Bayesian Dynamic Factor Models and Portfolio Allocation

Omar AGUILAR and Mike WEST

We discuss the development of dynamic factor models for multivariate financial time series, and the incorporation of stochastic volatility components for latent factor processes. Bayesian inference and computation is developed and explored in a study of the dynamic factor structure of daily spot exchange rates for a selection of international currencies. The models are direct generalisations of univariate stochastic volatility models, and represent specific varieties of models recently discussed in the growing multivariate stochastic volatility literature. We discuss model fitting based on retrospective data and sequential analysis for forward filtering and short-term forecasting. Analyses are compared with results from the much simpler method of dynamic variance matrix discounting that, for over a decade, has been a standard approach in applied financial econometrics. We study these models in analysis, forecasting and sequential portfolio allocation for a selected set of international exchange rate return time series. Our goals are to understand a range of modelling questions arising in using these factor models, and to explore empirical performance in portfolio construction relative to discount approaches. We report on our experiences and conclude with comments about the practical utility of structured factor models, and on future potential model extensions.

KEY WORDS: Dynamic Factor Analysis; Dynamic Linear Models; Exchange Rates Forecasting; Markov Chain Monte Carlo; Multivariate Stochastic Volatility; Portfolio Selection; Sequential Forecasting; Variance Matrix Discounting

1. INTRODUCTION

Since the mid-1980s, multivariate stochastic volatility models based on variance/covariance discounting (Quintana and West 1987, 88) have been used as components of applied Bayesian forecasting models in financial econometric settings (Quintana 1992; Putnam and Quintana 1994, 1995; Quintana and Putnam 1996; Quintana, Chopra and Putnam 1995). The success of such methods in portfolio construction is evidenced partly by the fact that they have been adopted and are in vigorous day-to-day use as components of global portfolio approaches in several major international banks. In more recent years, major developments in structured stochastic volatility (SV) modelling have led to the introduction of various approaches to modelling dependencies in volatility processes that, in principle, may lead to improvements in short-term forecasting of multiple financial and econometric time series. There is no doubt that these more complex SV models theoretically improve the description of several kinds of financial time series data, including exchange rate return series, and hold potential for improvements in practical short-term forecasting relative to discount models. Part of our interest here is to empirically explore this potential. We develop dynamic factor, multivariate stochastic volatility models to address:

- questions about their potential to provide practical improvements in short-term forecasting, and resulting dynamic portfolio allocations, of international exchange rates and other financial time series;
- issues of model structuring, implementation, Bayesian analysis and computation; and
- questions of comparison with the much simpler methods based on variance/covariance discounting.

We study these issues in connection with data analysis and portfolio construction using multiple series of returns on international exchange rates.

Variants of the basic method of variance matrix discounting (see above references to Quintana and coauthors) have formal theoretical bases in matrix-variate "random walks" (Uhlig 1994, 97). Further discussion is given below, and more background can be found in West and Harrison (1997, section 16.4.5). The basic discounting methods follow foundational developments for univariate series in Ameen and Harrison (1985) and Harrison and West (1987), and the formal multivariate models are direct generalisations of univariate models of Shephard (1994a). Related discussion appears in West and Harrison (1997, section 10.8.2). In the general multivariate context, the approach leads to the embedding of smoothed estimates of "local" variance/covariance structure within a Bayesian modelling framework, and so provides for adaptation to stochastic changes as time series data are processed. Modifications to allow for changes in discount rates in order to adapt to varying degrees of change, including marked/abrupt changes in volatility patterns, extend the basic approach. The resulting update equations for sequences of estimated volatility matrices have univariate components that relate

Mike West is the Arts and Sciences Professor of Statistics and Decision Sciences, and Director of the Institute of Statistics and Decision Sciences, at Duke University. The address is ISDS, Duke University, Durham, NC 27708-0251, USA, and the web site address is http://www.stat.duke.edu. Omar Aguilar is Vice President, Merrill Lynch Quantitative Research, Merrill Lynch, Mexico City. This research was partially supported by the National Science Foundation under grants DMS 9704432 and 9707914, and by CDC Investment Management Corporation, New York. The authors acknowledge useful discussions with Jose M Quintana, Neil Shephard and Hong Chang, and the comments and editorial efforts of the JBES Editor. This article is available in electronic form on the ISDS web site, http://www.stat.duke.edu

Journal of Business & Economic Statistics
???

closely to variants of ARCH and SV models, and so it is not surprising that they have proven useful in many applications. However, unlike these more formal models, discounting methods do not have real predictive capabilities, simply allowing for and estimating changes rather than anticipating them. Hence the interest in factor models that set out to explicitly describe changes through patterns of time-variation in parameters driving underlying latent processes. This is the key motivating concept underlying interest in SV models generally, and led has to various authors mentioning or developing multivariate SV models with dynamic factor structure. Key references include the initiating work of Harvey, Ruiz and Shephard (1994), and the later developments in the papers of Jacquier, Polson and Rossi (1994, 95), and Kim, Shephard and Chib (1998). Several authors have begun to develop and explore related factor models with ARCH/GARCH structure, as opposed to SV structure (e.g., Demos and Sentana 1998). Though we do not discuss this further here, some of the connections and comparisons with SV approaches are of general interest, as discussed particularly in the above referenced paper by Kim *et al*, and further work on comparative studies might be very worthwhile.

In the following section we detail the basic framework and notation for stochastic, time-varying variance matrices, followed by discussion of factor structure. We build on prior work in non-dynamic Bayesian factor analysis and develop MCMC methods of model fitting and computation in the chosen class of dynamic factor models. These models are essentially similar to those introduced in Harvey, Ruiz and Shephard (1994) and adopted by Jacquier, Polson and Rossi (1994, 95). Our empirical study also implements approximations to sequential analysis and updating using the particle filtering methods of Pitt and Shephard (1999a). We note that the work reported here was developed independently of, and in parallel to, that reported in Pitt and Shephard (1999b), and bears heavily on the technical and modelling contributions of those authors. Our computational approaches in model fitting differ somewhat from these authors, however. Furthermore, our perspective and objectives are fundamentally on questions of forecasting and portfolio decisions rather than exclusively methodological. This is reflected in our detailed analyses of international exchange rate time series that discuss model selection and specification, reports on our practical experiences with these models, and makes comparisons with variance discount approaches in analysis, forecasting and portfolio construction. We conclude with summary comments and pointers to future work.

2. TIME-VARYING VARIANCE MATRICES AND FACTOR STRUCTURE

2.1 Introduction and Notation

To introduce notation, consider a $q-$variate time series $\mathbf{y}_t$, $(t = 1, 2, \ldots,)$ as conditionally independent, Gaussian random vectors with means $\boldsymbol{\theta}_t$ and variance matrices $\boldsymbol{\Sigma}_t$, denoted by $N(\mathbf{y}_t|\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t)$ for each t. Throughout, our notation implicitly identifies the conditioning information: for any information set H, $p(\mathbf{y}_t|H)$ is the conditional density of $\mathbf{y}_t$ when H is known. Hence $p(\mathbf{y}_t|\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t)$ is the conditional density of $\mathbf{y}_t$ when the time t parameters $(\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t)$ are known, and, implicitly, $\mathbf{y}_t$ is conditionally independent of all other relevant quantities when $(\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t)$ are given. Similarly, if D_t represents all historical data and information available at any time t, then $p(\mathbf{y}_t|D_{t-1})$ is the conditional density of the one-step ahead predictive distribution at time $t-1$, when all uncertain quantities, usually including the parameters $(\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t)$, have been integrated out with respect to the relevant posterior distribution conditional on D_{t-1}. This is standard notation in the Bayesian statistics literature (e.g., West and Harrison 1997).

Though much applied interest resides in models with small time-variations in levels $\boldsymbol{\theta}_t$ predicted via dynamic regressions, we adopt a constant level model, $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ for all t, for our studies here. A primary interest here is in comparisons with discount methods in short-term forecasting, so that the same assumption will be made in discount model approaches. Our more recent research and studies in application have extended the framework to involve dynamic regression components in $\boldsymbol{\theta}_t$, and through these studies we have verified that, for the primary goal here of comparisons with discount methods, the restriction to a constant level here is immaterial.

Variance matrix discounting models produce sequences of posterior distributions for the matrices $\boldsymbol{\Sigma}_t$. Some details are provided in Appendix A, below. The analyses are naturally sequential, so that posterior distributions are revised, or updated, as more data is processed. By way of notation, we denote by $\mathbf{S}_t$ the resulting posterior estimate of $\boldsymbol{\Sigma}_t$ based on data up to time t, and by $\mathbf{S}_{t,n}$ the resulting posterior estimate of $\boldsymbol{\Sigma}_t$ based on the larger set of data up to a time $n > t$. Live analysis is generally sequential, with the "current" posterior for $\boldsymbol{\Sigma}_t$ at time t of key relevance in short-term forecasting; hence the interest in point estimates $\mathbf{S}_t$ as they are sequentially updated. Model exploration and development, on the other hand, involves retrospective data analysis; hence the interest in retrospectively revised point estimates $\mathbf{S}_{t,n}$.

2.2 Component and Factor Structure

From very early examples and applications of variance matrix discounting, principal component analyses of estimated sequences of $\boldsymbol{\Sigma}_t$ matrices have been used to explore the nature of changes over time in covariance patterns, and to provide insight into the latent mechanisms driving such changes. See, for example, the studies of monthly exchange rate time series in Quintana and West (1987), also reported in West and Harrison (1997, section 16.4.6), where the patterns of change over time in the principal component structure of the estimates $\mathbf{S}_{t,n}$ are explored. A standard principal component decomposition of $\mathbf{S}_{t,n}$ provides insight into the related decomposition of $\boldsymbol{\Sigma}_t$. Often, as in the above examples, this will yield a small number of dominant components representing latent factors contributing measurably to both total variability in the series and the covariance structure, together with additional residual components. This partly underlies the interest in dynamic factor models to more explicitly represent the latent structure and put the spot-light on inference on factor processes and their parameters. In line with Press and Shigemasu (1989), Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998), and (extrapolating to the stochastic volatility context here) Geweke and Zhou (1996), a basic $k-$factor dynamic model (with $k < q$,) for $\mathbf{\Sigma}_t$ is

$$\mathbf{\Sigma}_t = \mathbf{X}_t \mathbf{H}_t \mathbf{X}_t' + \mathbf{\Psi}_t = \sum_{j=1}^{k} \mathbf{x}_{tj} \mathbf{x}_{tj}' h_{tj} + \mathbf{\Psi}_t \qquad (1)$$

where

- $\mathbf{X}_t$ is the $q \times k$ factor loadings matrix at time t, with columns $\mathbf{x}_{tj}$,
- $\mathbf{H}_t = \text{diag}(h_{t1}, \ldots, h_{tk})$ is the diagonal matrix of instantaneous factor variances, and
- $\mathbf{\Psi}_t = \text{diag}(\psi_{t1}, \ldots, \psi_{tq})$ is the diagonal matrix of instantaneous, series-specific or "idiosyncratic" variances.

In terms of the time series $\mathbf{y}_t$ this is equivalent to the representation

$$\mathbf{y}_t = \boldsymbol{\theta}_t + \mathbf{X}_t \mathbf{f}_t + \boldsymbol{\epsilon}_t \qquad (2)$$

where

- $\mathbf{f}_t \sim N(\mathbf{f}_t | \mathbf{0}, \mathbf{H}_t)$ are conditionally independent realisations of the $k-$vector latent factor process,
- $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{\epsilon}_t | \mathbf{0}, \mathbf{\Psi}_t)$ are conditionally independent and series-specific quantities, and
- $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_s$ are mutually independent for all t, s.

This is the basic structure we adopt for the rest of the paper. Our analyses are based on choosing a specific number of factors k throughout. Discussion of choice of k values, and broader related issues of model uncertainty and specification, follow in section 2.5 and 6. First, we discuss further structure on the basic factor model with k fixed, and the stochastic volatility components for the factor and idiosyncratic variances.

2.3 Factor Model Constraints

The models analysed and applied in the studies of this paper are based on constant factor loadings, so that $\mathbf{X}_t = \mathbf{X}$ for all t. This provides a framework very similar to those mooted by the earlier authors, as referenced above, in which we aim to investigate the issues and difficulties in model implementation.

The $k-$factor model must be further constrained to define a unique model free from identification problems. A first constraint is that $\mathbf{X}$ be of full rank k to avoid identification problems arising through invariance of the model under location shifts of the factor loading matrix (e.g., Geweke and Singleton, 1980). Second, we must further constrain the factor loading matrix to avoid over-parametrisation – simply ensuring that the number of free parameters at time t in the factor representation does not exceed the $q(q+1)/2$ parameters in an unrestricted $\mathbf{\Sigma}_t$. Finally, we need to ensure invariance under invertible linear transformations of the factor vectors (Press 1985, chapter 10). On this latter issue, our work follows Geweke and Zhou (1996), among others, in adopting the "hierarchical" structural constraint in which the loadings matrix has the form

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ x_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ x_{k,1} & x_{k,2} & x_{k,3} & \cdots & 1 \\ x_{k+1,1} & x_{k+1,2} & x_{k+1,3} & \cdots & x_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{q,1} & x_{q,2} & x_{q,3} & \cdots & x_{q,k} \end{pmatrix}. \qquad (3)$$

This form immediately ensures that $\mathbf{X}$ is of full rank k. Further, if each of the idiosyncratic variances ψ_{tj} are non-zero, then, simply by counting the number of free parameters in the implied variance matrix $\mathbf{\Sigma}_t$, we deduce that the number of factors k is subject to an upper bound implied by the quadratic inequality $k^2 - (2q+1)k + q(q-1) \geq 0$; the practical implication is that k does not exceed the integer part of $q + 1/2 - \sqrt{1+8q}/2$. For example, with $q = 6$ or 7 we have $k \leq 3$, with $q = 15$ or 16 we have $k \leq 10$, while with $q = 30$ we have $k \leq 22$. For realistic values of q this bound is unlikely to be problematic, as practical interest will be in models with smaller numbers of factors. Hence we have full model identification in such cases. We note that, if one or more of the idiosyncratic variances are set to zero, the number of factors could be increased up to the absolute limit of q, though this is of little practical interest in problems with several or moderate numbers of series. Further, note that the identification issues are not complicated or extended at all in moving from a standard factor model to the dynamic factor model here, with time-varying factor and idiosyncratic variances, and even with time-varying factor loadings. Identifying constraints that are used in static model simply transfer to apply at each time t.

An important implication of the particular structure of $\mathbf{X}$ above is that it induces substantive identification of the factors as well as the purely technical model identification. That is, the chosen order of the univariate time series in the $\mathbf{y}_t$ vector serves to define the factors: the first series is the first factor plus a "noise" term, and so forth. This focuses attention on the choice of ordering in model specification, and provides interpretation. More discussion of this appears in the applied studies below, where, in particular, we report on experiences in repeat analyses that vary the series order. A critical comment to be borne in mind is that the order influences interpretation of the factors and may impact on model fit and the choice of k in particular, but has no impact whatsoever on forecasts: the variances and covariances between the series are quite independent of this modelling decision. Finally, note that the primary alternative to the hierarchical structure of $\mathbf{X}$ is to fix on an orthonormal loadings matrix (Press 1985, chapter 10).

We very much prefer the hierarchical structure primarily for the reasons of substantive interpretation of the factors, just discussed, and also as the resulting Bayesian analysis is much less technically complicated.

2.4 Stochastic Volatility Model Components

Multivariate generalisations of univariate SV models through dynamic factor models are mentioned by various authors, including Harvey, Ruiz and Shephard (1994), Shephard (1996), Kim, Shephard and Chib (1998), and have been investigated by Jacquier, Polson and Rossi (1995). The basic model of the latter authors assumes that the univariate factor series f_{ti} follow standard univariate SV models, but discuss possible extensions also mentioned in Kim, Shephard and Chib (1998). We adopt such an extension here, one in which the log volatilities of the factors follows a vector autoregression with correlated innovations, an extension that turns out to be of relevance in studies of exchange rate returns. We additionally adopt unrelated, univariate stochastic volatility models for the idiosyncratic variances.

For the factor variances, define $\lambda_{ti} = \log(h_{ti})$ for each $i = 1, \ldots, k$ and write $\boldsymbol{\lambda}_t = (\lambda_{t1}, \ldots, \lambda_{tk})'$. We assume a stationary vector autoregression of order one, VAR(1), centered around a mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and with individual AR parameters ϕ_i in the matrix $\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_k)$. That is, for $t = 1, 2 \ldots$, we have

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \Phi(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\omega}_t \quad (4)$$

with independent innovations

$$\boldsymbol{\omega}_t \sim N(\boldsymbol{\omega}_t|\mathbf{0}, \mathbf{U}) \quad (5)$$

for some innovations variance matrix $\mathbf{U}$. The implied marginal distribution for each $\boldsymbol{\lambda}_t$ is then

$$\boldsymbol{\lambda}_t \sim N(\boldsymbol{\lambda}_t|\boldsymbol{\mu}, \mathbf{W}) \quad (6)$$

where $\mathbf{W}$ satisfies $\mathbf{W} = \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi} + \mathbf{U}$ and has elements $W_{ij} = U_{ij}/((1-\phi_i)(1-\phi_j))$. The model allows dependencies across volatility series through non-zero off-diagonal entries in $\mathbf{U}$ and $\mathbf{W}$. Also, this marginal distribution defines the initial distribution for $\boldsymbol{\lambda}_0$.

For the idiosyncratic variances, define $\eta_{tj} = \log(\psi_{tj})$ for each $j = 1, \ldots, q$. For these series of log variances we assume standard univariate autoregressions of order one, namely

$$\eta_{tj} = \alpha_j + \rho_j(\eta_{t-1,j} - \alpha_j) + \xi_{tj}$$

with independent innovations $\xi_{tj} \sim N(\xi_{tj}|0, s_j)$. Unlike the factor volatilities, the η_{tj} processes are mutually independent series. Write $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)$ and $\mathbf{s} = (s_1, \ldots, s_q)$ for the vectors of parameters here. Our models assume stationarity and positive dependence within each series, so that $0 < \rho_j < 1$ for each j and the implied stationary marginal distributions are given by $N(\eta_{tj}|\alpha_j, s_j/(1-\rho_j^2))$, including the case $t = 0$ that provides the initial priors for the volatility series.

2.5 Model Specification and Practical Perspectives

Our discussion of model fitting and analyses, and applications below, are based on specific $k-$factor models. We do discuss some empirical experiences with varying the value of k in our exchange rate studies, though do not formally address inference on k here. In this connection, some general points to note are as follows. First, in analysing data consistent with a factor structure but with a model in which k is too large, we should expect to see multimodalities appearing in posterior densities for the elements of the factor loading matrix (Geweke and Singleton 1980; Lopes and West 1998). As illustrated in the latter reference, we have experienced this in standard (non-dynamic) factor models and the idea translates directly. Experiencing this would suggest that k be decreased. Second, and related to this, approaches to analysis using MCMC will tend to experience convergence difficulties in models with k too large; this has also been verified in ranges of studies of our exchange rate series with these dynamic factor models, as we discuss in the application sections below. By contrast, in a model in which k appears to be appropriate for the data under study, our experiences are that the MCMC algorithms converge rapidly and cleanly, again as in standard, non-dynamic factor models.

The broader question of inference on k is interesting and challenging. In Lopes and West (1998) we have explored a range of approaches to inference on k, Bayesian and likelihood-based. Translating this to the dynamic factor context is as yet unexplored.

From a serious practical viewpoint, we here outline what has been a useful exploratory perspective on choosing k – or a range of possible values of k to explore in parallel – in our applied work. In our studies of the exchange rate series we are interested in sequential forecasting and portfolio, and have a wealth of historical data at hand to use in model construction and the exploratory phases of analysis. Our examples below are founded on such an exploration of an initial stretch of historical data, followed by formal model fitting to the remaining data. In line with this, we choose both the value, or values, of k and hyperparameters of informative prior distributions for model parameters based on a rather informal look at some such initial data. A referee has referred to this as "ad hoc", and to some extent it is. It is also a scientifically sensible, permissible and practicable approach to initial model and prior specification, and perfectly formal and coherent from the viewpoints of statistical inference. In the exchange rate studies, we have a series of over 1,827 daily observations at hand. From an exploratory analysis of just the first 200 days, we settle on choices of k and prior distributions for the model to be used *from that time point onward*, fitting the so-specified model to the remaining data. In this initial analysis phase, we explore the early stretch of data using the variance matrix discounting method. This is simple, robust and trivially implemented to produce estimated sequences of $\boldsymbol{\Sigma}_t$ matrices over this first short section of historical data. As exemplified in various previous studies (e.g., Quintana and West 1987; West and Harrison

1997, chapter 16) using principal component decompositions of these sequences, such analysis provides insight into the underlying factor structure. In our current work, we use Cholesky decompositions of the estimated $\mathbf{\Sigma}_t$ matrices rather than principal components. The Cholesky decompositions map directly onto the factor model form of this paper, and so provides direct insight into plausible ranges of values for factor model parameters, including k, that can be used to develop informed (though still very diffuse) priors to feed into the formal analysis of the remaining data. This strategy is used below, where more details are given of the specific experiences with our chosen exchange rate series.

To be explicit technically, note that the Cholesky decomposition is $\mathbf{\Sigma}_t = \mathbf{L}_t\mathbf{L}_t'$ where $\mathbf{L}_t$ is a $q \times q$ lower triangular matrix with non-negative diagonal elements (all positive if $\mathbf{\Sigma}_t$ is of full rank). This can be expressed as $\mathbf{\Sigma}_t = \mathbf{X}_t\mathbf{H}_t\mathbf{X}_t'$ with $\mathbf{L}_t = \mathbf{X}_t{\mathbf{H}_t}^{1/2}$ where the $q \times q$ matrix $\mathbf{X}_t$ is lower triangular with diagonal elements of unity, and $\mathbf{H}_t$ is a $q \times q$ diagonal matrix whose diagonal elements are the squares of those of $\mathbf{L}_t$. Thus we have a full factor representation with $k = q$ and zero idiosyncratic variances. Assuming a restricted factor model with $k < q$ is appropriate, we simply identify the k largest values of $\mathbf{H}_t$ as the conditional variances of the k factors, and the first k columns of $\mathbf{X}_t$ as the corresponding factor loading matrix; then adding non-zero idiosyncratic variances provides a $k-$factor approximation.

3. BAYESIAN INFERENCE AND COMPUTATION

3.1 MCMC analysis

The model as specified so far comprises the basic factor structure (2) with supporting assumptions, specialised to fixed parameters $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ and $\mathbf{X}_t = \mathbf{X}$, and incorporating the SV models of Section 2.4. Model completion for Bayesian analysis requires prior distributions for the full set of parameters $\{\boldsymbol{\theta}, \mathbf{X};\ \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{U};\ \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{s}\}$. Bayesian inference for any specified prior requires the computation and summarisation of the implied posteriors for these model parameters, together with inferences on the factor processes $\mathbf{f}_t$ and the log-volatility sequences $\boldsymbol{\lambda}_t$ and η_{tj} (for each $j = 1, \ldots, q$), over the time window of $t = 1, 2, \ldots, n$ consecutive observations comprising the information set D_n. Computation via MCMC simulation builds on both the work of previous authors in the SV and factor modelling literature, and previous work by the current authors in quite different models with related technical structure (Aguilar and West 1998; West and Aguilar 1997).

To complete the model specification, we assume a prior specified in terms of conditionally independent components

$$p(\boldsymbol{\theta})p(\mathbf{X})p(\boldsymbol{\mu})p(\mathbf{\Phi})p(\mathbf{U})p(\boldsymbol{\alpha})p(\boldsymbol{\rho})p(\mathbf{s}) \qquad (7)$$

where the chosen marginal priors are either standard reference priors or proper priors that are chosen to be conditionally conjugate, as discussed below. The outlook here is to explore the use of reference priors to the extent possible to provide an initial analysis framework. Our prior specifications reflect this view, though, as discussed above, we do use relatively diffuse though proper priors for some model components. Further, specific applications may use alternative prior specifications, both in terms of informative priors on model components and in terms of prior dependencies between parameters, though we do not discuss other prior specifications here.

First, we assume standard reference priors for the univariate entries in the conditional mean $\boldsymbol{\theta}$ and the factor loading matrix $\mathbf{X}$, so that $p(\boldsymbol{\theta})p(\mathbf{X}) \propto$ constant. Note that the prior for $\mathbf{X}$ is, of course, subject to the specified $0/1$ constraints on values in the the upper triangle and diagonal in (3), so the constant prior density applies only to the remaining, uncertain elements. Second, we use independent normal priors for the univariate elements of $\boldsymbol{\mu}, \boldsymbol{\alpha}$, the diagonal elements of $\mathbf{\Phi}$ and the elements of $\boldsymbol{\rho}$. This allows for both reference priors, by setting the prior precisions to zero, and restriction of the values of each ϕ_j and ρ_j by adapting the prior to be truncated to $(0, 1)$. Third, we use an informative inverse Wishart prior for the VAR(1) innovations variance matrix $\mathbf{U}$ in the SV model for the factor volatilities. This is specified with hyperparameters based on prior discounting analysis of an initial, reserved section of data as discussed above. Notice that an improper reference prior on $\mathbf{U}$, could lead to problems as $\mathbf{U}$ and $\mathbf{\Psi}_t$ determine two separate sources of variability in the data that are confounded in the model. This point, rather critical to model implementation and resulting data analysis, is almost implicit in the prior work of Kim, Shephard and Chib (1998). These authors use informative proper priors for innovations variances that parallel our assumptions in their univariate SV models; though they present these priors without further discussion, the propriety is critical in overcoming otherwise potentially problematic confounding issues. Hence initial analysis of previous data, or some other prior elicitation activity, is needed. As mentioned above, our applied development uses variance discounting analyses in providing easy preliminary analysis of a reserved initial section of data as input to this. Though somewhat secondary, the fact that we choose to summarise this prior information in terms of a prior for $\mathbf{U}$ of conditionally conjugate inverse Wishart form does help in the following MCMC analysis of the factor model. Finally, we use a similar idea and strategy in specifying diffuse though proper inverse gamma priors for the elements of $\mathbf{s}$, assuming conditional independence across series.

Iterative posterior simulation uses an MCMC strategy that extends those in existing SV models (Jacquier, Polson and Rossi 1995; Kim, Shephard and Chib 1998) to the multivariate case, introduces elements of MCMC algorithms for Bayesian factor analysis as in Geweke and Zhou (1996), and adds novel components derived from models with latent VAR components developed by the current authors in a quite different context (Aguilar and West 1998; West and Aguilar 1997). We iteratively simulate values of all model parameters together with the full set of values of the latent processes $\mathbf{f}_t, \boldsymbol{\lambda}_t$ and the η_{tj} by sequencing through the set of conditional distributions detailed in

Appendix B. At some stages we have direct conditional simulations, at others we require the introduction of novel Metropolis-Hastings accept/reject steps. We note in passing that, from an algorithmic viewpoint, there are various possible extensions and alternative methods for components of the MCMC analysis, such as in utilising some of the ideas from Shephard and Pitt (1997) for example, though we have not explored such variants yet. We note that Pitt and Shephard (1999b) explore similar factor models and alternative computational schemes for model fitting and sequential analysis; our work was developed independently of, and in parallel to, that study, and some future comparisons of numerical methods will be of interest. Beyond the appendix material here, further technical details are available on request from the authors.

3.2 Sequential Filtering

Recent work in sequential methods of Bayesian analysis using simulation-based approaches has contributed significant methodology of use in broad classes of models, especially for problems of sequential learning on time-varying state parameters such as the stochastic volatility matrix $\mathbf{H}_t$ in our factor models. There is much current research concerned with developing such methods, which go back at least to West (1993) in the statistics literature, though they are not yet general enough to adequately handle larger scale problems that, in addition to evolving states, have several or many fixed model parameters. Simultaneous sequential inference on the factor model parameters $\{\boldsymbol{\theta}, \mathbf{X};\ \boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{U};\ \boldsymbol{\alpha}, \boldsymbol{\rho}, \mathbf{s}\}$ as well as all the time-evolving volatility processes $\{\mathbf{H}_t, \boldsymbol{\Psi}_t\}$ is an open research problem. For this reason, our current study explores sequential forecasting and portfolio allocations for a long section of the time series, based on chosen estimates of all fixed model parameters. The parameter estimates are taken from the MCMC analysis of a prior stretch of historical data. The strategy, which is "honest" from the viewpoint of sequential forecasting, is illustrated in the application to exchange rates below.

The sequential filtering methods adopted for updating posteriors on the time-evolving volatility processes are based on the auxiliary particle filtering (APF) method of Pitt and Shephard (1999a). The APF method is a proven technique for sequential updating of simulation-based summaries of posterior distributions for time-evolving states, and at each time t delivers a current sample of points from the prior $p(\mathbf{H}_t, \boldsymbol{\Psi}_t|D_{t-1})$ and then the resulting posterior $p(\mathbf{H}_t, \boldsymbol{\Psi}_t|D_t)$. These samples are trivially mapped into similar samples from the corresponding prior $p(\boldsymbol{\Sigma}_t|D_{t-1})$ and posterior $p(\boldsymbol{\Sigma}_t|D_t)$ by direct computation using $\boldsymbol{\Sigma}_t = \mathbf{X}\mathbf{H}_t\mathbf{X}' + \boldsymbol{\Psi}_t$ where $\mathbf{X}$ is estimated as just discussed. For portfolio allocations in this framework we require the one-step ahead forecast means and variance matrices of the returns, namely $\mathbf{g}_t = E(\mathbf{y}_t|D_{t-1})$ and $\mathbf{G}_t = V(\mathbf{y}_t|D_{t-1})$ at time $t-1$. These are easily evaluated. The mean $\mathbf{g}_t$ is constant, $\mathbf{g}_t = \mathbf{g}$, simply the estimate of $\boldsymbol{\theta}$ previously computed from the analysis of the initial data segment. The forecast variance $\mathbf{G}_t$ is computed as the sample mean of the Monte Carlo sample of variance matrices $\boldsymbol{\Sigma}_t$ representing $p(\boldsymbol{\Sigma}_t|D_{t-1})$.

We note and stress that, although this analysis ignores the fact that the model parameters are fixed over the course of the sequential portfolio allocation part of the study, the analysis nevertheless provides a coherent basis for model comparisons with a utility function directly measuring real-world performance in terms of cumulative financial returns. Any advances in statistical methods to adequately incorporate learning about the fixed model parameters together with the volatilities in the sequential context will only improve matters. Further discussion of this, and other practical issues and experiences, appear in the exchange rate studies below. Additional support for the efficacy of the APF method in factor models is given in Pitt and Shephard (1999b). More recent developments of this method, that extend the particle filtering techniques to incorporate fixed model parameters, are discussed in Liu and West (2000), with studies of their efficacy in this class of dynamic factor models.

4. STUDIES OF INTERNATIONAL EXCHANGE RATES

4.1 Data and Initial Discounting Analyses

Figure 1 displays time series graphs of the returns on weekday closing spot rates for several currencies relative to the US dollar during the period from January 1st, 1992 to December 31st 1998, with a total of 1827 data points in each series. The currencies are, in order, the German Mark (DEM), British Pound (GBP), Japanese Yen (JPY), French Franc (FRF), Canadian Dollar (CAD) and Spanish Peseta (ESP). The series were obtained from data vendor DataStream. We analyse the one-day-ahead returns as graphed, namely $y_{ti} = s_{ti}/s_{t-1,i} - 1$ for currency $i = 1, \ldots, q = 6$. In some of the graphs we have marked the point $t = 1000$, corresponding to October 31st, 1995, as a reference. We now discuss and summarise *retrospective model analyses* of the data prior to this time point, followed by studies of *sequential updating analyses* and out-of-sample forecasting beyond this time point. The series are ordered in the $\mathbf{y}_t$ vectors as listed above.

Initial exploratory analysis using variance matrix discounting with the controlling discount factor set at $\delta = 0.942$ is summarised in Figures 2. This analysis was chosen following a set of parallel analyses differing only through the value of δ, with δ values spanning a grid over $0.9 - 1$. At each δ value the marginal likelihood function is trivially computed from the discount analysis, resulting in a full marginal likelihood function for δ. This chosen value ($\delta = 0.942$) is the resulting MLE from such analyses of the first 1000 time points only. In Figures 2 we present some point estimates of aspects of the sequence of $\boldsymbol{\Sigma}_t$ matrices extracted from *(a)* the smoothed estimates $\mathbf{S}_{t,1000}$ over the period up to 10/31/95, i.e., $t = 1000$, and *(b)* the sequentially updated estimates $\mathbf{S}_t$ over the period from then until the end of the data at 12/31/98. The initial period presents smoother estimates than the latter, as a

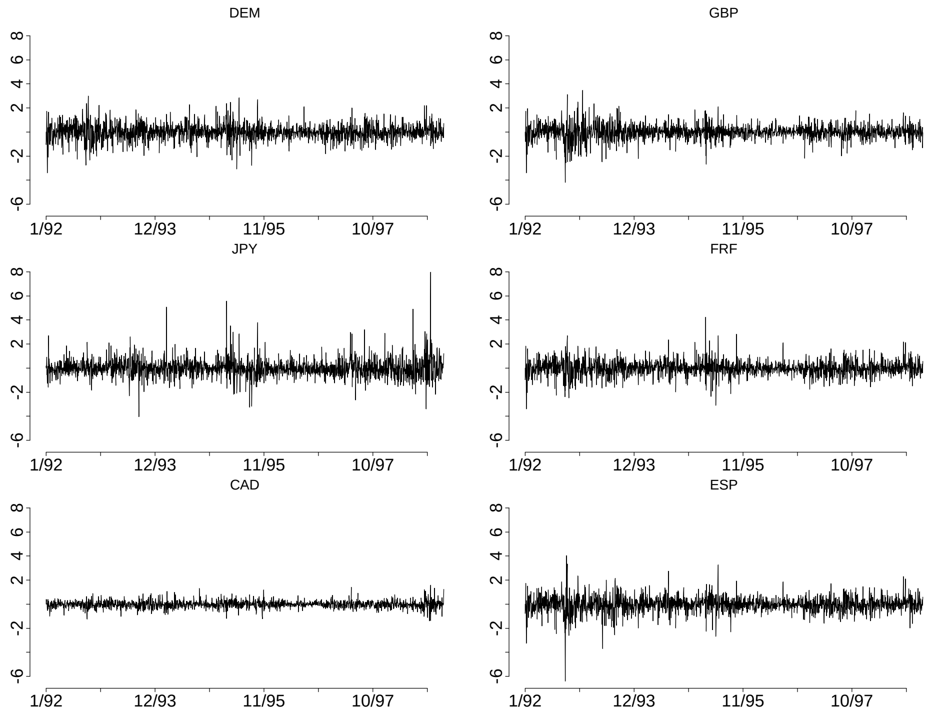

Figure 1. Exchange rate returns time series.

result; restricting to the "one-side" sequential approach from $t = 1001$ onwards leads into our predictive comparisons and is "honest" from the viewpoint of sequential forecasting. Figure 2(a) displays the estimated conditional standard deviations of the series, so computed. Figure 2(b) presents the corresponding estimates of "factor standard deviations" from a full Cholesky decomposition of the estimated matrices at each time point. Naturally, the trajectories of volatilities are smoother prior to $t = 1000$ than after: for $t \leq 1000$, the point estimate of a volatility at a time t is based on all the data up to time 1000, whereas the estimate at a point $t > 1000$ is based only on the data up to that point.

Marked and disparate patterns of volatility are evident in these figures. First, common patterns of volatility in the European currencies are evident in the trajectories in Figure 2(a). By comparing with the trajectories in Figure 2(b) it appears that much of this dependence structure is contributed by the first two "factor" processes – the DEM and GBP factors. Second, the JPY volatility series has features in common with those of the European currencies, but much of the structure in JPY appears to be specific to JPY. This is particularly evident in the highly volatile later period, 1997–1998. Third, the volatility of CAD is at much lower levels than the others, and appears to be almost wholly specific to Canada. This latter point is evidenced by the fact that the "factor" series for CAD is closely similar to the actual volatility series. Fourth, the "factor" trajectories for FRF and ESP are at very low levels compared to their conditional variances, the latter being apparently strongly determined by the DEM and GBP processes.

The suggestion is therefore for 3 factors with the series in the given order, and that we should expect the idiosyncratic components both JPY and CAD, in particular, to be of major importance in determining overall volatilities of those currencies. The specification of 3 factors with this order of currencies might have been anticipated on economic grounds, of course. Below we discuss fitting such a model. Additional commentary on models with different numbers of factors, and with currency series in various orders, is given in the concluding section.

4.2 Dynamic Factor Analysis

Taking $q = 6$ and $k = 3$ in the dynamic factor model (2) provides a maximal specification: under the assumed structure of the factor loadings matrix (3), and assuming each of the ψ_{tj} to be strictly positive for all t, the number of factors must necessarily be no greater than three. Hence, in addition to being suggested by the discounting analyses, this serves here as an encompassing model; if fewer than three factors are supported by the data, that fact will be reflected in posterior inferences about factor loadings and variances.

As discussed above, we specify proper though diffuse priors on some of the key model parameters, and develop these priors from a discounting analysis of a short section of 200 observations immediately prior to the first time point in January 1992 of the displayed data. Denote this time period by $t = -199, \ldots, -1, 0$. Using a discount model with $\delta = 0.942$ over this preliminary data, we compute Cholesky decompositions of estimated variance matrices, and extract corresponding estimates of the

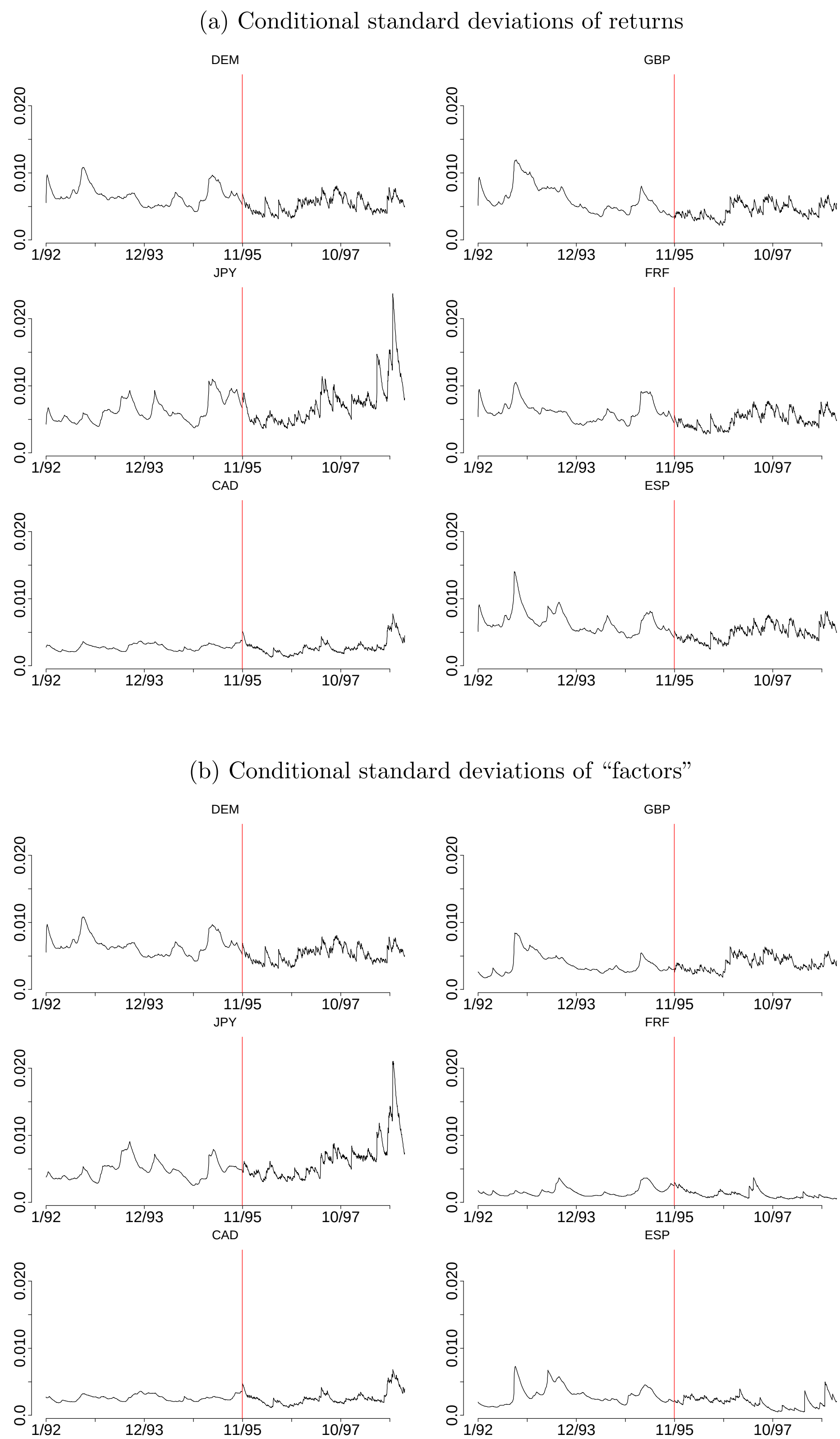

Figure 2. (a) Time trajectories of estimated conditional standard deviations of the returns time series from the discount analysis (square roots of the diagonal element of the estimated $\mathbf{\Sigma}_t$ sequence.). (b) Corresponding trajectories of the conditional standard deviations of the "factors" underlying the returns time series from the discount analysis (diagonal elements of the Cholesky decomposition of the estimated $\mathbf{\Sigma}_t$ sequence).

three largest variances – these are estimates over time of the λ_{tj} in our factor model over $t = -199, \ldots, -1, 0$. To assess priors for the parameters of VAR model for the factor volatilities, three separate AR(1) models were fitted to the log-volatilities so computed, using standard reference Bayesian analyses. This provides posteriors for the AR parameters and innovations variances, in each volatility series marginally, that we take as ball-park initial estimates to be used to specify an informative prior for $\mathbf{U}$ prior to analysis of the remaining data. This preliminary analysis gave approximate prior means of the three innovations variances around 0.001–0.002. With this in mind, we chose the prior for $\mathbf{U}$ in the factor model analysis to be inverse Wishart $W_{r_0}^{-1}(\mathbf{U}|\mathbf{R}_0)$ with $r_0 = 5$ degrees of freedom (a very low value that ensures a proper though diffuse prior) and $\mathbf{R}_0 = 0.0015\mathbf{I}$, appropriately centering the prior for $\mathbf{U}$. Note that the prior does not anticipate correlations across volatility processes, though this could easily be done. For the AR models for the idiosyncratic variances, the only proper prior components used are those for the innovations variances s_j. From the above Cholesky decomposition analyses we extract residual components by subtracting the estimates of the 3 factors, and simply assess the average level to provide ball-park estimates of levels of residual variability. This leads to inverse gamma priors centred at 0.003 in each case and also with 5 degrees of freedom. We note that reanalyses with more informed priors (e.g., with up to 100 degrees of freedom) have been explored, and these lead to closely similar summary inferences and portfolio results.

The MCMC analysis of this factor model involved a range of experiments with Monte Carlo sample sizes and starting values, and MCMC diagnostics. Our summary numerical and graphical inferences are based on over 20,000 simulations of posteriors, generated following a 5,000 burn-in period. We subsample a set of 1,000 spaced 20 apart so as to break correlations and record resulting samples for graphical display purposes. Many repeat studies confirm the convergence of the MCMC analysis and the adequacy of the posterior summaries – mainly point estimates – of factor model parameters. Standard MCMC diagnostics confirm our conclusions.

Summary graphs appear in Figures 3 and 4, and related numerical posterior summaries in Tables 1 and 2. We discuss major aspects of the analysis in connection with these figures and tables. First, we need to be clear that this MCMC analysis is applied only over times $t = 1, \ldots, 1000$, up to the end of October 1995. We discuss the later time period in the following section.

First consider the numerical summaries as tabulated. In each table we present posterior means for various parameters, with approximate posterior standard deviations in parentheses. Table 1 displays summaries for the factor loading matrix $\mathbf{X}$, the returns level $\boldsymbol{\theta}$, and the parameters of the idiosyncratic volatility models. For $\mathbf{X}$, note that the first column positively weights the first, DEM factor for all currencies but CAD. As is to be expected on economic grounds, the weights are very high for FRF and ESP, positive but lower in value for GBP, then lower again for JPY. The second column of $\mathbf{X}$ shows only minor negative weights for JPY and CAD on the GBP factor, and the third column gives similarly small negative weights to the JPY factor for FRF, CAD and ESP, other weights being essentially zero. Turning to the idiosyncratic volatility model parameters, it is clear that the idiosyncratic processes are very highly persistent, with autoregressive parameters in the $0.93 - 0.99$ range. The table also indicates the higher levels of variability evidenced in the JPY series, and the relatively low levels in CAD.

Table 2 presents summaries for the VAR factor volatility model parameters. All three factor volatilities are highly persistent processes, with ϕ_j parameters at or above 0.95. Base levels of volatility are similar in all three cases, as measured by the factors $\exp(\mu_j/2)$ on the volatility scales. Of the three, the GBP factor has somewhat lower levels of marginal volatility. The estimated correlations in $\mathbf{W}$ indicate high levels of positive dependencies among the innovations of the three factor SV processes. Upswings and peaks in volatilities related to, or driven by, international financial events tend to occur together, and hence the need for a model that allows for such dependencies. As perhaps might be expected, the correlation between the SV processes of the DEM and GBP factor has a higher posterior estimate than those between JPY and either of the EU factors.

Figure 3(a) displays posterior means of the conditional standard deviations of the returns series based on MCMC analysis of the data up to $t = 1000$. There are general similarities between these trajectories and the related trajectories from the discount analyses in Figure 2, as is to be expected. The major difference is that the trajectories are more appropriately peaked and variable in the factor model analysis. The discount method produces more heavily smoothed trajectories, pointing to one of the major known drawbacks of that method, and so obscures some of the evident peaks in volatility related to major economic changes and events. A key such event was Britain's withdrawal from the EU exchange rate agreement (the ERM), in late 1992, and this led to marked increases in volatilities across currencies, as reflected in the trajectory plots in Figure 3(a) around the end of 1992 and into 1993. A second period of increased volatility – again materially impacting several of the currencies – occurred in early 1995 with major changes in Japanese interest rate policies as a response to a weakening Yen and a move toward financial restructuring in Japan.

Figure 3(b) displays posterior means of the three factors in the dynamic factor model, together with their standard deviations based on MCMC analysis of data up to $t = 1000$. There are some general similarities between the volatility plots here and the related trajectories from the discount analyses in Figure 2, as is to be expected. Again, the major difference between the factor model and the discount model over this period is that the discount method produces more heavily smoothed trajectories. Notice that the three volatility processes in Figure 3(b) display commonalities – major peaks at similar times – consistent with

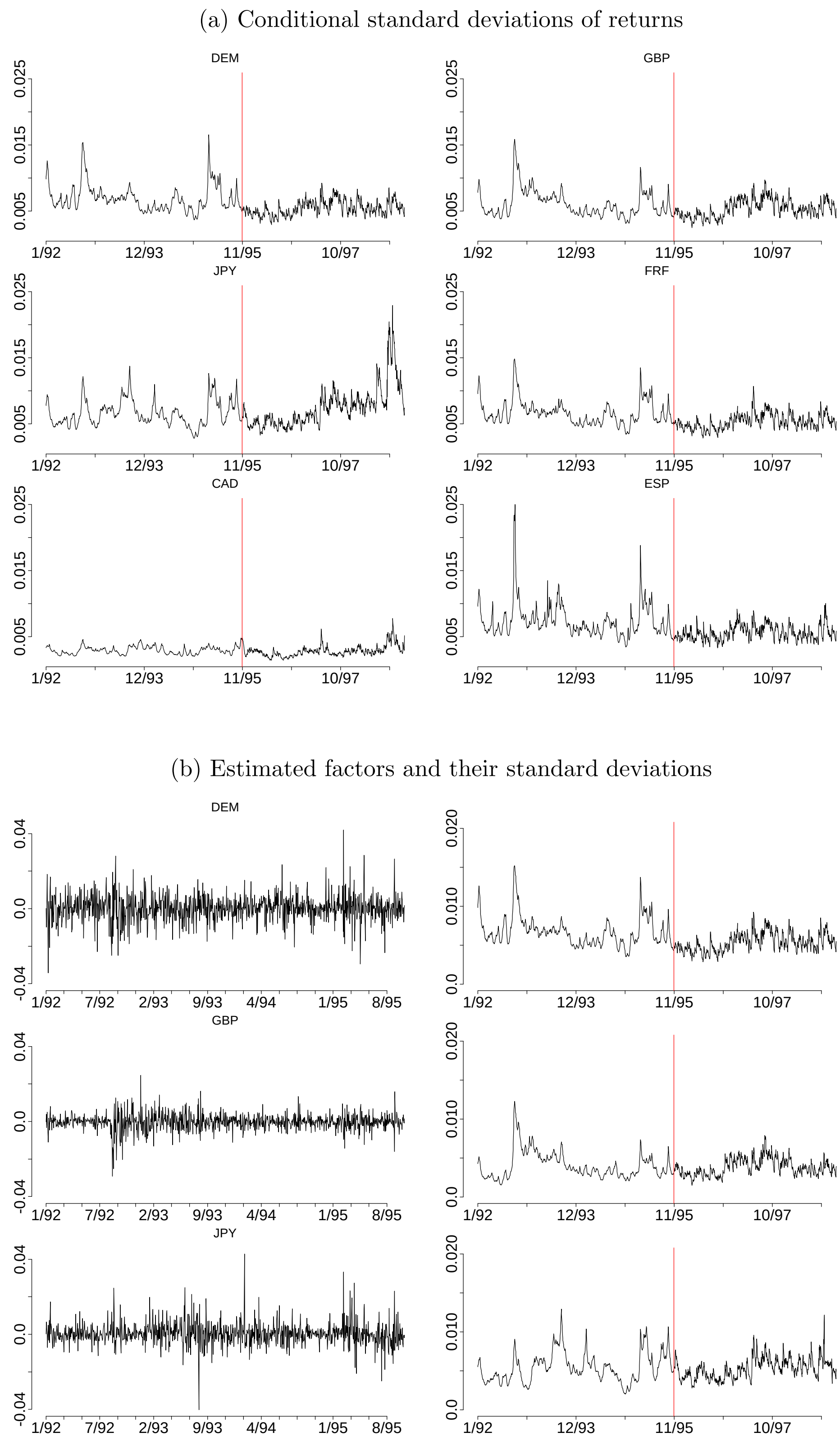

Figure 3. *(a) Posterior means of the conditional standard deviations of the returns series based on MCMC analysis of the data up to $t = 1000$, and then based on particle filtering from that point to the end of the time period. (b) Posterior means of the three factors and their standard deviations up to $t = 1000$ based on MCMC analysis of data to that time point. The volatility graphs also include the sequentially updated estimates from $t = 1001$ based on particle filtering from that point to the end of the time period.*

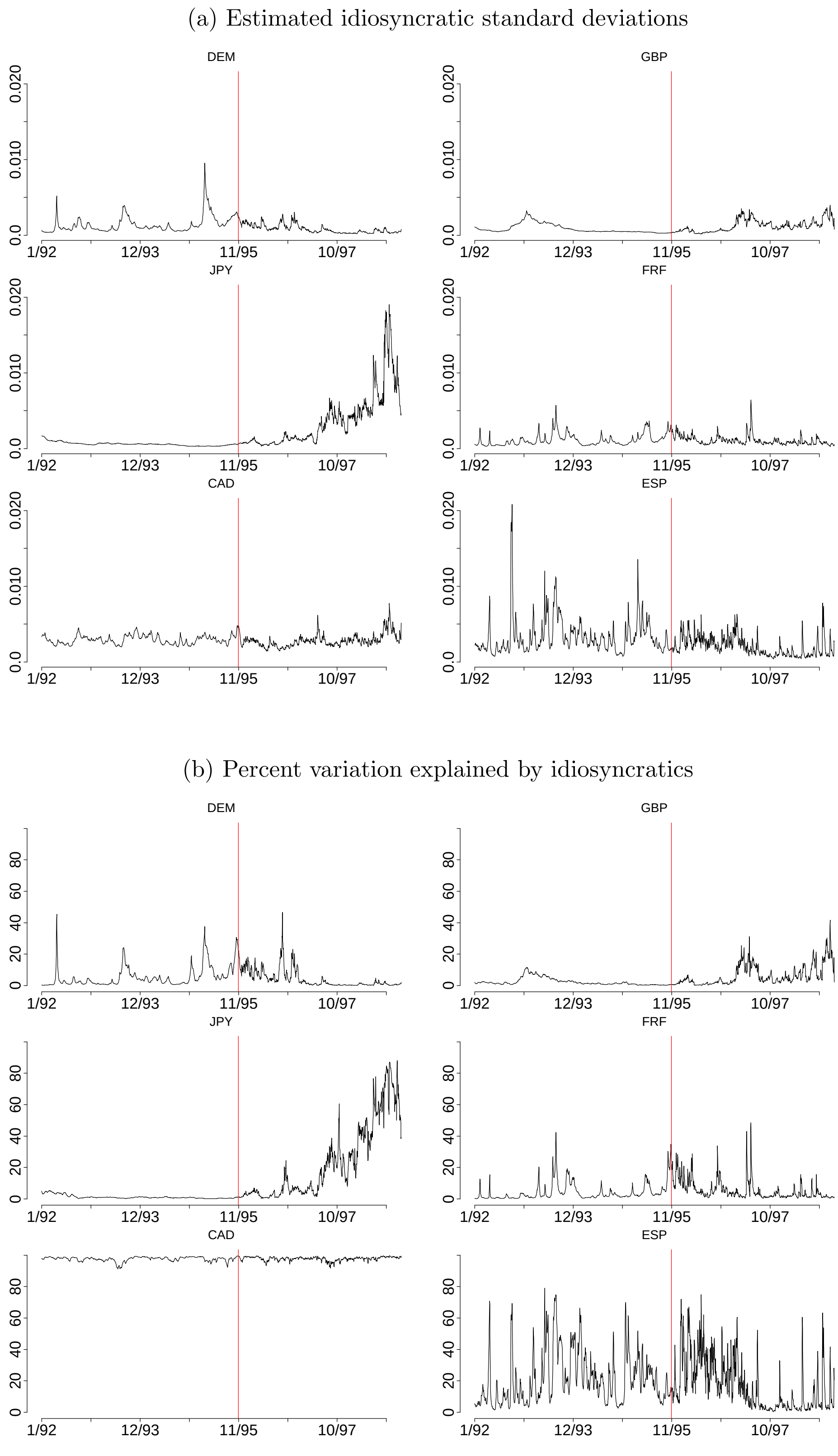

Figure 4. (a) Posterior means of the idiosyncratic standard deviations of the returns series based on MCMC analysis of the data up to $t = 1000$, and then based on particle filtering from that point to the end of the time period. (b) Percentage variation in the return series explained by the idiosyncratic variances, computed based on posterior means of all variances at each time point. As above, the estimates are based on MCMC analysis of the data up to $t = 1000$, and then based on particle filtering from that point to the end of the time period.

Table 1. Posterior means (SDs) for the factor loadings matrix, mean vector and parameters of the idiosyncratic SV models based on MCMC analysis of data up to 10/31/1995. (All truncated to 2 decimal places, so the zero SD entries indicate values smaller than 0.01.)

	Factor loadings matrix $\mathbf{X}$			$\boldsymbol{\theta}$ $(\times 10^{-4})$	ρ_j	$e^{\alpha_j/2}$ $(\times 10^{-3})$	$\sqrt{s_j/(1-\rho_j^2)}$
DEM	1	0	0	−1.50 (1.28)	0.99 (0.00)	0.52 (0.23)	1.45 (0.20)
GBP	0.67 (0.03)	1	0	0.17 (0.98)	0.99 (0.00)	0.57 (0.21)	1.54 (0.25)
JPY	0.53 (0.02)	−0.05 (0.03)	1.00 (0.00)	−2.33 (1.18)	0.99 (0.00)	0.74 (0.28)	1.59 (0.38)
FRF	0.97 (0.01)	−0.00 (0.01)	−0.02 (0.01)	−1.24 (1.25)	0.98 (0.03)	0.30 (0.24)	1.33 (0.21)
CAD	−0.03 (0.01)	0.04 (0.02)	−0.06 (0.01)	−1.23 (0.59)	0.94 (0.01)	0.26 (0.01)	0.68 (0.06)
ESP	0.95 (0.01)	−0.00 (0.01)	−0.02 (0.01)	−1.59 (1.26)	0.93 (0.01)	0.20 (0.02)	1.63 (0.12)

Table 2. Posterior means (SDs) for parameters of the factor SV model based on MCMC analysis of data up to 10/31/1995.

	ϕ_j	$e^{\mu_j/2}(\times 10^{-3})$	SDs and correlations in $\mathbf{W}$		
$j=1$	0.95 (0.01)	5.73 (0.32)	0.68 (0.07)		
$j=2$	0.96 (0.01)	3.68 (0.28)	0.52 (0.08)	0.81 (0.08)	
$j=3$	0.96 (0.01)	4.91 (0.37)	0.43 (0.08)	0.45 (0.09)	0.78 (0.09)

the positive dependencies across processes evident in the estimated correlations in the factor SV model.

Figure 4(a) displays posterior means of the idiosyncratic standard deviations of the returns series based on MCMC analysis of the data up to $t = 1000$. One of the key features evident here is the fact that the idiosyncratic variance process for JPY was almost negligible over this period up to 1995. Additional features of note are the very low levels of idiosyncratic variability for DEM and GBP, and, as expected, for CAD. Further, there is significant idiosyncratic variability in ESP as compared to FRF, reflecting the fact that FRF is more closely tied to the dominant European currencies, and DEM in particular through the ERM, than is ESP.

Figure 4(b) displays levels of variation in each of the currency return series contributed by their idiosyncratic variances, computed as percentages of the overall conditional variances. The estimates are computed by estimating all variances by their posterior means based on MCMC analysis of the data up to $t = 1000$. Of major note here is CAD; the trajectory indicates that volatility fluctuations in CAD are almost wholly idiosyncratic and unrelated to the European and Japanese factors.

5. SEQUENTIAL FORECASTING AND PORTFOLIO ALLOCATIONS

5.1 Perspective

Model comparisons are made with explicit focus on one-step forecast accuracy in the context of dynamic portfolio allocations, essentially following the perspective of Quintana (1992), Putnam and Quintana (1994), and Quintana and Putnam (1996). A similar perspective is adopted in Polson and Tew (1997) though with very different models. Our comparisons in this section are based on sequential updating and one-step ahead forecasting, with resulting portfolio allocation decisions implemented one-step ahead. At each time point $t-1$, we suppose that an existing investment in the various currencies under study may be reallocated according to a portfolio $\mathbf{a}_t$ for the next time point. The elements of $\mathbf{a}_t$ are the amounts invested in the corresponding currency. For this comparative analysis, we assume no transaction costs and that we may freely reallocate dollars instantaneously to long or short positions across the currencies. These allocation decisions are made sequentially; the choice of $\mathbf{a}_t$ is made at time $t-1$ based on the current, one-step ahead predictive distribution for $\mathbf{y}_t$ conditional on current and past data and information D_{t-1}. The realised portfolio return at time t is $r_t = \mathbf{a}_t'\mathbf{y}_t$, and models may be compared on the basis of cumulative returns over chosen time intervals.

Our study involves the general Markowitz mean-variance optimisation, applied at each time point one-step ahead. Thus the utility structures used for portfolio allocations require evaluation and sequential revision of the one-step ahead forecast means and variance matrices of the returns, denoted by $\mathbf{g}_t = E(\mathbf{y}_t|D_{t-1})$ and $\mathbf{G}_t = V(\mathbf{y}_t|D_{t-1})$ at time $t-1$. Computations in the variance matrix discounting analysis are simple and standard: these one-step ahead forecast distributions are multivariate T with easily updated parameters $\mathbf{g}_t$ and $\mathbf{G}_t$. Sequential computations and updating in dynamic factor models are quite non-standard, and challenging, and are here based on sequential methods of updating and generating Monte Carlo approximations to prior and posterior distributions. These deliver sets of samples representing the key distributions $p(\boldsymbol{\Sigma}_t|D_{t-1})$, and updates to these samples as t changes and new data are processed. These samples deliver Monte Carlo approximations to the required one-step ahead forecast means and variance matrices, and these are used in the portfolio allocation computations.

5.2 Sequential Analysis and Approximations

As discussed above, we explore sequential forecasting and portfolio allocations from $t = 1001$ onwards, using particle filtering methods to sequential update posteriors for volatility processes and so feed into the one-step-ahead portfolio decisions. We have earlier partitioned the data at time $t = 1000$, and applied the MCMC analysis to that initial time period. From this analysis, we compute the posterior means of all model parameters in the above set, and from that point $t = 1000$ onwards behave as if these parameters are all fixed at these estimates. Over the remaining time interval, a total of 827 days, we then apply the auxiliary particle filtering (APF) method of Pitt and Shephard (1999a) to sequentially revise posterior distributions for the volatility processes $\{\mathbf{H}_t, \mathbf{\Psi}_t\}$. As discussed in subsection 3.2, this easily delivers sequentially updated Monte Carlo estimates of one-step ahead forecast means and variance matrices of the returns, $\mathbf{g}_t = E(\mathbf{y}_t|D_{t-1})$ and $\mathbf{G}_t = V(\mathbf{y}_t|D_{t-1})$, as required. Recall that we assume a constant mean for our studies here, so that $\mathbf{g}_t = \mathbf{g}$, the estimate of $\boldsymbol{\theta}$ based on the first 1000 observations.

This use of model parameters at their posterior means computed only on data prior to time $t = 1000$ leads to extremely interesting results. One such is the fact that the resulting estimates of model parameters based on this initial data segment alone differ negligibly from the corresponding estimates from the MCMC analysis based on the entire time series. This suggests that the sequential analysis and portfolio results will be strongly indicative of the kinds of results we would achieve in extended analysis that includes learning on the parameters in the sequential updating phase, were future developments in methodology to lead to appropriate methods to include parameters (see Liu and West 2000 for some such developments).

Some of the results of this analysis are given in the four sets of graphs already discussed, now focusing on the time period after $t = 1000$. Figure 3(a) displays sequentially updated, particle filtering-based posterior means of the conditional standard deviations of the returns series from $t = 1000$ to the end of the time frame. Figure 3(b) displays sequentially updated, particle filtering-based posterior means of standard deviations of the three factors from $t = 1000$ to the end of the time frame. Figure 4(a) displays sequentially updated, particle filtering-based posterior means of the idiosyncratic standard deviations of the returns series from $t = 1000$ to the end of the time frame. Figure 4(b) displays levels of variation in each of the currency return series contributed by their idiosyncratic variances, computed as percentages of the overall conditional variances. The estimates are computed by estimating all variances by their sequentially updated, particle filtering-based posterior means from $t = 1000$ to the end of the time frame.

Many of the earlier comments about the patterns of volatility in the returns and the three factors over $t = 1, \ldots, 1000$ might be echoed here in connection with the following years. Perhaps the main additional point to highlight is the fact that, from 1995 onwards the fluctuations in the idiosyncratic component of JPY became quite substantial, with highly volatile fluctuations in the few years following financial restructuring in Japan.

5.3 Portfolio Allocation Rules

The decision context focuses on choosing the portfolio $\mathbf{a}_t$ with a specified utility function that balances the competing goals of high return and low risk, risk being here measured by variances. The portfolio $\mathbf{a}_t$ is so optimised at each time $t-1$ for implementation, and the resulting actual returns are then available at time t. The standard approach here adopts a specified *target* return m and considers only portfolios with that target as one-step ahead expectation; i.e., restrict to portfolios satisfying $\mathbf{a}_t'\mathbf{g}_t = m$. Among such portfolios, that chosen is the one that minimises the one-step ahead variance of returns, namely $\mathbf{a}_t'\mathbf{G}_t\mathbf{a}_t$. The optimal portfolio has an important dual property: it also maximises the one-step ahead expected return $\mathbf{a}_t'\mathbf{g}_t$ among all portfolios with common risk $\mathbf{a}_t'\mathbf{G}_t\mathbf{a}_t = \sigma^2$, where σ^2 and m are suitably related.

Two variants of this strategy are considered and compared. First, the traditional constrained portfolio fixes the total sum invested at each time point. Thus allocation vectors are additionally constrained at all times via $\mathbf{a}_t'\mathbf{1} = 1$. The solution is

$$\mathbf{a}_t^{(m)} = \mathbf{G}_t^{-1}(a_t\mathbf{g}_t - b_t\mathbf{1})$$

where $a_t = \mathbf{1}'\mathbf{e}_t$ and $b_t = \mathbf{g}_t'\mathbf{e}_t$ with

$$\mathbf{e}_t = \mathbf{G}_t^{-1}(\mathbf{1}m - \mathbf{g}_t)/d_t$$

and

$$d_t = (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{1})(\mathbf{g}_t'\mathbf{G}_t^{-1}\mathbf{g}_t) - (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{g}_t)^2.$$

By contrast, unconstrained portfolio allocation strategies allow values of $\mathbf{a}_t$ to vary more freely, subject only to the mean-variance constraints above. This means that we may choose each allocation without regard to resources, permitting arbitrary long or short positions across the currencies. This typifies the practical working context in global investments in large financial institutions, and is in line with recent work with discount models (Quintana and Putnam 1996). As in the constrained case, the optimal portfolio allocation $\mathbf{a}_t$ is the variance minimising portfolio among all portfolios with expected return equal to the target m. Also, again as in the constrained case, the optimal portfolio has the dual property that it maximises the one-step ahead expected return among all portfolios with common risk $\mathbf{a}_t'\mathbf{G}_t\mathbf{a}_t = \sigma^2$, where σ^2 and m are suitably related.

Under an unconstrained strategy, the optimum allocation at time t is given by

$$\mathbf{a}_t^{(*m)} = c_t\mathbf{G}_t^{-1}\mathbf{g}_t$$

where

$$c_t = m/\mathbf{g}_t'\mathbf{G}_t^{-1}\mathbf{g}_t.$$

5.4 Comparison of Models and Portfolios

The sequential analysis and portfolio allocations were implemented from the baseline $t = 1000$ at which model parameters were estimated. The portfolio allocations are based on implementing the above strategy with a fixed target return $m = 0.00016$ at each time step $t = 1001, \ldots,$ to the end of the time period. This study compares the portfolio returns under this strategy from the dynamic factor model and the optimal discount approach. A baseline comparison is also provided by the naive and trivial equally-weighted portfolio allocation $\mathbf{a}_t = \mathbf{1}/q$ for all t. Some summaries of the analysis appear in the figures as

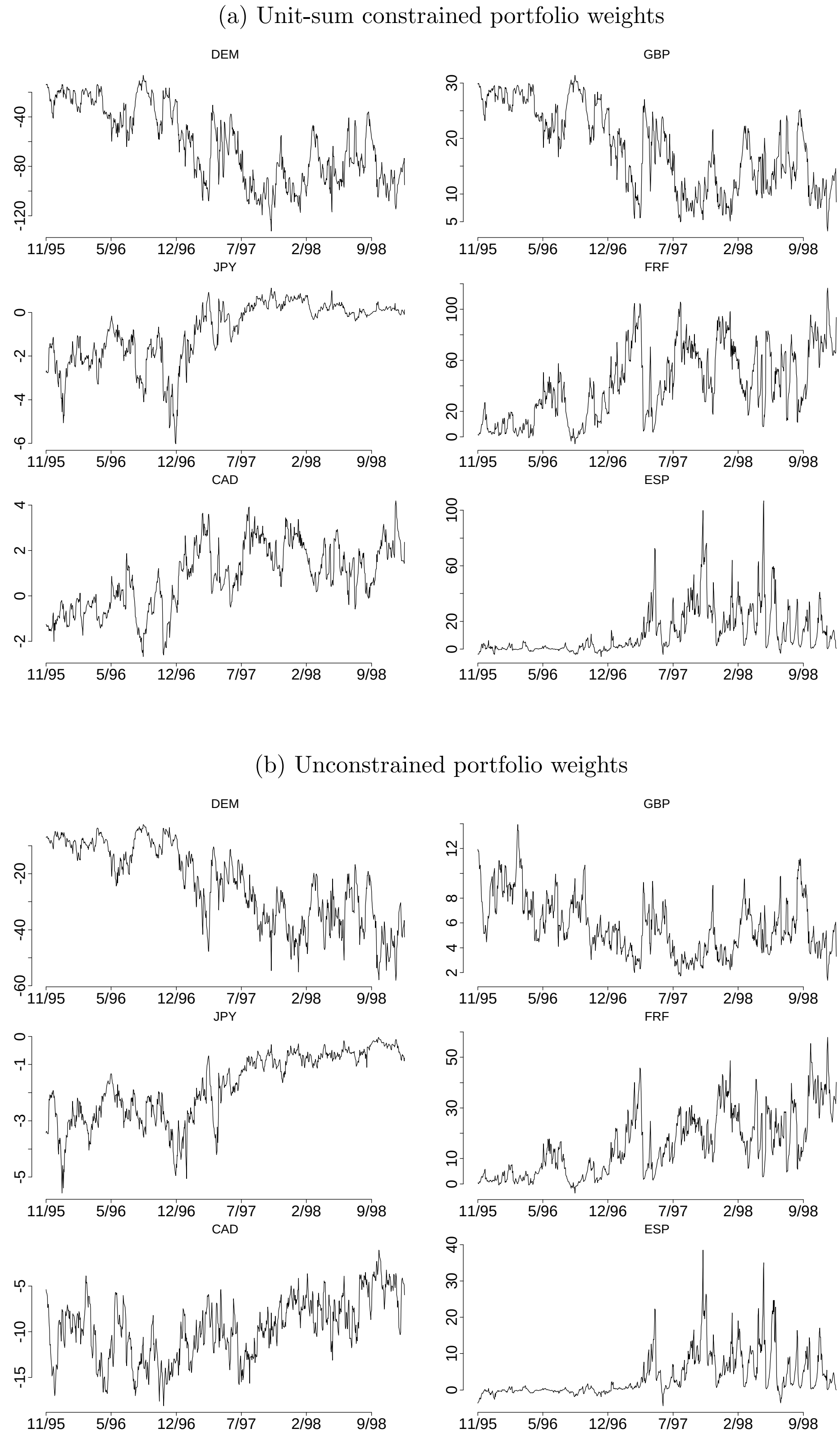

Figure 5. Sequential factor analysis results: portfolio weights in (a) the unit-sum, constrained portfolios, and (b) the unconstrained portfolios.

now detailed. Figure 5 graphs trajectories of the optimal portfolio weights in both the unit-sum, constrained portfolios (frames (a)) and the unconstrained portfolios (frames (b)). Figure 7 graphs trajectories of cumulative returns from both the constrained and unconstrained portfolio studies, compared with those from the optimal discount model. As a benchmark of poor performance, the returns from the equally weighted portfolio $\mathbf{a}_t = \mathbf{1}/6$ for each t are also graphed. The figure displays returns cumulated over moving 30- and 90-day periods, and also across the entire period from $t = 1001$ to the end of the data frame. In viewing the four frames in this figure, note the differing scales of the returns axis, chosen for clarity of presentations.

In viewing these graphs it is important to bear in mind that our study is focussed wholly on the impact on portfolios of changing patterns of volatility, and on comparisons of different models of volatility. We are not modelling dynamic structure in the means of returns, so the constant estimate of expected returns used across the time period is simply the posterior mean $\mathbf{g}$ of $\boldsymbol{\theta}$ based on the MCMC analysis up to $t = 1000$. In particular, the first element of $\mathbf{g}$ is negative, implying a negative expected return on DEM around the end of 1995. This is reflected in the portfolios across the time period up to the end of 1998 through negative portfolio weights on DEM, i.e., the portfolios are short on the German mark. More discussion on this appears below.

Among the notable features of the trajectories of constrained portfolio weights in Figure 5(a) are the fact that both JPY and CAD have generally very low weights across the period. The elements of $\mathbf{g}$ indicate a mean for CAD very close to zero, as is expected on economic grounds. Also, note that the weights on JPY move towards zero in the later periods of very high idiosyncratic volatility, reflecting increased risk aversion with respect to that currency. In connection with the short positions on DEM driven by the negative element of $\boldsymbol{\theta}$, note that, as it happens, both constrained and unconstrained portfolios adopt corresponding long positions on FRF, and the portfolio weights on DEM and FRF essentially offset each other.

Comparison of Figures 5(a) and 5(b) indicate very similar patterns in the trajectories of portfolio weights. What is more interesting is that the unconstrained portfolios adopt overall short positions across the period, as evidenced by the trajectory of $\mathbf{a}_t'\mathbf{g}$ in Figure 6. This is strongly influenced by the negative mean, and resulting short positions, for DEM, and reflects the response of the portfolio to the levels of volatility that, across the period, are generally high relative to the target mean return.

Comparison of these two factor model portfolios are sharpened in examination of the cumulative returns arising under each across this time period, in frames (a) and (b) of Figure 7. The full line in Figure 7(a) represents the cumulative return over the entire period using the constrained portfolio in the factor model; that in Figure 7(b) is the return using the unconstrained portfolio. The main point is clear: the lack of a resource constraint – in this case, implying access to unlimited short positions – leads to the unconstrained portfolio outperforming the constrained by a factor of nearly 4 over this period of 827 days.

Also graphed in Figure 7(a) are the cumulative returns resulting from the constrained portfolio used in the optimal discount model (as a dotted line), and from the naive, equal-weight portfolio (dashed line). It is evident that, when compared using the same constrained portfolio strategy, the use of sequential updating in the factor model clearly dominates the optimal discount approach. The corresponding trajectories in Figure 7(b) confirm this conclusion in the case of unconstrained portfolios.

Finally, frames (c) and (d) of Figure 7 display trajectories of cumulative returns over shorter investment horizons – 30 days and 90 days, respectively. In each, the full line represents the factor model with unconstrained portfolios, the dotted line represents the optimal discount model with unconstrained portfolios, and the dashed line represents

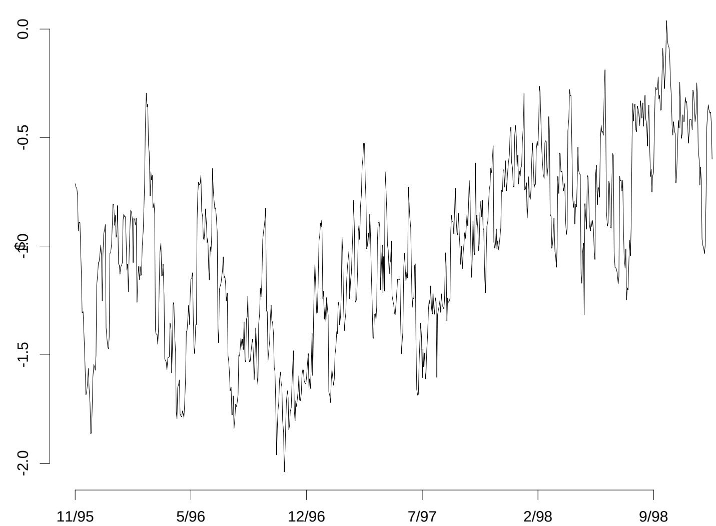

Figure 6. Sequential factor analysis results: Total implied investments in the unconstrained portfolios.

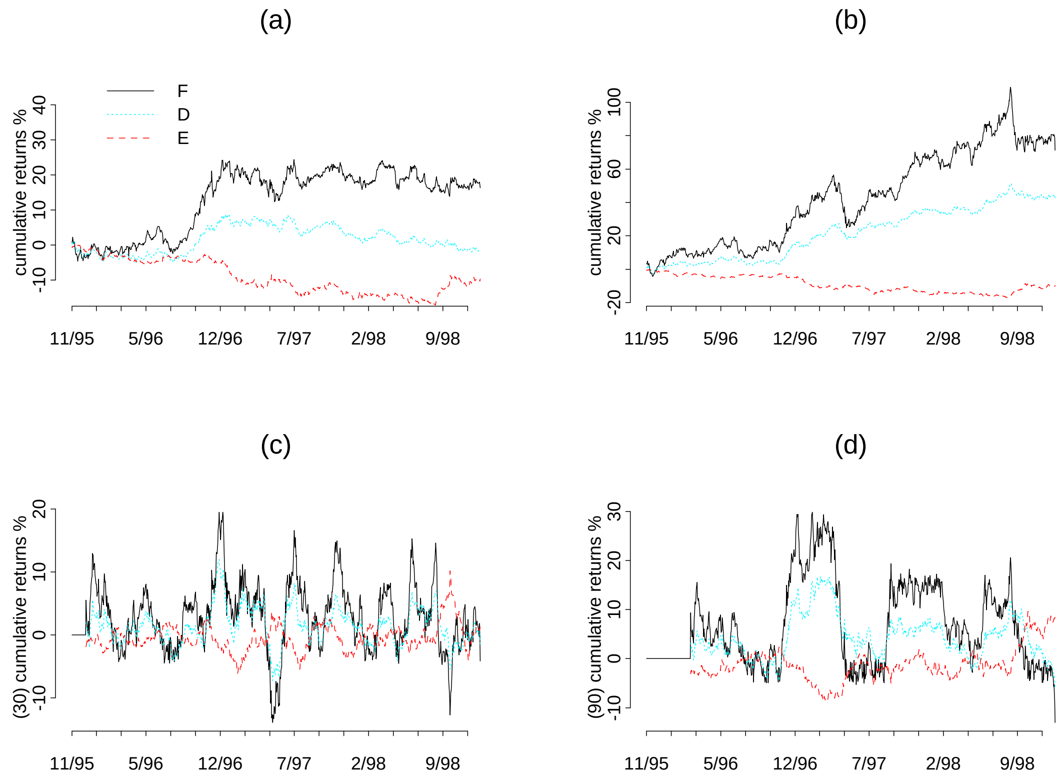

Figure 7. Sequential factor analysis results: Cumulative returns from factor ("F") and discount ("D") model-based portfolios, together with those from a basic equally weighted ("E") portfolio. (a) constrained portfolio, overall returns; (b) unconstrained portfolio, overall returns; (c) 30 day returns from unconstrained portfolios; (d) 90 day returns from unconstrained portfolios.

the naive, equal-weight portfolio. Here the similarities in portfolio performance between the discount methods and dynamic factor models are really very clear. The major differences arise in the periods of radically increased volatility, where the factor model is able to capitalise markedly in terms of short-term gains relative to the discount method. This shorter term responsiveness in the factor models leads to marked swings in portfolio structure that the unconstrained allocations significantly capitalise upon, and this has a persistent effect on overall cumulative returns thereafter.

6. ADDITIONAL DISCUSSION AND CONCLUDING COMMENTS

Our investigations indicate the feasibility of formal Bayesian analysis of structured dynamic factor models. The analysis is accessible computationally with nowadays moderate computational resources, and our empirical studies suggest that the analysis will be manageable with 20-30 dimensional time series and several factors. We are currently investigating more extensive applications in short-term forecasting and on-line portfolio allocations with higher dimensional models for longer-term exchange rate futures. The example here is suggestive of potential benefits, and supportive of the view that exploiting systematic volatility patterns via factor structuring may yield substantial improvements in short-term forecasting and decision making in dynamic portfolio allocation, especially in the unconstrained optimisation. The discount method does reasonably well at times, though is clearly eventually dominated in terms of cumulative return trajectories by the factor model. To partly offset this, however, we note that the shorter term focuses in Figure 7(c) and (d) indicate that the overall dominance of the factor model approach is partly based on its ability to adapt quickly to a small number of major changes in volatility patterns. For this reason, financial analysts might prefer the more easily implemented discount methods coupled with informed prospective interventions.

The dynamic factor models illustrated are amenable to direct implementation using our customised MCMC methods with the minimal/reference prior specifications we have used here. Our use of the variance discounting method on a reserved initial section of the data to provide input to informative priors is important in identifying "ball-park" scales for the $\mathbf{U}$ matrix of the VAR(1) SV model. Though not pursued here, other aspects of such preliminary analyses may be used to determine informative priors for other elements of the factor model. The established discounting methods are, relative to dynamic factor models, trivial to implement in the current context, a fact that is important in using discount methods to specify partial prior structure in the dynamic models. Our empirical findings indicate that, not surprisingly with this kind of data, moderately adaptive discount methods fare well in time of slow change in volatility levels and patterns, but are relatively uncompetitive in cases of more marked structural change. This is to be expected. Looking ahead, models and approaches that attempt to simplify the process of factor modelling, perhaps somehow integrating elements and concepts of variance matrix discounting into a specified factor structure, may be attractive from a computational/implementation viewpoint. We are currently investigating such a synthesis of approaches.

In the factor model context per se, one of the important features in our model is the use of correlated innovations in the multivariate AR model for the factor volatilities. The applied relevance of this is quite apparent from the exploratory analyses using discount models, and confirmed in the factor model analysis through the posterior distributions for both factor volatilities and the correlation parameters. Some important practical issues relate to the questions of the ordering of individual time series under the specific structure we adopt for factor loadings, and to the choice and assessment of the number of factors. On the first issue, it needs to be restressed that, from the viewpoint of forecasting and portfolio allocations, the ordering is irrelevant. The ordering is, of course, relevant in connection with model fitting and assessment, the interpretation of factors if such is desired, and the choice of the number of factors. It is often desirable to use a specific ordering to define and interpret the factors. In such cases, the ordering becomes a modelling decision to be made on substantive grounds, rather than an empirical matter to be addressed on the basis of model fit. This is of particular relevance in connection with the potential for factor models to improve sequential decisions by allowing informed interventions on specific factors. For example, having identified the "Japan" factor in our financial time series model, we now have opportunities for selective interventions. Anticipating a specific change in Japanese financial policy, for example, we may intervene to "current" prior distributions for parameters related only to that factor in the model. Such potential for selective interventions was, in fact, one of our key initial motivations for exploring factor models, and provides, we believe, a major incentive for practitioners to take interest.

In connection with the ordering and the number of factors, we report some initial results and experiences in static factor models – with constant volatilities – in Lopes and West (1998). That work extends MCMC analysis of factor models to include inference on the number of factors. Potential exists to develop such approaches to dynamic factor models, although that is currently unexplored. Of more direct interest here, perhaps, are empirical findings in a range of practical studies with the exchange rate series. First, running MCMC analyses in models with too many factors generally leads to problems in convergence of the simulations. For example, in a model with three factors when the data are really consistent with just two, the simulated values of the third factor will naturally be close to zero, although the MCMC will behave poorly and show evidence of non-convergence. We have experienced this in studies of the exchange rates prior to 1992 in an extended data set going to back to 1986. Before the withdrawal of the UK from the ERM in fall 1992, the GBP series generally tended to follow the European pattern, dominated by the DEM factor. The series prior to that point are much better explained with a two factor model – one EU factor and one Japan factor – than with three, and the above problem of non-convergence of the MCMC arose. After fall 1992, a three factor model is much more appropriate, with the third GBP factor explaining meaningful levels variation in the FRF and ESP series as well as those of the GBP series. Methodologically, we conclude that poor convergence characteristics can point to a model mis-specified with too many factors. The reverse of this, a model with too few factors, is rather more readily identified through patterns of co-movements in the estimated trajectories of the idiosyncratic variances, consistent with one or more additional factors. Again, we have experienced this in analysing the post-1992 series with a two factor model. This kind of exploratory analysis is critical and necessary to a full understanding of the use of these models, and will remain so even when more formal methods of learning the number of factors are available. Finally, these empirical studies point to a challenging and potentially most important issue – that the number of factors itself is truly dynamic, with different numbers of factors being relevant at different times.

Model extensions that are currently under investigation relax the assumptions of constancy of the factor loadings, and possible non-normal conditional distributions for factors. In this connection, we note very closely related developments by Pitt and Shephard (1999b), who explore similar factor models and related computational schemes for model fitting and sequential analysis (as mentioned earlier, the current work was developed independently of, and in parallel to, the work of Pitt and Shephard). Some of our recent work has extented the current framework to bring in dynamic regressions for the mean $\boldsymbol{\theta}_t$ in line with discount models in current implementations in major banks (Putnam and Quintana 1994; Quintana and Putnam 1996). Obviously, serious practical implementation of factor models demand such extensions. Particularly for forecasting and portfolio allocation with longer term horizons, such as 30-day exchange rate futures, we need extended models that incorporate dynamic regressions on relative interest rates and other possibly econometric indicators. Additional studies and empirical assessments of the methods on time series with larger numbers of univariate components and larger numbers of factors are also under investigation. Our experience to date leads us to believe that such investigations will be fruitful and support the preliminary conclusions reached in this report about the potential utility of factor models.

APPENDIX A: VARIANCE MATRIX DISCOUNTING

A brief summary of the basic method of variance matrix discounting are given here. As made clear by Uhlig (1994) these methods have formal theoretical bases in matrix-variate "random walks," though they had been in use by Quintana and coauthors (see references in the Introduction) and others for several years prior to that work. As summarised in West and Harrison (1997, chapter 16), the models involve sequential updating of posterior distributions for the sequences of variance matrices $\mathbf{\Sigma}_t$, with the following ingredients.

At a specific time t, the current posterior for $\mathbf{\Sigma}_t$ is an inverse Wishart form. Using the notation of West and Harrison (1997, chapter 16), the time t posterior is of the form $p(\mathbf{\Sigma}_t|D_t) = W_{n_t}^{-1}(\mathbf{\Sigma}_t|\mathbf{S}_t)$ where $D_t = \{D_0, \mathbf{y}_1, \ldots, \mathbf{y}_t\} = \{D_{t-1}, \mathbf{y}_t\}$ is the sequentially updated information set at time t. Here n_t is the degrees of freedom and $\mathbf{S}_t$ a posterior estimate of $\mathbf{\Sigma}_t$, the posterior harmonic mean. The notation $W_r^{-1}(\cdot|\mathbf{S})$ indicates the inverse Wishart distribution with r degrees of freedom and scale matrix $\mathbf{S}$ (see West and Harrison, as referenced). In the special case of zero-mean observations $\mathbf{y}_t$, the sequence of estimates $\mathbf{S}_t$ is trivially updated sequentially in time by the forward exponential moving average formula

$$\mathbf{S}_t = (1-a_t)\mathbf{S}_{t-1} + a_t\mathbf{y}_t\mathbf{y}_t' \tag{8}$$

where the weight a_t is given by $a_t = 1/(1+\delta n_{t-1})$ based on a discount factor δ. This discount factor lies in $(0,1)$, is typically between 0.9 and 1 and will be very close to unity for data at high sampling rates. Having analysed a fixed stretch of data $t = 1, \ldots, n$, the sequence of estimates $\mathbf{S}_t$ is revised by the related backward smoothing formula to incorporate the data at times $t+1, \ldots, n$ in inference on $\mathbf{\Sigma}_t$. Denoting the revised estimate of $\mathbf{\Sigma}_t$ by $\mathbf{S}_{t,n}$, the formula is given in terms of inverse variance matrices by the backward recursion

$$\mathbf{S}_{t,n}^{-1} = (1-\delta)\mathbf{S}_t^{-1} + \delta\mathbf{S}_{t+1,n}^{-1} \tag{9}$$

for each $t = n-1, n-2, \ldots, 1$, and starting with $\mathbf{S}_{n,n} = \mathbf{S}_n$. See West and Harrison (1997, pp608-609) for further details, and the references by Quintana and coauthors for development and applications in finance.

Extensions to models in which the time series has an unknown level $\boldsymbol{\theta}$ or, more generally, a dynamic regression component $\boldsymbol{\theta}_t$, are straightforward. Then the observation $\mathbf{y}_t$ is replaced by an appropriately scaled one-step ahead forecast error in the updating equation for $\mathbf{S}_t$. The details are standard and a side issue here, though the modification is critical in developing portfolio allocations. The resulting filtering and smoothing equations from Bayesian discounting analyses are modified following West and Harrison (1997, pp608-609).

APPENDIX B: MCMC DEVELOPMENT

In each of the following subsections we detail the conditional posteriors for various parameters and latent variables in turn. At each step it is implicit that we are conditioning on fixed values (previously simulated values) of all other variables. As noted in the text, further technical details are available on request from the authors. Our current implementation is in Fortran77 using standard random variate generation methods from RANDLIB.

Sampling the conditional mean

The basic model for $\mathbf{y}_t$ and the uniform prior for $\boldsymbol{\theta}$ immediately imply a multivariate normal conditional posterior for $\boldsymbol{\theta}$ given the values of $\mathbf{\Sigma}_t$. This is easily sampled.

Sampling the latent factors

The full conditional distribution of $\mathbf{f}_t$ is given by

$$N(\mathbf{f}_t|\mathbf{A}_t(\mathbf{y}_t - \boldsymbol{\theta}), \mathbf{H}_t - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t')$$

where $\mathbf{Q}_t = \mathbf{X}\mathbf{H}_t\mathbf{X}' + \mathbf{\Psi}_t$ and $\mathbf{A}_t = \mathbf{H}_t\mathbf{X}'\mathbf{Q}_t^{-1}$. The $\mathbf{f}_t$ are conditionally independent and so sample values are drawn independently from this set of normal distributions for $t = 1, 2\ldots, n$.

Sampling the factor loadings

From the model, the conditional likelihood function for the factor loading matrix $\mathbf{X}$ is $\prod_{t=1}^n N(\mathbf{y}_t - \boldsymbol{\theta}|\mathbf{X}\mathbf{f}_t, \mathbf{\Psi}_t)$. This is a standard form, log-quadratic in the uncertain elements of $\mathbf{X}$, and so combines with a normal or uniform reference prior to imply a multivariate normal conditional posterior, which is easily sampled.

Sampling the mean of the factor SV VAR(1) model

Under a normal prior $N(\boldsymbol{\mu}|\mathbf{m}_0, \mathbf{M}_0)$, the conditional posterior is normal $N(\boldsymbol{\mu}|\mathbf{m}, \mathbf{M})$ with

$$\mathbf{M}^{-1} = \mathbf{M}_0^{-1} + \mathbf{W}^{-1} + (n-1)(\mathbf{I} - \Phi)\mathbf{U}^{-1}(\mathbf{I} - \Phi)$$

and

$$\mathbf{M}\mathbf{m} = \mathbf{M}_0^{-1}\mathbf{m}_0 + \mathbf{W}^{-1}\boldsymbol{\lambda}_1 + (\mathbf{I} - \Phi)\mathbf{U}^{-1}\sum_{t=2}^{n}(\boldsymbol{\lambda}_t - \Phi\boldsymbol{\lambda}_{t-1}).$$

The case of a uniform reference prior is recovered by setting $\mathbf{M}_0^{-1} = \mathbf{0}$.

Sampling the VAR coefficients in the factor SV model

The structure of the conditional posterior for $\mathbf{\Phi}$, and the resulting Metropolis-Hastings strategy for simulation, is precisely as developed for component VAR models in a quite different context in West and Aguilar (1998) and Aguilar and West (1997). Writing $\boldsymbol{\gamma}_t = \boldsymbol{\lambda}_t - \boldsymbol{\mu}$, note that the full conditional posterior density for $\mathbf{\Phi}$ is proportional to

$$p(\mathbf{\Phi})N(\boldsymbol{\gamma}_1|\mathbf{0}, \mathbf{W})\prod_{t=2}^{n} N(\boldsymbol{\gamma}_t|\mathbf{\Phi}\boldsymbol{\gamma}_{t-1}, \mathbf{U})$$

where $\mathbf{W} = \mathbf{\Phi}\mathbf{W}\mathbf{\Phi} + \mathbf{U}$ is easily evaluated as a function of $\mathbf{\Phi}$ and $\mathbf{U}$. Write $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)'$ for the diagonal of $\mathbf{\Phi}$, and $\mathbf{E} = \text{diag}(\boldsymbol{\gamma}_{t-1})$. Then the conditional posterior may be written as proportional to

$$p(\boldsymbol{\Phi})c(\boldsymbol{\Phi})N(\boldsymbol{\phi}|\mathbf{b}, \mathbf{B})$$

where

$$\mathbf{B}^{-1} = \sum_{t=2}^{n} \mathbf{E}'\mathbf{U}^{-1}\mathbf{E} \quad \text{and} \quad \mathbf{Bb} = \sum_{t=2}^{n} \mathbf{E}'\mathbf{U}^{-1}\boldsymbol{\gamma}_t$$

and

$$c(\boldsymbol{\Phi}) = |\mathbf{W}|^{-I/2}\text{exp}(-\text{trace}(\mathbf{W}^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1')/2).$$

Under independent normal or uniform priors for the ϕ_j, the full conditional posterior distribution for $\boldsymbol{\Phi}$ is the multivariate normal $N(\boldsymbol{\phi}|\mathbf{b}, \mathbf{B})$ truncated to the $(0, 1)$ regions in each dimension, and then multiplied by the factor $c(\boldsymbol{\Phi})$. This may be sampled by several methods. We use a Metropolis Hastings algorithm that takes the truncated multivariate normal component as proposal distribution. That is, given a "current" value of $\boldsymbol{\phi}$, with corresponding matrices $\boldsymbol{\Phi}$ and $\mathbf{W}$, we sample a "candidate" vector $\boldsymbol{\phi}^*$ from this truncated normal, compute the corresponding diagonal matrix $\boldsymbol{\Phi}^*$ and variance matrix $\mathbf{W}^*$ such that $\mathbf{W}^* = \boldsymbol{\Phi}^*\mathbf{W}^*\boldsymbol{\Phi}^* + \mathbf{U}$, then accept this new $\boldsymbol{\phi}$ vector with probability

$$\min\{1, c(\boldsymbol{\Phi}^*)/c(\boldsymbol{\Phi})\}.$$

Sampling the $\mathbf{U}$ *matrix in the factor SV model*

Again, the structure of the conditional posterior for the innovations variance matrix $\mathbf{U}$ of the VAR(1) volatility model is closely related to developments in a quite different context in West and Aguilar (1998) and Aguilar and West (1997). Again using centred volatilities $\boldsymbol{\gamma}_t = \boldsymbol{\lambda}_t - \boldsymbol{\mu}$ for each t, we have a full conditional posterior for $\mathbf{U}$ proportional to

$$p(\mathbf{U})a(\mathbf{U})|\mathbf{U}|^{-(n-1)/2}\text{exp}(-\text{trace}(\mathbf{U}^{-1}\mathbf{G}))$$

where

$$\mathbf{G} = \sum_{t=2}^{n}(\boldsymbol{\gamma}_t - \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})(\boldsymbol{\gamma}_t - \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})'$$

and

$$a(\mathbf{U}) = |\mathbf{W}|^{-1/2}\text{exp}(-\text{trace}(\mathbf{W}^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1')/2)$$

with $\mathbf{W} = \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi} + \mathbf{U}$. Under a specified inverse Wishart prior $W_{r_0}^{-1}(\mathbf{U}|\mathbf{R}_0)$, this posterior density is proportional to

$$a(\mathbf{U})W_r^{-1}(\mathbf{U}|\mathbf{R})$$

where $r = r_0 + n - 1$ and $r\mathbf{R} = r_0\mathbf{R}_0 + (n-1)\mathbf{G}$. We use this inverse Wishart distribution as a proposal distribution in the Metropolis-Hastings algorithm. That is, given a "current" value of $\mathbf{U}$ and corresponding $\mathbf{W}$, we sample a "candidate" value $\mathbf{U}^*$ from $W_r^{-1}(\mathbf{U}|\mathbf{R})$, and accept it with probability

$$\min\{1, a(\mathbf{U}^*)/a(\mathbf{U})\}$$

where $\mathbf{W}^* = \boldsymbol{\Phi}\mathbf{W}^*\boldsymbol{\Phi} + \mathbf{U}^*$.

Sampling the factor volatility processes

Given currently imputed values for the full factor process $\mathbf{f}_t$ over time t, we follow Kim, Shephard and Chib (1998) in transforming each of the univariate time series of factor elements into a non-Gaussian linear model form. Specifically, for each factor i and time t define $z_{ti} = \log(f_{ti}^2)$ and note that

$$z_{ti} = \lambda_{ti} + \nu_{ti}$$

where the ν_{ti} terms are independent and distributed as $\log-\chi_1^2$. In vector form for all k factor series, we then have

$$\mathbf{z}_t = \boldsymbol{\lambda}_t + \boldsymbol{\nu}_t$$

where $\boldsymbol{\nu}_t = (\nu_{t1}, \ldots, \nu_{tk})'$. Based on the current values of the $\mathbf{z}_t$, for each t, this provides the set observation equations for the dynamic linear model with state equations

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\omega}_t$$

for each t. This is a direct multivariate extension of the univariate approach in Kim, Shephard and Chib (1998), whose ensuing analysis uses the now established approximation to the distribution of the error terms ν_{it} as a specified finite mixture of normals (Shephard 1994b). The multivariate extension is immediate:

- Introduce a set of indicator variables l_{ti} such that l_{ti} identifies the normal mixture component for ν_{ti}.
- Conditional on these indicators, we have a multivariate dynamic linear model for the sequence of log-volatility vectors. The forward-filtering, backward-sampling algorithm for state space models (Carter and Kohn 1994; Frühwirth-Schnatter 1994; West and Harrison 1997, chapter 15) now applies to directly simulate the full set of vectors $\{\boldsymbol{\lambda}_t, t = 1, \ldots, n\}$ from the implied conditional posterior. (Note that, in more elaborate models for the volatility processes, the alternative sampling method using the simulation smoother of de Jong and Shephard (1995) may have computational advantages not realised in this, the simplest VAR model.)
- Given these sampled values of the $\boldsymbol{\lambda}_t$, the complete conditional multinomial posterior probabilities over values of the indicators l_{it} are easily computed, the indicators being conditionally independent and so easily sampled.

Sampling idiosyncratic volatilities

Conditional on currently imputed values of $\boldsymbol{\theta}, \mathbf{X}$ and the $\mathbf{f}_t$, we have imputed values of each of the error vectors $\boldsymbol{\epsilon}_t = (\epsilon_{t1}, \ldots, \epsilon_{tq})'$ in equation (2). This defines a set of q independent and standard univariate SV models, as follows: with $v_{tj} = \log(\epsilon_{tj}^2)$ for each j, t,

$$v_{tj} = \eta_{tj} + \tau_{tj}$$

where the τ_{tj} terms are independent and distributed as $\log-\chi_1^2$, and $\eta_{tj} = \log(\psi_{tj})$ follow the AR(1) models of section 2.4, namely

$$\eta_{tj} = \alpha_j + \rho_j(\eta_{t-1,j} - \alpha_j) + \xi_{tj}$$

with independent innovations $\xi_{tj} \sim N(\xi_{tj}|0, s_j)$. The standard sampling schemes of Kim, Shephard and Chib (1998), and with minor modifications from West and Aguilar (1997), now apply to each of these $j = 1, \ldots, q$ univariate SV models. This produces imputed values for η_{tj}, hence the volatilities ψ_{tj}, for each $j = 1, \ldots, q$ and for all t.

Sampling the means in the idiosyncratic SV models

Conditional on currently imputed values of each of the η_{tj}, the univariate SV models imply standard likelihood functions for each of the means α_j, independently across series j. Under independent normal priors, or standard reference uniform priors, these lead to independent normal posteriors that are trivially sampled.

Sampling the $\boldsymbol{\rho_{tj}}$ in the idiosyncratic SV models

Conditional on currently imputed values of each of the η_{tj}, the univariate SV models imply likelihood functions for each of the ρ_j, independently across series j. These are precise univariate analogues of the likelihood functions for the VAR parameter matrix $\mathbf{\Phi}$ discussed above, and the simulation strategy there is applied, in its univariate analogue, to each of these q models to impute new values for each of the ρ_j.

Sampling the $\boldsymbol{s_j}$ in the idiosyncratic SV models

Conditional on currently imputed values of each of the η_{tj}, the univariate SV models imply likelihood functions for each of the s_j, independently across series j. These are again precise univariate analogues of the likelihood functions for the VAR parameter matrix $\mathbf{U}$ discussed above. The structure simplifies in that the inverse Wisharts in the multivariate case are replaced by inverse gamma distributions for the s_j, but otherwise the general strategy is the same. This is applied, in each of the q models in parallel, to impute new values for each of the s_j.

REFERENCES

Aguilar, O., and West M. (1998), "Analysis of Hospital Quality Monitors using Hierarchical Time Series Models," in *Case Studies Bayesian Statistics in Science and Technology: Case Studies 4,* (eds: C. Gatsonis et al), Springer-Verlag, New York, 287-302.

Ameen, J.R., and Harrison, P.J. (1985), "Normal Discount Bayesian Models," in *Bayesian Statistics 2*, (eds: J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), North Holland, Amsterdam, and Valencia University Press, 271-298.

Carter, C., and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Demos, A., and Sentana, E. (1998), "An EM Algorithm for Conditionally Heteroskedastic Factor Models," *Journal of Business and Economic Statistics,* **16**, 357-361.

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis,* **15**, 183–202.

de Jong, P., and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika,* **82**, 339-50.

Geweke, J.F., and Singleton, K.J. (1980), "Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small," *Journal of the American Statistical Association,* **75**, 133-137.

Geweke, J.F., and Zhou, G. (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The Review of Financial Studies*, **9**, 557-587.

Harrison, P.J., and West, M. (1987), "Practical Bayesian Forecasting," *The Statistician,* **36**, 115-125.

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

Jacquier, E., Polson, N.G., and Rossi, P. (1994), "Bayesian Analysis of Stochastic Volatility Models (with Discussion)," *Journal of Business and Economic Statistics*, **12**, 371-388.

Jacquier, E., Polson, N.G. and Rossi, P. (1995), "Models and Priors for Multivariate Stochastic Volatility," *Discussion Paper,* Graduate School of Business, University of Chicago.

Kim, S., Shephard, N. and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies,* **65,** 361-393.

Liu, J., and West, M. (2000), "Combined Parameter and State Estimation in Simulation-based Filtering," in *Sequential Monte Carlo Methods in Practice*, (eds: A. Doucet, J. F. G. De Freitas, and N.J. Gordon), New York: Springer-Verlag *(forthcoming).*

Lopes, H.F., and West, M. (1998), "Model Uncertainty in Factor Analysis," *Discussion Paper #98-38,* Institute of Statistics and Decision Sciences, Duke University.

Pitt, M.K., and Shephard, N. (1999a), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association,* **94**, 590-599.

Pitt, M.K., and Shephard, N. (1999b), "Time Varying Covariances: A Factor Stochastic Volatility Approach.," (with Discussion), in *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 547-570.

Polson, N.G., and Tew, B.V. (1997), "Bayesian Portfolio Selection: An Empirical Analysis of the SP500 Index 1970-1996," *Discussion Paper,* Graduate School of Business, University of Chicago.

Press, S.J. (1985), *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference,* California: Krieger.

Press, S.J., and Shigemasu, K. (1989), "Bayesian Inference in Factor Analysis," in *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin*, (eds: L.J. Gleser, M.D. Perlman, S.J. Press and A.R. Sampson,), New York: Springer Verlag, 271-287.

Putnam, B.H., and Quintana, J.M. (1994), "New Bayesian Statistical Approaches to Estimating and Evaluating Models of Exchange Rates Determination," In *Proceedings of the ASA Section on Bayesian Statistical Science, 1994 Joint Statistical Meetings,* American Statistical Association.

Putnam, B.H., and Quintana, J.M. (1995), "The Evolution of Bayesian Forecasting Models," in *Asset Allocation: Applying Quantitative Discipline to Asset Allocation,* (ed: B.H. Putnam), London: Global Investor, Euromoney Publications, 139-146.

Quintana, J.M., and West, M. (1987), "An Analysis of International Exchange Rates using Multivariate DLMs," *The Statistician,* **36**, 275-281.

Quintana, J.M., and West, M. (1988), "Time Series Analysis of Compositional Data," in *Bayesian Statistics 3*, (eds: J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), Oxford University Press, 747-756.

Quintana, J.M. (1992), "Optimal Portfolios of Forward Currency Contracts," in *Bayesian Statistics 4*, (eds: J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith), Oxford University Press.

Quintana, J.M., Chopra, V.K., and Putnam, B.H. (1995), "Global Asset Allocation: Stretching Returns by Shrinking Forecasts," in *Proceedings of the ASA Section on Bayesian Statistical Science, 1995, Joint Statistical Meetings*, American Statistical Association.

Quintana, J.M., and Putnam, B.H. (1996), "Debating Currency Markets Efficiency using Multiple-Factor Models," in *Proceedings of the ASA Section on Bayesian Statistical Science, 1996 Joint Statistical Meetings,* American Statistical Association.

Shephard, N. (1994a), "Local Scale Models: State Space Alternatives to Integrated GARCH Processes," *Journal of Econometrics,* **6**, 333-364.

Shephard, N. (1994b), "Partial Non-Gaussian State Space," *Biometrika,* **81**, 115-132.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, (eds: D.R. Cox, O.E. Barndorff-Nielson and D.V. Hinkley), London: Chapman & Hall, 1-67.

Shephard, N., and Pitt, M.K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika,* **84**, 653-67.

Uhlig, H. (1994), "On Singular Wishart and Singular Multivariate Beta Distributions," *Annals of Statistics,* **22**, 395-405.

Uhlig, H. (1997), "Bayesian Vector-Autoregressions with Stochastic Volatility," *Econometrica*, **65**, 59-73.

West, M. (1993), "Mixture models, Monte Carlo, Bayesian Updating and Dynamic Models," *Computing Science and Statistics,* **24,** 325-333.

West, M., and Aguilar, O. (1997), "Studies of Quality Monitor Time Series: The V.A. Hospital System," report for the VA Management Science Group, Bedford, MA., Institute of Statistics and Decision Sciences, Discussion Paper #96-02.

West, M., and Harrison, P.J. (1997), *Bayesian Forecasting and Dynamic Models*, (2nd Edn.), New York: Springer Verlag.