AN INDEX OF CO-MOVEMENTS IN FINANCIAL TIME SERIES

by

Enrique Sentana and
Mushtaq Shah

DISCUSSION PAPER NO. 193

LSE FINANCIAL MARKETS GROUP

DISCUSSION PAPER SERIES

July 1994

Enrique Sentana is Professor of Economics at the Centro de Estudios Monetarios y Financieros (CEMFI) and an Associate Member of the Financial Markets Group. Mushtaq Shah is a Researcher with Goldman Sachs International (London) and was a Research Assistant at the Financial Markets Group. Any opinions expressed here are those of the authors and not necessarily those of the Financial Markets Group.

AN INDEX OF CO-MOVEMENTS IN FINANCIAL TIME SERIES

Enrique Sentana

(CEMFI, LSE Financial Markets Group and CEPR)

Mushtaq Shah

(Goldman Sachs International, London)

October 1992

revised: July 1994

We would like to thank Manuel Arellano, John Campbell, Greg Connor, Frank Diebold, Steve Durlauf, Rob Engle and Danny Quah for useful discussions, as well as seminar participants at CEMFI, LSE and the 1993 European Congress of the Psychometric Society. Of course the usual caveat applies. Financial support from the ESRC and the LSE Financial Markets Group is gratefully acknowledged. Correspondence to Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain

ABSTRACT

Financial indices are constructed to capture the strong common variation in a large number of financial time series. Often, these measures are also of interest themselves since they can be related to important underlying economic concepts. We employ multivariate statistical theory to define indices that exploit the comovement within the data. Specifically, we advocate the use of the GLS representing portfolio obtained from the best (in the maximum likelihood sense) one factor exact approximation to the covariance matrix of the series. We show that an index constructed in this way is robust to distribution assumptions and the true underlying structure of the data. We apply our techniques to construct a sensible new summary measure of exchange rate comovements.

1. Introduction

Casual observation suggests that many financial series move closely together over time, at least in the short run. Bilateral exchange rates against one specified currency, interest rates on bonds with different maturities, and share prices for different companies trading on the same stock market constitute obvious examples. In many cases of interest the number of series involved is typically rather large, and summary measures are usually provided by means of index variables in an attempt to capture their common variation (e.g. stock market indices, bond indices or effective exchange rates).

Often, these measures are also of interest themselves since they can be related to important underlying economic concepts. For instance, value weighted stock market indices are usually employed to represent "the market" in empirical work related to the Capital Asset Pricing Model (CAPM). Similarly, trade weighted effective exchange rates are usually analysed in relation to the effect of the currency movements on the country's balance of payments. Nevertheless, it has long been recognised that these indices are not without disadvantages. First, and from a merely descriptive point of view, they tend to be rather sensitive to the weightings of their components, and do not always adequately reflect the perceived comovements in the series. For this reason, alternative index measures are also computed, e.g. equally weighted stock price indices. Second, and from an economic point of view, they are only approximations to the theoretical concepts that they are supposed to represent, and in some cases not necessarily very good ones. For instance, the weights used to compute effective exchange rates are estimates derived from a macroeconometric model, which typically concentrate on the current account and ignore the very important effects of international capital movements. In the case of the stock market, the Roll (1977) critique that the set of assets available to investors includes many others besides stocks also applies. Thirdly, if the dynamic structure characterising the prices of many financial assets can be represented by a full-rank system of integrated series, as the empirical evidence seems to suggest, alternative indices, with however similar weights, will not be cointegrated, and hence would diverge in the long run.

Given this background, it is worth thinking carefully about what the best summary measure of the movements in the different series is, in some well defined sense. Our contribution in this paper is to employ standard multivariate statistical theory, and in particular principal components and factor analysis, in order to answer this question on the basis of the contemporaneous unconditional covariance matrix of the series. The rationale comes from the fact that even though such a matrix contains little information about serial correlation or higher order dynamics, it has much to say about cross-sectional correlation. Besides, in the empirically not irrelevant case in which changes in the series are a constant plus multivariate white noise, the contemporaneous covariance matrix contains all the relevant information about unconditional second moments.

Specifically, we advocate the use of the first Generalised Least Squares (GLS) based factor representing portfolio obtained from the best (in the Maximum likelihood sense) one factor exact approximation to the variance-covariance matrix of the series. We show that this index is similar in spirit to the first principal component of the data, but argue that it is more adequate for capturing the common variation in the series since a factor model always explains variances perfectly, and hence can concentrate on covariances. Importantly, our results do not impose any specific assumptions on moments or distributions other than covariance stationarity. For instance, we do not assume that the data is generated as a realization of a i.i.d. multivariate normal variate with constant mean and an exact one factor representation for its covariance matrix. This is important since nowadays most researches accept that financial data is somewhat skewed, mildly serially correlated, substantially leptokurtic, and with strong nonlinear dependence as measured e.g. by the autocorrelations for the squares (see e.g. Boothe and Glassman (1987)).

The paper is divided as follows. In section 2 we discuss principal components based methods, while in section 3 we look at the statistical underpinnings of our proposed factor analytic method. In section 4 we carry out empirical applications to nominal bilateral exchange rates against the US dollar and the British pound. Finally, section 5 concludes. Proofs are gathered in the appendix.

2. Principal Component Methods for Measuring Co-movements in Financial Time Series

Let's consider a vector of M random variables, $\mathbf{y}_t$ (t=1,2,...) with M potentially large, whose first differences, $\Delta\mathbf{y}_t=\mathbf{x}_t$, follow a covariance stationary multivariate stochastic process with mean α, full-rank covariance matrix Σ and spectral density matrix $\mathbf{f}(\omega)$. For simplicity, we assume that $\mathbf{y}_0$ is fixed and equal to 0. An empirically not irrelevant example of such a process is a multivariate random walk with drift, which is obtained when $(\mathbf{x}_t-\alpha)$ is multivariate white noise.

Our objective is to find a linear combination of $\mathbf{x}_t$ such that it explains most of its (co-)movements. In the time series literature there is a standard solution to this dimension reduction problem based on the spectral representation of $\mathbf{x}_t$:

$$x_t = \int_{-\pi}^{\pi} e^{it\omega}\, dX(\omega)$$

where

$$E[dX(\omega)d\bar{X}(\omega)] = \mathbf{f}(\omega)$$

If our objective function is to approximate $E[(x_t-\alpha)'(x_t-\alpha)]=tr(\Sigma)$ as closely as possible by means of $V(v_t)$, where v_t is some linear combination of $\mathbf{x}_t$, then we should choose

$$v_t^S = \int_{-\pi}^{\pi} e^{it\omega}\, q_1'(\omega)dX(\omega)$$

where $q_1(\omega)$ is the first eigenvector of $\mathbf{f}(\omega)$ (see e.g. Priestley (1981))[1].

[1] In the time domain, we can construct v_t^S as $\sum_{r=1}^{M}\sum_{s=-\infty}^{\infty} p_{rs}x_{rt-s}$, where

Then we can obtain a cumulative index z_t^S to represent y_t as:

$$z_t^S = \sum_{s=1}^{t} v_s^S$$

But since v_t^S is generally a linear combination of present, past and <u>future</u> values of $\mathbf{x}_t$ (see footnote 1), it cannot be used on an on-line basis to describe the comovements of the series as they happen. For financial market participants, this is a major disadvantage.

The cointegration literature suggests an alternative related procedure, based solely on the spectral density matrix of $\mathbf{x}_t$ at frequency zero, f(0). The rationale is that if $\mathbf{y}_t$ is cointegrated with M-1 cointegrating vectors, then the only common trend will be given by:

$$z_t^{CT} = q_1'(0)\mathbf{y}_t = \sum_{s=1}^{t} q_1'(0)\mathbf{x}_s$$

There are two potential problems with this measure. First, it is not clear why the zero frequency should always be the right one to concentrate on in order to obtain an index, since z_t^{CT} is usually extremely smooth and ignores by construction all the short run movements. Second, the empirical evidence suggests that for financial variables **f**(0) may well have full rank , so that the series do not seem to be cointegrated despite the fact that they move very closely together in the "short run"[2]. For example, Diebold, Gardeazabal and Yilmaz (1992) find that this is the case for exchange rates.

Incidentally, note that the lack of cointegration makes our task even more complicated as it implies that any two linear combinations of the y_t's, however similar, are not going to be cointegrated, and would diverge in the long run. However, it may well be the case that the correlation between changes in the two indices is very high, so they could both provide very good measures of the changes in the series.

As a third alternative, we propose here cumulative indices related to the first principal component of $\mathbf{x}_t$, appropriately standardised so that the weights add up to one. The rationale comes from the fact that even though Σ contains little information about serial correlation, it has much to say about cross-sectional correlation. Besides, in the empirically not irrelevant case in which $(\mathbf{x}_t-\alpha)$ is multivariate white noise, $\mathbf{f}(\lambda)=0$ for $\lambda\neq 0$, and $\mathbf{f}(0)=\Sigma$, and all three measures coincide.

If we write the spectral decomposition of Σ as:

$$\Sigma = QMQ' = (q_1 \ Q_2)\begin{bmatrix} \mu_1 & 0 \\ 0 & M_1 \end{bmatrix}\begin{bmatrix} q_1' \\ Q_2' \end{bmatrix}$$

with M diagonal and Q orthonormal, and call $v_s^P=(q_1'\ell)^{-1}q_1'\mathbf{x}_s$, where ℓ is a vector of M ones, then the principal component based index will be given by $z_t^P = \sum_{s=1}^{t} v_s^P$[3].

There are two well known complementary interpretations of the first principal component. First, $q_1'\mathbf{x}_t$ maximises $V(\mathbf{w}'\mathbf{x}_t)$ subject to $\mathbf{w}'\mathbf{w}=1$. Hence, z_t^P is the index variable that explains most of the variances and covariances in the changes of the series.

An alternative interpretation of v_s^P can be obtained if we project $\mathbf{x}_t$ on (a constant and) $\lambda'\mathbf{x}_t$, where λ is a Mx1 vector. From least squares

---

$p_{rs} = 1/2\pi \int_{-\pi}^{\pi} q_{1r}(\omega)e^{i\omega s}d\omega.$

[2] This possibility is behind the concept of common features recently introduced by Engle and Korizky (1993).

[3] We could also work with the first re-scaled principal component of $\mathbf{y}_t$ directly, but the problem is that $V(\mathbf{y}_t)$ is O(t) unless f(0)=0. Hence, to avoid unnecessary complications when the sample size goes to infinity, we prefer to define our indices as cumulative ones. Note that if $(\mathbf{x}_t-\alpha)$ is white noise, $V(\mathbf{y}_t)=t\Sigma$, and it makes no difference whether we work with $\mathbf{y}_t$ or $\mathbf{x}_t$.

regression theory we can decompose $\mathbf{x}_t$ into:

$$\mathbf{x}_t = \hat{\mathbf{x}}_t(\lambda) + \hat{\varepsilon}_t(\lambda)$$

where:

$$\hat{\mathbf{x}}_t(\lambda) = \alpha + b(\lambda)(\lambda'\mathbf{x}_t - \lambda'\alpha)$$
$$b(\lambda) = \Sigma\lambda(\lambda'\Sigma\lambda)^{-1}$$
$$V[\hat{\mathbf{x}}_t(\lambda)] = \Sigma_1(\lambda) = \Sigma\lambda(\lambda'\Sigma\lambda)^{-1}\lambda'\Sigma$$
$$V[\hat{\varepsilon}_t(\lambda)] = \Sigma_2(\lambda) = \Sigma - \Sigma\lambda(\lambda\Sigma\lambda)^{-1}\lambda\Sigma$$
$$\text{cov}[\hat{\mathbf{x}}_t(\lambda), \hat{\varepsilon}_t(\lambda)] = 0$$

In particular, $b(q_1)=q_1$, $\Sigma_1(q_1)=\mu_1 q_1 q_1'$ and $\Sigma_2(q_1)=Q_2 M_2 Q_2'$. Since $\Sigma_1(q_1)$ is the rank 1 symmetric matrix that best approximates Σ in the Frobenius norm[4] (see Magnus and Neudecker (1988)), what the first principal component does is to find the linear combination of $\mathbf{x}_t$ that minimises the (norm of the) residual variance $\Sigma_2(\lambda)$, so that if we replace $\mathbf{x}_t$ by its fitted values from the above regression, the covariance matrix of the fitted values is closest to Σ.

The first principal component, however, may have some drawbacks as a measure of the <u>co</u>movements between the variables. This is especially true if we are more interested in explaining covariances than variances. As an extreme example consider the case in which the changes in M-1 of the variables are highly correlated with each other but the M-th one is uncorrelated with the rest. As σ_{MM} grows, the principal component, v_t^P, converges to x_{Mt}, since most of the variation in the data will be given by this variable. This problem is well known, and to alleviate it many researches suggest the use of the principal components obtained from the spectral decomposition of the correlation matrix $\mathbf{R}=dg^{-1}(\Sigma)\Sigma dg^{-1}(\Sigma)$. The reason is that principal components are scale-dependent, and hence they differ as one changes the scaling of the variables. Notice, though, that the (reweighted) principal component, v_t^{RP}, still will have a non-zero weight on the M-th variable, but not a disproportionate one. In fact, if the number of series $M \to \infty$, then the weight on the M-th variable will

[4] That is, $\Sigma_1(q_1)=\mu_1^{1/2}q_1(\mu_1^{1/2}q_1)'$ minimises $tr(\Sigma-\delta\delta')^2 = \sum_{i=1}^{M}\sum_{j=1}^{M}(\sigma_{ij}-\delta_i\delta_j)^2$.

converge to zero.

3. Factor Analytic Methods for Measuring Co-movements in Financial Time Series

Nevertheless, there is an alternative index variable that in such a situation will indeed impose a weight of zero on the M-th variable for any finite M>2. It is given by the cumulative index obtained from the first Generalised Least Squares based factor representing portfolio (appropriately scaled so that the weights add up to one) obtained from the best (in the Maximum likelihood sense) one factor exact approximation to Σ. We shall define this measure in section 3.3.

Its rationale, though, depends on several properties of pseudo maximum likelihood estimates of factor models, which we previously discuss in section 3.2. Importantly, such properties do not depend on any specific distributional assumptions on $\mathbf{x}_t$. In particular, they do not require that $\mathbf{x}_t$ is i.i.d. Gaussian or that Σ has an exact one factor representation. This feature of our proposed index is particularly important in our case since most empirical studies using financial data find that such data is far from being normal or i.i.d. For instance, it is often found to be somewhat skewed, mildly serially correlated, substantially leptokurtic, and with strong dependence in its squares or absolute values (see e.g. Boothe and Glassman (1987)) .

3.1. Numerical Properties of Pseudo-Maximum Likelihood Estimators of Exact Zero Factor Models

Although our interest lies in models with one factor, it is convenient to consider first the case of an exact zero factor model in order to illustrate the different issues involved. Suppose that we have a sample of T observations on the Mx1 random vector $\mathbf{x}_t$, and would like to fit by pseudo-maximum likelihood the model $\mathbf{x}_t \sim$ i.i.d. $N(a,\Gamma)$, where a is a Mx1 vector and Γ a diagonal MxM matrix with typical element γ_j. For such a model to make sense, Γ has to be positive semi-definite (i.e. $\gamma_j \geq 0$ $\forall j$).

Hence, the admissible parameter space is $\mathbb{R}^M \times \mathbb{R}^M_+$, where $\mathbb{R}^M_+$ is the non-negative orthant of $\mathbb{R}^M$. For those parameter values for which Γ is not singular, our objective function will be proportional to:

$$L_T(a,\Gamma|x_1,..,x_T) = -M/2 \log 2\pi - 1/2 \log|\Gamma| -1/2T \sum_{t=1}^{T} (x_t-a)'\Gamma^{-1}(x_t-a) =$$
$$= -M/2 \log 2\pi - 1/2 \log|\Gamma| -1/2 \, tr\{\Gamma^{-1}[S_T+(\bar{x}_T-a)(\bar{x}_T-a)']\} = L(a,\Gamma|\bar{x}_T,S_T)$$

where $\bar{x}_T=1/T \sum_{t=1}^{T} x_t$ is the sample mean and $S_T = 1/T \sum_{t=1}^{T} (x_t-\bar{x}_T)(x_t-\bar{x}_T)'$ the unrestricted pseudo ML estimate of Σ, the covariance matrix of x_t.

Given that $\bar{x}_T$ would be the pseudo-ML estimate of a, we can concentrate the log-likelihood function to:

$$L^c(\Gamma|S_T) = -M/2 \log 2\pi - 1/2 \log|\Gamma| - 1/2 \, tr[\Gamma^{-1}S_T]$$

In view of the diagonality of Γ, the concentrated log-likelihood function can also be written as:

$$L^c(\Gamma|S_T) = \sum_{j=1}^{M} [-1/2 \log 2\pi - 1/2 \log|\gamma_j| - 1/2 \, s_{jjT}/\gamma_j]$$

where $s_{jjT}=1/T \sum_{t=1}^{T} (x_{jt}-\bar{x}_{jT})^2$. From the first order conditions, it is clear that $\tilde{\gamma}_{jT}=s_{jjT}\geq 0$, as expected.

For completeness, we will analyse under what circumstances the pseudo-ML estimator involves a singular Γ. Although $L^c(\Gamma|S_T)$ breaks down for Γ singular, its value does not automatically become $+\infty$. Actually, it is straightforward to prove that $- 1/2 \log|\gamma_j| - 1/2 \, s_{jjT}/\gamma_j$ goes to $-\infty$ when γ_j tends to 0 for $s_{jjT}\neq 0$. Hence, $\tilde{\gamma}_{jT}=0$ is only plausible if $s_{jjT}=0$. Therefore, S_T singular is a necessary condition for $\tilde{\Gamma}_T$ singular, but it is far from sufficient. In fact, only those singularities in the data for which $dg(S_T)$ does not have full rank will produce a singular $\tilde{\Gamma}_T$. In particular, $\tilde{\Gamma}_T$ will still have full rank when the number of series exceeds the number of observations.

Thus, the pseudo-ML estimators of the exact zero factor model will be $\tilde{a}_T=\bar{x}_T$ and $\tilde{\Gamma}_T=dg(S_T)$, regardless of x_t being Gaussian, independent or identically distributed, regardless of whether Σ is diagonal, and regardless of S_T being singular.

Using the analogy principle (see Manski (1988)), we can understand $\tilde{a}_T$ and $\tilde{\Gamma}_T$ as estimators of population characteristics $\tilde{a}$ and $\tilde{\Gamma}$ defined as:

$$\arg\max_{a,\Gamma} E[-M/2 \log 2\pi -1/2 \log|\Gamma| -1/2 \, (x_t-a)'\Gamma^{-1}(x_t-a)] = \arg\max_{a,\Gamma} L(a,\Gamma|\alpha,\Sigma)$$

where the expectation is taken with respect to the true distribution of x_t. It is then clear that $\tilde{a}=\alpha$ and $\tilde{\Gamma}=dg(\Sigma)$, and this will be valid even if $dg(\Sigma)$ is singular.

Under very mild regularity conditions on the serial dependence and heterogeneity of x_t (e.g. under ergodicity; see White (1984)), it is possible to prove that plim $\tilde{a}_T=\tilde{a}=\alpha$ and plim $\tilde{\Gamma}_T=\tilde{\Gamma}=dg(\Sigma)$, as expected.

3.2 Numerical Properties of Pseudo-Maximum Likelihood Estimators of Exact One Factor Models

Suppose now that we want to fit by pseudo-maximum likelihood the model x_t ~ i.i.d. $N(a,bb'+\Gamma)$, where a and b are Mx1 vectors and Γ a diagonal MxM matrix with typical element γ_j. Since $bb'+\Gamma$ is positive semi-definite if and only if Γ is also positive semi-definite, the admissible parameter space is now $\mathbb{R}^{2M} \times \mathbb{R}^M_+$. Note that if $rank(\Gamma)=M$, then $rank(bb'+\Gamma)=M$. Similarly, if $rank(\Gamma)<M-1$, $rank(bb'+\Gamma)\leq M-1$. However, if there is only one Heywood case, i.e. $\gamma_{j^\circ}=0$ for some j°, $rank(bb'+\Gamma)=M$ unless b_{j° is also zero.

For those parameter values that imply $rank(bb'+\Gamma)=M$, the pseudo log-likelihood function of the sample is proportional to:

$$L_T(a,b,\Gamma|x_1,..,x_T) = -M/2 \log 2\pi -1/2 \log|bb'+\Gamma| -1/2T \sum_{t=1}^{T} (x_t-a)'(bb'+\Gamma)^{-1}(x_t-a) =$$

$$=-M/2 \log 2\pi - 1/2 \log|bb'+\Gamma| - 1/2 \, tr\{(bb'+\Gamma)^{-1}[S_T+(\bar{x}_T-a)(\bar{x}_T-a)']\} = L(a,b,\Gamma|\bar{x}_T,S_T)$$

Once more, $\bar{\mathbf{x}}_T$ would be the pseudo-ML estimate of a, and we can concentrate the log-likelihood function to:

$$L^c(b,\Gamma|S_T) = -M/2 \log 2\pi - 1/2 \log|bb'+\Gamma| - 1/2 \, tr[(bb'+\Gamma)^{-1}S_T]$$

In the appendix we derive the first order Kuhn-Tucker conditions for the maximization of $L^c(b,\Gamma|S_T)$ subject to $\Gamma \geq 0$ (see A1.1-4). The following lemma characterises those solutions to the first order conditions which are in the interior of the parameter space[5]

Lemma 1: If $\tilde{\Gamma}_T$ has full rank, the first order conditions (A1.1), (A1.3) and an equality version of (A1.2) are satisfied by the following estimators:

$$\tilde{b}_T = \tilde{\Gamma}_T^{1/2} \tilde{p}_{1T} (\tilde{\nu}_{1T}-1)^{1/2} \quad (1a)$$

$$\tilde{\Gamma}_T = dg(S_T - \tilde{b}_T \tilde{b}'_T) \quad (1b)$$

where

$$\tilde{P}_T \tilde{N}_T \tilde{P}'_T = (\tilde{p}_{1T} \;\; \tilde{P}_{2T}) \begin{bmatrix} \tilde{\nu}_{1T} & 0 \\ 0 & \tilde{N}_{2T} \end{bmatrix} \begin{bmatrix} \tilde{p}'_{1T} \\ \tilde{P}'_{2T} \end{bmatrix} \quad (1c)$$

is the spectral decomposition of $\tilde{\Gamma}_T^{-1/2} S_T \tilde{\Gamma}_T^{-1/2}$.

Lemma 1, which gives us an explicit solution for $\tilde{\Gamma}_T$ in terms of $\tilde{b}_T$ and vice versa, provides the basis for computational procedures that obtain pseudo ML estimates of b and Γ (see Magnus and Neudecker (1988) for two such procedures). It assumes, though, that the solution is in the interior of the admissible parameter space. However, idiosyncratic variances often become zero during estimation. The following lemma states

---

[5] See also chapter 17, sections 12 and 13 of Magnus and Neudecker (1988) for equivalent results in optimization problems that do not take into account the inequality restrictions.

under what circumstances a solution with one Heywood case, i.e. $rank(\tilde{\Gamma}_T)=M-1$, will satisfy the first order Kuhn-Tucker conditions.

Lemma 2: If we reorder the variables so that the Heywood case corresponds to the last variable, and partition S_T accordingly as:

$$\begin{bmatrix} S_{11T} & s_{1MT} \\ s'_{1MT} & s_{MMT} \end{bmatrix} \begin{matrix} (M-1 \\ (1 \end{matrix}$$

the following estimators:

$$\tilde{b}_T = \begin{bmatrix} \tilde{b}_{1T} \\ \tilde{b}_{MT} \end{bmatrix} = \begin{bmatrix} s_{1MT} s_{MMT}^{-1/2} \\ s_{MMT}^{1/2} \end{bmatrix} \begin{matrix} (M-1 \\ (1 \end{matrix} \quad (2a)$$

$$\tilde{\Gamma}_T = \begin{bmatrix} \tilde{\Gamma}_{1T} & 0 \\ 0' & 0 \end{bmatrix} = \begin{bmatrix} dg(S_{11T} - s_{MMT}^{-1} s_{1MT} s'_{1MT}) & 0 \\ 0' & 0 \end{bmatrix} \begin{matrix} (M-1 \\ (1 \end{matrix} \quad (2b)$$

will satisfy the first order conditions (A1.1-4) provided that

$$s'_{1MT} \tilde{\Gamma}_{1T}^{-1} [\tilde{\Gamma}_{1T} - (S_{11T} - s_{MMT}^{-1} s_{1MT} s'_{1MT})] \tilde{\Gamma}_{1T}^{-1} s_{1MT} \geq 0 \quad (2c)$$

If (2c) is not satisfied, lemma 2 only proves that (2a-b) constitutes a local maximum of a restricted one factor model in which one idiosyncratic variance is fixed to zero. But if it is satisfied, unrestricting that idiosyncratic variance in the admissible positive direction will decrease the log-likelihood function. Hence, (2a-b) will also be a local maximum of the inequality restricted optimization problem.

But even if (2c) does not hold, lemma 2 may still explain why Heywood cases arise during estimation. In particular, since (2a-b) satisfy the condition $dg(\tilde{b}_T \tilde{b}'_T + \tilde{\Gamma}_T) = dg(S_T)$, the zig-zag routine proposed in Magnus and Neudecker (1988) based on alternating between (1a) and (1b) will get stuck at these parameter values. Similarly, the EM algorithm of Rubin and Thayer (1982) will also get stuck at the same values, because the corresponding factor estimate is $s_{MMT}^{-1/2}(x_{Mt} - \bar{x}_{Mt})$ with an estimated mean square error of zero, and the OLS estimates in the regression of $\mathbf{x}_t$ on a constant and this factor estimate are the sample mean and (2a) respectively, with residual variances given by (2b).

From a practical point of view, lemma 2 implies that on top of interior solutions, there are M potential boundary solutions that may be local maxima. Hence, condition (2c) has to checked M times. For those variables that satisfy it, the next step is to evaluate the likelihood function to see whether they constitute the global maximum.

It is possible to come up with examples for which (2c) is not satisfied, and others for which it is. As an extreme case, if M=2, $(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})$ is a scalar equal to $\tilde{\Gamma}_{1T}$, and (2c) will always hold with equality. Besides, it is not difficult to prove that (2a-b) will also be a global maximum. The intuition is that an unrestricted exact one factor model is underidentified for M=2, but a Heywood case makes it exactly identified.

Note that lemmas 1 and 2 also characterise local extrema when S_T is singular. Such a singularity may occur for two reasons: either rank(Σ)<M (see Rao (1973)) or rank(Σ)=M but the number of series, M, is greater than the number of observations, T, in which case rank(S_T)=T. This second situation may arise in applications to stock prices, as the number of individual shares traded in major stock markets can reach several thousands. Nevertheless, in some very special instances when rank(S_T)<M, the global pseudo-MLE estimator implies a parameter configuration with (bb'+Γ) singular.

The assumed covariance matrix (bb'+Γ) will not have full rank if either rank(Γ)<M-1, so that there is more than one Heywood case, or if rank(Γ)=M-1 but the corresponding b is also zero. A limiting argument similar to the one employed for the zero factor model shows that the second case constitutes a global maximum of the likelihood function if and only if dg(S_T) has less than full rank. In the empirically relevant case in which all variables have positive variances, the following lemma characterises when a solution with more than one Heywood case will be a global maximum of the pseudo log-likelihood function.

Lemma 3: Suppose that k≥2 variables are all proportional to each other. If we reorder the variables so that the proportional ones are last, and partition S_T accordingly as:

$$\begin{bmatrix} S_{11T} & S_{12T} & s_{1MT} \\ S'_{12T} & S_{22T} & s_{2MT} \\ S'_{1MT} & s'_{2MT} & s_{MMT} \end{bmatrix} \begin{matrix} (M-k \\ (k-1 \\ (1 \end{matrix}$$

the following estimators:

$$\tilde{b}_T = \begin{bmatrix} \tilde{b}_{1T} \\ \tilde{b}_{2T} \\ \tilde{b}_{MT} \end{bmatrix} = \begin{bmatrix} s_{1MT}s_{MMT}^{-1/2} \\ s_{2MT}s_{MMT}^{-1/2} \\ s_{MMT}^{1/2} \end{bmatrix} \begin{matrix} (M-k \\ (k-1 \\ (1 \end{matrix}$$

$$\tilde{\Gamma}_T = \begin{bmatrix} \tilde{\Gamma}_{1T} & 0 & 0 \\ 0' & 0 & 0 \\ 0' & 0' & 0 \end{bmatrix} = \begin{bmatrix} dg(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT}) & 0 & 0 \\ 0' & 0 & 0 \\ 0' & 0' & 0 \end{bmatrix} \begin{matrix} (M-k \\ (k-1 \\ (1 \end{matrix}$$

yield an unbounded pseudo-log likelihood function.

Lemma 3 shows that a solution with more than one Heywood case will be obtained when several variables are proportional to each other. In this case, the log-likelihood function is unbounded. Again, S_T singular is a necessary condition for $(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)$ singular, but it is far from sufficient. In fact, as it happened for the zero factor model, only those singularities in the data that the assumed covariance matrix can exactly replicate will be pseudo-ML solutions.

Although lemma 3 looks more like an oddity, it could be potentially relevant for exchange rate data. If a subset of currencies in our sample maintain a totally unadjustable system of fixed parities amongst themselves, their movements will be exactly proportional to each other, and any of them will be chosen as the factor. Note, though, that lemma 3 is only valid in the limit, and does not apply to target zone contexts such as the European Exchange Rate Mechanism.

Using the analogy principle again, we can understand $\tilde{a}_T$, $\tilde{b}_T$ and $\tilde{\Gamma}_T$ as estimators of population characteristics $\tilde{a}$, $\tilde{b}$ and $\tilde{\Gamma}$ defined as:

$$\arg\max_{a,b,\Gamma} E[\ -M/2 \log 2\pi - 1/2 \log|bb'+\Gamma| - 1/2\ (x_t-a)'(bb'+\Gamma)^{-1}(x_t-a)\ ] = \arg\max_{a,b,\Gamma} L(a,b,\Gamma|\alpha,\Sigma)$$

It is then clear that $\tilde{a}=\alpha$, and also that lemmas 1 to 3 apply to the population characteristics $\tilde{b}$ and $\tilde{\Gamma}$ if we replace S_T by Σ. Furthermore, under mild regularity conditions, it is also possible to prove that they are the plims of $\tilde{a}_T$, $\tilde{b}_T$ and $\tilde{\Gamma}_T$.

Finally, note that if we scale the variables $\mathbf{x}_t$ by premultiplying with a diagonal non-singular matrix **D**, the pseudo maximum likelihood estimates of an exact one factor model based on the scaled variables $\mathbf{D}x_t$ are given by $\mathbf{D}\tilde{a}_T$, $\mathbf{D}\tilde{b}_T$ and $\mathbf{D}\tilde{\Gamma}_T\mathbf{D}$, and the same will be true of the population characteristics. This is the closely related to the invariance property of ML estimates

3.3 Generalised Least Squares Factor Representing Portfolios

If $\tilde{\Gamma}$ has full rank, the population first Generalised Least Squares factor representing portfolio is defined as

$$v_t^G=(\tilde{b}'\tilde{\Gamma}^{-1}\ell)^{-1}\tilde{b}'\tilde{\Gamma}^{-1}\mathbf{x}_t$$

On the other hand, if rank($\tilde{\Gamma}$)=M-1, v_t^G will coincide with the variable whose idiosyncratic variance is zero, Finally, if rank($\tilde{\Gamma}$)<M-1, the factor representing portfolio can be obtained as any linear combination of the variables that are proportional to each other with weights that add up to one. In what follows, we make the assumption that $\tilde{\Gamma}$ is regular.

The rationale for using the first factor representing portfolio even though a one factor model is possibly misspecified stems from the following propositions that relate it to principal components theory:

Proposition 1: Let $f_t^G=(\tilde{b}'\tilde{\Gamma}^{-1}\tilde{b})^{-1}\tilde{b}'\tilde{\Gamma}^{-1}x_t \propto v_t^G$ be the GLS first factor representing portfolio. Then v_t^G is proportional to the (theoretical) first principal component of the scaled variables $\mathbf{x}_t^*=\tilde{\Gamma}^{-1/2}\mathbf{x}_t$.

Again, there is an alternative way of looking at this result, which is given in the following proposition:

Proposition 2: The covariance matrix of the possibly misspecified one factor model $\tilde{\Gamma}^{-1/2}(\tilde{b}'\tilde{b}+\tilde{\Gamma})\tilde{\Gamma}^{-1/2}=(\nu_1-1)p_1p_1'+I$ best approximates $\Sigma^*=\tilde{\Gamma}^{-1/2}\Sigma\tilde{\Gamma}^{-1/2}$ in the Frobenius norm among all symmetric matrices of the form $\delta\delta'+I$.

Corollary 1: $\Sigma_1^*=(\nu_1-1)p_1p_1'$ minimises $\sum_{i\neq j}^{M}(\sigma_{ij}^*-\sigma_{1ij}^*)^2$ among all symmetric matrices with rank 1.

Corollary 2: If one component of $\mathbf{x}_t$, say the M-th one, is uncorrelated with all the others, then v_t^G has a zero weight on it.

The difference between the first principal component based on the standardized variables and the first factor representing portfolios is therefore a subtle one. Instead of scaling the variables by their standard deviations, what the pseudo-maximum likelihood procedure is doing is to chose the scaling so that the unexplained variance is the same across assets. The advantage of this scaling solution is that since we fit the variances perfectly well all the time, we can concentrate on the covariances. This becomes particularly obvious if some variables have zero correlations with all the others. Corollary 2 shows that they will have no weight in our index measure.

Another look at these indices can be obtained if we follow the inverse route, and relate principal components to GLS portfolios. Propositions 3 and 4 make it clear, but first we need another preliminary result:

Lemma 4: If rank(Σ)>1, the pseudo maximum likelihood parameter estimates of an exact one factor model with scalar idiosyncratic variance $x_t \sim$ i.i.d. $N(a, bb'+\gamma I)$ satisfy $\hat{a}_T=\bar{x}_T$ and $\hat{b}_T=\hat{\gamma}^{1/2}q_1(\hat{\gamma}^{-1}\mu_1-1)^{1/2}$ [6].

Proposition 3: The (theoretical) first principal component based on the covariance matrix of x_t is proportional to the first GLS based factor representing portfolio $f_t^P=(\hat{b}'\hat{b})^{-1}\hat{b}'x_t$ computed on the basis of (pseudo) maximum likelihood parameters estimates obtained by fitting the restricted model $x_t \sim$ i.i.d. $N(a, bb'+\gamma I)$.

Proposition 4: The (theoretical) first principal component based on the correlation matrix of x_t is proportional to the first GLS based factor representing portfolio computed on the basis of (pseudo) maximum likelihood parameters estimates obtained by fitting the restricted model $x_t \sim$ i.i.d. $N(a, bb'+\gamma dg(\Sigma))$.

The nesting of the assumed models of propositions 3 and 4 in the unrestricted one factor model of proposition 1 provides another reason for preferring the (unrestricted) GLS factor representing portfolios. For if the latter is misspecified, the former are even worse.

Much stronger results can be obtained if we strengthen the assumptions about the data generating process. Going to the other extreme, if we assume that x_t is serially uncorrelated and Σ has indeed an exact one factor structure, i.e. that

$$x_t = \alpha + bf_t + \varepsilon_t \tag{3}$$

where f_t can be interpreted as the true underlying index, then our suggested procedure is optimally designed to extract the common component, f_t, from the observed variables. If we define $f_t^G=(b'\Gamma^{-1}b)^{-1}b'\Gamma^{-1}x_t$, then f_t^G is the representing portfolio with highest correlation with the underlying factor for any finite number of assets M. If we assume that M is actually unbounded, then we can also derive large M results which show

[6] If rank(Σ)=1, then $\hat{\gamma}_T=0$ and $\hat{b}_T=\mu_{1T}^{1/2}q_{1T}$.

that both the GLS and the principal components estimates of f_t are mean square error convergent for the true underlying index (see Sentana (1994)). For finite samples, though, the alternative Kalman filter estimate $f_t^K=(\lambda+b'\Gamma^{-1}b)^{-1}b'\Gamma^{-1}x_t$, where $\lambda=V(f_t)$, is the most efficient in the mean square error sense among all portfolios with constant weightings. But since f_t^K is proportional to f_t^G, they both coincide once re-scaled so that their weights add up to one.

Furthermore, f_t^G remains relatively efficient in models with more complex mean and variance dynamics. For instance, if the true model is a conditionally heteroskedastic one factor model of the form:

$$x_t | I_{t-1} \sim D(a, b\lambda_{t|t-1}b'+\Gamma_{t|t-1})$$

(see e.g. Diebold and Nerlove (1989) and Harvey, Ruiz and Sentana (1992)), where I_{t-1} contains the past values of the series, $\lambda_{t|t-1}$ is the conditional variance of the factor and $\Gamma_{t|t-1}$ the conditional variance of the idiosyncratic terms, the factor representing portfolio with highest conditional correlation is given by the conditional GLS portfolio $f_t^{CG}=(b'\Gamma_{t|t-1}^{-1}b)^{-1}b'\Gamma_{t|t-1}^{-1}x_t$, which will generally have time-varying weights even when re-scaled. In this context, the results in Sentana (1994) indicate that f_t^{CG} is significantly more correlated than f_t^{CG} <u>only if</u> there is very substantial variation over time in $\lambda_{t|t-1}$ and $\Gamma_{t|t-1}$.

Similarly, if the data generating process is a dynamic one factor model of the form given by (3) augmented with a nondegenerate transition equation for f_t such as:

$$f_t = \rho f_{t-1} + u_t \tag{4}$$

(see e.g. Engle and Watson (1981) or Peña and Box (1987)), our proposed factor estimate, which incorrectly assumes that ρ=0, may not be too bad even for large $|\rho|$ as compared to the updated Kalman filter estimate, f_t^{DK}, obtained recursively from equations (3) and (4). As an illustration, Figure 1 plots the efficiency of f_t^K relative to that of f_t^{DK} for different values of ρ, and a parameter configuration which corresponds to our estimates for US dollar exchange rates in section 4. Quah and Sargent (1993) have recently suggested the use of measures closely related to f_t^{DK} as indices of business cycle movements in real economic variables like sectoral employment. Figure 1 indicates that modelling the dynamics to obtain better indices may not necessarily be worth the extra effort, especially if the time series model for f_t is unknown.

Note that in both examples the unconditional covariance matrix of x_t, which has the exact one factor form, contains a substantial proportion of the information on cross-correlations, and no potentially confounding information on dynamic correlations.

In practice, one has to work with the sample variance-covariance matrix of the data or its ML alternative, S_T. The analysis in this section can be trivially repeated for these sample analogues. One potential disadvantage of our method versus principal components is that it requires the computation of the pseudo ML estimates of the factor model. However, the EM algorithm of Dempster, Laird and Rubin (1977) and Rubin and Thayer (1982) provides a cheap and reliable estimation method. As a matter of fact, this method performs even better for large M (see Demos and Sentana (1992)).

4. Empirical Application

We apply the techniques outlined in the previous sections to exchange rate data for the US and the UK. Agents are interested in a single summary measure of a country's exchange rate for at least two reasons. First, movements in exchange rates can have a significant impact on trade and inflation. Second, it is of some interest, especially for financial market participants, to know whether the strength of a particular currency reflects its own "intrinsic strength" or just the weakness of one particular currency with a large weight in its basket. Finding the appropriate weights to construct an effective exchange rate index is by no means a straightforward task. The standard practice is to use trade weights constructed by looking at the bilateral trade flows between two

countries. More sophisticated variants allow for third country effects. However, clearly such procedures are an approximation to the "true" unobserved weights. Moreover, as we have already mentioned, different weights may well give very different index levels in the long run. Table 1 compares the US Federal Reserve's effective exchange rate weights with those of the Bank of England, and they are different. Still, while these indices may give a fair indication of the trade competitiveness of a currency, they may not be so good at measuring a currency's "intrinsic" strength, since they do not take into account the massive capital flows that tend to dominate trade flows.

We use monthly data for the period 1973:06 to 1991:04 of bilateral exchange rates against the US Dollar for 23 countries to compute our alternative indices. The currencies are the Australian Dollar, the Austrian Schilling, the Belgian Franc, the Canadian Dollar, the Danish Krone, the Finnish Markka, the French Franc, the Deustchemark, the Greek Drachma, the Hong Kong Dollar, the Irish Punt, the Italian Lira, the Japanese Yen, the Dutch Guilder, the New Zealand Dollar, the Norwegian Krona, the Portuguese Escudo, the Singapore Dollar, the South African Rand, the Spanish Peseta, the Swedish Krone, the Swiss Franc and the Pound Sterling.

We start by comparing the exchange rate indices constructed from the first principal component of the correlation and covariance matrix of the log differences of the bilateral exchange rates. We use the following procedure to construct the indices. First, we take the first principal component of the differenced data and find the weights on each currency. We then standardised the weights so that they added up to unity, and rescaled the principal component. The new rescaled principal component was cumulated to construct the indices. We plot the principal component series with the Federal Reserve trade weighted exchange rate (in logs) in Figure 2. Note that although changes in the three series are very closely correlated (correlation coefficient 0.98), the principal component based indices suggest that the dollar rose to much higher levels, and is still at a much higher level than implied by the Federal Reserve's series at the end of the sample. This is because the principal component measures give too high a weight to certain currencies that were even weaker against the dollar. Most would argue that the series that one gets out of the principal component approach are just too implausible.

Next we proceed to construct an index based on the pseudo-MLE procedure discussed in section 3.3. In our most simple model we assume no drift (i.e. $\alpha=0$) and a constant variance. We go on to relax these restrictions below. The results are plotted in Figure 3[7]. The levels of the two series tend to be much closer together, and they peak at roughly the same level. Besides, the pseudo-MLE index does appear to identify commonly perceived periods of dollar weakness and strength. As one would expect, though, there is still variation between the two indices. The third column of table 1 present the weights on the bilateral currencies for our proposed index. As can be seen these weights are different from both the Bank of England's and the Federal Reserve's. Notice that our approach allows us to consider more series than are in the other indices. We also computed new indices assuming the existence of a drift and a time-varying variance for the common factor that we assume of the GARCH(1,1) form (Bollerslev (1986)). Remarkably, the results were very similar (see figure 4).

We then went on to repeat the whole exercise for UK data. The findings were very similar, in that the indices based on the first principal component of the covariance or correlation matrices were very different from the official Bank of England trade-weight index (see figure 5). These indices were rather implausible and did not appear to identify commonly perceived periods of sterling strength and weakness. In contrast, the index based on the pseudo-MLE factor analytic approach seemed to be far more sensible (see figure 6). The average weights of the Bank of England and our index are presented in table 2.

[7] The pseudo ML estimates used with US and UK data correspond to an interior optimum. No Heywood case solution was found to be a local optimum in either dataset, as condition (2c) was never satisfied.

Notice that since the changes in the (log) nominal exchange rates against sterling are a linear transformation of the changes in the (log) nominal exchange rates against the US dollar, an exact factor model cannot hold for both the US dollar and Sterling unless the factor loadings for all currencies are exactly one (see Manhieu and Schotman (1992)). We therefore emphasise the pseudo-maximum likelihood nature of our approach. Nevertheless our findings are rather remarkable in that despite having necessarily a misspecified model, the estimated factors are very plausible.

As a crude indicator of the plausibility of the 230 overidentifying covariance restrictions implicit in our exact one factor model, we computed a likelihood ratio test against the alternative that Σ is fully unrestricted. The test statistics for the US and UK datasets are 816.1 and 1364.4 respectively, which would be highly significant at conventional χ^2 levels. One should not take these results literally, though, because the LR test in misspecified models is not often robust (see White (1982)). In any case, it is worth emphasising once more that we still get sensible indices for both exchange rates, which confirms that models can be useful even when they are wrong.

One candidate theory to explain our results is a model in which shocks to each currency are only mildly correlated. More formally, suppose that there is a countably infinite collection of countries, and that changes in the values of their currencies against some unspecified bundle of goods, due to say money supply shocks as in Domowitz and Hakkio (1984), are generated according to an <u>approximate</u> factor model (cf. Chamberlain and Rothschild (1983)) with zero factors. In this context, it is easy to prove that changes in M bilateral nominal exchange rates against any specified currency would have a one factor approximate factor model, in which the common factor could be interpreted as shocks to the chosen numeraire currency[8]. Although we estimate one factor exact factor models,

[8] Manhieu and Schotman (1992) have recently considered a special case of this model in which shocks to each currency are completely uncorrelated,

the cross-sectional consistency results in Chamberlain and Rothschild (1983) and Sentana (1994) imply that a relatively large M allows us to obtain plausible estimates of the common factors even though we ignore the mild correlations in idiosyncratic terms.

5. Conclusions

Financial indices are often constructed in order to capture the common variation of a large number of financial time series. However, if these series are not cointegrated, then alternative indices with different weighting schemes will diverge in the long run. Since a "true" weighting scheme may not even exist, let alone be observed, we advocate a weighting scheme that maximises the comovement within the data. An index constructed in this view can be viewed as a useful complement to more standard measures.

We successfully apply our techniques to construct a summary measure of exchange rate comovement, but they can be readily applied to other time series. Two obvious examples spring to mind. First, since many researchers in finance use both value-weighted and equally-weighted indices to measure the co-movements in stock prices, our proposed measure would provide an obvious complement. Second, many central bankers target several measures of money, on the grounds that the definition of "true" money supply is unknown. Using our technique a money supply index could be usefully constructed which maximises the co-variation in the different measures, and would be a weighted average of all the components.

---

but it implies a variance covariance matrix in which all covariances coincide.

Appendix

First order Kuhn-Tucker Conditions for Exact One Factor Models

Magnus and Neudecker (1988) show that the differential of $L^c(b,\Gamma|S_T)$ is:

$$dL^c(b,\Gamma|S_T) = -1/2\ \mathrm{tr}\{[(bb'+\Gamma)^{-1}-(bb'+\Gamma)^{-1}S_T(bb'+\Gamma)^{-1}][(db)b'+b(db')+d\Gamma]\}$$

Taking into account that γ_j cannot be negative, the first order Kuhn-Tucker conditions will be:

$$(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}\tilde{b}_T - (\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}S_T(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}\tilde{b}_T=0 \qquad \text{(A1.1)}$$

$$dg[(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}-(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}S_T(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}]\geq 0 \qquad \text{(A1.2)}$$

$$\tilde{\Gamma}_T\geq 0 \qquad \text{(A1.3)}$$

$$dg[(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}-(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}S_T(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}]\cdot\tilde{\Gamma}_T=0 \qquad \text{(A1.4)}$$

Proof of Lemma 1

By virtue of the Woodbury formula

$$(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}=\tilde{\Gamma}_T^{-1}-\tilde{\Gamma}_T^{-1}\tilde{b}_T\tilde{b}_T'\tilde{\Gamma}_T^{-1}/(1+\tilde{b}_T'\tilde{\Gamma}_T^{-1}\tilde{b}_T)$$

If we now substitute (1a) in this expression, we obtain:

$$(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}-(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1}S_T(\tilde{b}_T\tilde{b}_T'+\tilde{\Gamma}_T)^{-1} =$$

$$= \tilde{\Gamma}_T^{-1/2}[I-(\tilde{\nu}_{1T}-1)\tilde{\nu}_{1T}^{-1}\cdot\tilde{p}_{1T}\tilde{p}_{1T}']\tilde{\Gamma}_T^{-1/2} - \tilde{\Gamma}_T^{-1/2}(\tilde{\nu}_{1T}^{-1}\cdot\tilde{p}_{1T}\tilde{p}_{1T}'+\tilde{P}_{2T}\tilde{N}_{2T}\tilde{P}_{2T}')\tilde{\Gamma}_T^{-1/2} =$$

$$= \tilde{\Gamma}_T^{-1/2}(I-\tilde{p}_{1T}\tilde{p}_{1T}'-\tilde{P}_{2T}\tilde{N}_{2T}\tilde{P}_{2T}')\tilde{\Gamma}_T^{-1/2}$$

It is then straightforward to check that (A1.1) is satisfied. To prove (A1.2), note that the above expression can be written as

$$\tilde{\Gamma}_T^{-1/2}(I+(\tilde{\nu}_{1T}-1)\tilde{p}_{1T}\tilde{p}'_{1T}-\tilde{\nu}_{1T}\tilde{p}_{1T}\tilde{p}'_{1T}-\tilde{P}_{2T}\tilde{N}_{2T}\tilde{P}'_{2T})\tilde{\Gamma}_T^{-1/2}=\tilde{\Gamma}_T^{-1}(\tilde{\Gamma}_T+\tilde{b}_T\tilde{b}'_T-S_T)\tilde{\Gamma}_T^{-1}$$

which has a zero diagonal in view of (1b). Finally note that since:

$$S_T-\tilde{b}_T\tilde{b}'_T = \tilde{\Gamma}_T^{1/2}\cdot[\tilde{\Gamma}_T^{-1/2}S_T\tilde{\Gamma}_T^{-1/2}-(\tilde{\nu}_{1T}-1)\tilde{p}_{1T}\tilde{p}'_{1T}]\cdot\tilde{\Gamma}_T^{1/2}=$$

$$\tilde{\Gamma}_T^{1/2}\cdot(\tilde{p}_{1T}\ \tilde{P}_{2T})\begin{bmatrix} 1 & 0 \\ 0 & \tilde{N}_{2T}\end{bmatrix}\begin{bmatrix}\tilde{p}'_{1T} \\ \tilde{P}'_{2T}\end{bmatrix}\cdot\tilde{\Gamma}_T^{1/2}$$

such a $\tilde{\Gamma}_T$ will indeed be positive.

Proof of Lemma 2

First of all note that $\tilde{\Gamma}_{1T}$ in (2b) is always nonnegative because S_T is positive semi-definite. For the parameter values in (2a) and (2b), the assumed covariance matrix can be written as:

$$(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T) = \begin{bmatrix} dg(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})+s_{MMT}^{-1}s'_{1MT}s'_{1MT} & s_{1MT} \\ s'_{1MT} & s_{MMT}\end{bmatrix}$$

which, using the partitioned inverse formula, yields:

$$(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)^{-1} = \begin{bmatrix} \tilde{\Gamma}_1^{-1} & -s_{MMT}^{-1}\tilde{\Gamma}_1^{-1}s_{1MT} \\ -s_{MMT}^{-1}s'_{1MT}\tilde{\Gamma}_1^{-1} & s_{MMT}^{-1}+s_{MMT}^{-2}s'_{1MT}\tilde{\Gamma}_1^{-1}s_{1MT}\end{bmatrix}$$

Then, straightforward algebra shows that:

$$S_T(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)^{-1}\tilde{b}_T = \tilde{b}_T$$

so that (A1.1) will indeed be satisfied. Similarly, it is easy to check that the (1,1) block of $(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)^{-1}-(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)^{-1}S_T(\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T)^{-1}$ is

$$\tilde{\Gamma}_{1T}^{-1}[\tilde{\Gamma}_{1T}-(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})]\tilde{\Gamma}_{1T}^{-1}$$

whereas

$$s_{MMT}^{-2}s'_{1MT}\tilde{\Gamma}_{1T}^{-1}[\tilde{\Gamma}_{1T}-(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})]\tilde{\Gamma}_{1T}^{-1}s_{1MT}$$

is its M,M element. Hence, the M-1 first conditions in (A1.2) will be satisfied with equality, whereas the last one will be satisfied by (2c).

Proof of Lemma 3

If we define δ as the k-1 vector of proportionality factors for $\mathbf{x}_{2t}$ (i.e. $\mathbf{x}_{2t}=\delta \mathbf{x}_{Mt}$), we can write S_T as

$$S_T = \begin{bmatrix} S_{11T} & s_{1MT}\delta' & s_{1MT} \\ \delta s'_{1MT} & s_{MMT}\delta\delta' & s_{MMT}\delta \\ s'_{1MT} & s_{MMT}\delta' & s_{MMT}\end{bmatrix}$$

For $\varepsilon\geq 0$, let's define the sequence of estimators

$$\tilde{b}_T(\varepsilon) = \begin{bmatrix} s_{1MT}s_{MMT}^{-1/2} \\ \delta s_{MMT}^{1/2} \\ s_{MMT}^{1/2}\end{bmatrix} = \tilde{b}_T$$

$$\tilde{\Gamma}_T(\varepsilon) = \begin{bmatrix} \tilde{\Gamma}_{1T} & 0 & 0 \\ 0' & \varepsilon I & 0 \\ 0' & 0' & 0\end{bmatrix} \quad \text{with } \tilde{\Gamma}_{1T}=dg(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})$$

Since $\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T(\varepsilon)$ has full rank unless $\varepsilon=0$, $[\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T(\varepsilon)]^{-1}S_T$ can be computed for any $\varepsilon>0$. Tedious algebra shows that it will be given by:

$$\begin{bmatrix} \tilde{\Gamma}_{1T}^{-1}(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT}) & 0 & 0 \\ 0' & 0 & 0 \\ s_{MMT}^{-1}s'_{1MT}-s_{MMT}^{-1}s'_{1MT}\tilde{\Gamma}_{1T}^{-1}(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT}) & \delta' & 1\end{bmatrix}$$

whose trace is always M-k+1. On the other hand, $\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T(\varepsilon)$ is increasingly singular as $\varepsilon\to 0$, so $|\tilde{b}_T\tilde{b}'_T+\tilde{\Gamma}_T(\varepsilon)|$ goes to 0, and the log-likelihood function goes to $+\infty$. The optimality of this solution is confirmed by the fact that in this case the left hand side of (2c) is proportional to

$$s_{MMT}^{-2}s'_{1MT}\tilde{\Gamma}_{1T}^{-1}[\tilde{\Gamma}_{1T}-(S_{11T}-s_{MMT}^{-1}s_{1MT}s'_{1MT})]\tilde{\Gamma}_{1T}^{-1}s_{1MT}+\delta'\delta\varepsilon^{-1}$$

which goes to $+\infty$ as ε goes to 0.

Proof of Proposition 1:

From the population version of lemma 1, $(\tilde{b}'\tilde{\Gamma}^{-1}\tilde{b})^{-1}\tilde{b}'\tilde{\Gamma}^{-1}\mathbf{x}_t = (\nu_1-1)^{-1/2}p_1'\tilde{\Gamma}^{-1/2}\mathbf{x}_t$, and $V(\tilde{\Gamma}^{-1/2}\mathbf{x}_t)=\tilde{\Gamma}^{-1/2}\Sigma\tilde{\Gamma}^{-1/2}$.

Proof of Proposition 2:

Straightforward if we notice that the spectral decomposition of $\tilde{\Gamma}^{-1/2}\Sigma\tilde{\Gamma}^{-1/2}-I$ is simply $P(N-I)P'$.

Proof of Corollary 1:

From proposition 2, $tr(\Sigma^*-I-\Sigma_1^*)^2=\sum_{i=1}^{M}(\sigma^*_{ii}-1-\sigma^*_{1ii})^2 + \sum_{i\neq j}^{M}(\sigma^*_{ij}-\sigma^*_{1ij})^2$. But from Lemma 1 and the invariance property of ML estimates, $\sigma^*_{ii}-1-\sigma^*_{1ii}=0$.

Proof of Corollary 2:

We just need to show that $\tilde{b}_M=0$. But from Corollary 1, it is clear that since $\sigma^*_{iM}=0$ for $i\neq M$, this must be indeed the case.

Proof of Lemma 4:

A trivial modification of lemma 1, since the spectral decomposition of $\hat{\gamma}_T^{-1}S_T$ is simply $Q_T\hat{\gamma}_T^{-1}M_TQ_T'$.

Proof of Proposition 3:

Straightforward from the population version of lemma 4.

Proof of Proposition 4:

Straightforward from the population version of lemma 4, and the invariance property of ML estimates if we notice that $\mathbf{x}_t$~i.i.d. $N(a,bb'+\gamma dg(\Sigma))$ is equivalent to $dg^{-1}(\Sigma)\mathbf{x}_t$~i.i.d. $N(c,dd'+\gamma I)$, where $c=dg^{-1}(\Sigma)a$ and $d=dg^{-1}(\Sigma)b$.

References

Bollerslev, T. (1986): "Generalized ARCH", Journal of Econometrics 31, 307-327.

Boothe, P. and Glassman, D. (1987): "The Statistical Distribution of Exchange Rates: Empirical Evidence and Economic Implications", Journal of International Economics 22, 297-319.

Chamberlain, G. and Rothschild, M. (1983): "Arbitrage, Factor Structure, and Mean Variance Analysis on Large Asset Markets", Econometrica 51, 1281-1304.

Demos, A. and Sentana, E. (1992): "An EM-based Algorithm for Conditionally Heteroskedastic Factor Models", LSE Financial Markets Group Discussion Paper 140.

Dempster, A.; Laird, N. and Rubin, D. (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society 39, 1-38.

Diebold, F. X.; Gardeazabal, J. and Yilmaz, K. (1992): "On Cointegration and Exchange Rate Dynamics", mimeo, Universidad del Pais Vasco.

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

Domowitz, I; Hakkio, C.S. (1985): "Conditional Variance and the Risk Premium in the Foreign Exchange Market", Journal of International Economics 19, 47-66.

Engle, R.F. and Korizky, S. (1993): "Testing for Common Features", Journal of Business and Economic Statistics 11, 369-380.

Engle, R.F. and Watson, M. (1981): "A One-factor Multivariate Time Series Model of Metropolitan Wage Rates", JASA 76, 774-781.

Harvey, A.C.; Ruiz, E. and Sentana, E. (1992): "Unobserved Components Time Series Models with ARCH Disturbances", Journal of Econometrics 52, 129-157.

Magnus, J. and Neudecker, H. (1988) Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester.

Manhieu, R. and Schotman, P. (1992): "Neglected Common Factors in Exchange Rate Volatility", mimeo, University of Limburg.

Manski, C.F. (1988): Analog Estimation Methods in Econometrics, Chapman and Hall, New York.

Peña, D. and Box, G.E.P.: "Identifying a Simplifying Structure in Time Series", JASA 82, 836-843.

Priestley, M.B. (1981): Spectral Analysis and Time Series, Academic Press, London.

Quah, D. and Sargent, T. (1993): "A Dynamic Index Model for Large Cross Sections", LSE Centre for Economic Performance Discussion Paper 132.

Rao, C.R. (1973): Linear Statistical Inference and its Applications, 2nd edn., John Wiley, New York.

Roll, R.W. (1977): "A Critique of the Asset Pricing Theory's Tests - Part I: On Past and Potential Testability of the Theory", Journal of Financial Economics 4, 129-176.

Rubin, D. and Thayer, D. (1982): "EM Algorithms for ML Factor Analysis", Psychometrica 47, 69-76.

Sentana, E. (1994): "Factor Representing Portfolios in Large Asset Markets", mimeo, CEMFI.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", Econometrica 50, 1-26.

White, H. (1984): Asymptotic Theory for Econometricians, Academic Press, London.

Table 1: US Effective Exchange Rate Weights

	Federal Reserve Weights (%) (1)	Bank of England Weights (%) (2)	MLE Weights (%) (3)	Correlation of bilateral US $ Rates with PMLE Factor Estimates
Australia			0.1	0.230
Austria		0.7	20.8	0.990
Belgium	6.4	2.8	3.5	0.933
Canada	9.1	18.7	0.3	0.235
Denmark		0.7	10.0	0.977
Finland		0.5	2.2	0.863
France	13.1	8.3	3.6	0.930
Germany	20.8	14.5	26.6	0.993
Greece			0.7	0.702
Hong Kong			0.4	0.320
Ireland		0.5	2.1	0.879
Italy	9.0	5.6	1.2	0.801
Japan	13.6	25.9	0.6	0.645
Netherlands	8.3	2.8	18.2	0.988
New Zealand			0.2	0.361
Norway		0.6	2.8	0.896
Portugal			0.8	0.776
Singapore			0.4	0.444
South Africa			0.2	0.377
Spain		1.6	0.8	0.713
Sweden	4.2	2.1	1.6	0.831
Switzerland	3.6	2.8	2.2	0.898
UK	11.9	12.0	0.7	0.688
Correlation with Weights for (1)	1.0	0.81	0.39	
Correlation with Weights for (2)	0.81	1.0	0.10	

Table 2: UK Effective Exchange Rate Weights

	Bank of England Weights (%) (1)	MLE Weights (%) (2)	Correlation of Bilateral Sterling Rates with PMLE Factor Estimates
Australia		0.2	0.334
Austria	1.2	22.1	0.986
Belgium	5.3	3.3	0.895
Canada	1.9	0.4	0.430
Denmark	1.5	10.3	0.966
Finland	1.5	2.0	0.788
France	11.8	3.5	0.896
Germany	20.0	28.1	0.990
Greece		0.6	0.616
Hong Kong		0.4	0.494
Ireland	2.4	2.2	0.794
Italy	7.7	1.1	0.726
Japan	8.8	0.5	0.533
Netherlands	5.0	16.2	0.978
New Zealand		0.2	0.290
Norway	1.3	2.5	0.839
Portugal		0.8	0.672
Singapore		0.5	0.494
South Africa		0.2	0.368
Spain	2.0	0.8	0.609
Sweden	3.8	1.5	0.763
Switzerland	5.5	2.0	0.838
UK	20.4	0.4	0.423
Correlation with Weights for (1)	1.0	0.37	

US $ Effective Exchange Rate
Principal Component based Measures

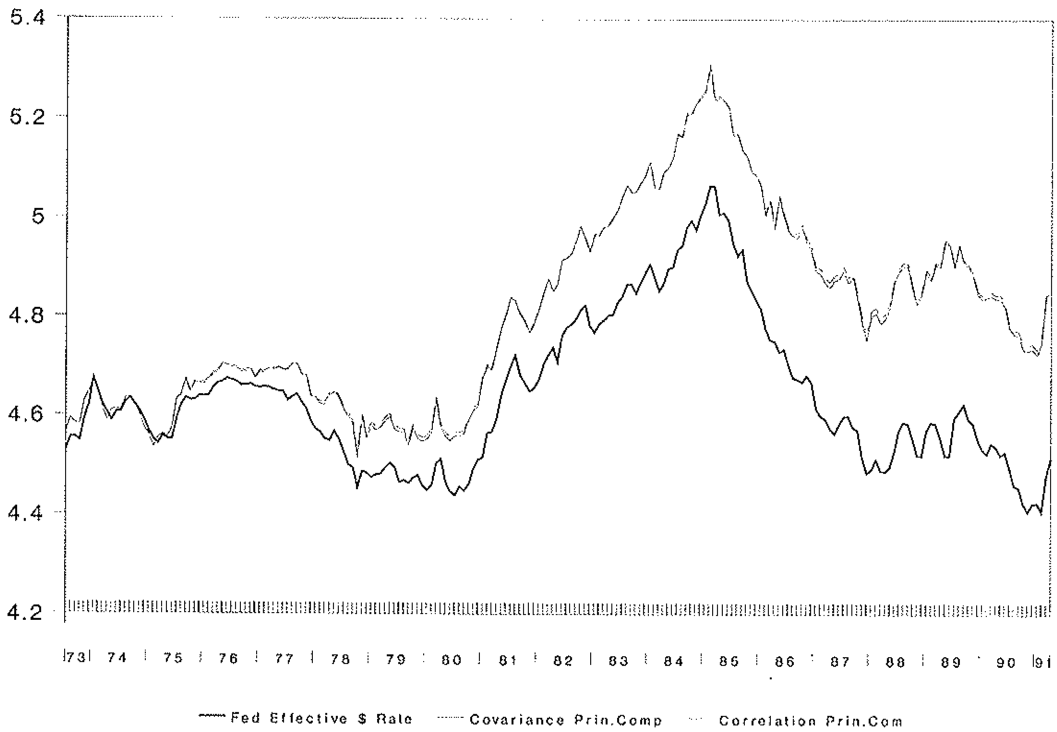

Figure 2

Relative Efficiency of Static vs Dynamic
Factor Estimates

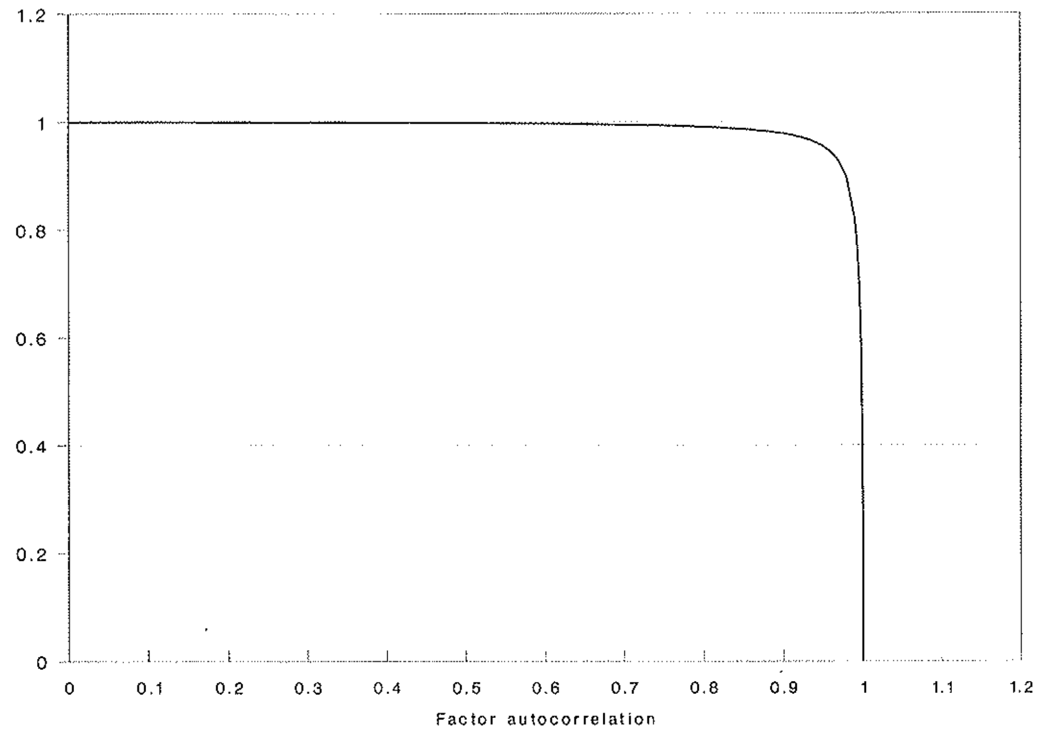

Figure 1

US $ Effective Exchange Rate
Other PMLE Factor Analysis Measures

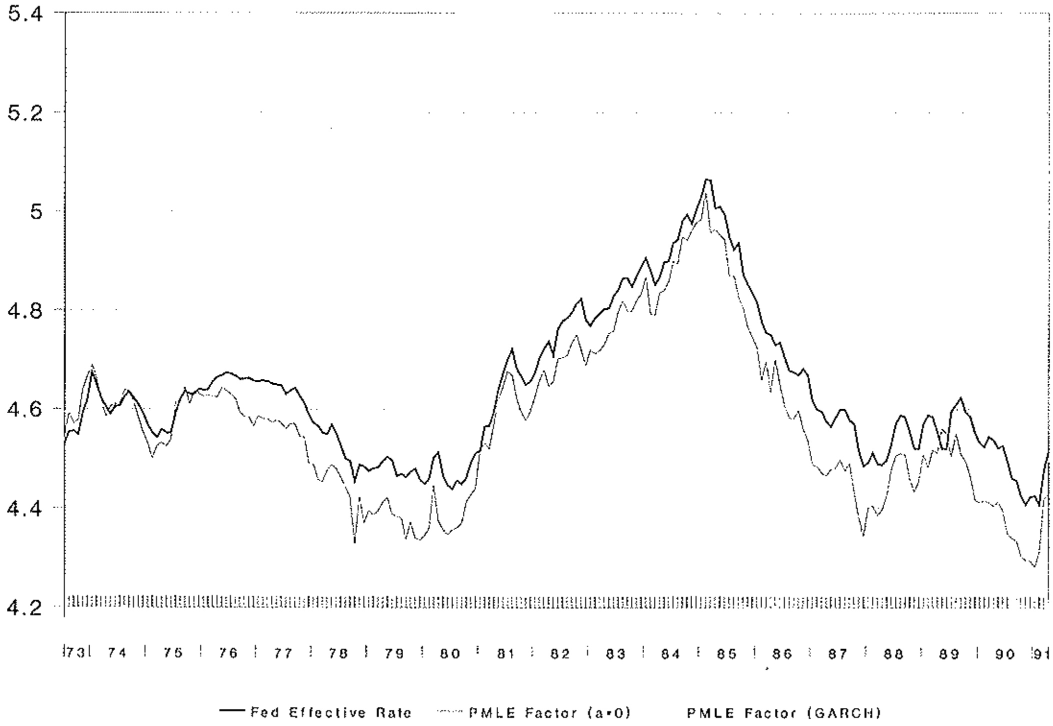

Figure 4

US $ Effective Exchange Rate
PMLE Factor Analysis Measure

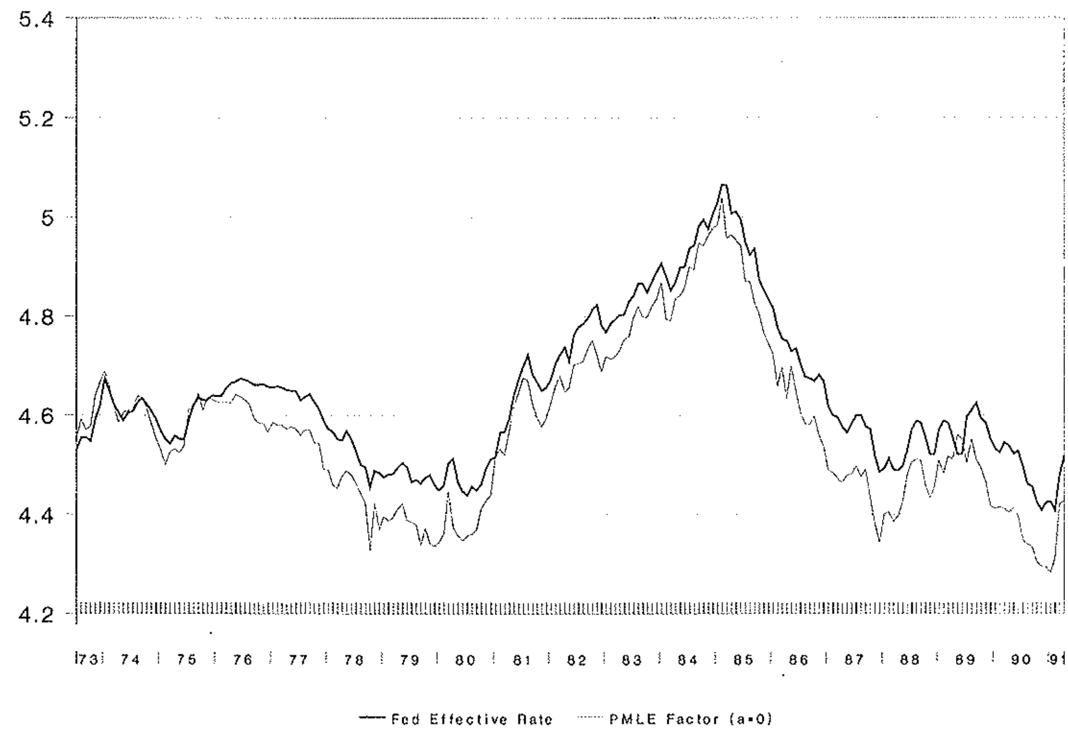

Figure 3

Sterling Effective Exchange Rate
PMLE Factor Analysis Measure

Figure 6

Sterling Effective Exchange Rate
Principal Component based Measures

Figure 5