.

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

Vahid, F. and J.V. Issler (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
(During the period from 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

Notes:

* "GICS" stands for Global Industry Classification Standard.

** WMC has been recently changed to WMR.

*** MIM and NCP are no longer traded on the Australian Stock Exchange.

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

Stock	Mean	St.Dev.	Skewness	Kurtosis	ARCH	$\alpha+\beta$	β
Mining							
AMC	0.0023	0.0348	-0.5674	6.0796	0.0150	0.8581	0.7115
BHP	0.0000	0.0351	0.6086	4.8252	0.0011	0.9560	0.8741
CSR	0.0002	0.0358	0.1862	3.1434	0.7322		
MIM	-0.0018	0.0543	0.0395	4.1165	0.0204	0.9102	0.8460
RIO	0.0015	0.0386	0.3014	3.3472	0.0107	0.9858	0.9579
WMC	-0.0005	0.0439	-0.4216	5.9406	0.1993		
Energy							
STO	0.0017	0.0350	0.2989	3.5078	0.9981		
WPL	0.0028	0.0364	0.2438	3.5483	0.9467		
Banks							
ANZ	0.0031	0.0332	-0.2895	3.9104	0.6044		
CBA	0.0040	0.0280	-0.4134	3.6174	0.6988		
NAB	0.0032	0.0306	-0.3692	4.0922	0.9239		
SGB	0.0023	0.0029	-0.1383	3.9486	0.0011	0.7400	0.6434
WBC	0.0031	0.0316	-0.4216	3.9681	0.0059	0.2293	0.1155
Food & Bev							
FGL	0.0028	0.0294	0.3047	3.5074	0.2533		
SRP	0.0018	0.0393	-0.4515	6.2919	0.0637		
Other							
BIL	0.0039	0.0337	0.2522	4.3419	0.8063		
LLC	-0.0007	0.0567	-6.7069	76.010	0.9998		
MAY	-0.0001	0.0433	-0.1888	6.1116	0.6088		
NCP	0.0025	0.0542	0.3309	5.6748	0.0001	0.9921	0.9437
QAN	0.0017	0.0407	-0.0648	4.0001	0.4054		
WOW	0.0036	0.0299	0.2084	3.3449	0.6091		

Notes:

* Entries in the 'ARCH' column are p-values of the LM test for the null hypothesis of no conditional heteroskedasticity against an ARCH(4) alternative.

** $\alpha+\beta$ is the sum of estimated ARCH and GARCH parameters in a GARCH(1,1) specification. β is the GARCH parameter. These estimates are only 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., $r_t^2 = \omega + (\alpha+\beta)\, r_{t-1}^2 - \beta \upsilon_{t-1} + \upsilon_t$, where υ_t is the expectation error, that is $\upsilon_t = r_t^2 - E\left(r_t^2 \mid \mathcal{I}_{t-1}\right)$.

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

Stock	P-values of serial correlation LM test RV	$Jump$	BV	$\ln\left(\sqrt{BV}\right)$	ϕ	θ	d
Mining							
AMC	0.9721	0.9999	<0.0001	<0.0001	0.9491	0.6875	0.171
BHP	<0.0001	0.1305	<0.0001	<0.0001	0.9092	0.5955	0.161
CSR	0.0004	0.5050	<0.0001	<0.0001	0.8977	0.5405	0.142
MIM	<0.0001	0.0427	<0.0001	<0.0001	0.9770	0.7714	0.149
RIO	0.0011	0.0552	0.0017	<0.0001	0.9652	0.8042	0.118
WMC	<0.0001	0.0705	<0.0001	<0.0001	0.8982	0.5632	0.128
Energy							
STO	0.0399	0.7001	0.0002	<0.0001	0.8701	0.6266	0.127
WPL	<0.0001	0.2676	<0.0001	<0.0001	0.9588	0.7006	0.138
Banks							
ANZ	0.0119	0.4183	0.1184	<0.0001	0.7149	0.3470	0.118
CBA	0.0334	0.9848	0.0033	<0.0001	0.6962	0.4038	0.154
NAB	<0.0001	0.0003	<0.0001	<0.0001	0.8454	0.5113	0.177
SGB	0.0872	0.8453	0.0011	<0.0001	0.7240	0.4642	0.157
WBC	<0.0001	0.3286	<0.0001	<0.0001	0.8655	0.6044	0.146
Food & Bev.							
FGL	0.0951	0.9541	<0.0001	<0.0001	0.7518	0.4478	0.123
SRP	0.9977	0.9722	0.9925	0.0037	0.7589	0.5996	0.122
Other							
BIL	0.0068	0.9994	<0.0001	<0.0001	0.8822	0.6685	0.145
LLC	0.9999	0.9999	0.0086	<0.0001	0.9302	0.7336	0.105
MAY	0.7525	0.9989	0.0107	<0.0001	0.9639	0.7970	0.145
NCP	0.0140	0.5652	<0.0001	<0.0001	0.9743	0.8063	0.150
QAN	0.3494	0.5106	0.1332	<0.0001	0.8516	0.6441	0.112
WOW	0.0048	0.4436	<0.0001	<0.0001	0.8494	0.6477	0.132

Notes:

* Entries in columns 2 to 5 are p-values of the LM test for the null hypothesis of no serial correlation against an AR(4) alternative for the realized variance (RV), the jump component ($Jump$), the realized bi-power variation (BV) and the logarithm of the square root of bi-power variation ($\ln\left(\sqrt{BV}\right)$) respectively.

** ϕ and θ are the estimated autoregressive and moving average parameters of the ARMA(1,1) model $\ln\left(\sqrt{BV_t}\right) = c + \phi \ln\left(\sqrt{BV_{t-1}}\right) - \theta\epsilon_{t-1} + \epsilon_t$.

*** d is the estimated degree of fractional integration in $\ln\left(\sqrt{BV}\right)$. The 5% and 1% critical values for a test of H_0: $d > 0$ are respectively $d = 0.092$ and $d = 0.144$ for a sample of 260.

Table 4: Out-of-sample performance of univariate models in forecasting $\ln\left(\sqrt{BV}\right)$

	ARMA		SES		Pooled
Stock	p, q	$RMSE$	α	$RMSE$	$RMSE$
Mining					
AMC	1,1	0.2216	0.270	0.2265	0.2225
BHP	1,1	0.2844	0.320	0.2887	0.2824
CSR	1,1	0.2919	0.330	0.3002	0.2963
MIM	2,2	0.1673	0.224	0.1738	0.1683
RIO	1,1	0.2352	0.168	0.2383	0.2329
WMC	1,1	0.2672	0.322	0.2736	0.2685
Energy					
STO	1,1	0.2652	0.112	0.2720	0.2670
WPL	1,1	0.3886	0.262	0.4024	0.3864
Banks					
ANZ	1,1	0.2747	0.282	0.3057	0.2906
CBA	1,0	0.2529	0.176	0.2832	0.2531
NAB	1,1	0.2550	0.344	0.2724	0.2633
SGB	3,0	0.2679	0.232	0.2740	0.2724
WBC	1,1	0.2694	0.300	0.2839	0.2731
Food & Bev.					
FGL	1,1	0.2468	0.140	0.2467	0.2450
SRP	1,0	0.2765	0.030	0.2930	0.2903
Other					
BIL	1,1	0.4611	0.206	0.4668	0.4681
LLC	1,1	0.2932	0.198	0.3005	0.2843
MAY	2,0	0.3352	0.150	0.3614	0.3454
NCP	1,1	0.2685	0.174	0.2738	0.2610
QAN	1,0	0.4000	0.226	0.4062	0.3969
WOW	1,0	0.2686	0.192	0.2659	0.2610

Notes:

* p and q are the autoregressive and moving average order of ARMA models.

** The SES column reports the smoothing parameter (α) and the RMSE of the single exponential smoothing model.

*** In the "pooled" model, all autoregressive parameters are restricted to be the equal across equations.

Table 5: Out-of-sample performance of factor models in forecasting $\ln\left(\sqrt{BV}\right)$

Stock	RMSE of EqW	PC	IVLI	CC
Mining				
AMC	0.1956	0.1976	0.2025	0.2010
BHP	0.2724	0.2716	0.2660	0.2622
CSR	0.2792	0.2800	0.2811	0.2814
MIM	0.1647	0.1659	0.1637	0.1653
RIO	0.2268	0.2291	0.2353	0.2340
WMC	0.2548	0.2605	0.2615	0.2626
Energy				
STO	0.2533	0.2717	0.2691	0.2711
WPL	0.3838	0.4172	0.3856	0.3898
Banks				
ANZ	0.2645	0.2700	0.2725	0.2731
CBA	0.2498	0.2609	0.2539	0.2532
NAB	0.2480	0.2480	0.2561	0.2560
SGB	0.2650	0.2580	0.2767	0.2666
WBC	0.2622	0.2560	0.2609	0.2560
Food & Bev.				
FGL	0.2367	0.2496	0.2395	0.2443
SRP	0.2875	0.2916	0.2919	0.2905
Other				
BIL	0.4477	0.4433	0.4419	0.4438
LLC	0.3292	0.3091	0.3233	0.3331
MAY	0.3196	0.3261	0.3296	0.3265
NCP	0.2556	0.2607	0.2599	0.2608
QAN	0.3866	0.3879	0.3835	0.3898
WOW	0.2526	0.2544	0.2506	0.2500

Notes:

* EqW are models that incorporate an equally weighted index of all log-volatilities. PC are models that incorporate the first two principal components. IVLI are models that incorporate a leading index estimated from a reduced rank regression using principal component analysis. CC are models that incorporate two leading indices corresponding to two largest canonical correlations. For more information refer to the text.