Factor-augmented Error Correction Models*

Anindya Banerjee† Massimiliano Marcellino‡

15 July 2008

Abstract

This paper brings together several important strands of the econometrics literature: error-correction, cointegration and dynamic factor models. It introduces the Factor-augmented Error Correction Model (FECM), where the factors estimated from a large set of variables in levels are jointly modelled with a few key economic variables of interest. With respect to the standard ECM, the FECM protects, at least in part, from omitted variable bias and the dependence of cointegration analysis on the specific limited set of variables under analysis. It may also be in some cases a refinement of the standard Dynamic Factor Model since it allows us to include the error correction terms into the equations, and by allowing for cointegration prevent the errors from being non-invertible moving average processes. In addition, the FECM is a natural generalization of factor augmented VARs (FAVAR) considered by Bernanke, Boivin and Eliasz (2005) inter alia, which are specified in first differences and are therefore misspecified in the presence of cointegration. The FECM has a vast range of applicability. A set of Monte Carlo experiments and two detailed empirical examples highlight its merits in finite samples relative to standard ECM and FAVAR models. The analysis is conducted primarily within an in-sample framework, although the out-of-sample implications are also explored.

Keywords: Dynamic Factor Models, Error Correction Models, Cointegration, Factor-augmented Error Correction Models, VAR, FAVAR

JEL-Codes: C32, E17

*We thank the Research Council of the EUI for supporting this research. We also thank Igor Masten for many useful discussions and help with the simulations reported in this paper, Katarzyna Maciejowska provided excellent research assistance. We are also grateful to two anonymous referees, Jennifer Castle, Luca Sala, Neil Shephard, James Stock, and seminar participants at the EUI, Bocconi University and at the Hendry Festschrift Conference in Oxford for helpful comments on a previous draft. Responsibility for any errors remains with us.

†Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, e-mail: a.banerjee@bham.ac.uk

‡Bocconi University and European University Institute, Via della Piazzuola 43, 50133 Firenze, Italy, e-mail: massimiliano.marcellino@eui.eu

1 Introduction

Our paper is an exploration of a few of the many themes studied by David Hendry, whom this conference volume honours. Starting at least with Davidson, Hendry, Srba and Yeo (1978), Hendry has argued in favour of the powerful role of error-correction mechanisms (ECM) in modelling macroeconomic data. While originally undertaken in an environment with supposedly stationary data, the subsequent development of cointegration served to renew emphasis on the long-run co-movement of macroeconomic variables. Models lacking such information are likely to be misspecified both within-sample and out-of-sample (or forecasting context).

Breaks in the structure of models pose additional challenges for forecasting since models well specified within sample may not provide any guide for the forecasting performance of such models. Key references for this observation include Clements and Hendry (1995) where an interesting finding is that including reduced-rank or cointegrating information may not have beneficial effects on the forecasting performance of models except in small sample sizes. However, unrestricted vector autoregressions will be dominated by models which incorporate cointegration restrictions for larger systems of equations where cointegration relations impose a large number of restrictions. This is important background for the analysis conducted here, since we focus precisely and very largely on the implications of modelling cointegration in very large systems of equations.

Yet more pertinently from the point of view of our analysis, the fact that in large datasets much of the cointegration information may be unknown or difficult to model, will lead to a dependence of the performance of the macroeconomic models on exactly how successfully the cointegration information is extracted from the data. This is by no means a trivial problem, especially if the dimension of the system N is large. Clements and Hendry (1995) explore this issue using alternative criteria for assessing forecasting accuracy including the trace mean squared forecast error criterion (TMSFE) and their preferred invariant generalised forecast error second moment (GFESM) criterion. More recent analysis by Hendry (2006) has argued in favour of using a differenced vector error correction model (DVECM) which introduces error-correction information into a double-differenced-VAR (DDVAR). Particularly in an environment with structural change, a DVECM retains information relating to the change in the equilibrium in the system.

The main contributions of our paper are (a) to bring together two important recent strands of econometric literature on modelling co-movement that have a common origin but, in their implementations, have remained apart, namely, cointegration and dynamic factor models [1] and (b) to evaluate the role of incorporating long-run information in modelling, within the framework of both simulation exercises (where the emphasis is on evaluating efficiency within-sample) and empirical examples (where we look at both within-sample and out-of sample performance). It is important, in our view, to consider factor models since a significant issue, as in Clements and Hendry (1995), is the modelling of large systems of equations in which the complete cointegrating space may either be difficult to identify or it may not be necessary to do so since we may be interested in only a sub-system as our variables of interest. In such circumstances, as we shall see, proxying for the missing cointegrating information, may turn out to be extremely useful.

Our evaluations are based both on in-sample measures of model fit, including R^2 and adjusted-R^2 (which, in our simulation exercises, is equivalent to one-step ahead mean squared forecast error

[1] Our focus here is on the widespread application of these methods in econometrics to model macroeconomic variables. Factor models have of course been used in a large number of other contexts for a much longer period.

(MSFE) since here the models may be taken to be correctly specified and the fitted value of the modelled variable can be interpreted as its forecast), as well as on a number of other criteria such as AIC and BIC, in circumstances (such as in our empirical examples) where the cointegrating information needs to be estimated and correct specification can therefore no longer be assumed to hold.

Our attempt here is to develop a manageable approach to the problems posed by large datasets where there is cointegration and where such cointegration should be taken into account in modelling the data. In particular, in this paper we study the relationship between dynamic factor models and error correction models. We introduce the Factor-augmented Error Correction Model (FECM), where the factors extracted from a dynamic factor model for a large set of variables in levels are jointly modelled with a limited set of economic variables of main interest.

The FECM represents an improvement with respect to the standard ECM for the subset of variables, since it protects, at least in part, from omitted variable bias and the dependence of cointegration analysis on the specific limited set of variables under analysis. The FECM is also a refinement of dynamic factor models, since it allows us to include the error correction terms into the equations for the key variables under analysis, preventing the errors from being non-invertible MA processes. The FECM can also be considered as a natural generalization of factor-augmented VARs (FAVAR) considered by Bernanke, Boivin and Eliasz (2005), Favero, Marcellino and Neglia (2005) and Stock and Watson (2005). The FAVARs in all of these papers are specified in first differences, so that they are misspecified in the presence of cointegration.

The FECM may be expected to have a vast range of applicability. Therefore, in order to evaluate its relative merits in small, medium and large samples, we conduct a set of Monte Carlo experiments, while to illustrate its use in practice we present two empirical applications with economic data. The first empirical example studies the relationships among four US interest rate series (at different maturities), and proceeds to analyze the relationships among these interest rate series and other macroeconomic variables. The second example reconsiders the famous article by King, Plosser, Stock and Watson (1991) on stochastic trends and economic fluctuations in the US economy. In both examples, the factors are estimated from a large set of 110 monthly US macroeconomic variables, extracted from the dataset in Stock and Watson (2005).

The simulation and empirical results show systematic gains in terms of explanatory power from the use of the FECM with respect to both an ECM and a FAVAR model.

The rest of the paper is organized as follows. In Section 2 we introduce the FECM. In Section 3 we discuss a simple analytical example. In Section 4 we present the design and results of the Monte Carlo experiments to evaluate the finite sample performance of the FECM. In Section 5 we discuss the empirical examples. Finally, in Section 6 we summarize and conclude.

2 The Factor-augmented Error Correction Model

Let us assume that the N $I(1)$ variables x_t evolve according to the $VAR(p)$ model

$$x_t = \Pi_1 x_{t-1} + ... + \Pi_p x_{t-p} + \epsilon_t, \tag{1}$$

where ϵ_t is $i.i.d.(0, \Omega)$ and, for simplicity, the starting values are fixed and equal to zero.

The $VAR(p)$ can be reparametrized into the Error Correction Model (ECM)

$$\Delta x_t = \alpha\beta' x_{t-1} + \nu_t, \tag{2}$$

or in the so-called common trend specification

$$x_t = \Psi f_t + u_t, \tag{3}$$

see, e.g., Johansen (1995, p.49). In particular,

$$\Pi = \sum_{s=1}^{p} \Pi_s - I_n = \underset{N\times N-r}{\alpha} \underset{N-r\times N}{\beta'},$$

$$v_t = \Gamma_1 \Delta x_{t-1} + ... + \Gamma_{p-1}\Delta x_{t-p+1} + \epsilon_t, \quad \Gamma_i = -\sum_{s=i+1}^{p} \Pi_s, \quad \Gamma = I - \sum_{s=1+1}^{p-1} \Gamma_i,$$

$$\underset{N\times r}{\Psi} = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}, \quad \underset{r\times 1}{f_t} = \alpha'_\perp \sum_{s=1}^{t} \epsilon_s, \quad u_t = C(L)\epsilon_t.$$

β' is the $N-r \times N$ matrix of cointegrating vectors with rank $N-r$, where $N-r$ is the number of cointegrating vectors. r is therefore the number of $I(1)$ common stochastic trends (or factors), $0 < r \leq N$, gathered in the $r \times 1$ vector f_t and the matrix $\alpha'_\perp \Gamma \beta_\perp$ is invertible since each variable is $I(1)$. α is the so-called loading matrix, which also has reduced rank $N-r$ and determines how the cointegrating vectors enter into each individual element x_{it} of the $N \times 1$ vector x_t.[2] u_t is an $N-$dimensional vector of stationary errors. We also assume that there are no common cycles in the sense of Engle and Kozicki (1993), i.e., no linear combinations of the first differences of the variables that are correlated of lower order than each of the variables (in first differences), although adding such cycles (as in the analytical example below) poses no significant complications and is assumed here only for convenience.[3]

From equation (3), it is possible to write the model for the first differences of x_t, Δx_t, as

$$\Delta x_t = \Psi \Delta f_t + \Delta u_t, \tag{4}$$

where Δu_t and ν_t can be correlated over time and across variables.

The literature on dynamic factor models has relied on a specification similar to (4) and has focused on the properties of the estimators of the common factors Δf_t, or of the common components $\Psi \Delta f_t$, under certain assumptions on the idiosyncratic errors, when the number of variables N becomes large. See, for example, Stock and Watson (2002a, 2002b) and Forni, Hallin,Lippi and Reichlin (2000). A few papers have also analyzed the model in (3) for the divergent N case, most notably Bai and Ng (2004) and Bai (2004).[4] We shall make use of both specification (3) and (4) when discussing factor models in what follows.

By contrast, the literature on cointegration has focused on (2) and has studied the properties of

[2] Note that as $N \to \infty$, and the number of factors r remains fixed, the number of cointegrating relations $N-r \to \infty$.

[3] Common cycles are associated with reduced rank of (some of) the coefficient matrices in $C(L)$, where we remember that the errors in the stochastic trend representation (3) are $u_t = C(L)\epsilon_t$. Therefore, the presence of common cycles is associated with stationary common factors driving x_t, in addition to the I(1) factors.

[4] Bai and Ng (2004) also allow for the possibility that some elements of the idiosyncratic error u_t are I(1). We will not consider this case and assume instead that all the variables under analysis are cointegrated, perhaps after pre-selection. We feel that this is a sensible assumption from an economic point of view.

tests for the cointegrating rank $(N-r)$ and estimators of the cointegrating vectors (β'), see e.g. Engle and Granger (1987) or Johansen (1995). A few papers have attempted to extend the analysis to the large N case, generating the so-called panel cointegration tests, where a factor structure is employed to explore issues relating to the dependence across the variables. See e.g. Banerjee, Marcellino and Osbat (2004) and Banerjee and Carrion-i-Silvestre (2007), where the latter paper uses techniques used by Bai and Ng (2004) in developing their PANIC tests for unit roots in panels.[5] The extension of PANIC techniques to study cointegration is complicated by the curse of dimensionality which makes the modelling of cointegration - particularly when N is large and there are multiple cointegrating vectors, i.e. $N-r>1$- extremely difficult and often subject to criticism.

To continue with our analysis, let us impose, without any loss of generality, the identifying condition

$$\underset{N-r\times N}{\beta'} = \left(\underset{N-r\times r}{\beta^{*\prime}} \; : \; \underset{N-r\times N-r}{I} \right).$$

This is standard practice in this literature, as also implemented by Clements and Hendry (1995, page 129, lines 1 - 5) and ensures that the transformation from the levels x_t which are $I(1)$ to $I(0)$-space (involving taking the cointegrated combinations and the differences of the $I(1)$ variables) is scale preserving.

From (3), partitioning u_t into

$$u_t = \begin{pmatrix} \underset{r\times 1}{u_{1t}} \\ \\ \underset{N-r\times 1}{u_{2t}} \end{pmatrix},$$

the model for the error correction terms can be written as

$$\beta' x_t = \beta' u_t = \beta^{*\prime} u_{1t} + u_{2t}. \tag{5}$$

In this model each of the $N-r$ error correction terms depends on a common component that is a function of only r shocks, u_{1t}, and on an idiosyncratic component, u_{2t}. Different normalizations of the cointegrating vectors change the exact shocks that influence each error correction term, but its decomposition into a common component driven by r shocks and an idiosyncratic component remains valid. This is also in line with the stochastic trend representation in (3), where the levels of the variables are driven by r common trends.

Let us now partition the N variables in x_t into the N_A of major interest, x_{At}, and the $N_B = N - N_A$ remaining ones, x_{Bt}. We can partition the common trends model in (3) accordingly as

$$\begin{pmatrix} x_{At} \\ x_{Bt} \end{pmatrix} = \begin{pmatrix} \Psi_A \\ \Psi_B \end{pmatrix} f_t + \begin{pmatrix} u_{At} \\ u_{Bt} \end{pmatrix}, \tag{6}$$

where Ψ_A is of dimension $N_A \times r$ and Ψ_B is $N_B \times r$. Notice that when the number of variables N increases, the dimension of Ψ_A is fixed, while the number of rows of Ψ_B increases correspondingly. Therefore, for (6) to preserve a factor structure asymptotically, driven by r common factors, it is necessary that the rank of Ψ_B remains equal to r. Instead, the rank of Ψ_A can be smaller than r, i.e., x_{At} can be driven by a smaller number of trends, say $r_A \leq r$.

From the specification in (6), it is evident that x_{At} and f_t are cointegrated, while from (3) the f_t are uncorrelated random walks. Therefore, from the Granger representation theorem, there must

[5] Other papers in this area include Breitung and Das (2005, 2007), Pesaran (2006), Bai, Kao and Ng (2007).

exist an error correction specification of the type

$$\begin{pmatrix} \Delta x_{At} \\ \Delta f_t \end{pmatrix} = \begin{pmatrix} \gamma_A \\ \gamma \end{pmatrix} \delta' \begin{pmatrix} x_{At-1} \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} e_{At} \\ e_t \end{pmatrix}. \tag{7}$$

In practice, correlation in the errors of (7) is handled by adding additional lags of the differenced dependent variables, so that the model becomes

$$\begin{pmatrix} \Delta x_{At} \\ \Delta f_t \end{pmatrix} = \begin{pmatrix} \gamma_A \\ \gamma \end{pmatrix} \delta' \begin{pmatrix} x_{At-1} \\ f_{t-1} \end{pmatrix} + A_1 \begin{pmatrix} \Delta x_{At-1} \\ \Delta f_{t-1} \end{pmatrix} + ... + A_q \begin{pmatrix} \Delta x_{At-q} \\ \Delta f_{t-q} \end{pmatrix} + \begin{pmatrix} \epsilon_{At} \\ \epsilon_t \end{pmatrix}. \tag{8}$$

We label (8) as a Factor-augmented Error Correction Model (FECM).

Since there are $N_A + r$ dependent variables in the FECM model (8), x_{At} is driven by f_t or a subset of them, and the f_t are uncorrelated random walks, there must be N_A cointegrating relationships in (8). Moreover, since Ψ_A is of dimension $N_A \times r$ but can have reduced rank r_A, there are $N_A - r_A$ cointegrating relationships that involve the x_A variables only, say $\delta_A^{'} x_{At-1}$, and the remaining r_A cointegrating relationships involve x_A and the factors f.

The cointegrating relationships $\delta_A^{'} x_{At-1}$ would also emerge in a standard ECM for Δx_{At} only, say

$$\Delta x_{At} = \alpha_A \delta_A^{'} x_{At-1} + v_{At}. \tag{9}$$

However, in addition to these $N_A - r_A$ relationships, in the FECM there are r_A cointegrating relationships that involve x_{At} and f_t, and that proxy for the potentially omitted $N - N_A$ cointegrating relationships in (9) with respect to the equations for Δx_{At} in the full ECM in (2).[6] Moreover, in the FECM there appear lags of Δf_t as regressors in the equations for Δx_{At}, that proxy for the potentially omitted lags of Δx_{Bt} in the standard ECM for Δx_{At} in (9). Therefore, the FECM provides an improved representation for the variables of interest x_{At}, in terms of modelling both the long-run and short-run evolution of these variables.

It is also important to point out that in the dynamic factor models à la Stock and Watson (2002a, 2002b) and in FAVAR specifications the error correction terms never appear, i.e., $\gamma_A = 0$ is imposed in (8). Therefore, the FECM also represents an improvement for the specification of dynamic factor models and FAVAR models. Moreover, in our context where the data generating process is the common trends specification in (3), standard factor and FAVAR models have two additional substantial problems. First, the error process Δu_t in (4) has a non-invertible moving average component that prevents, from a theoretical point of view, the approximation of each equation of the model in (4) with an AR model augmented with lags of the factors. Second, and perhaps even more problematically, in (4) Δf_t and Δu_t are in general not orthogonal to each other, and in fact they can be highly correlated. This feature disrupts the factor structure and, from an empirical point of view, can require a large number of factors to summarize the information contained in Δx_t.

Notice that if the starting model is given by (3) but the shocks driving the integrated factors are orthogonal to u_t, so that Δf_t and Δu_t are also orthogonal, then the model in (4) is a proper factor model, but with a non-invertible moving average component. This feature does not pose any additional complications for the estimation of the common component $\Psi \Delta f_t$ either with the

[6] In the full ECM model (2), there would be up to $N - r_A$ cointegrating relationships in the equations for Δx_{At}, while in (9) there are only $N_A - r_A$ cointegrating relationships, so that there are $N - N_A$ potentially omitted long run relationships in the ECM for Δx_{At} only.

static principal component approach of Stock and Watson (2002a,b) or with the dynamic principal component method of Forni et al. (2000, 2005). However, the presence of a unit root in the moving average component still prevents the approximation of each equation of the model in (4) with an AR model augmented with lags of the factors, while factor augmented AR models have become a standard tool for forecasting.

The FECM also has its problems. In particular, the approach may find it difficult to handle situations where there is a large number of error correction terms affecting each equation, or when the cointegrating relationships include all the variables in x_t and not just the subset x_{At}.

An additional complication for the FECM is that in practice the common stochastic (integrated) factors, f_t, are not known. However, the principal components of x_t are a consistent estimator for (the space spanned by) f_t when N diverges, see e.g. Stock and Watson (1988) and Bai (2004). Moreover, Bai (2004) and Bai and Ng (2006) have shown that, when $\sqrt{T}/N$ is $o_p(1)$, the estimated factors can be used in subsequent analyses without creating any generated regressors problems. Therefore, the estimated factors can be used in the FECM instead of the true factors, assuming that the available dataset is large enough to satisfy the condition $\sqrt{T}/N$ is $o_p(1)$. The role of the use of estimated versus true factors in finite sample is one of the issues explored in the simulation exercise.

3 An analytical example

Before proceeding to the simulations, we first consider a simple analytical example to illustrate the relationships between the ECM representation, the FECM, and the FAVAR. Let us assume that the N variables are generated by the ECM model

$$\Delta x_t = \alpha\beta' x_{t-1} + \epsilon_t, \tag{10}$$

with $\epsilon_t \sim i.i.d.(0, I_N)$, one common stochastic trend ($r = 1$), and

$$\beta' = \begin{pmatrix} -1 & 1 & 0 & 0 & ... & 0 & 0 \\ -1 & 0 & 1 & 0 & ... & 0 & 0 \\ -1 & 0 & 0 & 1 & & & \\ ... & & & & & & \\ -1 & 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 0 & 0 & 0 & ... & 0 \\ -1 & 0 & 0 & ... & 0 \\ -1 & -1 & 0 & ... & 0 \\ -1 & 0 & -1 & & \\ ... & & & & \\ -1 & 0 & 0 & ... & -1 \end{pmatrix}.$$

Therefore, the equations of the ECM are

$$\begin{aligned} \Delta x_{1t} &= \epsilon_{1t} \\ \Delta x_{2t} &= -(-x_{1t-1} + x_{2t-1}) + \epsilon_{2t} \\ \Delta x_{3t} &= -(-x_{1t-1} + x_{2t-1}) - (-x_{1t-1} + x_{3t-1}) + \epsilon_{3t} \\ \Delta x_{4t} &= -(-x_{1t-1} + x_{2t-1}) - (-x_{1t-1} + x_{4t-1}) + \epsilon_{4t} \\ &... \\ \Delta x_{Nt} &= -(-x_{1t-1} + x_{2t-1}) - (-x_{1t-1} + x_{Nt-1}) + \epsilon_{Nt}. \end{aligned} \tag{11}$$

The stochastic trend representation becomes

$$
\begin{aligned}
x_{1t} &= \sum_{s=1}^{t} \epsilon_{1s} \\
x_{2t} &= x_{1t-1} + \epsilon_{2t} \\
x_{3t} &= x_{1t-1} - \epsilon_{2t-1} + \epsilon_{1t-1} + \epsilon_{3t} \\
x_{4t} &= x_{1t-1} - \epsilon_{2t-1} + \epsilon_{1t-1} + \epsilon_{4t} \\
&\ldots \\
x_{Nt} &= x_{1t-1} - \epsilon_{2t-1} + \epsilon_{1t-1} + \epsilon_{Nt}.
\end{aligned} \tag{12}
$$

From this representation it clearly emerges that the variables are driven by an I(1) common factor, $\sum_{s=1}^{t} \epsilon_{1s}$, and by an I(0) common factor, $\epsilon_{1t} - \epsilon_{2t}$. If we write the model in (12) in a compact notation as

$$
x_t = \nu \sum_{s=1}^{t-1} \epsilon_{1s} + \epsilon_t + C_1 \epsilon_{t-1},
$$

where $\nu = (1, 1, ..., 1)'$, it clearly emerges that C_1 has reduced rank (equal to one), i.e., there are common cycles in the sense of Engle and Kozicki (1993).

From the stochastic trend representation in (12), we can easily derive that the specification for the error correction terms (cointegrating relationships) $\beta' x_{t-1}$ is given by

$$
\begin{aligned}
x_{2t} - x_{1t} &= -(\epsilon_{1t} - \epsilon_{2t}) \\
x_{3t} - x_{1t} &= \epsilon_{1t-1} - \epsilon_{2t-1} - \epsilon_{1t} + \epsilon_{3t} \\
x_{4t} - x_{1t} &= \epsilon_{1t-1} - \epsilon_{2t-1} - \epsilon_{1t} + \epsilon_{4t} \\
&\ldots \\
x_{Nt} - x_{1t} &= \epsilon_{1t-1} - \epsilon_{2t-1} - \epsilon_{1t} + \epsilon_{Nt}.
\end{aligned} \tag{13}
$$

Therefore, the error correction terms are driven by two common I(0) factors, one is the same as for the levels of the variables, $\epsilon_{1t} - \epsilon_{2t}$, the other is the first difference of the common I(1) factor, $\Delta \sum_{s=1}^{t} \epsilon_{1s} = \epsilon_{1t}$.

Substituting the expression in (13) for $\beta' x_{t-1}$ into the ECM in (10), the representation for Δx_t corresponding to (4) is

$$
\begin{aligned}
\Delta x_{1t} &= \epsilon_{1t} \\
\Delta x_{2t} &= \epsilon_{1t-1} - \epsilon_{2t-1} + \epsilon_{2t} \\
\Delta x_{3t} &= \epsilon_{1t-1} - \epsilon_{2t-1} - (\epsilon_{1t-2} - \epsilon_{2t-2}) + \epsilon_{1t-1} + \epsilon_{3t} - \epsilon_{3t-1} \\
\Delta x_{4t} &= \epsilon_{1t-1} - \epsilon_{2t-1} - (\epsilon_{1t-2} - \epsilon_{2t-2}) + \epsilon_{1t-1} + \epsilon_{4t} - \epsilon_{4t-1} \\
&\ldots \\
\Delta x_{Nt} &= \epsilon_{1t-1} - \epsilon_{2t-1} - (\epsilon_{1t-2} - \epsilon_{2t-2}) + \epsilon_{1t-1} + \epsilon_{Nt} - \epsilon_{Nt-1}.
\end{aligned} \tag{14}
$$

A few features of the model in (14) are worth noting. First, the common factors are the same as those in the model for $\beta' x_{t-1}$, namely, $\epsilon_{1t} - \epsilon_{2t}$ and ϵ_{1t}. Second, the common factors have a dynamic impact on the variables. Therefore, the number of static factors à la Stock and Watson (2002a, 2002b) in (14) would be larger than that of dynamic factors à la Forni et al. (2000, 2005). The difference

can be substantial in models with more dynamics. Third, the idiosyncratic errors are non-invertible MA(1) in almost all the equations, given by $\epsilon_{it} - \epsilon_{it-1}$. This feature remains valid in models with a more complex dynamics and suggests, as mentioned, that AR approximations to the equations of (14), namely FAVAR models, are inappropriate, at least from a theoretical point of view, when the factor model structure is (at least in part) due to cointegration. Finally, in this example the common factors driving the error correction terms, namely $\epsilon_{1t} - \epsilon_{2t}$ and ϵ_{1t}, are orthogonal to most of the errors ϵ_{1t},ϵ_{2t},...,ϵ_{Nt}, which makes (14) a proper factor model. However, as mentioned in the previous Section, typically the model for Δx_t no longer has a factor structure due to correlation between the driving forces of the error correction terms and the errors in the equations for the components of Δx_t.

Let us now assume that we are particularly interested in $x_{At} = (x_{2t}, x_{3t}, x_{4t})'$ and derive the subset ECM model for Δx_{At}. Since the three variables are driven by one stochastic trend, there will be two cointegrating relationships, whose parameters can be set equal to

$$\beta_A' = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

It can be shown that the pseudo-true values of the loadings of the cointegrating relationships are

$$\alpha_A = \begin{pmatrix} -1/7 & -1/7 \\ 6/7 & -1/7 \\ -1/7 & 6/7 \end{pmatrix}.$$

Hence, the ECM for Δx_{At} is

$$\Delta x_{At} = \alpha_A \beta_A' x_{t-1} + u_t, \tag{15}$$

where the errors follow a complex MA(2) process. Therefore, with respect to the equations for Δx_{At} in the ECM (11) for the whole vector Δx_t, there is a bias both in the long-run and short-run dynamics.

The FECM in this context requires modelling the variables $x_{ft} = (f_{1t}, x_{2t}, x_{3t}, x_{4t})'$, where the stochastic trend model in (12) implies that $f_{1t} = x_{1t-1}$. Therefore, the relevant equations of the FECM are

$$\begin{aligned} \Delta x_{2t} &= -(-f_{1t-1} + x_{2t-1}) + \epsilon_{2t} + \epsilon_{1t-1} \\ \Delta x_{3t} &= -(-f_{1t-1} + x_{2t-1}) - (-f_{1t-1} + x_{3t-1}) + \epsilon_{3t} + 2\epsilon_{1t-1} \\ \Delta x_{4t} &= -(-f_{1t-1} + x_{2t-1}) - (-f_{1t-1} + x_{4t-1}) + \epsilon_{4t} + 2\epsilon_{1t-1}. \end{aligned} \tag{16}$$

Comparing (16) with the subset of equations for Δx_{At} in the ECM (11), we see that α and β are unaffected, and the errors remain uncorrelated over time. It is worth recalling that both these properties no longer necessarily hold in more complex specifications, e.g., if the variables in x_{At} depend on more than three cointegrating relationships or on the lags of other variables in x_t. Moreover, the standard deviation of the errors in (16) increases with respect to (11), and the errors become correlated across equations. With respect to the corresponding equations in (14), the standard deviation of the errors is larger for Δx_{3t} and Δx_{4t}. It can instead be shown that the standard deviation of the errors of the FECM is smaller than that of the subset ECM in (15).

Finally, it is worth considering the equation for Δf_{1t}. From, (10), it can be written as either

$$\Delta f_{1t} = \epsilon_{1t-1}, \tag{17}$$

or

$$\Delta f_{1t} = -(-f_{1t-1} + x_{2t-1}) - \epsilon_{2t-1}. \tag{18}$$

The two representations are observationally equivalent. The former is in line with the theoretical model (7), and indicates that the changes in the factors should be weakly exogenous for the parameters of the cointegration relationships. However, standard econometric packages for VAR and cointegration analysis will use the latter representation, where Δf_{1t} is instead affected by the error correction term.

4 Monte Carlo experiments

In this section we conduct a set of simulation experiments to evaluate in finite samples the performance of the FECM, relative to that of an ECM and a FAVAR for the same small subset of variables of interest. An important feature to consider in the Monte Carlo design, is the way in which error-correcting or cointegrating information enters into the system for the variables of interest, i.e. whether the cointegrating vectors are common to each variable, or are idiosyncratic, or are a combination of the two. Another important aspect to bear in mind is how much cointegrating information needs to be incorporated, when looking at a sub-system of interest, from outside this sub-system. In the terminology established above, FECM should not in theory be able to handle well situations where there is a large number of error correction terms affecting each equation, or when the cointegrating relationships include all the variables in x_t and not just the subset x_{At}. However, in these cases, which are likely encountered in practical empirical situations, ECM and FAVAR would also experience serious problems. It is therefore worthwhile studying the performance of the alternative estimation methods using both simulations and empirical examples.

4.1 Design of the Monte Carlo

The basic data generating process (DGP) is the error correction mechanism

$$\Delta x_t = \alpha\beta' x_{t-1} + \epsilon_t, \tag{19}$$

where x_t is N-dimensional, α and β are of dimension $N \times N - r$, r is the number of common stochastic trends, and $\epsilon_t \sim N(0, I)$. We fix $r = 1$, set the cointegrating vectors equal to

$$\beta' = \begin{pmatrix} -1 & 1 & 0 & 0 & ... & 0 & 0 \\ -1 & 0 & 1 & 0 & ... & 0 & 0 \\ -1 & 0 & 0 & 1 & & & \\ ... & & & & & & \\ -1 & 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix},$$

and assume that we are particularly interested in the variables $x_{At} = (x_{2t}, x_{3t}, x_{4t})'$.

We then consider three versions of this DGP, which differ according to the shape of the matrix of

loadings, α. In DGP1, α is given by

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & ... & 0 \\ -1 & 0 & 0 & ... & 0 \\ 0 & -1 & 0 & ... & 0 \\ ... & & & & \\ 0 & 0 & 0 & ... & -1 \end{pmatrix},$$

so that each cointegrating relationship affects a single variable. This is a simplified version of the analytical example in the previous section. Using techniques similar to those used in the analytical example, it can be shown that the subset ECM for x_{At} leads to biases in α and β, and to correlated errors with a larger variance than those from the FECM. The ranking of the FAVAR and of the FECM should also favour the latter, since the model for Δx_t has a proper factor structure but the errors are non-invertible processes.

The loading matrix for DGP2 is

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & ... & 0 \\ -1 & 0 & 0 & ... & 0 \\ -1 & -1 & 0 & ... & 0 \\ -1 & 0 & -1 & & \\ ... & & & & \\ -1 & 0 & 0 & ... & -1 \end{pmatrix},$$

as in the analytical example in the previous section, so that one cointegrating relationship is common while the remaining $N-2$ relationships are idiosyncratic.

Finally, in DGP3 we set

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & ... & ... & ... & ... & ... & 0 & 0 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & ... & 0 & 0 \\ 0 & -1 & -1 & -1 & -1 & -1 & 0 & ... & ... & 0 \\ 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & ... & 0 \\ . & . & . & . & . & . & . & . & . & \\ . & . & . & . & . & . & . & . & . & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & -1 \end{pmatrix}.$$

This is a case where the ranking of the ECM and FECM is less clear-cut for two reasons. First, the FECM equations should depend on as many error correction terms as modelled variables, four, while at most three error correction terms can be included in the FECM. Second, some of the error correction terms depend on variables not modelled in the FECM, such as x_5 and x_6.

For all three DGPs, we consider the following configurations for T and N: $T \in (50, 100, 200, 500)$ and $N \in (50, 100, 200)$.

The comparisons between ECM, FECM and FAVAR are based on the residual variances for each estimated equation/variable in x_{At} normalized on the variance obtained from estimating the ECM.

The residual variances reported in the tables below are computed using the sample size T as the scaling term for the sum of squared residuals. We also have results, not reported here, where the residual variances are computed using a degrees-of-freedom correction for the number of parameters estimated under each method. While these results, available from us upon request, are quantitatively altered, the relative rankings of ECM versus FECM versus FAVAR are not. Rankings based on the adjusted-R^2 of each equation are also qualitatively similar and not reported to save space.

As discussed above, under correct specification as in most of our simulation experiments, the residual variance criterion yields a ranking equivalent to that resulting from a comparison of one-step ahead MSFEs. Instead, the equivalence does not necessarily hold in empirical applications, and therefore we also report the one-step-ahead MSFEs in our empirical examples in Section 5.

In the FECM, the number of cointegrating relationships is taken as given, although the cointegrating vectors and the loading matrix are estimated using maximum likelihood, see Johansen (1995). The factors are estimated from the levels of the data using the methods proposed by Bai (2004). His information criterion IPC_2 is used to select the number of factors.

In the ECM, the number of cointegrating relationships is taken as known. The cointegrating vectors and the loading matrix are again estimated.

Finally, in the FAVAR, the factors are estimated from the first differences of the data using the methods proposed by Stock and Watson (2002a, 2002b). Wherever the number of factors needs to be estimated, i.e. they are not imposed, the choice is based on the PC_2 criterion of Bai and Ng (2002).

4.2 Results

The results of the comparisons are reported in Tables 1 to 3 below. Each table contains, in its sub-panels, the results for each of the three equations (x_{2t}, x_{3t}, x_{4t}), the different methods, and the different combinations of N and T. Table 1 reports the results for DGP 1, where in panel A the number of factors is assumed known and is imposed while in panel B it is chosen according to Bai's (2004) IPC_2 information criterion when analyzing data in levels and Bai and Ng's (2002) PC_2 criterion for data in differences. Since in practice correlation of the errors of the dynamic system is handled by including additional lags of the differenced dependent variable (labelled q) (see equation (8)), we report the results for this lag length fixed to 1 and 3, $q = 1$ and $q = 3$ respectively, and also where it is chosen by using the Hannan-Quinn criterion, labelled HQ in the tables. We will refer to the two panels of Table 1 as Tables 1A and 1B. Tables 2A and 2B and Tables 3A and 3B report the corresponding results for DGP 2 and DGP 3 respectively. The average number of estimated factors (k_estim) is also reported in each table.

The results emerge with a great deal of clarity. For DGP 1, which is the simplest DGP, Table 1A indicates that FECM clearly dominates ECM and FAVAR, with gains typically in the range of 12% to 17%. In some cases the gain is even higher (in the order of 23%) especially where $T = 50$ (for almost all values of N) and the model includes 3 lags of the differenced dependent variable. FAVAR is often not better than ECM, except when T is large and selection of the lag length is by the Hannan-Quinn criterion which favours a lag-length selection of 3 or more in cases where the gains of FAVAR over ECM is the largest (for example where $T = 200$ or 500). The dominance of FECM however remains, and even when there are gains of FAVAR over ECM these are much smaller than those observed for FECM, i.e. in the neighbourhood of 5% in the best cases. Losing the long-run, by estimating the model in differences, is a major loss for the fit of the equations which is a finding which matches

quite precisely our predictions from the theory above. These results hold both when the number of factors is assumed to be known and when they are estimated (Table 1B), since, given that the data are generated with 1 factor imposed, the k_estim panel shows that the number of factors is chosen with high precision. When lag-length selection is undertaken using HQ, FECM and ECM are typically estimated without lags while FAVAR models include two or three lags. In this sense, by allowing for richer dynamics, HQ selection favours the FAVAR approach in comparison to the FECM, but is nevertheless insufficient to dilute the dominance of FECM.

For DGP 2, where the system for the first four variables is still self-contained (in the sense of there not being any extra cointegrating information coming from the rest of the system) but there is idiosyncratic cointegration, FECM continues to dominate FAVAR with the best results arising when q is set to 3. However, the gains are systematically smaller than for DGP1. As in DGP 1, except when favoured by generous lag-length selection, FAVAR is again worse than ECM for a sizeable fraction of cases.

For DGP 3, where each model under comparison is misspecified and our theoretical analysis leads us to anticipate difficulties, there is an interesting dominance of FECM over the other models. However, the size of the gains from FECM depends on the equation estimated. These are noticeable particularly for equation 1, where the gains are seen to be as large as 28%, although gains around 15% are more typical. For equations 2 and 3, the gains are in the neighbourhood of 5% other than in exceptional cases. The comparison between FAVAR and ECM is generally much more ambiguous, with FAVAR occasionally performing approximately 10% worse than ECM, for example when lag-length is fixed at 1,while HQ as usual favours a greater lag length and hence improved performance of FAVAR in terms of fit. The panel labelled k_estim now shows much less precision concerning the estimation of the number of factors especially for FAVAR where the more complicated cointegrating information in the DGP leads to the choice of 3 (dynamic) factors. This is in marked contrast to the corresponding results for DGP 1 and 2 presented previously.

Overall, the simulation results suggest that the FECM provides an excellent modelling choice, even in difficult circumstances. However, the best test of its performance is with real economic data which we consider in the next section of the paper.

5 Empirical examples

In this section we present two empirical examples as illustrations of the theoretical and simulation results presented above. The first example analyzes the relationships among US interest rates at different maturities, and among them and macroeconomic variables, an issue that is receiving increasing attention in the literature, see e.g. Diebold, Rudebusch and Arouba (2006) and the references therein. The second example reconsiders the famous article by King et al. (1991) on stochastic trends and economic fluctuations in the US economy.

In both examples, the factors are estimated from a large set of 110 monthly US macroeconomic variables, extracted from the dataset given in Stock and Watson (2005). The time span of the data series is 1959:1 to 2003:12, although for our examples we look only at a smaller interval, starting in the 1985. We focus on the post-1985 period, both to consider a homogenous monetary policy regime and to avoid the potentially problematic effects of the great moderation on factor estimation. The data series as well as the transformations implemented are listed in Table 4.

The number of factors is estimated using the criteria in Bai (2004) for the I(1) case, and in Bai

and Ng (2002) for the stationary case. Specifically, as in the simulations, we use their $IPC2$ and $PC2$ criteria respectively, which seem to have better finite sample properties.

Note that it is not the purpose of the estimation methodology proposed to identify the factors (which are incorporated in the FECM), since the estimated factors are not invariant to rotations of the space of factors. Instead, the factors proxy for and provide independent information on common trends, missing from both the standard ECM and the FAVAR. In particular, since the factors are orthogonal to each other they cannot be cointegrated - i.e. the additional cointegrating relations cannot simply be I(0) combinations of the factors being added, since such combinations are by construction impossible.

For each model, we report the standard R^2, the adjusted R^2 (denoted $\overline{R}^2$) and also the AIC and BIC criteria, in order to provide sufficient information for a comparison of the within-sample performance of each model. In addition, in order to assess the performance of these models in a forecasting context, we also report the MSFE and mean absolute error (MAE) for 1-step-ahead forecasts over the evaluation sample 1999:1 - 2003:12.

We provide a summary of the results in the two panels of Table 5, which will be called Tables 5A and 5B, with further details available from us upon request.

5.1 Interest rates at different maturities

We focus on four interest rates: the fed-fund, the 3-month t-bill rate, and the 1- and 10-year bond rates. Thus, in the notation of Section 2, $N_A = 4$. Over the sample under analysis, the variables tend to move closely together, with some more persistent deviations for the 10-year bond rate.

Empirically, the hypothesis of a unit root cannot be rejected for any series, using a standard ADF test with AIC or BIC lag-length selection. The interesting issue is whether and how many cointegrating relationships there are among the four rates. From a theoretical point of view, the expectational theory of the term structure implies the existence of 3 cointegrating vectors. However, when cointegration is tested with the Johansen (1988) trace statistic in a VAR with AIC or BIC lag-length selection, only two cointegrating vectors are detected (more formally, the hypothesis of at most one cointegrating vector is rejected), at the conventional 10% level. This result, $r_A = 2$ in the notation of Section 2, does not change either with the addition of a lag in the VAR to capture possible serial correlation in the residuals, or when using the maximum eigenvalue version of the test.

The fit of the resulting ECM model, which corresponds to equation (9), is summarized in the first row of the first panel of Table 5A.

A possible rationale for the finding of two cointegrating vectors among the four rates is that the interest rate spreads are likely driven by the evolution of the economic fundamentals, and omitting these variables from the analysis can spuriously decrease the number of cointegrating vectors. To evaluate whether this is the case, we have enlarged the information set with the estimated factors from the non-stationary large dataset (that includes the 110 variables less the 4 rates, i.e. $N = 110$ and $N_B = 106$), and jointly modelled the rates and the factors with a FECM, which corresponds to equation (8).

The Bai (2004) criterion suggests a single factor is sufficient to summarize the information in the whole dataset, but since it instead indicates the need for four factors for the subset of real variables (one for the nominal variables), and omitting relevant variables in the FECM is problematic, we prefer to proceed with four factors. In this case, the AIC and BIC criteria for lag-length determination indicate

either 3 or 1 lags in the VAR for the rates and the estimated factors, and again we prefer the less parsimonious specification to protect from omitted variable bias and serial correlation in the residuals.

For the FECM, the Johansen trace test indicates 4 cointegrating vectors. This is in line with the theoretical prediction of Section 2 that we should find in the FECM a cointegrating rank equal to N_A. The fit of the resulting FECM is summarized in the second row of Table 5A. There is a systematic increase both in R^2 and in $\overline{R}^2$ with respect to the ECM and, interestingly, the gains increase with the maturity.

Finally, we evaluate a FAVAR model, where the changes in the variables are regressed on their own lags and on lags of estimated factors, using two lags of each regressor as suggested by the information criteria. More precisely, the $N_B = 106$ macroeconomic variables plus the $N_A = 4$ interest rates are assumed to depend on a set of common factors and on an idiosyncratic error. Each variable is properly transformed to achieve stationarity; in particular, the interest rates are first differenced. The factors are estimated as the principal components of the (stationary) variables, while we recall that the factors in the FECM are extracted from the variables in levels. The Bai and Ng (2002) criterion indicates six factors.

From the third row of Table 5A it may be seen that both the R^2 and the $\overline{R}^2$ of the FAVAR are lower than those of the FECM for each of the four interest rates (even though the FECM uses only four factors). The strongest gains from the FECM arise from looking at the 10-year bond rate, which is in some sense an intuitive result given that long-run movements of the stock market are likely to be very relevant for this variable.

The second panel of Table 5A provides information on the computed AIC and BIC for the three models. The AIC ranking is very coherent with that reported in the first panel, while the BIC, which puts a stronger penalty on over-parameterization, prefers the more parsimonious ECM for 3-month and 10-year maturities.

The findings so far confirm empirically that it is important to take cointegration into account. Moreover, we recall that in the presence of cointegration the errors of the model for Δx_t are not invertible, so that they cannot be approximated by an AR process, as in the FAVAR, at least from a theoretical point of view.

The results reported in the third panel of Table 5A are, as expected, more ambiguous with respect to the efficacy of FECM models in a forecasting context. Comparisons of the (one-step ahead) MSFE and MAE criteria show that either the standard ECM and FECM provide better forecasts than FAVAR for each maturity. The comparison between the ECM and FECM is more mixed, attributable perhaps to the fact that the factor space is estimated and may thus be susceptible to the presence of structural breaks (which are of course important for forecasting and are not taken account of here). In future research it would be interesting to consider modifications of the FECM model to take account of structural breaks - along the lines of a differenced FECM model (DFECM) to correspond to the Hendry (2006) formulation of a DVECM model described briefly in the introduction, in order to allow for change in the cointegrating or equilibrium information that may have occurred.

5.2 Stochastic trends and economic fluctuations

As a second example, we consider an updated and slightly simplified version of the model in King et al. (1991) henceforth KPSW. KPSW analyzed a system with 6 variables at the quarterly level, over the period 1949-1988: per capita real consumption, per capita gross private fixed investment,

per capita "private" gross national product, money supply, inflation and a short term interest rate. They detected three cointegrating vectors, which they identified as a money demand function (where real money depends on GNP and the interest rate), a consumption equation (where the ratio of consumption to GNP depends on the real interest rate), and an investment equation (where the ratio of investment to GNP depends on the real interest rate).

Since we have monthly time series, we focus on four variables ($N_A = 4$): real consumption (C), real personal income (PI), real money (M), and real interest rate (Ri), where the first three variables are expressed in logs. We consider again the sample 1985-2003, and focus on three models: ECM, FECM, and FAVAR.[7]

The AIC and BIC criteria select 2 lags in the VAR, and in this case the Johansen trace test detects two cointegrating vectors, i.e. $r_A = 2$ (more formally, the hypothesis of at most one cointegrating vector is rejected), at the conventional 10% level. The cointegrating vectors are similar to the money demand and consumption equations of KPSW, except that personal income appears not to matter in the former. The fit of the resulting ECM model (the counterpart of the theoretical equation (9)) is summarized in the first row of the first panel of Table 5B.

We then enlarge the information set with the estimated factors from the non-stationary large dataset (that includes the $N = 110$ variables less the $N_A = 4$ variables included in the ECM), and jointly model the four variables and the factors with a FECM (equation (8)). As in the previous example, and not surprisingly since the data are mostly the same, the Bai (2004) criterion suggests a single factor but it indicates four factors for the subset of real variables. Therefore, we proceed with four factors. In this case, the AIC and BIC criteria for lag-length determination indicate either 3 or 2 lags in the extended VAR and, as in the previous example, we prefer the less parsimonious specification to protect from omitted variable bias and serial correlation in the residuals. In this case, the Johansen trace test suggests 4 cointegrating vectors, two more than the standard ECM. This result is again in line with the theoretical prediction of rank equal to N_A. The fit of the resulting FECM is summarized in the second row of Table 5B.

As in the previous example, the performance of the FAVAR is slightly but systematically worse than that of the FECM, which also dominates the ECM in terms of fit.[8] This further reinforces the message that it is important to take cointegration between the variables and the factors explicitly into consideration.

The results reported in the second panel of Table 5B show that the ranking of the models is virtually unaltered according to the AIC, while, as in the case of the previous empirical example, the BIC prefers the more parsimonious ECM in most cases. For each variable, the FAVAR performs worst according to both AIC and BIC.

The final panel of Table 5B reports more mixed results when the models are used for one-step ahead forecasting. In particular, the FAVAR is best for the real interest rate, the ECM for real consumption, and the FECM for personal income and real money. Also in this case the mixed results could be related to the presence of structural breaks, and as above, research into robustifying the FECM to the presence of such breaks is an important element of our future research.

[7] Comparable results are obtained in a five variable system where the real interest rate is split into the nominal rate and the inflation rate.

[8] The Bai and Ng (2002) criteria indicate again six factors (extracted from the 106 macroeconomic variables plus the four variables under analysis in this example, after a proper transformation of each variable to achieve stationarity).

6 Conclusions

In this paper we study the case where a large set of variables are linked by cointegration relationships, which is a very important topic both from a theoretical point of view and for empirical applications. Early studies, such as Stock and Watson (1988), show that (the levels of) each cointegrated variable is driven by a limited number of common integrated trends plus an idiosyncratic stationary error term. Therefore, the variables in levels can be represented as a factor model, where orthogonality between the common and the idiosyncratic components is guaranteed by the fact that the former is integrated while the latter is stationary by construction.

A first result of this paper is to notice that, in general, the factor structure is lost when the differences of the variables are modelled. In fact, even though the first differences of the factors are driving all the variables, they are no longer necessarily orthogonal to the "idiosyncratic" errors. Moreover, even when the factors are orthogonal to the idiosyncratic errors, the latter are non-invertible processes. While this is not a problem for factor estimation, the presence of non-invertible errors does not allow autoregressive approximations of the factor model, FAVAR, which are instead commonly used in the literature.

The presence of the non-invertible errors in the model for the variables in differences is related to the omission of the error correction terms. Hence, we introduce the FECM which requires us to summarize the information in the (levels of the) large set of variables with a limited number of factors, and then to model jointly the factors and the variables of interest with a cointegrated VAR.

The FECM improves upon the standard small scale ECM by protecting from omitted variable bias both in the long run and in the short run. It also improves upon the FAVAR model by taking long run restrictions into explicit account. However, the FECM remains an approximation, which is expected to work well only under certain conditions, in particular when the few variables of interest are influenced by a limited number of error correction terms.

Both Monte Carlo experiments and empirical analyses show that the FECM performs often better than ECM and FAVAR models.

To conclude, we believe that the FECM represents an interesting modelling approach, and a natural generalization of the FAVAR (to include long run information) and ECM (to include information from a large set of *cointegrated* variables). Because of this, the FECM is of potential usefulness in a wide range of empirical analyses.

References

[1] Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics*, **122**, 137-183.

[2] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191-221.

[3] Bai, J, and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, **72**, 1127-1177.

[4] Bai, J, and S. Ng (2006). Confidence intervals for diffusion index forecasts with a large number of predictors and inference for factor-augmented regressions. *Econometrica,* **74**, 1133-1150.

[5] Bai, J, Kao, C. and S. Ng (2007).Panel cointegration with global stochastic trends. Center for Policy Research Working Paper No. 90, Syracuse University.

[6] Banerjee, A. and J. Carrion-i-Silvestre (2007). Cointegration in panel data with breaks and cross-section dependence, mimeo.

[7] Banerjee, A., M. Marcellino and C. Osbat (2004). Some cautions on the use of panel methods for integrated series of macroeconomic data. *Econometrics Journal*, **7**, 322-340.

[8] Bernanke, B.S., J. Boivin and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, **120**, 387-422.

[9] Breitung, J. and S. Das (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica,* **59**, 414–433.

[10] Breitung, J. and S. Das (2007). Testing for unit roots in panels with a factor structure. Forthcoming *Econometric Theory.*

[11] Clements, M.P. and D.F. Hendry (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127-146.

[12] Davidson, J.E.H., D.F. Hendry, F. Srba and J. S. Yeo (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661-692.

[13] Diebold, F.X., G. Rudebusch and S.B. Arouba (2005). The macroeconomics and the yield curve: a dynamic latent variable approach. Forthcoming *Journal of Econometrics.*

[14] Engle, R.F. and C.W. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, **55**, 257-276.

[15] Engle, R.F. and S. Kozicki (1993). Testing for common features. *Journal of Business and Economic Statistics*, **11**, 369-390.

[16] Favero, C., M. Marcellino and F. Neglia (2005). Principal components at work: the empirical analysis of monetary policy with large data sets. *Journal of Applied Econometrics*, **20**, 603-620.

[17] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000). The generalized dynamic-factor model. *Review of Economics and Statistics*, **82**, 540-554.

[18] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2005). The generalized dynamic factor model. *Journal of the American Statistical Association*, **100**, 830-840.

[19] Hendry, D.F. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, **135**, 399-426.

[20] Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231-254.

[21] Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press, Oxford and New York.

[22] King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991). Stochastic trends and economic fluctuations. *American Economic Review*, **81**, 819-840.

[23] Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, **74**, 967–1012.

[24] Stock, J.H. and M.W. Watson (1998). Testing for common trends. *Journal of the American Statistical Association*, **83**, 1097-1107.

[25] Stock, J.H. and M.W. Watson (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, **97**, 1167-1179.

[26] Stock, J.H. and M.W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. J*ournal of Business and Economic Statistics*, **20**, 147-162.

[27] Stock, J.H. and M.W. Watson (2005). Implication of dynamic factor models for VAR analysis. NBER Working Paper 11467.

Table 1: Results for DGP1
Ratios of Residual Variances

A: Number of factors imposed							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.829	1.177	0.829	1.177	0.831	1.179
	q = 3	0.771	1.030	0.771	1.033	0.773	1.032
	HQ	0.841	1.019	0.838	1.026	0.839	1.018
100	q = 1	0.853	1.137	0.852	1.138	0.853	1.137
	q = 3	0.833	1.009	0.834	1.011	0.833	1.007
	HQ	0.840	0.979	0.841	0.982	0.841	0.980
200	q = 1	0.864	1.122	0.863	1.121	0.864	1.121
	q = 3	0.856	1.004	0.854	1.003	0.856	1.004
	HQ	0.846	0.946	0.847	0.943	0.846	0.945
500	q = 1	0.869	1.113	0.869	1.113	0.869	1.113
	q = 3	0.866	1.001	0.866	1.002	0.867	1.001
	HQ	0.861	0.942	0.860	0.943	0.861	0.943

B: Number of factors estimated							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.833	1.165	0.831	1.162	0.831	1.162
	q = 3	0.770	0.968	0.771	0.971	0.770	0.969
	HQ	0.839	0.967	0.839	0.965	0.838	0.969
100	q = 1	0.854	1.138	0.853	1.136	0.853	1.138
	q = 3	0.833	1.008	0.832	1.007	0.833	1.009
	HQ	0.845	0.981	0.842	0.978	0.841	0.976
200	q = 1	0.865	1.121	0.864	1.122	0.865	1.122
	q = 3	0.855	1.003	0.855	1.004	0.855	1.003
	HQ	0.848	0.946	0.848	0.946	0.849	0.947
500	q = 1	0.869	1.112	0.869	1.113	0.869	1.112
	q = 3	0.867	1.002	0.867	1.002	0.867	1.001
	HQ	0.859	0.942	0.860	0.942	0.860	0.943

A: Number of factors imposed							
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.836	1.178	0.836	1.179	0.835	1.178
	q = 3	0.774	1.026	0.773	1.025	0.772	1.026
	HQ	0.841	1.019	0.842	1.031	0.843	1.024
100	q = 1	0.858	1.136	0.858	1.135	0.858	1.135
	q = 3	0.837	1.004	0.840	1.011	0.837	1.005
	HQ	0.847	0.981	0.847	0.980	0.846	0.979
200	q = 1	0.868	1.118	0.868	1.119	0.868	1.119
	q = 3	0.860	1.001	0.858	0.998	0.860	1.000
	HQ	0.852	0.941	0.853	0.945	0.852	0.942
500	p = 1	0.875	1.111	0.874	1.111	0.874	1.111
	p = 3	0.872	0.998	0.872	0.998	0.872	0.999
	HQ	0.864	0.938	0.864	0.938	0.865	0.939

B: Number of factors estimated							
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.834	1.176	0.835	1.172	0.835	1.175
	q = 3	0.776	1.012	0.775	1.009	0.774	1.009
	HQ	0.845	1.003	0.845	1.001	0.846	0.999
100	q = 1	0.858	1.136	0.859	1.138	0.858	1.134
	q = 3	0.837	1.006	0.838	1.007	0.838	1.006
	HQ	0.850	0.980	0.849	0.981	0.849	0.977
200	q = 1	0.868	1.118	0.868	1.119	0.868	1.119
	q = 3	0.859	0.999	0.860	1.000	0.859	1.000
	HQ	0.852	0.941	0.852	0.942	0.853	0.944
500	p = 1	0.874	1.110	0.874	1.110	0.874	1.110
	p = 3	0.871	0.998	0.872	0.998	0.872	0.999
	HQ	0.865	0.939	0.865	0.939	0.865	0.939

A: Number of factors imposed							
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.837	1.176	0.838	1.178	0.838	1.174
	q = 3	0.774	1.033	0.774	1.026	0.777	1.033
	HQ	0.846	1.022	0.846	1.012	0.843	1.014
100	q = 1	0.860	1.134	0.861	1.135	0.860	1.133
	q = 3	0.838	1.005	0.840	1.003	0.840	1.003
	HQ	0.848	0.974	0.848	0.976	0.847	0.973
200	q = 1	0.871	1.119	0.871	1.117	0.871	1.117
	q = 3	0.862	1.000	0.861	0.998	0.861	0.998
	HQ	0.855	0.943	0.856	0.941	0.855	0.941
500	q = 1	0.876	1.110	0.876	1.109	0.876	1.110
	q = 3	0.874	0.998	0.874	0.997	0.873	0.996
	HQ	0.867	0.938	0.867	0.938	0.867	0.937

B: Number of factors estimated							
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.837	1.178	0.837	1.178	0.837	1.178
	q = 3	0.778	1.026	0.778	1.029	0.778	1.028
	HQ	0.843	1.013	0.845	1.013	0.844	1.012
100	q = 1	0.859	1.133	0.860	1.135	0.860	1.135
	q = 3	0.840	1.005	0.839	1.004	0.840	1.004
	HQ	0.848	0.974	0.848	0.973	0.850	0.977
200	q = 1	0.871	1.117	0.871	1.118	0.871	1.118
	q = 3	0.862	0.999	0.862	0.999	0.862	0.999
	HQ	0.852	0.940	0.852	0.939	0.852	0.938
500	q = 1	0.876	1.109	0.877	1.110	0.876	1.110
	q = 3	0.874	0.997	0.874	0.997	0.874	0.997
	HQ	0.868	0.938	0.868	0.939	0.868	0.938

k_estim								
N=50	FECM	FAVAR	N=100	FECM	FAVAR	N=200	FECM	FAVAR
T	SW	SW	T	SW	SW	T	SW	SW
50	1	1.595	50	1	1.188	50	1	1.008
100	1	1.001	100	1	1	100	1	1
200	1	1	200	1	1	200	1	1
500	1	1	500	1	1	500	1	1

Notes: Each cell of the table (i.e. for each equation, estimation method and (N,T) configuration) in the panel labelled 'Ratios of Residual Variances' records the residual variance of the equation relative to the residual variance obtained from estimating, for the same configuration, the subset ECM consisting of (x(2t), x(3t), x(4t)) only.
The results are reported where the number of lagged differenced variables is fixed to 1 (q=1), where it is fixed to 3 (q=3) and where it is chosen by the HQ criterion (HQ).
Equation 1 refers to the equation for x(2t), Equation 2 to x(3t) and Equation 3 to x(4t). FECM-SW estimates the factor error-correction model with the factors extracted in levels according to Bai (2004). FAVAR-SW estimates factor augmented VAR models with factors extracted from differences of the data according to Stock and Watson (2002).
The results are based on 10,000 replications for q=1 and 2,000 replications for q=3 and HQ.
The panel labelled k_estim records the (average) number (across replications) of estimated factors.

Table 2: Results for DGP2
Ratios of Residual Variances

A: Number of factors imposed							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.893	1.206	0.867	1.071	0.867	1.067
	q = 3	0.832	1.077	0.806	0.991	0.808	0.995
	HQ	1.006	1.075	1.001	0.953	0.998	0.951
100	q = 1	0.921	1.172	0.892	1.036	0.892	1.037
	q = 3	0.902	1.065	0.871	0.975	0.871	0.974
	HQ	0.942	1.092	0.920	0.977	0.922	0.979
200	q = 1	0.932	1.160	0.902	1.026	0.902	1.025
	q = 3	0.924	1.061	0.895	0.972	0.894	0.974
	HQ	0.933	1.045	0.903	0.951	0.902	0.952
500	q = 1	0.938	1.153	0.907	1.020	0.907	1.020
	q = 3	0.936	1.062	0.906	0.974	0.906	0.974
	HQ	0.939	1.022	0.908	0.944	0.908	0.944

B: Number of factors estimated							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.894	1.190	0.868	1.050	0.868	1.049
	q = 3	0.829	1.012	0.801	0.934	0.802	0.932
	HQ	1.003	1.017	0.999	0.908	0.995	0.912
100	q = 1	0.921	1.173	0.892	1.036	0.891	1.036
	q = 3	0.900	1.063	0.871	0.975	0.870	0.973
	HQ	0.942	1.099	0.918	0.982	0.917	0.981
200	q = 1	0.932	1.160	0.902	1.026	0.902	1.026
	q = 3	0.924	1.062	0.895	0.973	0.894	0.973
	HQ	0.933	1.049	0.903	0.953	0.903	0.952
500	q = 1	0.938	1.154	0.907	1.021	0.907	1.020
	q = 3	0.937	1.062	0.906	0.974	0.906	0.973
	HQ	0.940	1.022	0.909	0.945	0.909	0.945

A: Number of factors imposed							
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.897	1.208	0.873	1.068	0.872	1.066
	q = 3	0.834	1.074	0.808	0.993	0.807	0.989
	HQ	1.023	1.069	1.015	0.947	1.016	0.945
100	q = 1	0.925	1.173	0.898	1.036	0.898	1.036
	q = 3	0.903	1.062	0.875	0.972	0.876	0.973
	HQ	0.946	1.094	0.924	0.978	0.924	0.979
200	q = 1	0.936	1.159	0.908	1.023	0.908	1.025
	q = 3	0.929	1.063	0.900	0.971	0.900	0.971
	HQ	0.937	1.047	0.909	0.952	0.908	0.952
500	q = 1	0.943	1.153	0.914	1.020	0.914	1.020
	q = 3	0.941	1.061	0.912	0.972	0.912	0.972
	HQ	0.943	1.022	0.913	0.942	0.913	0.943

B: Number of factors estimated							
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.897	1.201	0.873	1.063	0.872	1.062
	q = 3	0.834	1.055	0.808	0.973	0.808	0.972
	HQ	1.013	1.057	1.011	0.938	1.011	0.941
100	q = 1	0.923	1.171	0.897	1.035	0.896	1.035
	q = 3	0.903	1.062	0.876	0.973	0.876	0.973
	HQ	0.945	1.093	0.924	0.979	0.923	0.978
200	q = 1	0.937	1.160	0.908	1.025	0.908	1.025
	q = 3	0.927	1.060	0.899	0.971	0.899	0.972
	HQ	0.938	1.046	0.909	0.951	0.910	0.952
500	q = 1	0.943	1.154	0.913	1.020	0.913	1.019
	q = 3	0.940	1.060	0.912	0.972	0.912	0.973
	HQ	0.943	1.020	0.914	0.942	0.913	0.942

A: Number of factors imposed							
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.900	1.212	0.874	1.067	0.874	1.068
	q = 3	0.834	1.075	0.814	0.994	0.814	0.992
	HQ	1.012	1.072	1.009	0.953	1.006	0.949
100	q = 1	0.926	1.174	0.900	1.035	0.900	1.034
	q = 3	0.904	1.060	0.875	0.967	0.876	0.972
	HQ	0.946	1.088	0.925	0.974	0.925	0.974
200	q = 1	0.938	1.158	0.910	1.024	0.911	1.024
	q = 3	0.930	1.059	0.903	0.971	0.903	0.970
	HQ	0.937	1.043	0.909	0.949	0.910	0.948
500	q = 1	0.945	1.154	0.917	1.019	0.917	1.019
	q = 3	0.942	1.060	0.915	0.971	0.916	0.972
	HQ	0.944	1.019	0.915	0.941	0.915	0.941

B: Number of factors estimated							
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.900	1.206	0.876	1.067	0.875	1.067
	q = 3	0.836	1.075	0.810	0.990	0.810	0.990
	HQ	1.008	1.061	1.004	0.941	1.006	0.945
100	q = 1	0.926	1.172	0.900	1.035	0.900	1.036
	q = 3	0.905	1.062	0.879	0.973	0.879	0.972
	HQ	0.950	1.095	0.929	0.980	0.929	0.980
200	q = 1	0.938	1.159	0.911	1.024	0.911	1.025
	q = 3	0.929	1.060	0.902	0.970	0.903	0.971
	HQ	0.938	1.043	0.910	0.949	0.911	0.950
500	q = 1	0.944	1.152	0.916	1.019	0.916	1.019
	q = 3	0.942	1.061	0.915	0.972	0.916	0.972
	HQ	0.944	1.020	0.917	0.942	0.916	0.942

k_estim								
N=50	FECM	FAVAR	N=100	FECM	FAVAR	N=200	FECM	FAVAR
T	SW	SW	T	SW	SW	T	SW	SW
50	1	1.607	50	1	1.203	50	1	1.011
100	1	1.002	100	1	1	100	1	1
200	1	1	200	1	1	200	1	1
500	1	1	500	1	1	500	1	1

Notes: See notes to Table 1

Table 3: Results for DGP3
Ratios of Residual Variances

A: Number of factors imposed							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.850	1.027	0.919	1.103	0.934	1.125
	q = 3	0.825	1.009	0.858	1.029	0.865	1.041
	HQ	0.988	0.947	1.060	1.003	1.072	1.019
100	q = 1	0.869	0.999	0.944	1.076	0.962	1.096
	q = 3	0.894	0.991	0.934	1.022	0.945	1.038
	HQ	0.935	0.927	1.003	0.984	1.020	1.005
200	q = 1	0.880	0.994	0.956	1.070	0.975	1.087
	q = 3	0.919	0.991	0.961	1.024	0.972	1.037
	HQ	0.938	0.940	0.996	0.987	1.011	1.004
500	q = 1	0.887	0.992	0.963	1.068	0.983	1.084
	q = 3	0.932	0.991	0.974	1.026	0.985	1.039
	HQ	0.944	0.965	0.992	1.003	1.003	1.016

B: Number of factors estimated							
		Equation 1		Equation 2		Equation 3	
N=50	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.814	0.964	0.881	1.006	0.914	1.010
	q = 3	0.717	0.798	0.749	0.802	0.760	0.802
	HQ	0.974	0.804	1.038	0.823	1.115	0.825
100	q = 1	0.850	0.966	0.920	1.007	0.965	1.008
	q = 3	0.844	0.905	0.883	0.910	0.900	0.911
	HQ	0.931	0.908	0.993	0.936	1.075	0.942
200	q = 1	0.877	0.971	0.952	1.001	0.977	1.001
	q = 3	0.911	0.943	0.953	0.941	0.966	0.939
	HQ	0.933	0.928	0.991	0.935	1.017	0.936
500	q = 1	0.887	0.975	0.963	0.981	0.983	0.975
	q = 3	0.932	0.964	0.975	0.947	0.985	0.936
	HQ	0.943	0.959	0.991	0.929	1.003	0.916

A: Number of factors imposed							
		Equation 1		Equation 2		Equation 3	
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.856	1.028	0.922	1.104	0.936	1.126
	q = 3	0.823	1.006	0.859	1.040	0.865	1.054
	HQ	0.999	0.950	1.064	1.008	1.077	1.025
100	q = 1	0.875	1.001	0.945	1.080	0.963	1.100
	q = 3	0.898	0.988	0.935	1.025	0.944	1.040
	HQ	0.938	0.922	1.003	0.984	1.020	1.004
200	q = 1	0.885	0.995	0.957	1.077	0.976	1.094
	q = 3	0.923	0.986	0.961	1.025	0.972	1.041
	HQ	0.942	0.936	0.997	0.987	1.011	1.005
500	q = 1	0.892	0.993	0.965	1.074	0.984	1.091
	q = 3	0.937	0.988	0.976	1.027	0.986	1.042
	HQ	0.949	0.961	0.993	1.003	1.004	1.018

B: Number of factors estimated							
		Equation 1		Equation 2		Equation 3	
N=100	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.831	0.971	0.896	1.014	0.926	1.017
	q = 3	0.743	0.811	0.774	0.809	0.784	0.812
	HQ	0.984	0.791	1.046	0.814	1.114	0.815
100	q = 1	0.871	0.972	0.942	1.017	0.964	1.019
	q = 3	0.888	0.908	0.925	0.915	0.937	0.918
	HQ	0.941	0.904	1.004	0.938	1.033	0.944
200	q = 1	0.886	0.978	0.957	1.020	0.976	1.020
	q = 3	0.924	0.948	0.962	0.954	0.972	0.956
	HQ	0.941	0.923	0.997	0.944	1.011	0.949
500	q = 1	0.892	0.981	0.964	1.008	0.984	1.005
	q = 3	0.938	0.967	0.976	0.963	0.986	0.959
	HQ	0.949	0.945	0.993	0.941	1.004	0.937

A: Number of factors imposed							
		Equation 1		Equation 2		Equation 3	
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.857	1.030	0.921	1.108	0.934	1.131
	q = 3	0.828	1.007	0.860	1.038	0.867	1.057
	HQ	1.000	0.946	1.064	1.010	1.075	1.029
100	q = 1	0.875	1.000	0.945	1.083	0.962	1.104
	q = 3	0.902	0.987	0.937	1.026	0.946	1.044
	HQ	0.941	0.921	1.004	0.986	1.021	1.009
200	q = 1	0.887	0.995	0.958	1.079	0.976	1.097
	q = 3	0.925	0.984	0.962	1.024	0.972	1.042
	HQ	0.945	0.935	0.998	0.988	1.012	1.007
500	q = 1	0.895	0.993	0.966	1.077	0.984	1.094
	q = 3	0.940	0.986	0.977	1.027	0.986	1.044
	HQ	0.951	0.959	0.993	1.002	1.004	1.019

B: Number of factors estimated							
		Equation 1		Equation 2		Equation 3	
N=200	lags	FECM	FAVAR	FECM	FAVAR	FECM	FAVAR
T		SW	SW	SW	SW	SW	SW
50	q = 1	0.855	0.981	0.919	1.027	0.934	1.038
	q = 3	0.813	0.842	0.844	0.851	0.851	0.855
	HQ	1.002	0.820	1.060	0.847	1.081	0.853
100	q = 1	0.876	0.977	0.946	1.030	0.963	1.035
	q = 3	0.899	0.917	0.935	0.930	0.944	0.935
	HQ	0.945	0.904	1.008	0.938	1.022	0.945
200	q = 1	0.888	0.981	0.958	1.028	0.976	1.028
	q = 3	0.925	0.948	0.962	0.956	0.972	0.958
	HQ	0.945	0.923	0.998	0.945	1.011	0.949
500	q = 1	0.894	0.984	0.965	1.024	0.984	1.022
	q = 3	0.939	0.968	0.977	0.973	0.986	0.973
	HQ	0.951	0.952	0.993	0.952	1.004	0.951

k_estim								
N=50	FECM	FAVAR	N=100	FECM	FAVAR	N=200	FECM	FAVAR
T	SW	SW	T	SW	SW	T	SW	SW
50	1.9952	3.046	50	1.8387	2.9871	50	1.1446	2.6085
100	1.9638	3	100	1.1969	2.995	100	1.0004	2.7457
200	1.1718	3	200	1	3	200	1	2.9752
500	1	3	500	1	3	500	1	3

Notes: See notes to Table 1

Table 4: Dataset for the empirical examples

Code	Short Descrip.	Nom	Real	Fin	Tcode I(0) dataset	Tcode I(1) dataset
a0m052	PI	0	1	0	5	4
A0M051	PI less transfers	0	1	0	5	4
A0M224_R	Consumption	0	1	0	5	4
A0M057	M&T sales	0	1	0	5	4
A0M059	Retail sales	0	1	0	5	4
IPS10	IP: total	0	1	0	5	4
IPS11	IP: products	0	1	0	5	4
IPS299	IP: final prod	0	1	0	5	4
IPS12	IP: cons gds	0	1	0	5	4
IPS13	IP: cons dble	0	1	0	5	4
IPS18	iIP:cons nondble	0	1	0	5	4
IPS25	IP:bus eqpt	0	1	0	5	4
IPS32	IP: matls	0	1	0	5	4
IPS34	IP: dble mats	0	1	0	5	4
IPS38	IP:nondble mats	0	1	0	5	4
IPS43	IP: mfg	0	1	0	5	4
IPS307	IP: res util	0	1	0	5	4
IPS306	IP: fuels	0	1	0	5	4
PMP	NAPM prodn	0	1	0	1	1
A0m082	Cap util	0	1	0	2	1
LHEL	Help wanted indx	0	1	0	2	1
LHELX	Help wanted/emp	0	1	0	2	1
LHEM	Emp CPS total	0	1	0	5	4
LHNAG	Emp CPS nonag	0	1	0	5	4
LHUR	U: all	0	1	0	2	1
LHU680	U: mean duration	0	1	0	2	1
LHU5	U < 5 wks	0	1	0	5	4
LHU14	U 5-14 wks	0	1	0	5	4
LHU15	U 15+ wks	0	1	0	5	4
LHU26	U 15-26 wks	0	1	0	5	4
LHU27	U 27+ wks	0	1	0	5	4
A0M005	UI claims	0	1	0	5	4
CES002	Emp: total	0	1	0	5	4
CES003	Emp: gds prod	0	1	0	5	4
CES006	Emp: mining	0	1	0	5	4
CES011	Emp: const	0	1	0	5	4
CES015	Emp: mfg	0	1	0	5	4
CES017	Emp: dble gds	0	1	0	5	4
CES033	Emp: nondbles	0	1	0	5	4
CES046	Emp: services	0	1	0	5	4
CES048	Emp: TTU	0	1	0	5	4
CES049	Emp: wholesale	0	1	0	5	4
CES053	Emp: retail	0	1	0	5	4
CES088	Emp: FIRE	0	1	0	5	4
CES140	Emp: Govt	0	1	0	5	4
A0M048	Emp-hrs nonag	0	1	0	5	4
CES151	Avg hrs	0	1	0	1	1
CES155	Overtime: mfg	0	1	0	2	1
aom001	Avg hrs: mfg	0	1	0	1	1
PMEMP	NAPM empl	0	1	0	1	1
HSFR	HStarts: Total	0	1	0	4	4
HSNE	HStarts: NE	0	1	0	4	4
HSMW	HStarts: MW	0	1	0	4	4
HSSOU	HStarts: South	0	1	0	4	4
HSWST	HStarts: West	0	1	0	4	4

Code	Short Descrip.	Nom	Real	Fin	Tcode I(0) dataset	Tcode I(1) dataset
HSBR	BP: total	0	1	0	4	4
HSBNE	BP: NE	0	1	0	4	4
HSBMW	BP: MW	0	1	0	4	4
HSBSOU	BP: South	0	1	0	4	4
HSBWST	BP: West	0	1	0	4	4
PMI	PMI	0	1	0	1	1
PMNO	NAPM new or	0	1	0	1	1
PMDEL	NAPM vendor	0	1	0	1	1
PMNV	NAPM Invent	0	1	0	1	1
A0M008	Orders: cons g	0	1	0	5	4
A0M007	Orders: dble g	0	1	0	5	4
A0M027	Orders: cap gc	0	1	0	5	4
A1M092	Unf orders: db	0	1	0	5	4
A0M070	M&T invent	0	1	0	5	4
A0M077	M&T invent/s	0	1	0	2	1
FM1	M1	1	0	0	6	5
FM2	M2	1	0	0	6	5
FM3	M3	1	0	0	6	5
FM2DQ	M2 (real)	1	0	0	5	4
FMFBA	MB	1	0	0	6	5
FMRRA	Reserves tot	1	0	0	6	5
FMRNBA	Reserves nont	1	0	0	6	5
FCLNQ	C&I loans	1	0	0	6	5
FCLBMC	C&I loans	1	0	0	1	1
CCINRV	Cons credit	1	0	0	6	5
A0M095	Inst cred/PI	1	0	0	2	1
FYFF	FedFunds	0	0	1	2	1
FYGM3	3 mo T-bill	0	0	1	2	1
FYGT1	1 yr T-bond	0	0	1	2	1
FYGT10	10 yr T-bond	0	0	1	2	1
PWFSA	PPI: fin gds	1	0	0	6	5
PWFCSA	PPI: cons gds	1	0	0	6	5
PWIMSA	PPI: int mat'ls	1	0	0	6	5
PWCMSA	PPI: crude mat	1	0	0	6	5
PSCCOM	Commod: spot	1	0	0	6	5
PSM99Q	Sens mat'ls pr	1	0	0	6	5
PMCP	NAPM com pr	1	0	0	1	1
PUNEW	CPI-U: all	1	0	0	6	5
PU83	CPI-U: apparel	1	0	0	6	5
PU84	CPI-U: transp	1	0	0	6	5
PU85	CPI-U: medica	1	0	0	6	5
PUC	CPI-U: comm.	1	0	0	6	5
PUCD	CPI-U: dbles	1	0	0	6	5
PUS	CPI-U: services	1	0	0	6	5
PUXF	CPI-U: ex food	1	0	0	6	5
PUXHS	CPI-U: ex shelt	1	0	0	6	5
PUXM	CPI-U: ex med	1	0	0	6	5
GMDC	PCE defl	1	0	0	6	5
GMDCD	PCE defl: dlbes	1	0	0	6	5
GMDCN	PCE defl: nonc	1	0	0	6	5
GMDCS	PCE defl: servi	1	0	0	6	5
CES275	AHE: goods	1	0	0	6	5
CES277	AHE: const	1	0	0	6	5
CES278	AHE: mfg	1	0	0	6	5
HHSNTN	Consumer exp	0	1	0	2	1

Notes:

Transformation codes: 1 no transformation; 2 first difference; 3 second difference;
4 logarithm; 5 first difference of logarithm; 6 second difference of logarithm.
Dataset extracted from Stock and Watson (2005). Sample is 1985:1-2003:12

Table 5. Empirical analyses

A - Alternative models for interest rates

	R-squared				Adjusted R-squared			
	FF	3m	1y	10y	FF	3m	1y	10y
ECM (1 lag, 2 coint.)	0.41	0.31	0.23	0.11	0.40	0.29	0.21	0.09
FECM (2 lags, 4 facs-lev, 4 coint.)	0.49	0.42	0.40	0.31	0.44	0.36	0.34	0.24
FAVAR (2 lags, 6 facs)	0.46	0.41	0.37	0.25	0.41	0.35	0.31	0.17
	AIC				**BIC**			
	FF	3m	1y	10y	FF	3m	1y	10y
ECM (1 lag, 2 coint.)	-0.42	-0.52	0.09	0.14	-0.29	-0.39	0.22	0.27
FECM (2 lags, 4 facs-lev, 4 coint.)	-0.65	-0.65	-0.18	0.00	-0.33	-0.33	0.14	0.32
FAVAR (2 lags, 6 facs)	-0.59	-0.63	-0.14	0.09	-0.27	-0.32	0.18	0.41
	MSFE				**MAE**			
	FF	3m	1y	10y	FF	3m	1y	10y
ECM (1 lag, 2 coint.)	0.016	0.031	0.043	0.069	0.098	0.135	0.162	0.211
FECM (2 lags, 4 facs-lev, 4 coint.)	0.033	0.023	0.037	0.101	0.146	0.119	0.155	0.249
FAVAR (2 lags, 6 facs)	0.024	0.032	0.046	0.087	0.133	0.143	0.171	0.239

Note:
FF is the federal fund rate while 3m, 1y and 10y are, respectively, three month, 1 year and 10 year treasury bill rates
Information criteria are defined as minus log likelihood plus penalty function, hence should be minimized
MSFE and MAE are for 1-step ahead forecasts (for interest rates in levels) over the sample 1999:1-2003:12.

B - Alternative models for the KPSW example

	R-squared				Adjusted R-squared			
	C	PI	M	Ri	C	PI	M	Ri
ECM (1 lag, 2 coint.)	0.16	0.13	0.32	0.38	0.13	0.10	0.30	0.36
FECM (2 lags, 4 facs-lev, 4 coint.)	0.28	0.18	0.50	0.47	0.21	0.10	0.45	0.41
FAVAR (2 lags, 6 facs)	0.26	0.18	0.40	0.37	0.19	0.10	0.34	0.31
	AIC				**BIC**			
	C	PI	M	Ri	C	PI	M	Ri
ECM (1 lag, 2 coint.)	-7.78	-7.54	-8.65	04:33	-7.68	-7.43	-8.55	4.44
FECM (2 lags, 4 facs-lev, 4 coint.)	-7.81	-7.49	-8.85	4.31	-7.49	-7.17	-8.53	4.62
FAVAR (2 lags, 6 facs)	-7.79	-7.47	-8.66	4.47	-7.47	-7.15	-8.33	4.79
	MSFE				**MAE**			
	C	PI	M	Ri	C	PI	M	Ri
ECM (1 lag, 2 coint.)	0.180	0.338	0.246	27.010	0.332	0.506	0.324	3.985
FECM (2 lags, 4 facs-lev, 4 coint.)	0.309	0.124	0.216	34.906	0.427	0.279	0.322	4.464
FAVAR (2 lags, 6 facs)	0.243	0.141	0.224	9.363	0.376	0.295	0.316	2.369

Notes:
C is per capita real consumption, PI per capita real personal income, M real money, and Ri real interest rate
Information criteria are defined as minus log likelihood plus penalty function, hence should be minimized
MSFE and MAE are for 1-step ahead forecasts of growth in C, PI, M and change in Ri over the sample 1999:1-2003:12.
MSFEs for C, PI and M are multiplied by 10000, MAE by 100