HIGH DIMENSION DYNAMIC CORRELATIONS

September 14, 2007

PREPARED FOR A FESTSCHRIFT FOR DAVID HENDRY

August 23, 2007

Oxford

ABSTRACT

This paper develops time series methods for forecasting correlations in high dimensional problems. The Dynamic Conditional Correlation model is given a new convenient estimation approach called the MacGyver method. It is compared with the FACTOR ARCH model and a new model called the FACTOR DOUBLE ARCH model. Finally the latter model is blended with the DCC to give a FACTOR DCC model. This family of models is estimated with daily returns from 18 US large cap stocks. Economic loss functions designed to form optimal portfolios and optimal hedges are used to compare the performance of the methods. The best approach invariably is the FACTOR DCC and the next best is the FACTOR DOUBLE ARCH.

I. Introduction

David Hendry is perhaps the best. When the debate over general to simple specification searches gets most heated, the approach is often called the Hendry method. Critics say that only David can really do this. Everyone agrees that in his hands, economic time series yield up reasonable models that are coherent with both data and theory. To quiet the critics, much of David's recent research has been to automate the methodology so that even a computer can do it. Econometricians might be skeptical of such methods, yet their practical relevance for empirical work is hard to dismiss.

At the same time, David has carefully described what he calls "predictive failure" of economic time series models. Even the best models appear to frequently have forecast

errors larger than can be expected. Perhaps the economy is not as stationary as we think. Perhaps there are "Black Swans" everywhere, to use a recent metaphor by Taleb(2007) which refers to the observation that since every swan seen in Europe was white, there was no way statistically to be prepared for the fact that there were black swans in Australia. This tension between model selection and predictive failure lies at the heart of all empirical econometrics.

Nowhere is this tension more apparent than in Finance where large sums are invested on the assumption that history will repeat itself in some fashion. Many such investments are well rewarded until there is some unexpected event where they go rapidly in the wrong direction. The summer of 2007 is unfolding as such an event and it will be interesting to look back at it.

Many financial models are designed to measure risks. Once risks are measured, then investments can be structured to maximize return and minimize risk. However, if the risks are not well represented by historical experience, then it will turn out that investors frequently take risks they did not realize they were taking and consequently have, on average, inferior outcomes.

The goal of this paper is to develop time series measures of risks in a highly multivariate framework. In this multidimensional context, the risk is often summarized not only by the volatility of the components but also the correlation among them. Since the world is hopelessly multidimensional, the forecasting of correlations is a central feature of financial planning. The dynamics must quickly adapt to new types of risks yet be unresponsive to random shocks. The structure must evolve but still allow the possibility of "Black Swan" events without letting such events have undue effect on the

performance. This is very similar to the model selection vs. predictive failure dichotomy emphasized by David Hendry.

In this paper I will take a forecasting point of view in modeling correlations between asset returns. The paper will discuss the Dynamic Conditional Correlation model and its general approach to estimating correlations. For large systems, a new improved estimation method will be presented, called the MacGyver method. Then to increase accuracy, a factor structure will be incorporated into the DCC model. This paper will introduce the FACTOR DCC model that has the potential to forecast correlations in high dimensional systems of asset returns.

II. Dynamic Conditional Correlation

The search for multivariate models that can effectively estimate volatilities and correlations for a large class of assets continued for more than 20 years. The menu of choices is immense and has been surveyed recently by Bauwens Laurent and Rambouts(2006) and less recently by Bollerslev Engle and Nelson(1994). Only a few of these models are amenable to estimating correlations for more than half a dozen assets.

One of the most practical and simple is the Dynamic Conditional Correlation (DCC) model introduced by Engle(2002). This model uses a sequential estimation scheme and a very parsimonious parameterization to enable it to estimate models with fifty or more assets rather easily. It is a simple generalization of Bollerslev's Constant Conditional Correlation (CCC) model.

Consider an nx1 vector of returns, r_t , with conditional covariance matrix H_t.

The conditional covariance matrix must be positive definite if there are no redundant assets. Such a covariance matrix can always be decomposed into a diagonal matrix D_t which has the conditional standard deviation of each asset along the main diagonal, and a correlation matrix R_t which has ones on the diagonal and correlations off the diagonal.

$$V_{t-1}\left(r_t\right) = H_t = D_t R_t D_t, \quad D_t \sim diagonal, \; R_t \sim correlation \tag{1}$$

Clearly, the correlation matrix is the same as the covariance matrix of standardized returns, s_t.

$$s_t \equiv D_t^{-1} r_t, \; V_{t-1}\left(s_t\right) = R_t \tag{2}$$

Hence models to estimate the conditional correlations would naturally use standardized returns as inputs. These standardized returns are sometimes called volatility adjusted returns or standardized residuals. As residuals will have a slightly different meaning in the factor model, I will use the expression, standardized returns.

The heart of the DCC model is the parameterization of the updating or forecasting equation. The correlation matrix is expressed as a function of past observables and this matrix must indeed be a correlation matrix, i.e. be positive definite and have ones on the diagonal. Two versions can be presented simply. Many others are obvious generalizations. The two are called “integrated” and “mean reverting”. They are defined more precisely by

$$R_t = diag\left(Q_t\right)^{-1/2} Q_t diag\left(Q_t\right)^{-1/2} \tag{3}$$

where Q is defined either by the integrated model

$$Q_t = \left(1-\lambda\right) s_{t-1} s_{t-1}' + \lambda Q_{t-1} \tag{4}$$

or the mean revering model

$$Q_t = \Omega + \alpha s_{t-1} s_{t-1}' + \beta Q_{t-1}. \tag{5}$$

There are advantages and disadvantages of each specification. Both (4) and (5) will generate *Q* matrices that are positive definite as long as the initial condition is positive definite and the matrix intercept of (5) is positive definite. If *Q* is positive definite, then *R* will be a correlation matrix. The integrated model assumes all changes in correlations to be permanent. This may be a reasonable assumption, although it appears that many correlation processes are mean reverting. The integrated model is not a satisfactory description of the data in another way. It implies that asymptotically correlations go to plus or minus one. This can be verified by simulating (2),(3) and (4). Nevertheless, it may be a good filter in the sense of Nelson and Foster(1994).

The mean reverting model assumes that all changes in correlations are transitory although they can last quite a long time if the sum of alpha and beta is close to unity. This model has $n(n-1)/2+2$ parameters while the integrated model has only 1. The solution to this set of extra parameters is to introduce another set of estimating equations. These equations are moment conditions that can be used with the FOC of the likelihood function. Letting the sample correlation of the standardized returns be $\bar{R}$, a second relation can be obtained among the unknowns.

$$\bar{R} = \frac{1}{T}\sum_{t=1}^{T} s_t s_t', \quad \bar{Q} = \frac{1}{T}\sum_{t=1}^{T} Q_t \cong \Omega + \alpha \bar{R} + \beta \bar{Q} \tag{6}$$

Finally, adding the assumption that on average, $\bar{Q} = \bar{R}$, the intercept can be expressed as

$$\Omega = (1 - \alpha - \beta)\bar{R} \tag{7}$$

and equation (5) can be rewritten as

$$Q_t = \bar{R} + \alpha\left(s_{t-1}s_{t-1}' - \bar{R}\right) + \beta\left(Q_{t-1} - \bar{R}\right) \tag{8}$$

Hence, *Q* is mean reverting to the average correlation. Equation (8) only has two parameters no matter how big the system is. The assumption (7) is called "correlation targeting" and is an estimator of the omega parameters that is different from maximum likelihood. As a consequence, this is necessarily an asymptotically inefficient estimator although it may be relatively robust to some forms of misspecification. This is a generalization of the "variance targeting" approach of Engle and Mezrich(1996). It has been analyzed both theoretically and empirically in Engle and Sheppard(2005).

Many other specifications have been introduced for DCC models. Asymmetric correlation models were introduced in Cappiello, Engle and Sheppard(2008) who find that correlations become larger when two returns are both negative than if they are equally positive and all other factors are the same. They also introduce a more generous parameterization of the DCC process called Generalized DCC. Engle(2002) discusses additional lags. Many formulations are included in the forthcoming book Engle(200?).

Maximum likelihood estimation of such systems must include a distributional assumption and multivariate normality is common although not very accurate. Fortunately, most multivariate GARCH models are QMLE estimators so that they will be consistent as long as the covariance equations are correctly specified even if the normality assumption is incorrect. See for example Bollerslev and Wooldridge(1992).

In the case of multivariate normality, the average log likelihood function plus some unimportant constants, becomes

$$
\begin{aligned}
L &= -\frac{1}{2T}\sum_{t=1}^{T}\left[\log|H_t| + r_t' H_t^{-1} r_t\right] \\
&= -\frac{1}{2T}\sum_{t=1}^{T}\left[2\log|D_t| + r_t' D_t^{-2} r_t\right] - \frac{1}{2T}\sum_{t=1}^{T}\left[\log|R_t| + s_t' R_t^{-1} s_t\right] + \frac{1}{2T}\sum_{t=1}^{T} s_t' s_t
\end{aligned}
\tag{9}
$$

This log likelihood can be maximized with respect to all the parameters in the volatilities, which are inside the matrix *D*, and correlations which are inside *R*. This means simply maximizing the first line of (9). Alternatively, the likelihood can be approximately maximized by finding the maximum of the first square bracket terms with respect to the volatility parameters, and then the maximum of the second square bracket terms with respect to the correlation parameters. As long as the first problem gives a consistent estimate of the volatility parameters, then the maximum of the second part will be consistent under standard regularity conditions and the third term will converge to a constant. Engle and Sheppard(2005) gives this argument in more detail. The log likelihood for the correlation estimation is called L_2 and is given by

$$
L_2 = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|R_t| + s_t' R_t^{-1} s_t\right] \tag{10}
$$

This estimation method is very appealing. Univariate models are estimated for each of the volatilities and typically these are simple GARCH models. They can be estimated separately as long as there are no common parameters. Then the correlations are estimated by MLE where the data are the standardized returns. In the two cases discussed above there are only one or two parameters respectively in this estimation stage regardless of n. Thus a very parsimonious parameterization can be used to estimate arbitrarily large systems.

III. The MacGyver Method

The estimation of correlation matrices for large systems might appear solved from the previous section. However, there are three reasons to believe that this is not a full solution. First, the evaluation of the log likelihood function requires inversion of matrices, R_t , which are full nxn matrices, for each observation. To maximize the likelihood function, it is necessary to evaluate the log likelihood for many parameter values and consequently invert a great many nxn matrices. Convergence is not guaranteed and sometimes it fails or is sensitive to starting values. These numerical problems can surely be alleviated but ultimately for very large n, the numerical issues will dominate. Secondly, Engle and Sheppard(2005a) show that in correctly specified models with simulated data, there is a downward bias in α when n is large. Thus the correlations are estimated to be smoother and less variable when a large number of assets are considered than when a small number of assets are considered. Thirdly, there may be structure in correlations which is not incorporated in this specification. This of course depends upon the economics of the data in question but the introduction of the FACTOR DCC model in section IV is a response to this issue.

In this section I will introduce a new estimation method which is designed to solve the first two problems and a few others as well. I call this a MacGyver method after the old TV show which showed MacGyver using whatever was at hand to cleverly solve his problem. The show was a triumph of brain over brawn.

The MacGyver method is based on bivariate estimation of correlations. It assumes that the selected DCC model is correctly specified between every pair of assets i and j. Hence the correlation process is simply

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}} \qquad (11)$$
$$q_{i,j,t} = \bar{R}_{i,j}\left(1-\alpha-\beta\right) + \alpha s_{i,t-1} s_{j,t-1} + \beta q_{i,j,t-1},$$

and the log likelihood function for this pair of assets is simply extracted from (10). It is given by

$$L_{2,i,j} = -\frac{1}{2}\sum_t \left(\log\left[1-\rho_{i,j,t}^2\right] + \frac{s_{i,t}^2 + s_{j,t}^2 - 2\rho_{i,j,t} s_{i,t} s_{j.t}}{\left[1-\rho_{i,j,t}^2\right]} \right) \qquad (12)$$

Because the high dimension model is correctly specified, so is the bivariate model. The estimates are MLE, at least conditional on the volatility parameters. The only parameters are (α, β) and the estimates should be consistent using only data on one pair of assets. Obviously however, information is being ignored that could yield more efficient estimates. Thus improved estimation should combine parameter estimates from these bivariate models. The combined parameters are then used with equation (11) to calculate the correlations.

An analytical solution to optimal combination of bivariate parameter estimates seems extremely difficult. The data are dependent from one pair to another but the dependence is a function of the parameters. Knowing the dependence of the data does not lead easily to measures of the dependence of the parameter estimates. At some point perhaps an analytical solution to this problem will arise. In the meantime, I will develop the estimator based on Monte Carlo performance.

A variety of simulation environments is postulated. In each case all bivariate pairs are estimated and then simple aggregation procedures such as means or medians are applied. Several issues immediately arise. What should be done about cases where the

estimation does not converge or where it converges to a value outside the region of stationarity? When averaging parameters that have constrained ranges, it is easy to introduce bias.

Six estimators will be considered. These are shown in Table 1 below

Table 1

MacGyver Estimators

	Unrestricted MLE	Restricted MLE
Mean	MEAN	_R MEAN
Median	MED	_R MED
5%Trimmed Mean	MEANT05	_R MEANT05

The trimmed means are computed by deleting the largest and smallest 5% of the estimates and then taking the mean of the remaining ones. The unrestricted MLE simply maximizes the log likelihood (12) without restriction. If it does not converge in a finite number of iterations, then the final value of the estimate is taken. Obviously, this estimation can and does occasionally lead to some very bizarre parameter estimates. The restricted MLE reparameterizes the log likelihood using a logistic functional form so that both parameters must lie in the interval (0,1). Their sum was not restricted in this case. The model is expressed as

$$q_{i,j,t} = \bar{R}_{i,j} + \frac{e^{\theta}}{1+e^{\theta}}\left(s_{i,t-1}s_{j,t-1} - \bar{R}_{i,j}\right) + \frac{e^{\phi}}{1+e^{\phi}}\left(q_{i,j,t-1} - \bar{R}_{i,j}\right) \tag{13}$$

The optimizer chooses (θ, ϕ) but the estimated values of (α, β) are passed back to be averaged across bivariate pairs.

Ten experiments are run with various parameter values and dimensions. All have a time series sample size of 1000 observations and 100 replications of each experiment. The dimensions range from n=3 to n=50. The ten experiments are defined in Table A1. The true correlation matrix has all correlations equal to Rhobar. In each case the parameters are estimated by bivariate MLE or restricted bivariate MLE and then the summary measures are computed according to each expression in Table 1. The ultimate result is a table of root mean squared errors and a table of biases across the simulations for each of the two parameters, alpha and beta.

The tables of RMS errors and biases are presented in the appendix as tables A2 and A3. The net result is that the smallest errors are achieved by the median estimator. For beta the best estimator for each experiment is either the median or the median of the restricted estimator. On average the median of the unrestricted estimator is the smallest. For alpha, the medians are best in most experiments and the median of the unrestricted bivariate parameter estimates has the smallest rms error. This estimator effectively ignores all the non-convergent and non-stationary solutions and gives parameter estimates which are very close to the true value.

The biases of these estimators are also of interest. In all experiments the bias in beta is negative and the bias in alpha is positive. This is not surprising in a context where the beta is truncated from above (at one) while alpha is truncated below (at zero). Notice however that this bias is in the opposite direction from the bias observed by Sheppard and Engle(2005a) who found alpha too small and beta too big for large systems. Notice also that the bias is very small. On average over the experiments the bias in alpha is .001 and the bias in beta is -.008. Since these biases result from bivariate estimation, there is no

large system bias as there is for MLE estimation of DCC. In fact, the RMSE's are smallest for the largest systems.

In addition to the computational simplification and bias reduction, there are several other advantages to this MacGyver method of estimating a DCC model. When there are 50 assets, there are 1225 bivariate pairs. When there are 100 assets, there are 4950 asset pairs. Hence the number of bivariate estimations increases as well. However, since only the median of all these estimations is needed, there is little loss of efficiency if some are not run. This opens the possibility of estimating a subset of the bivariate pairs. While it is not clear how to select a good subset, it is clear that there is little advantage to doing all of them. When new assets are added to the collection, it may not be necessary to reestimate at all if the investigator is confident that the specification is adequate.

A second advantage is that the data sets for each bivariate pair need not be of the same length. Thus, an asset with only a short history can be added to the system without requiring the shortening of all other series. This is particularly important when examining large asset classes and cross country correlations as there are many assets which are newly issued, merged or otherwise associated with short time histories.

A potential third advantage which will not be explored in this paper, is that there may be evidence in these bivariate parameter estimates that the selected DCC model is not correctly specified. Presumably, the bivariate models would show less dispersion if the model is correctly specified than if it is incorrect.

IV. FACTOR DCC

The most popular approach to estimating large covariance matrices in finance is the use of factor models. By specifying a small number of factors that summarize all the dependence between returns, a complete matrix of correlations can be estimated. This is a simple strategy in principle but in practice it is difficult to select factors and it is difficult to estimate the factor loadings and other parameters of the process. Nevertheless, this class of models can incorporate directly some effects which the DCC model can only indirectly replicate.

Consider first, the very simple static one factor model that is the centerpiece of the Capital Asset Pricing Model or CAPM. Measuring returns in excess of the risk free rate and letting r_m be the market return, the model is most simply expressed as

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t} \tag{14}$$

From theoretical arguments, we expect the alphas to be zero in an efficient market. And we expect the idiosyncratic returns to be uncorrelated across assets.

$$\begin{aligned} V\left(r_{i,t}\right) &= \beta_i^2 V\left(r_{m,t}\right) + V\left(\varepsilon_{i,t}\right) \\ Cov\left(r_{i,t}, r_{j,t}\right) &= \beta_i \beta_j V\left(r_{m,t}\right) \end{aligned} \tag{15}$$

Thus the correlation between two assets can be expressed as

$$\rho_{i,j} = \frac{\beta_i \beta_j V\left(r_{m,t}\right)}{\sqrt{\left(\beta_i^2 V\left(r_{m,t}\right) + V\left(\varepsilon_{i,t}\right)\right)\left(\beta_j^2 V\left(r_{m,t}\right) + V\left(\varepsilon_{j,t}\right)\right)}} \tag{16}$$

These expressions insure that the correlation matrix will be positive definite. They however do not provide any measures of time varying variances, covariances or correlations.

The simplest approach to formulating a dynamic version of this one factor model is to follow Engle, Ng and Rothschild(1990,1992). In this case the factor has time

varying volatility and can be modeled with some form of ARCH model. Consequently, the expressions in (15) and (16) can be rewritten in terms of conditional variances. The conditional correlation then becomes:

$$\rho_{i,j,t} = \frac{\beta_i \beta_j V_{t-1}\left(r_{m,t}\right)}{\sqrt{\left(\beta_i^2 V_{t-1}\left(r_{m,t}\right) + V_{t-1}\left(\varepsilon_{i,t}\right)\right)\left(\beta_j^2 V_{t-1}\left(r_{m,t}\right) + V_{t-1}\left(\varepsilon_{j,t}\right)\right)}} \tag{17}$$

The model used by ENR assumed that the idiosyncratic volatilities were not changing over time. They called this the FACTOR ARCH model and I will use that name here. Letting $\beta,$ and r_t be nx1 vectors, the statistical specification is

$$V_{t-1}\begin{pmatrix} r_t \\ r_{m,t} \end{pmatrix} = \begin{pmatrix} \beta\beta' h_{m,t} + D^2 & \beta h_{m,t} \\ \beta' h_{m,t} & h_{m,t} \end{pmatrix}, \quad D \sim diagonal \tag{18}$$

The conditional correlation between each pair of assets would be time varying only because the market volatility is changing. From an examination of (17) it is clear that the conditional correlation in this model is a monotonic function of market volatility ranging from zero to one as market volatility ranges from zero to infinity.

In the ENR or FACTOR ARCH model, there would always be portfolios of assets which would have no ARCH. Engle and Susmel(1993) looked for such portfolios and found in an international context, that they unlikely. Almost all portfolios have time varying volatility, even if they have a zero beta on the market. Hence, there must either be more factors or time varying idiosyncratic volatility.

Assuming normal errors, the statistical model is

$$r_t \left| r_{m,t}, \mathrm{F}_{t-1} \sim N\left(\beta r_{m,t}, D^2\right), \; r_{m,t} \right| \mathrm{F}_{t-1} \sim N\left(0, h_{m,t}\right) \tag{19}$$

This model can simply be estimated by MLE by regressing each asset return on the market return with OLS, and then estimating the market volatility with GARCH.

The natural extension of this model is to allow the idiosyncrasies to follow a GARCH process as well as the market return. Thus there are two GARCH processes for an asset. For convenience we will call this model a FACTOR DOUBLE ARCH. The model is expressed as:

$$V_{t-1}\begin{pmatrix} r_t \\ r_{m,t} \end{pmatrix} = \begin{pmatrix} \beta\beta' h_{m,t} + D_t^2 & \beta h_{m,t} \\ \beta' h_{m,t} & h_{m,t} \end{pmatrix}, \quad D_t \sim diagonal\left\{ garch\, std \right\} \tag{20}$$

Where D_t is a diagonal matrix with GARCH standard deviations on the diagonal.

Assuming conditionally normal returns this can be rewritten as

$$r_t \left| r_{m,t}, \mathrm{F}_{t-1} \sim N\left(\beta r_{m,t}, D_t^2\right), \; r_{m,t} \right| \mathrm{F}_{t-1} \sim N\left(0, h_{m,t}\right) \tag{21}$$

This model is still easy to estimate by MLE. The return on an asset is regressed on the market return with disturbances that follow a GARCH. Then the GARCH for the market is only estimated once. To see that this two step estimator is MLE, express the likelihood for this problem as the density of asset returns conditional on the market return times the marginal density of market returns. Ignoring irrelevant constants, the log likelihood is:

$$L\left(r, r_m\right) = -\sum_{t=1}^{T} \log\left|D_t\right| - \frac{1}{2}\sum_{t=1}^{T}\left(r_t - \beta r_{m,t}\right) D_t^{-2}\left(r_t - \beta r_{m,t}\right) - \frac{1}{2}\sum_{t=1}^{T}\left[\log\left(h_t\right) + \frac{r_{m,t}^2}{h_t}\right] \tag{22}$$

This model satisfies the weak exogeneity conditions of Engle Hendry and Richard(1983) which allow for separate estimation of the conditional and marginal models. As long as the parameters are distinct or *variation free* so that no information from the marginal model would affect inference in the conditional model, then market returns can be

considered weakly exogenous and the MLE of the system is the same as the MLE done in two steps.

There are many reasons to believe that the ONE FACTOR DOUBLE ARCH model just described, will still be too simple to accurately forecast correlations. The correlations between stock returns in the same industry are typically higher than for stocks across industries and these correlations will rise if the industry volatility rises. These are essentially additional factors with impacts on correlations that vary over time. Even more interesting are factors that have zero variance some of the time and a large variance other times. Energy prices might be in this category. It would be impossible to identify this factor until it is active but then it may be too late. Finally, the model assumes that the factor loadings or betas are constant over time, yet whenever a firm changes its line of business, its sensitivity to various factors will naturally change.

Ideally, the model should allow correlations among idiosyncrasies and between idiosyncrasies and market shocks and these correlations should be time varying. In this way the statistical model will recognize the changing correlation structure when a new factor emerges or factor loadings change.

The FACTOR DCC model is designed to do just this. It proceeds exactly as described above for the FACTOR DOUBLE ARCH and then estimates a DCC model on the residuals. More precisely, the FACTOR DCC model has the specification

$$r_t = \beta r_{m,t} + D_t \varepsilon_t, \;\; r_{m,t} = \sqrt{h_t}\,\varepsilon_{m,t}, \begin{pmatrix} \varepsilon_t \\ \varepsilon_{m,t} \end{pmatrix} \sim N\left(0, R_t\right) \tag{23}$$

The specification of the correlation matrix can be the same as (3) coupled with either (4) or (5) or more general versions of DCC. Partitioning the (n+1)x(n+1) correlation matrix into its conformal parts as

$$R_t = \begin{pmatrix} R_{11,t} & R_{1,m,t} \\ R_{m,1,t} & 1 \end{pmatrix} \tag{24}$$

the covariance matrix of returns is given by

$$V_{t-1}\begin{pmatrix} r_t \\ r_{m,t} \end{pmatrix} = \begin{pmatrix} \beta\beta' h_{m,t} + D_t R_{1,1,t} D_t + \sqrt{h_{m,t}}\beta R_{m,1,t} D_t + \sqrt{h_{m,t}} D_t R_{1,m,t}\beta' & \beta h_{m,t} + \sqrt{h_{m,t}} D_t R_{1,m,t} \\ \beta' h_{m,t} + \sqrt{h_{m,t}} R_{m,1,t} D_t & h_{m,t} \end{pmatrix} \tag{25}$$

When new factors arise, the correlations between some stocks will increase. These are captured by the second term in the upper left hand block. When betas change, these effects will be captured by the third and fourth terms.

The model in equation (23) is only a small generalization of the basic DCC model. The data in this case are not just standardized returns but standardized idiosyncratic returns. If either the FACTOR ARCH or the FACTOR DOUBLE ARCH are correctly specified, then the DCC should find zero correlations both conditionally and unconditionally. Estimation is naturally done in two steps again where the first step estimates both the static factor loading and the idiosyncratic GARCH. The second step estimates the DCC parameters. Here joint estimation is possible or the MacGyver method can be used. In this paper the MacGyver method will be employed.

The conditional correlations are again defined as the conditional covariance divided by the product of the conditional standard deviations using the expression for the conditional covariance matrix of returns in (25). In each case there are now four terms and the last three depend upon the DCC estimated correlations.

V. Performance

To examine the properties of these correlation estimators a set of 18 daily US large cap equity returns will be examined. The data range from 1994 through 2004 for

2771 observations. The tickers are {aa, axp, ba, cat, dd, dis, ge, gm, ibm, ip, jnj, jpm, ko, mcd, mmm, mo, mrk, msft } which are all components of the Dow. The S&P500 is taken as the market return.

1. MacGyver Estimates

The MacGyver method is applied to this data set to estimate all the correlations with DCC. Although it is not necessary to use the same GARCH model for each series, in this investigation I do. To account for the asymmetry in volatility the GJR or Threshold GARCH model is used. It is specified by

$$r_{i,t} = \sqrt{h_{i,t}}\varepsilon_{i,t},\ h_{i,t} = \omega_i + \theta_i r_{i,t-1}^2 + \gamma_i r_{i,t-1}^2 I_{r_{i,t-1}<0} + \phi h_{i,t-1} \tag{26}$$

The standardized returns from these models are saved and used as inputs for the DCC estimation by MacGyver. For 18 returns there are 18*17/2=153 bivariate models. For most of these, the results are quite standard. For a few, they are completely unsatisfactory. For example, all of the alphas are estimated to be between zero and .05 except for one that is just over 2. Similarly, most of the betas are less than one but for the same bivariate estimate, beta is over four thousand. A few of the betas are quite small or negative. Nevertheless, the medians are very close to general experience. The median for alpha is 0.0157 and the median for beta is .9755 so that the sum is just over .99 leading to a good degree of persistence in correlations. The actual plots are given in the appendix.

The DCC estimation produces 153 time series of correlations based on these two parameters and the unconditional correlations. It is difficult to examine so many time series at once. Some clear patterns can easily be seen by looking at the average correlations. These will establish the stylized facts of correlations in the US equity

market. In Figure 1, the mean correlation is plotted from the 100 day historical method and the DCC method with TARCH volatilities.

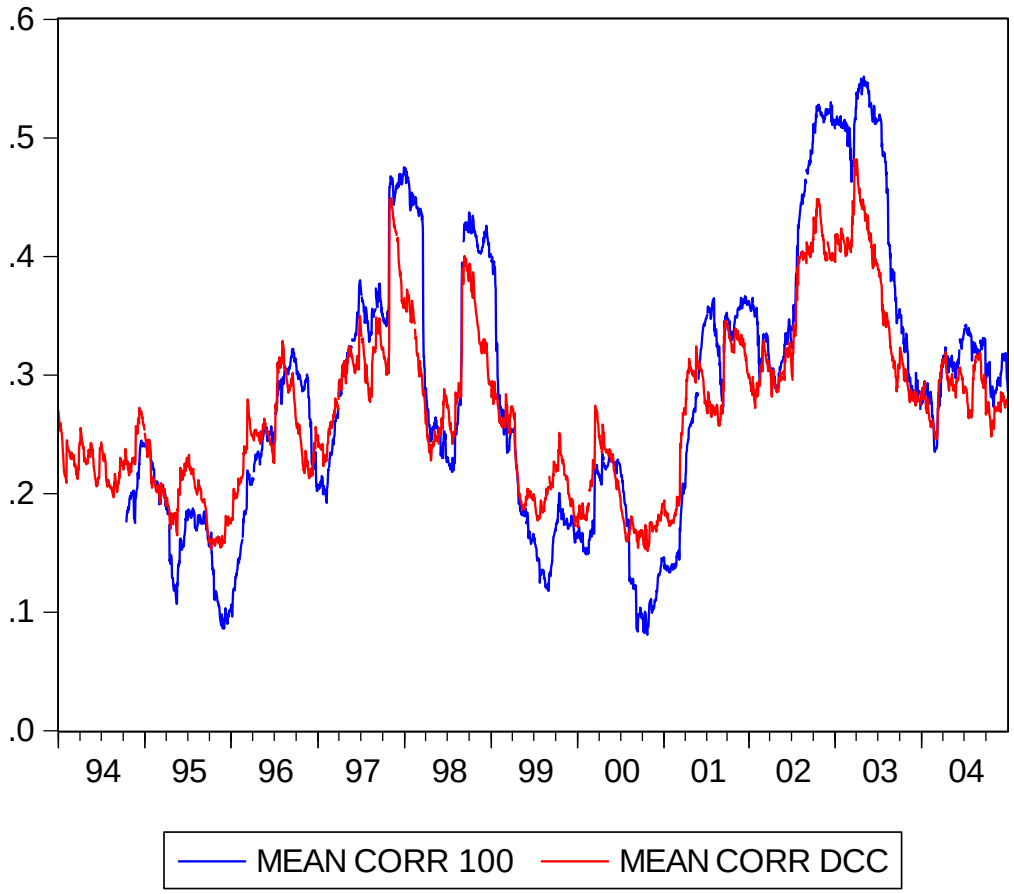

Figure 1

Mean Historical and DCC Correlations

The historical correlations and the DCC correlations trace out very much the same pattern. The range of the historical correlations is a little greater but this may be a result of the choice of smoothing. A 200 day correlation would move substantially less. The

historical correlations also have wider peaks making the correlation estimate somewhat slower to respond to news. A plot of the cross sectional standard deviation of the 153 bivariate correlations reveals that the historical correlations are more varied across pairs than the DCC.

It is clear that these correlations have changed substantially over the 10 year period. The highest correlations are during the recession in 2002 and the first part of 2003. Correlations are low during the internet bubble and the subsequent bursting of the bubble. They rise in 2001 and abruptly increase further after 9/11. In 2003 correlations fall as the economy and stock market recover. There are two episodes of spiking correlations in the late 90's which can be associated with the LTCM/Russian Default and the Asian currency crisis. In fact the proximate cause of the second spike is the "Anniversary Crash" on October 27, 1997 when the market fell 7% and then recovered 5% the next day. These events are plotted with the correlations in Figure 2.[1] It certainly appears that economic crises lead to rising correlations.

[1] The LTCM dummy is defined for August 1998 through 25 September 1998, the Asia Crisis dummy is defined for May 14 1997 through July 31, 1997, the Anniversary Crash dummy is defined for October 27 and 28, 1997.

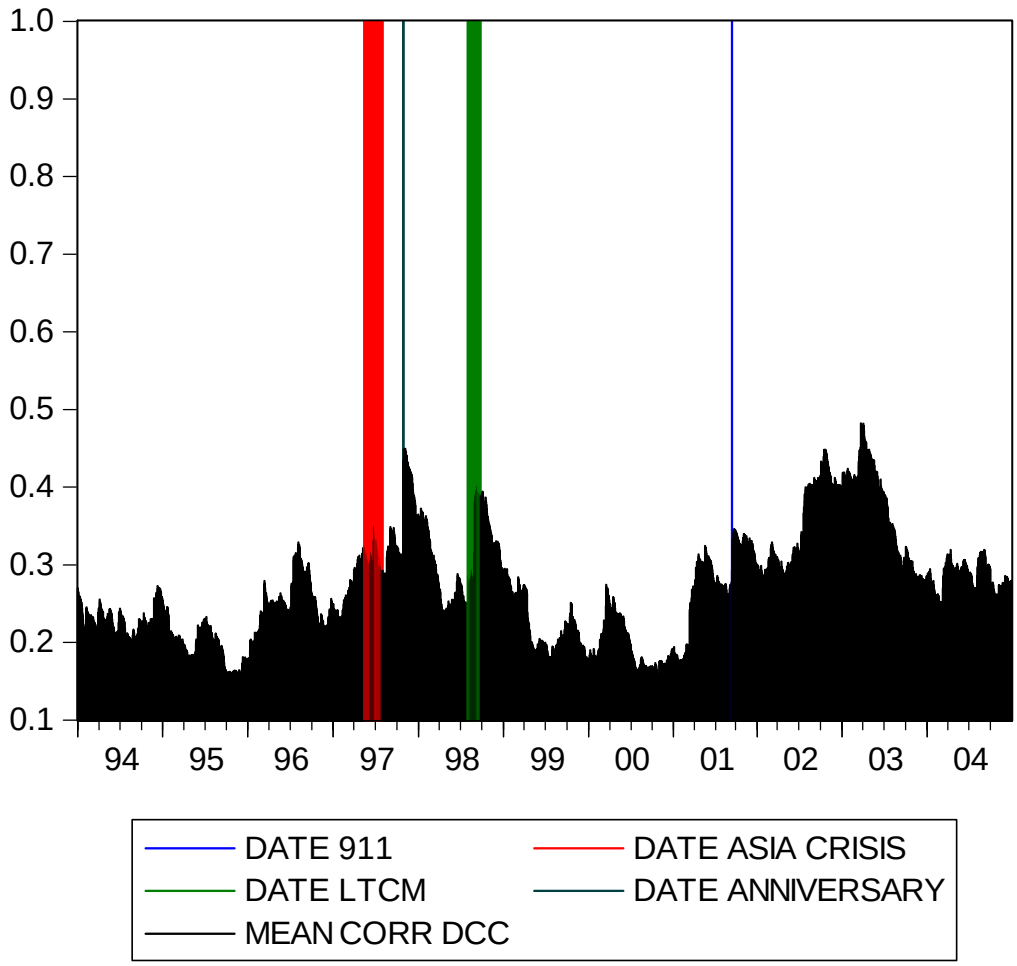

Figure 2

Mean Correlations and Significant Dates

2. FACTOR ARCH, FACTOR DOUBLE ARCH

The sharp movements in correlations that are associated with movements in the S&P itself suggest the usefulness of a factor model. The FACTOR ARCH and the FACTOR Double ARCH are now calculated. They follow the specification (18) and

(20) . The betas are estimated by OLS for the FACTOR ARCH and by GLS with GARCH errors for the FACTOR Double ARCH and consequently are slightly different. From Figure 3, it is clear that these differences are small for all 18 stocks.

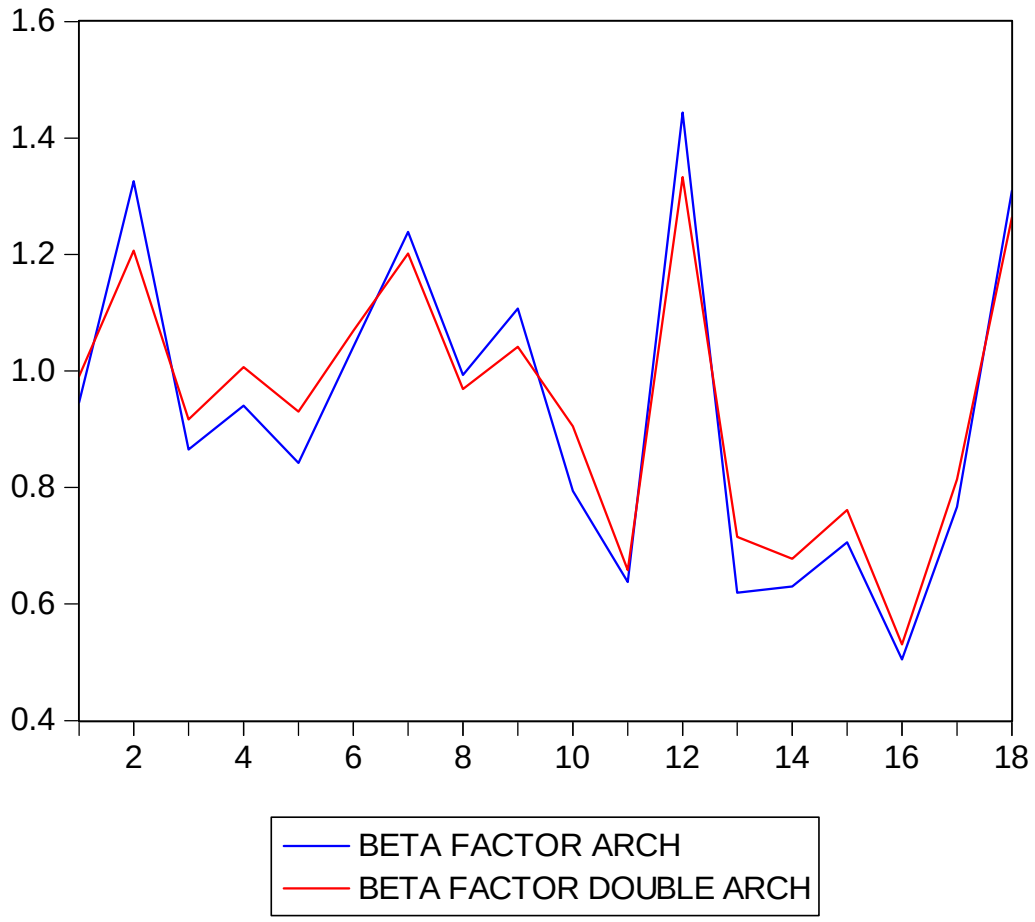

Figure 3

Betas

The correlations from each of these models can be calculated using (17). The average across all pairs is again a useful measure. This is shown in Figure 4.

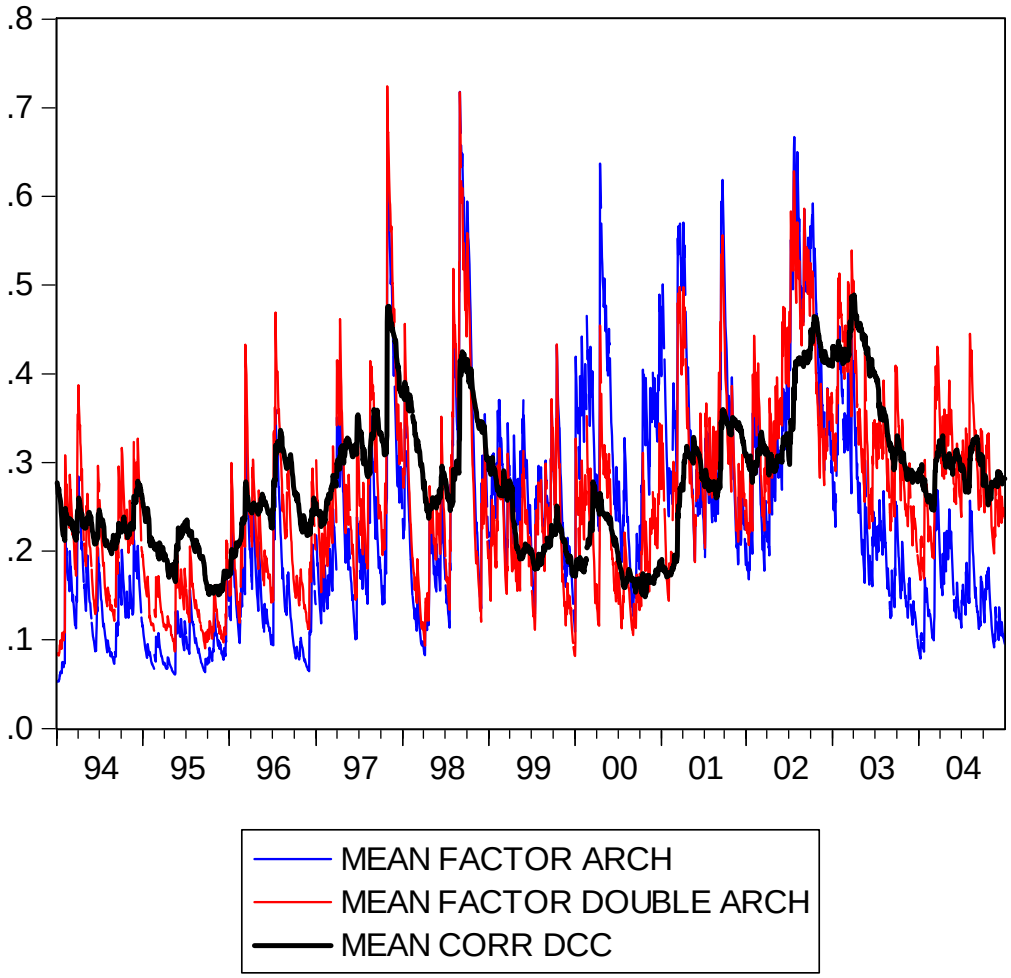

Figure 4.

Mean Correlations of FACTOR MODELS

The average correlation from the FACTOR DOUBLE ARCH model is very similar in level to the average DCC. It differs primarily in that the FACTOR DOUBLE ARCH correlations are more volatile. When the correlations spike up because of some market event, they rise up to .7 in several cases, and when the correlations fall, they fall further. It is not clear whether the higher volatility is a good or bad aspect of this estimator as we do not know what the true conditional correlations are at any point in time.

The patterns of the FACTOR ARCH are however different in several important ways. Over the last two years of the sample, the FACTOR ARCH correlations fall much lower than any of the other correlation estimators. This is also the case in the middle nineties. The opposite however occurs in 1999 and 2000 when the FACTOR ARCH correlations are higher than DCC and FACTOR DOUBLE ARCH. These differences are easy to understand. The monotonic relation between average correlation and market volatility in the FACTOR ARCH model implies that the correlations should be at their lowest in the middle nineties and since 2003 since market volatility is lowest then. However, the idiosyncratic volatilities also change in much the same direction so that the more accurately estimated correlations either from the DCC or the FACTOR Double ARCH model mitigate these movements. In the internet bubble, the opposite effect is observed. The market volatility is high but so are idiosyncratic volatilities so the correlations are low. The FACTOR ARCH model cannot do this. The observation of Campbell et al( ). that idiosyncratic volatilities are rising, should not be interpreted as a trend but rather as a process that ultimately reverses in about 2002.

The cross sectional standard deviations of these three estimators are interesting. They are shown in Figure 5.

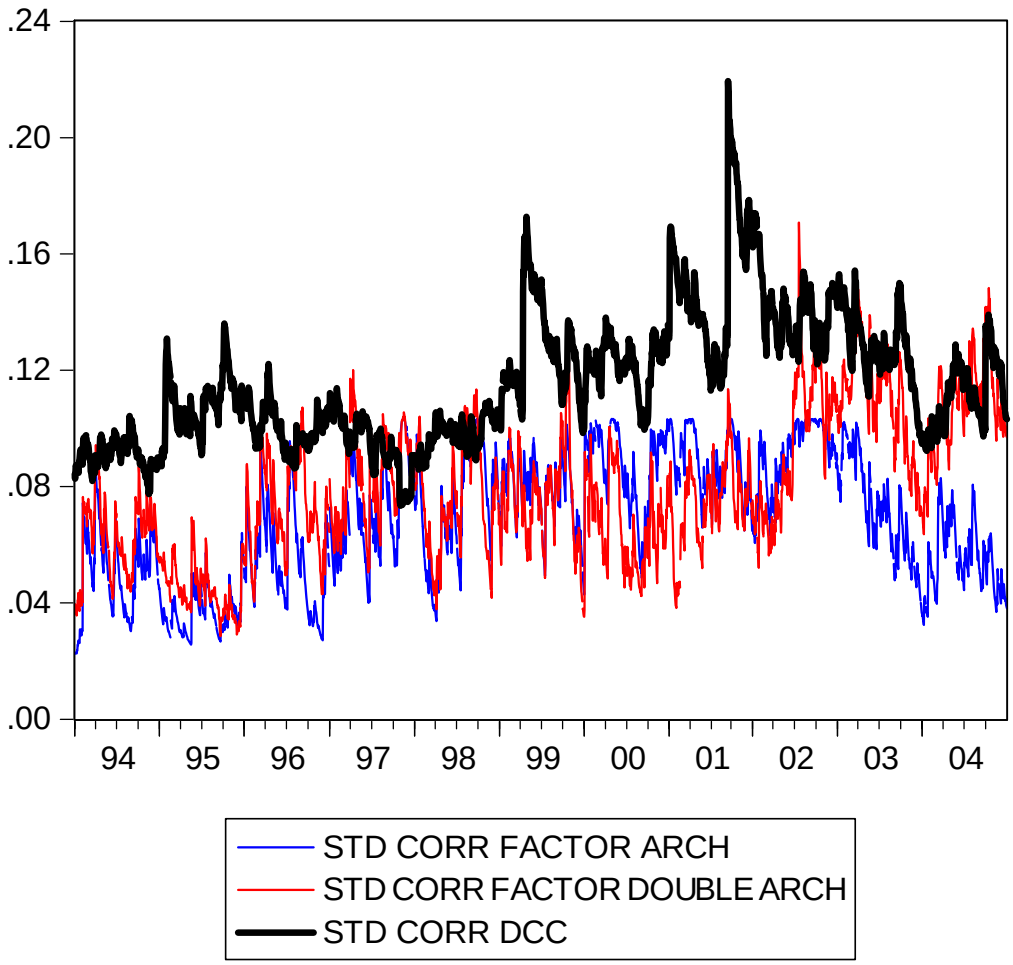

Figure 5

Cross Sectional Standard Deviations of Factor Correlations

From the figure it is clear that the DCC has correlations that differ more across pairs than the two factor models. Perhaps this is not surprising as the component due to the factor is the same for all pairs in the Factor models, whereas each pair has its own time series in the DCC. Thus the FACTOR DOUBLE ARCH model is more volatile over time but the DCC is more volatile cross sectionally. It remains to be seen whether this is a good or a bad feature of the models.

3. FACTOR DCC

As discussed above, FACTOR DCC simply estimates a DCC model from the residuals of the FACTOR DOUBLE ARCH model following the specification in (23). The MacGyver method is used to estimate the parameters of this DCC. The median alpha =.009 while the median beta =.925. The sum of these two numbers is much farther from unity than the DCC estimates on the simple returns; hence the correlation process is less persistent. Because alpha is smaller, it is also lest volatile.

The residuals from the FACTOR DOUBLE ARCH model should be uncorrelated both conditionally and unconditionally of the one factor model is correct, thus the DCC on these residuals might not find anything. The average residual correlation and its cross sectional standard deviation are shown in figure 6.

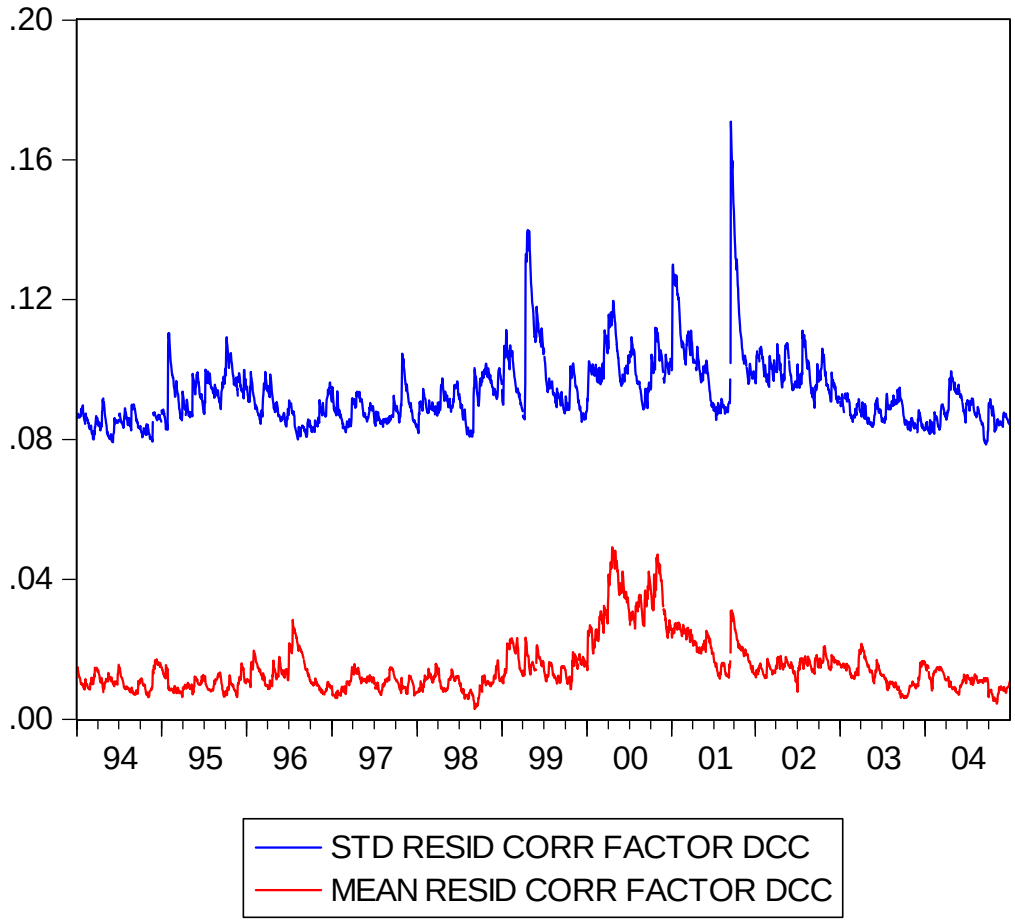

Figure 6

Mean and Standard Deviation of Residual Correlations from FACTOR DCC

The average correlation of the residuals is quite small. It averages .01 over time and cross sectional pairs. It does rise in the middle of the sample but only to .04. The cross sectional standard deviation is however of the same order of magnitude as the cross sectional standard deviation of the DCC although it does not rise quite so much. Thus the average correlation is small but it has substantial cross sectional variability.

When these residual correlations are incorporated into the calculation of the conditional correlations, the result is a substantial change for some pairs and very little for many others. In fact, the average correlation looks almost identical to the FACTOR DOUBLE ARCH. However the cross sectional dispersion is now greater. The cross sectional standard deviation of FACTOR DCC is shown in Figure 7.

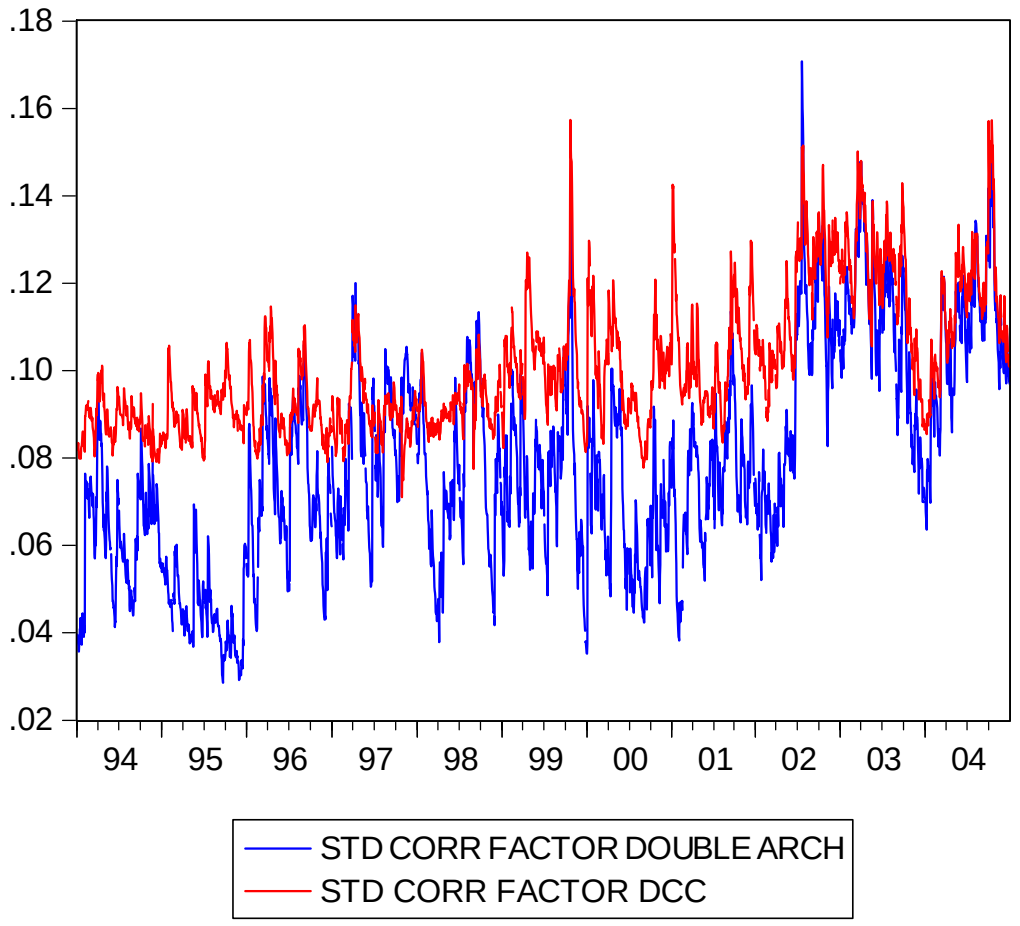

Figure 7

Cross Sectional Standard Deviation of FACTOR Correlations.

The reasons for these differences are easily seen in a few examples shown in the appendix. Stocks in the same industry have idiosyncratic shocks that are correlated. The FACTOR DCC method incorporates these idiosyncratic correlations into the correlation estimates. If these residual correlations are constant, the correction is static but if it is dynamic, then a time varying correction is automatically generated by the FACTOR DCC method. The appendix shows the correlations estimated between International Paper and Caterpillar, between Merck and Johnson and Johnson, and between Coke and Phillip Morris.

4. Hedging Experiment

To establish which of these models does a better job of forecasting correlations, an economic criterion is desirable since we never know what the correlations truly were. A natural criterion is based on portfolio optimization or hedging. This is an example of the methodology introduced by Engle and Colacito(2006). The optimal portfolio of two stocks with equal expected return, is to choose the minimum variance combination. For example, the minimum variance combination of assets (i,j) is given by

$$\begin{aligned} r_{port,t} &= w_t r_{i,t} + (1-w_t) r_{j,t} \\ w_t &= \frac{h_{j,j,t} - h_{i,j,t}}{h_{i,i,t} + h_{j,j,t} - 2h_{i,j,t}}, \; V_{t-1}(r_t) = H_t \end{aligned} \tag{27}$$

Thus the optimal proportion of each asset to hold is changing over time based on the forecast of the covariance matrix. To achieve this optimal holding, the investor would forecast the next day covariance matrix just before the close and then adjust his portfolio to have the weights given in (27). The criterion for success is that the portfolio indeed has a smaller variance than if the weights had not been changed. More generally, this

benefit should cover transaction costs. For the purpose here, we simply want to know which method of forecasting the covariance matrix achieves the lowest variance.

A closely related problem is holding a position in one stock because it has an abnormal expected return and hedging the position with a second stock. Typically this would mean shorting the either the first or second stock to obtain a hedge portfolio with the minimum variance. Although the problem is different, the same approach can be used to solve it. The optimal hedge is given by

$$r_{port,t} = r_{i,t} - \beta_{i,j,t} r_{j,t}, \quad \beta_{i,j,t} = h_{i,j,t} / h_{j,j,t} \tag{28}$$

The criterion for success again is simply the smallest variance of the portfolio.

These two criteria are applied for each of the models we have discussed, to all the pairs of stocks in the data set, and on all the dates in the data set. The average volatility for each pair over time is averaged over all pairs to obtain a single number for the performance of a particular correlation estimator. The results are in Table 1 and two figures in the appendix.

Table 1

Average Volatility of Optimized Portfolios

	CONST	HIST100	DCC	FACTOR ARCH	DOUBLE ARCH	FACTOR DCC
Min Variance	0.015764	0.015889	0.015632	0.015743	0.015628	0.015604
Hedge	0.018989	0.019072	0.018844	0.019026	0.018883	0.018804

The results show that for both criteria, the FACTOR DCC model produces the best hedge portfolio. For the Hedging problem, the DCC is next followed by the FACTOR DOUBLE ARCH while for the minimum variance criterion, the order is the opposite. All except the 100 day historical volatility outperform the optimal constant set of weights and

in the hedging problem, the FACTOR ARCH. The differences are however very small. It appears that the gains from a better model may only be 1% reduction in volatility. This does not however mean that for other problems the gains will also be small. See for example Engle and Colacito(200) for discussion of this.

To determine whether these differences are systematic or not, I looked at how many of the pairs preferred one estimator to another. These winning percentages tell a much stronger story. In the appendix, tables A2 and A3 show the fraction of times the row method beats the column method. The best method has the largest fractions in the labeled row. For example, in hedging, the FACTOR DCC is superior to a constant hedge for 88% of the pairs, and superior to the DCC for 74%. It beats the historical hedge 98%, the FACTOR ARCH 99% and the DOUBLE ARCH 91%. Although the differences are small, they are systematic.

V. Conclusion

Forecasts of correlations are of prime importance for financial decision making. This paper has introduced several new models and compared them with existing models. The use of FACTOR models combined with time series methods such as ARCH/GARCH and DCC provides a rich class of estimators that can approximate a flexible correlation structure. Models called FACTOR ARCH, FACTOR DOUBLE ARCH and FACTOR DCC are introduced and compared. The paper discusses estimation of these models in the context of large systems. A new method for estimating DCC models in large systems

is called the MacGyver method. It first estimates all the bivariate models and then selects the median of all these parameter estimates to use in creating and forecasting correlations.

An economic loss function is used to compare these models on a data set of large cap US stocks. Dynamic portfolio strategies designed to create minimum variance portfolios from pairs of stocks have different performance depending on what model is used for correlations. By averaging over time and pairs of stocks, it is found that the FACTOR DCC method has the best performance. This is followed either by the DCC or the FACTOR DOUBLE ARCH in different circumstances. The standard one factor model FACTOR ARCH is not as good and a 100 day Historical correlation is worse than a constant hedge.

REFERENCES

1. Bauwens,Luc Sébastien Laurent, and Jeroen Rombouts (2006)"Multivariate GARCH models: a survey" *Journal of Applied Econometrics*, 21/1, 79-109**)**

2. Bollerslev, Tim, Robert Engle and Dan Nelson, "ARCH Models," in *Handbook of Econometrics, Volume IV,* ed. R. Engle and D. McFadden (Amsterdam: North Holland, 1994), 2959-3038.

3. Bollerslev, Tim and Jeffrey M. Wooldridge "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances, (1992), *Econometric Reviews* 11.2 143-172.

4. Campbell John, Martin Lettau, Burton Malkiel, and Yexiao Xu, (2001), "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk", *Journal of Finance*, LVI(1)

5. Cappiello, Lorenzo, Robert Engle and Kevin Sheppard,(2007) "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *forthcoming Journal of Financial Econometrics*

6. Engle, Robert, (2002), "Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models, *Journal of Business and Economic Statistics*, July: 339-350*,* V20N3

7. Engle, Robert, *Anticipating Correlations,* Princeton University Press. Forthcoming

8. Engle, Robert and Riccardo Colacito,(2006) Testing and Valuing Dynamic Correlations for Asset Allocation, *Journal of Business and Economic Statistics,* Vol 24, N2, pp 238-253.

9. Engle, Robert David F. Hendry and Jean-Francois Richard,(1983), "Exogeneity," *Econometrica* 51 (1983): 277-304.

10. Engle, Robert and Ken Kroner, (1995) "Multivariate Simultaneous GARCH," *Econometric Theory* 11 122-150.

11. Engle, Robert and Joe Mezrich "GARCH for Groups," Risk (1996): 36-40.

12. Engle, Robert, Victor Ng and Michael Rothschild(1990)"Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics* 45: 213-237.

13. Engle, Robert, Victor Ng and Michael Rothschild,(1992) "A Multi-Dynamic Factor Model for Stock Returns," *Journal of Econometrics* 52 : 245-266.

14. "Time Varying Volatility and the Dynamic Behavior of the Term Structure," (with V. Ng), *Journal of Money, Credit and Banking 25* (1993): 336-349.

15. Engle, Robert and Kevin Sheppard, (2005)Theoretical properties of Dynamic Conditional Correlation Multivariate GARCH, manuscript

16. Engle, Robert and Kevin Sheppard, (2005a) Evaluating the Specification of Covariance Models for Large Portfolios, manuscript

17. Engle, Robert and R. Susmel(1993),"Common Volatility in International Equity Markets," *Journal of Business and Economic Statistics* 11: 167 - 176.
18. Nelson, Dan and Dean Foster,(1994) “Asymptotic Filtering Theory for Univariate ARCH Models”, *Econometrica,* 1, 1-41.
19. Taleb, Nassim,(2007) *The Black Swan: The Impact of the Highly Improbable,* Random House

APPENDIX

Table A1
Experiments for MacGyver Simulation

	Number	Alpha	Beta	Rhobar
Exp1	3.00	0.05	0.90	0.50
Exp2	5.00	0.05	0.90	0.50
Exp3	10.00	0.05	0.90	0.50
Exp4	20.00	0.05	0.90	0.50
Exp5	30.00	0.05	0.90	0.50
Exp6	50.00	0.05	0.90	0.50
Exp7	10.00	0.05	0.94	0.50
Exp8	10.00	0.02	0.97	0.50
Exp9	10.00	0.05	0.90	0.20
Exp10	10.00	0.05	0.90	0.80

Table A2

RMS Errors from MacGyver Method

Alpha

	amean	a_rmean	ameant05	a_rmeant05	amed	a_rmed
exp1	**0.01092**	0.01233	0.01150	0.01386	0.01150	0.01386
exp2	0.00720	0.00807	0.00710	0.00854	**0.00682**	0.00843
exp3	0.02550	0.00543	0.00528	0.00504	**0.00491**	0.00510
exp4	0.01509	0.00419	0.00438	0.00383	0.00410	**0.00357**
exp5	5.63324	0.00403	0.00389	0.00361	0.00358	**0.00321**
exp6	0.01088	0.00352	0.00301	0.00308	**0.00274**	0.00275
exp7	0.00578	0.00380	0.00430	**0.00371**	0.00424	0.00379
exp8	471.66375	0.00552	0.00377	0.00391	**0.00283**	0.00295
exp9	10.58470	0.00469	0.00424	0.00432	**0.00402**	0.00441
exp10	0.00554	0.00498	0.00542	**0.00483**	0.00539	0.00516
average	48.79626	0.00566	0.00529	0.00547	**0.00501**	0.00532

Beta

	bmean	b_rmean	bmeant05	b_rmeant05	bmed	b_rmed
exp1	0.03743	0.05134	0.03417	0.03584	**0.03417**	0.03584
exp2	0.03693	0.03630	0.02716	0.02910	**0.02186**	0.02438
exp3	1.20351	0.02736	0.02080	0.02040	0.01629	**0.01515**
exp4	0.21512	0.02543	0.01796	0.01729	0.01289	**0.01179**
exp5	0.07398	0.02388	0.01542	0.01592	**0.01041**	0.01047
exp6	0.14473	0.02400	0.01490	0.01578	**0.00969**	0.01022
exp7	0.01311	0.00819	0.00705	0.00747	**0.00578**	0.00703
exp8	88.71563	0.05939	0.06618	0.03728	**0.01015**	0.01178
exp9	15.72034	0.03100	0.02297	0.02220	0.01510	**0.01449**
exp10	0.02063	0.01894	0.01585	0.01453	0.01361	**0.01177**
average	10.61814	0.03058	0.02425	0.02158	**0.01499**	0.01529

Table A3
Bias from MacGyver Method

Alpha

	amean	a_rmean	ameant05	a_rmeant05	amed	a_rmed
exp1	0.00071	0.00110	0.00024	0.00094	0.00024	0.00094
exp2	0.00124	0.00285	0.00077	0.00210	0.00079	0.00153
exp3	-0.00036	0.00261	0.00184	0.00195	0.00082	0.00106
exp4	0.00352	0.00254	0.00222	0.00188	0.00142	0.00088
exp5	-0.55754	0.00248	0.00183	0.00180	0.00099	0.00095
exp6	0.00096	0.00224	0.00152	0.00152	0.00071	0.00067
exp7	0.00272	0.00158	0.00214	0.00134	0.00134	0.00092
exp8	-48.11060	0.00428	0.00158	0.00298	0.00142	0.00174
exp9	-1.05017	0.00251	0.00208	0.00192	0.00072	0.00113
exp10	0.00247	0.00233	0.00215	0.00196	0.00183	0.00170
average	-4.97070	0.00245	0.00164	0.00184	**0.00103**	0.00115

Beta

	bmean	b_rmean	bmeant05	b_rmeant05	bmed	b_rmed
exp1	-0.01710	-0.02087	-0.01165	-0.01022	-0.01165	-0.01022
exp2	-0.01945	-0.02216	-0.01450	-0.01674	-0.00942	-0.01079
exp3	0.09886	-0.02275	-0.01615	-0.01605	-0.01036	-0.00933
exp4	-0.01104	-0.02333	-0.01524	-0.01518	-0.00951	-0.00900
exp5	-0.02231	-0.02257	-0.01333	-0.01429	-0.00741	-0.00811
exp6	-0.00020	-0.02264	-0.01344	-0.01414	-0.00765	-0.00789
exp7	0.00175	-0.00661	-0.00062	-0.00577	-0.00252	-0.00479
exp8	11.18001	-0.05195	-0.01487	-0.02937	-0.00859	-0.00998
exp9	1.54597	-0.02653	-0.01931	-0.01814	-0.01033	-0.00994
exp10	-0.01524	-0.01525	-0.01106	-0.01087	-0.00803	-0.00666
average	1.27412	-0.02347	-0.01302	-0.01508	**-0.00855**	-0.00867

Table A4

Fraction of Minimum Variance Portfolios Where Row Beats Column

	CONST	HIST 100	DCC	FACTOR ARCH	DOUBLE ARCH	FACTOR DCC
CONST	0.000	0.778	0.235	0.451	0.196	0.163
HIST 100	0.222	0.000	0.000	0.196	0.013	0.007
DCC	0.765	1.000	0.000	0.791	0.464	0.261
FACTOR ARCH	0.549	0.804	0.209	0.000	0.137	0.137
DOUBLE ARCH	0.804	0.987	0.536	0.863	0.000	0.183
FACTOR DCC	0.837	0.993	0.739	0.863	0.817	0.000

Table A5

Fraction of Hedges Where Row Beats Column

	CONST	HIST 100	DCC	FACTOR ARCH	DOUBLE ARCH	FACTOR DCC
CONST	0.000	0.660	0.154	0.634	0.258	0.114
HIST 100	0.340	0.000	0.013	0.379	0.121	0.013
DCC	0.846	0.987	0.000	0.899	0.569	0.255
FACTOR ARCH	0.366	0.621	0.101	0.000	0.059	0.007
DOUBLE ARCH	0.742	0.879	0.431	0.941	0.000	0.088
FACTOR DCC	0.886	0.987	0.745	0.993	0.912	0.000

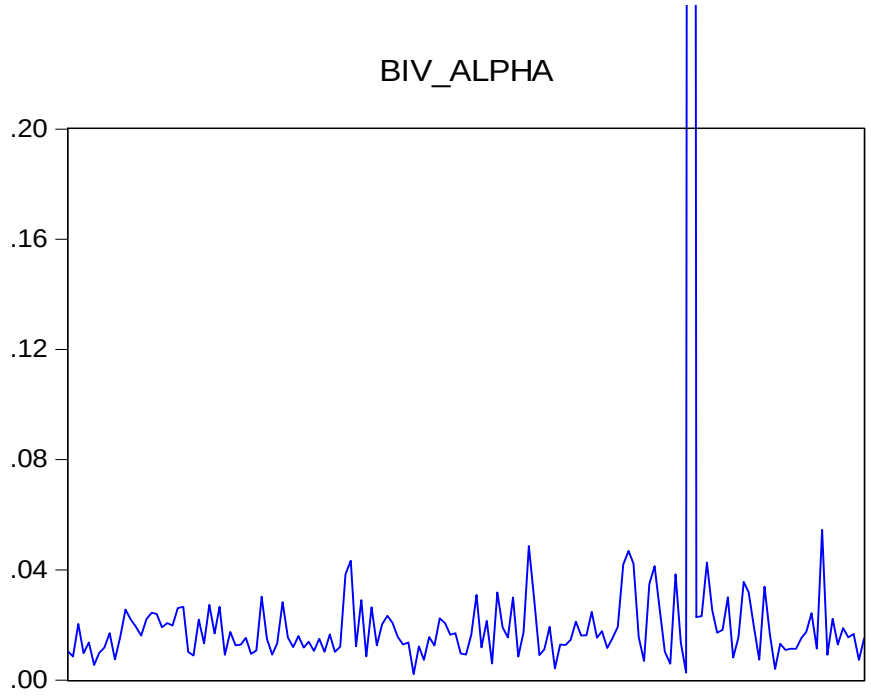

Figure A1
Estimated Alphas from Bivariate estimates in MacGyver Method

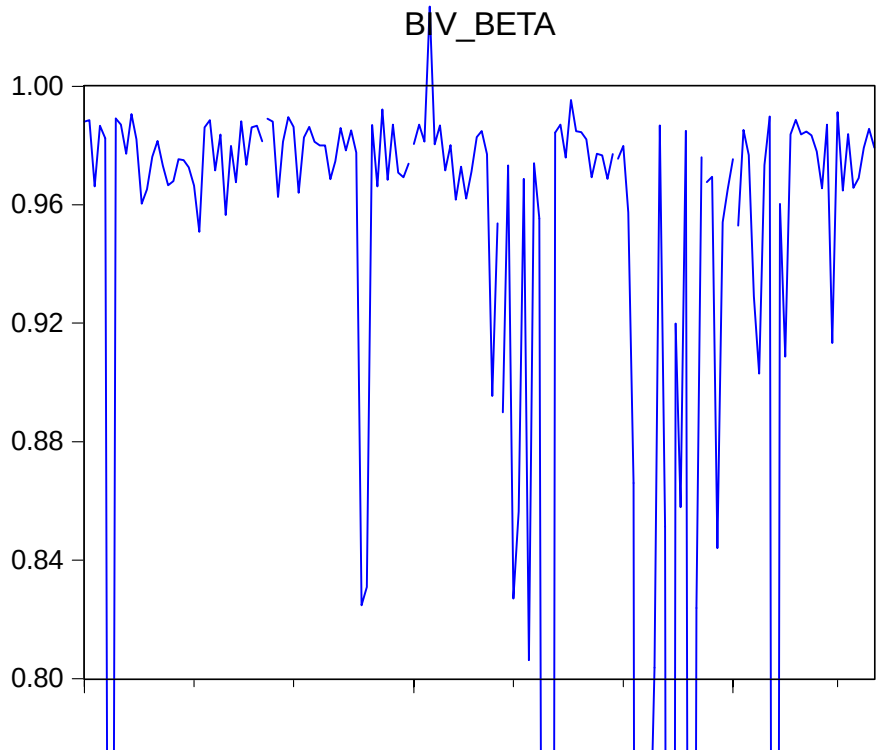

Figure A2

Estimated Betas from Bivariate estimates in MacGyver Method

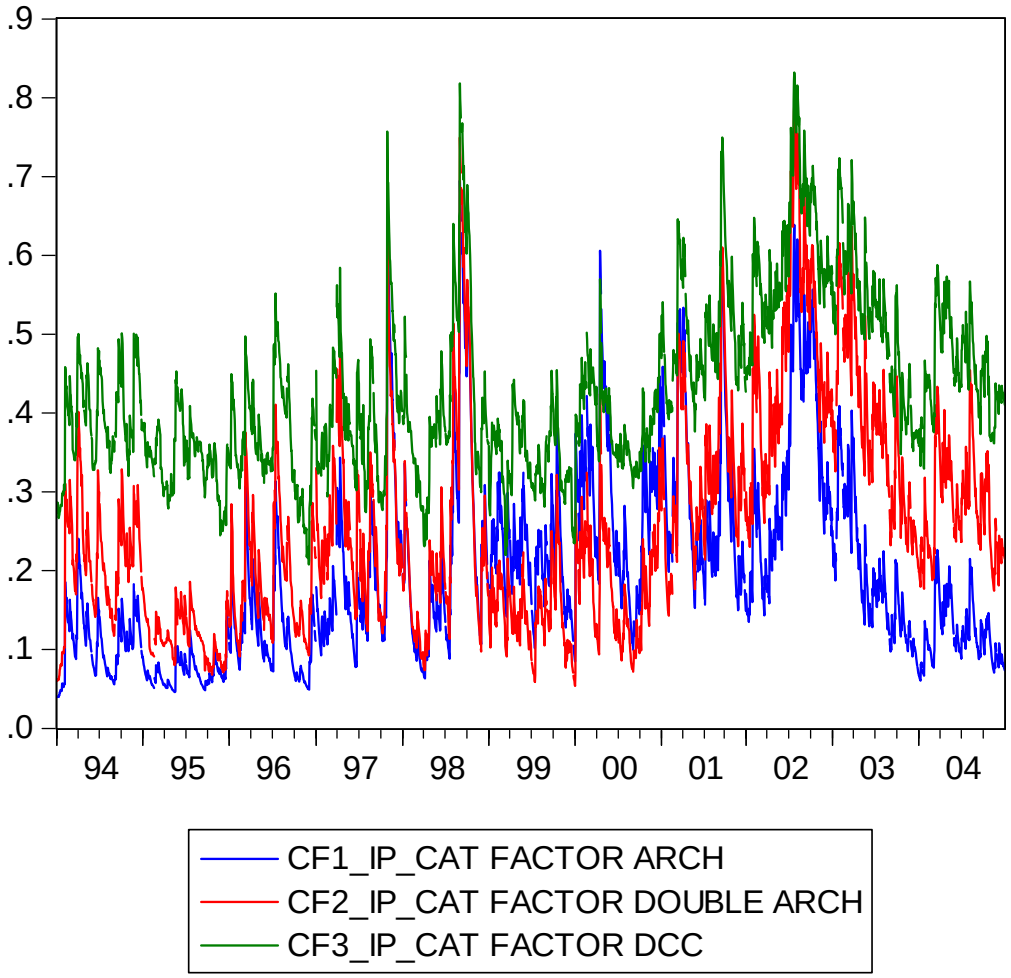

Figure A3

Correlations between International Paper Catepillar by several Methods

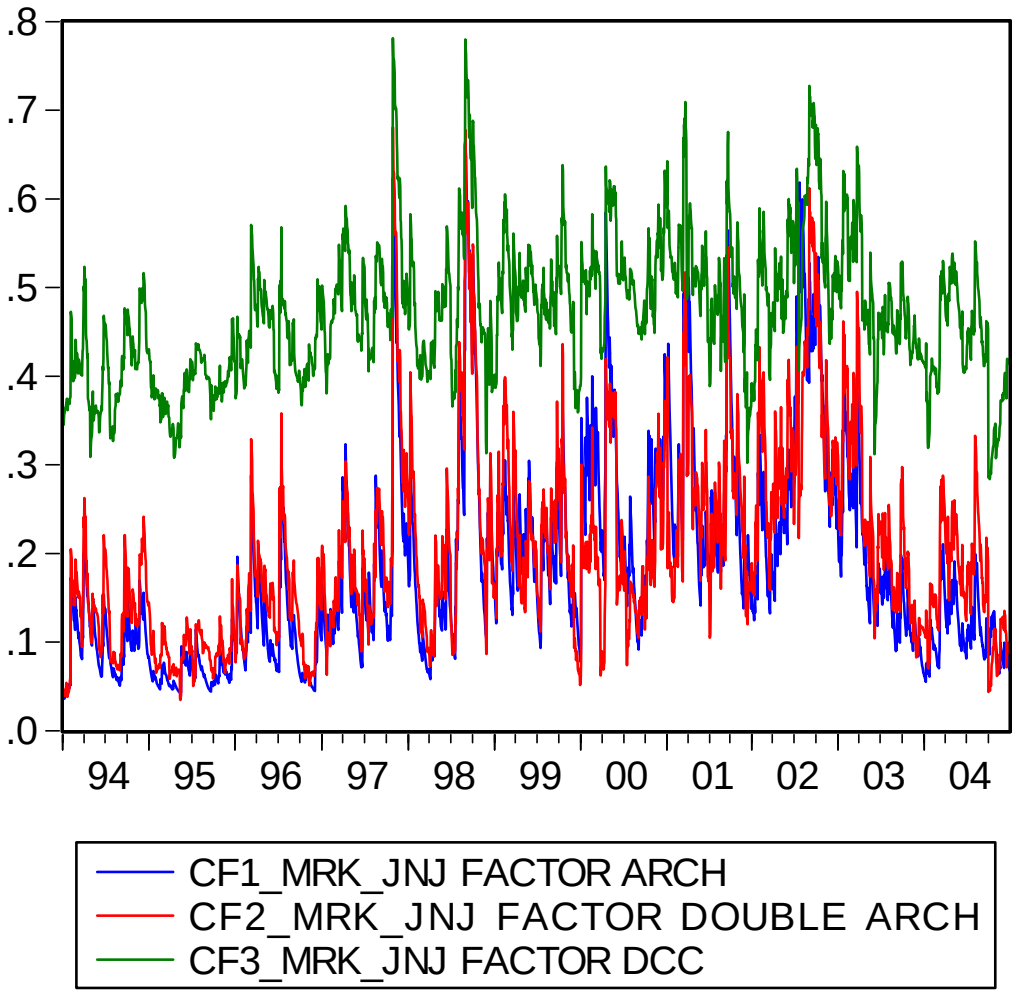

Figure A4

Correlations between Merck and Johnson and Johnson by various methods

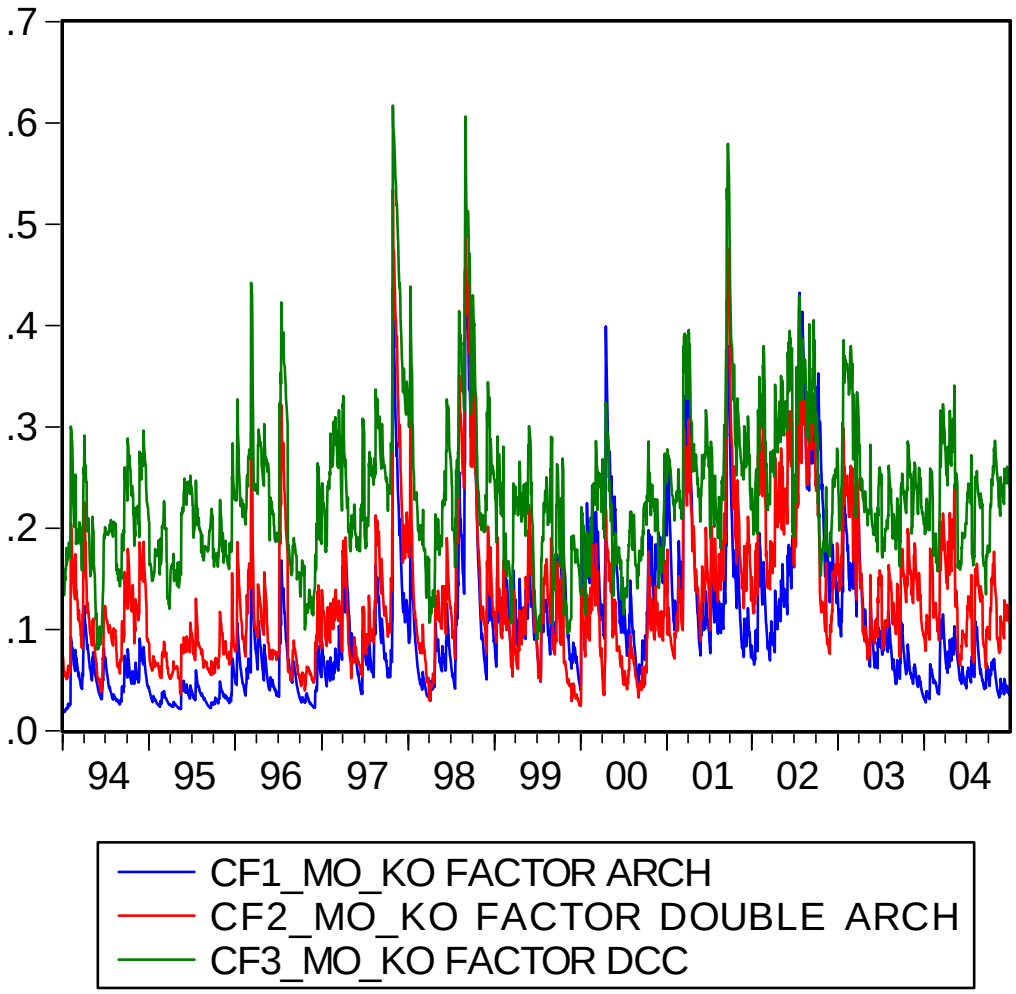

Figure A5

Correlations between Phillip Morris and Coca Cola by various Methods

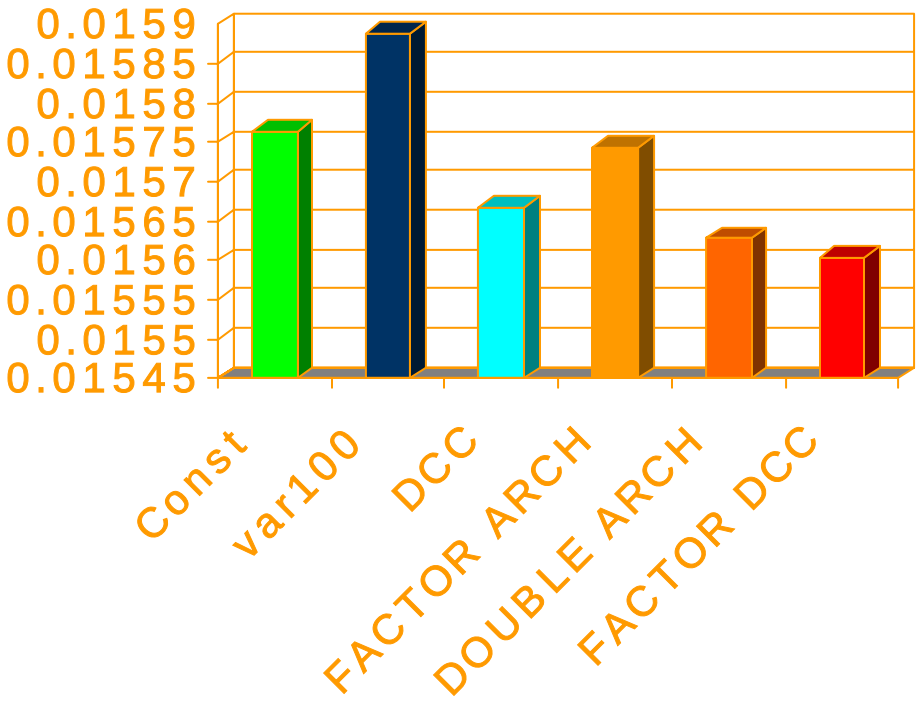

Figure A6

Average Performance of Minimum Variance Portfolios

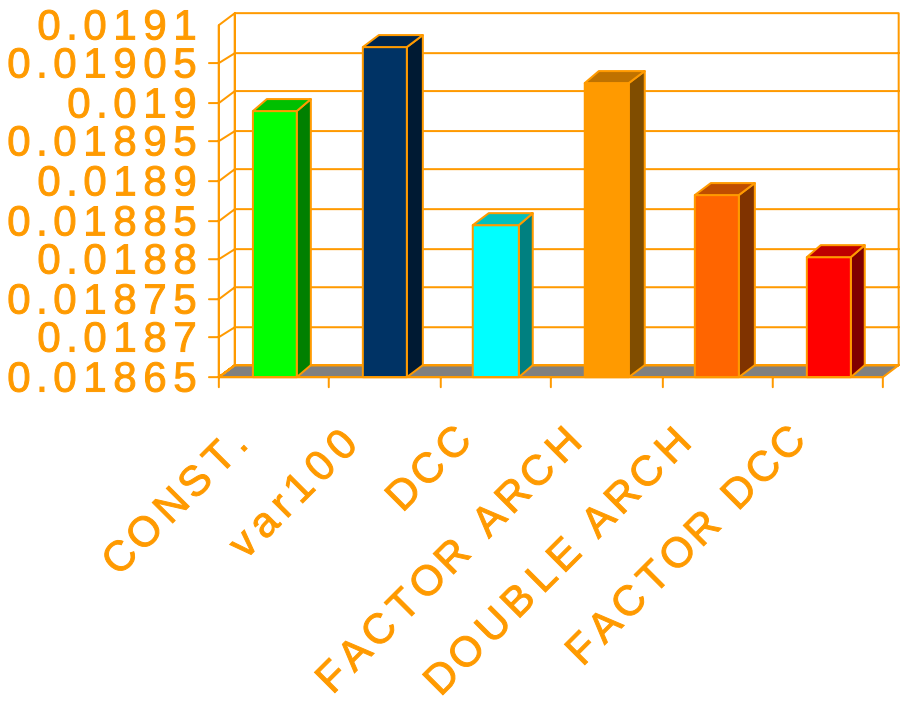

Figure A7

Average Volatility of Optimal Hedge Portfolios