ISSN 1745-8587

Birkbeck
UNIVERSITY OF LONDON

BCAM 2403

An analytical approach to estimating dynamic factor models with macroeconomic application

Tom Spencer
Birkbeck,University of London

November 2024

BCAM

Birkbeck Centre for Applied Macroeconomics

An analytical approach to estimating dynamic factor models with macroeconomic application

Tom Spencer*

November 24, 2024

Until recently literature on dynamic factor models invariably assumed that the cross sectional error covariance matrix is diagonal. This is due to the many parameters which need to be estimated for large data sets, which is known as the curse of dimensionality. Spencer (2024) introduced a novel approach which relaxes the diagonal assumption. This is achieved by concentrating out the parameters with the coefficients being based on maximum likelihood estimates. Importantly, the cross sectional covariance matrix is concentrated out so does not need to be explicitly estimated. The approach was implemented by numerical optimisation. Out of sample tests using Monte Carlo simulations showed the general covariance approach to perform well, with smaller prediction errors overall compared to a range of existing diagonal approaches and a diagonal equivalent of the general covariance approach. The present paper introduces an analytical implementation of the general covariance approach which avoids the need for numerical optimisation. The factors are instead estimated by solving the first order conditions with respect to the factors. The approaches are demonstrated on a large macroeconomic data set so the results and methodology should therefore be of interest to macro-economists and policy makers. The results are positive for the general covariance approaches, which indicates that the assumption of diagonal cross sectional covariance matrix is likely invalid for the large macroeconomic data set.

1 Introduction

Factor models have grown in popularity along with interest in modelling big data. This has been made possible with ever larger datasets and computing power. Factor models aim to pick out unobserved or latent variables in the data, for example the business cycle in macroeconomics or the Big Five personality traits in psychology. The estimated factors should explain the observable variables well.

Traditional factor models do not model the factors as being dynamic. In many cases, data are not time dependent so this makes sense. Principal components (PC) is the most common approach to this. The PC approach is a straightforward linear transformation, which doesn't account for any dynamics in the factors. However for many economic applications the data is time series, in which case it is usually more appropriate to model the factors as being dynamic. This is where dynamic factor models (DFM) are useful. The most common way to model the factors as being dynamic is to model them as being autoregressive.

Most recent approaches in the literature assume a diagonal cross sectional covariance matrix, except for a relatively small number of approaches which assume a sparse non-diagonal matrix, e.g. via shrinkage.

*Birkbeck, University of London (tspenc02@mail.bbk.ac.uk)

None of these are able to model a full general cross sectional covariance matrix for big data. The diagonal assumption is adopted because in these approaches the cross sectional covariance matrix is explicitly modelled by parameters which need to be estimated. It currently seems the most popular sophisticated approach to estimating large scale (i.e. large N) dynamic factor models is by quasi-maximum likelihood (QML) estimation such as Doz et al. (2012).

Spencer (2024) (herein referred to as S24) introduced a way to improve on QML by modelling the full general cross sectional covariance matrix. This is achieved by concentrating out the model coefficients using maximum likelihood estimates. The cross sectional covariance matrix is concentrated out so it does not require explicit estimation. The approach was shown to produce good forecasts in an out of sample experiment using Monte Carlo simulations.

The present paper applies the S24 approach to a standard large macroeconomic data set. The results and surrounding discussion should therefore be of interest to macroeconomists and policy makers. The individual series are grouped according to typology. The results reveal a significant structure in the cross sectional error covariance matrix. This presents mainly as diagonal blocks of high correlations for within-group correlations which is to be expected. There are also significant inter-group correlations, for example, as we would expect, the labour market group errors are correlated with the output group errors. The present paper also introduces an analytical version of the S24 approach which avoids the need for numerical optimisation. This is obtained by solving the first order conditions with respect to the factors. The analytical approach should be faster than numerical optimisation in most practical situations but should be particularly useful for extremely large data sets.

The remainder of this paper is as follows. Sections 2 and 3 explain the notation and background on DFMs as well as the new analytical approach. Section 4 discusses the test set up using a standard macroeconomic data set and Section 5 provides the results and discussion. Section 6 gives concluding remarks.

2 Background

2.1 Overview of DFMs

This section reviews the existing literature on DFMs, with particular regard to the diagonal cross sectional error covariance assumption. For simplicity, it is assumed that the factor dynamics are of autoregressive order 1 and that there are no lags of the factors in the cross sectional equation. The data generating process is assumed to be such that the cross section of N observed individuals x_t load onto the K unobserved factors f_t as follows:

$$x_t = \Lambda f_t + \eta_t, \tag{1}$$

for $t = 1, ..., T$ where x_t and η_t are $N \times 1$, Λ is $N \times K$ and f_t is $K \times 1$. η_t is assumed to be i.i.d., normally distributed $\eta_t \sim N(0, \Sigma)$ and independent of f_t. x_t is usually standardised. Estimation of Λ is feasible for large N from a practical point of view since only a few factors are used in the equation for each individual time series. An identifying normalisation is required in order to identify Eq. (3), since rotating the factors should have no impact on the resulting common component $F\Lambda'$.

The factors are assumed to be dynamic and follow a vector autoregression (VAR):

$$f_t = \Phi f_{t-1} + \varepsilon_t, \tag{2}$$

where Φ is $K \times K$ and such that $\{f_t\}_{t=1,\dots,T}$ is stable, i.e. that the roots r of $|I_K - r\Phi| = 0$ lie outside the unit circle. ε_t is $K \times 1$, i.i.d. normally distributed as $\varepsilon_t \sim N(0, \Sigma_f)$ and uncorrelated with f_{t-1}. Σ and Σ_f are assumed to be positive definite. η_t and ε_t are assumed to be independent as standard. Eqs. (1) and (2) are referred to in this paper as the measurement and state equation respectively to align with state space modelling terminology.

Transposing and stacking Eqs. (1) and (2) gives:

$$X = F\Lambda' + \eta \tag{3}$$

and:

$$F = F^{-}\Phi' + \varepsilon, \tag{4}$$

where $X = (x_1, x_2 \dots x_T)'$and $(\eta_1, \eta_2 \dots \eta_T)'$ are $T \times N$, $\varepsilon = (\varepsilon_1, \varepsilon_2 \dots \varepsilon_T)'$, $F = (f_1, f_2 \dots f_T)'$ and F^{-} is the lagged F, i.e. $F^{-} = (f_0, f_1 \dots f_{T-1})'$. ε, F and F^{-} are all $T \times K$. An identifying normalisation is required in order to identify Eq. (3), since rotating the factors should have no impact on the resulting common component $F\Lambda'$. Identifying normalisations are common in factor estimation in general and there are many available (see e.g Bai and Li (2012) & Bai and Li (2016)).

2.2 Estimation

Prior to S24, it seems the most accurate way of estimating large scale (i.e. large N) dynamic factor models was by quasi-maximum likelihood (QML) estimation such as Doz et al. (2012). The precursor to the QML approach is the two step approach of Doz et al. (2011). The two step approach uses initial estimates of the factors obtained by principal components (PCs). During the first step the parameters (including covariances) of Eqs. (1) and (2) are estimated by OLS, and then the second step uses the Kalman filter & smoother to obtain better estimates of the factors given the parameters. The QML approach extends this by iterating between calculating the factors given the parameters, then calculating the parameters given the factors, in an Expectation Maximisation (EM) algorithm. The EM algorithm in this context obtains better and better estimates of factors given parameters (E step) and parameters given factors (M step) until convergence.

The QML approach assumes that the covariance Σ of the idiosyncratic term η_t is diagonal (see assumption R2 of Doz et al. (2012)). Nonetheless, the QML approach is consistent for large N and T along any path of N and T tending to infinity. The rate of consistency for estimating the factors is $min(\sqrt{T}, \frac{T}{\log(N)})$(see Doz et al. (2012)). PCs are also consistent however the QML approach can provide efficiency improvements over PCs in finite samples.

2.3 Current reasoning on the diagonal assumption

For small N DFMs, maximum likelihood estimation using a general cross sectional covariance matrix is not an issue (see e.g. Diebold et al. (2021)). DFMs using a general cross sectional covariance matrix however are thought to suffer from the curse of dimensionality, i.e. run into problems for large N. In summary, none of the methods in the literature before the S24 approach are feasible for large N and a general Σ matrix.

There is a wealth of literature that mentions that the diagonal assumption is adopted due to the very high number of parameters which need to be estimated, for example Poncela et al. (2021). However, in depth technical reasons for this are not easy to find. Overall, the literature indicates that the curse of dimensionality in DFMs is because firstly the QML type approaches need to invert $N \times N$ covariance matrices in the Kalman

filter, and secondly that estimating the covariance matrices adds to the number parameters making estimation unfeasible. Some examples where this is discussed in the literature are given in S24 and are not repeated here.

The S24 general covariance approach relaxes the assumption of diagonal Σ. The cross sectional covariance matrix is concentrated out so does not need to be explicitly estimated, thus avoiding the curse of dimensionality. The approach is implemented by numerical optimisation. Having general Σ is desirable intuitively as some of the correlation in x_t may be attributed to the dynamic factors and some may be attributed to the measurement equation error term, which is not dynamic. The non-diagonal Σ aims at full maximum likelihood since fewer assumptions are required compared to existing approaches such as the QML approach. Non-diagonal Σ should lead to more optimal estimates of the factors, having a higher log-likelihood and leading to better forecasts in finite samples. The approach is explained further in the next section, since there is a lot of overlap between it and the analytical approach introduced in the present paper.

3 New general and diagonal covariance approaches

3.1 Likelihood derivation for the general approach

The S24 general covariance approach relaxes the assumption of diagonal Σ. The model assumptions are the same as Section 2 except that the approach here relaxes the assumption that Σ is diagonal.

The log-likelihood for the full sample can be calculated as follows. The likelihood is based on the innovations in x_t (i.e. $x_t - E(x_t|x_{t-1}, x_{t-2}, ..., x_1)$) which is common in the literature, see e.g. Watson and Engle (1983) and Doz et al. (2012). Substituting Eq. (2) into Eq. (1) gives:

$$x_t = \Lambda(\Phi f_{t-1} + \varepsilon_t) + \eta_t = \Lambda\Phi f_{t-1} + \Lambda\varepsilon_t + \eta_t, \tag{5}$$

so can be written:

$$x_t = \Lambda\Phi f_{t-1} + w_t, \tag{6}$$

where $w_t = \Lambda\varepsilon_t + \eta_t \sim N(0, \Omega)$ is $N \times 1$ and Ω is $N \times N$. Given initial estimates of the parameters, the covariance matrices can be estimated as

$$\hat{\Sigma} = \frac{1}{T}\hat{\eta}'\hat{\eta} = \frac{1}{T}(X - \hat{F}\hat{\Lambda}')'(X - \hat{F}\hat{\Lambda}') \tag{7}$$

and

$$\hat{\Sigma_f} = \frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{T}(\hat{F} - \hat{F}^{-}\hat{\Phi})'(\hat{F} - \hat{F}^{-}\hat{\Phi}). \tag{8}$$

Given the parameter and factor estimates, $\hat{w}_t$ is estimated as $\hat{w}_t = \hat{\Lambda}\hat{\varepsilon}_t + \hat{\eta}_t$ where $\hat{\eta}_t = x_t - \hat{\Lambda}\hat{f}_t$ and $\hat{\varepsilon}_t = \hat{f}_t - \hat{\Phi}\hat{f}_{t-1}$. The log-likelihood can then be written:

$$L(\hat{F}, \hat{\Lambda}, \hat{\Phi}) = \frac{1}{2}(T\log|\hat{\Omega}^{-1}| - TN\log(2\pi) - \sum_{t=1}^{T}\hat{w}_t'\hat{\Omega}^{-1}\hat{w}_t), \tag{9}$$

where $\hat{\Omega}$ is the the estimated $N \times N$ covariance matrix of w_t, i.e. $\hat{\Omega} = \frac{1}{T}\hat{w}'\hat{w}$ where $\hat{w} = \hat{\varepsilon}\hat{\Lambda}' + \hat{\eta}$ is $T \times N$. $\hat{\Omega}$ is assumed to be positive definite, hence the present paper only the covers the case where $T > N$.

It is assumed that ε and η are independent, i.e. orthogonal to each other. They could in principle be

allowed to be correlated but the benefit of doing so is unclear and existing literature invariably assumes they are uncorrelated as standard.

Under the assumption that ε and η are orthogonal to each other, Ω is estimated by

$$\hat{\Omega} = \frac{1}{T}\hat{w}'\hat{w} = \hat{\Sigma} + \hat{\Lambda}\hat{\Sigma}_f\hat{\Lambda}' + \hat{\Lambda}\hat{\Sigma}_{\varepsilon\eta} + (\hat{\Lambda}\hat{\Sigma}_{\varepsilon\eta})' = \hat{\Sigma} + \hat{\Lambda}\hat{\Sigma}_f\hat{\Lambda}', \tag{10}$$

where $\hat{\Sigma}_{\varepsilon\eta} = \frac{1}{T}\hat{\varepsilon}'\hat{\eta}$ $(K \times N)$ is a matrix of the covariances between ε and η which is zero under the assumption that ε and η are orthogonal to each other. The last term of Eq. (9) can be simplified (see Hamilton (1994) or S24):

$$\sum_{t=1}^{T} \hat{w}_t'\hat{\Omega}^{-1}\hat{w}_t = TN \tag{11}$$

This simplification is well known in the context of vector autoregression, however it is rare for dynamic factor models. Noting that $\log|\hat{\Omega}^{-1}| = -\log|\hat{\Omega}|$ the log-likelihood then simplifies to:

$$L(\hat{F}, \hat{\Lambda}, \hat{\Phi}) = \frac{1}{2}(-T\log|\hat{\Omega}| - TN\log(2\pi) - TN), \tag{12}$$

where $\hat{\Omega} = \hat{\Sigma} + \hat{\Lambda}\hat{\Sigma}_f\hat{\Lambda}'$. Note that the calculation of the log-likelihood in this way avoids the need to invert $\hat{\Omega}$.

3.2 Likelihood derivation for the diagonal approach

The parameters can also be concentrated out using a diagonal cross sectional covariance matrix to form a restricted version of the general covariance approach above. This restriction is the only difference between the new general covariance approach and the new diagonal covariance approach. In the same way that the general covariance case assumes zero correlation between η and ε, so too does the diagonal equivalent.

Recall from section 3.1 above, we had (see Eq. (12)) for the full log-likelihood:

$$L(\hat{F}, \hat{\Lambda}, \hat{\Phi}) = \frac{1}{2}(-T\log|\hat{\Omega}| - TN\log(2\pi) - TN), \tag{13}$$

where Ω was estimated by

$$\hat{\Omega} = \hat{\Sigma} + \hat{\Lambda}\hat{\Sigma}_f\hat{\Lambda}'. \tag{14}$$

In the same way as the cross-equation error correlations were set to zero to implement the independent η and ε assumption in Eq. (10) for the general approach, the off diagonal elements of Ω can be set to zero to implement the diagonal assumption, i.e. Ω is estimated by

$$\hat{\Omega} = \hat{D} + \hat{\Lambda}\hat{\Sigma}_f\hat{\Lambda}', \tag{15}$$

where D is a diagonal matrix containing the diagonal elements of $\hat{\Sigma}$. The log-likelihood simplification (see Eq. (11)) works out in a similar manner under this assumption.

3.3 Concentrating out the parameters and solving for the factors

For both the new general covariance and diagonal covariance approaches, the coefficients are estimated by maximum likelihood. This enables concentrating out the parameters from the log-likelihood so that it only depends on the factors.

The formulae for the coefficients for the general covariance approach are:

$$\hat{\Lambda} = X'\hat{F}(\hat{\varepsilon}'\hat{\varepsilon} + \hat{F}'\hat{F})^{-1} \tag{16}$$

and:

$$\hat{\Phi} = \hat{F}'\hat{F}^{-}(\hat{F}^{-\prime}\hat{F}^{-})^{-1} \tag{17}$$

So the formula for $\hat{\Phi}$ is the same as the OLS formula, but $\hat{\Lambda}$ includes an additional $\hat{\varepsilon}'\hat{\varepsilon}$ term which is an interaction term coming from the measurement and state equation being linked through the log-likelihood. The diagonal model also uses $\hat{\Phi} = \hat{F}'\hat{F}^{-}(\hat{F}^{-\prime}\hat{F}^{-})^{-1}$ but is more complicated for $\hat{\Lambda}$ which is solved for iteratively. See S24 for derivations and further details. Based on the estimated coefficients, the covariance matrices are also concentrated out using Eqs. (7) & (8)

Concentrating out is well known in optimisation problems in general and involves rewriting one or more of the parameters to be optimised as a function of a different variable. In the case here, the coefficient matrices are concentrated out as functions of the factors, by maximum likelihood as explained above. The covariance matrices are also concentrated out through the resulting residuals, so Ω itself is also concentrated out.

Mathematically,

$$L(\Omega(F, \Lambda, \Phi)) = L(\Omega(F, \Lambda(F), \Phi(F))) = l(\Omega(F)) = l(F), \tag{18}$$

where $l(F)$ is now the concentrated log-likelihood. The task at hand is to solve:

$$\hat{F}^{updated} = \arg\max_{F}(l(F)), \tag{19}$$

where $l(F)$ is as per Eq. (12) and $\hat{F}^{updated}$ is the new value of F given the initial (e.g. PCs) factors. The parameters can be recovered once the new factors have been estimated, using the maximum likelihood coefficients explained above. The S24 diagonal and general approaches adopt the common normalisation $F'F/T = I_k$, likewise with the analytic approach introduced in the present paper.

The S24 general covariance approach uses numerical optimisation to estimate the factors according to Eq. (19), for both the S24 general covariance and S24 diagonal approaches. The present paper introduces an analytical equivalent of the general covariance approach, as opposed to numerical optimisation. This is described in the next section.

3.4 Solving the first order conditions of the likelihood with respect to the factors

Introduced in the present paper is a new approach, based on the same assumptions as the S24 general cross sectional covariance approach outlined above, but instead estimates the factors (i.e. finding $\arg\max_F(l(F))$) analytically. This is achieved by solving the first order conditions of the likelihood with respect to the factors. Hence the approach introduced in the present paper is an analytical version of the S24 general covariance approach. The remainder of this section provides the derivation of the analytical approach. The hat accents denoting the initial estimates are omitted for notational convenience.

3.4.1 Derivative of the log-likelihood with respect to the factors

In calculating the derivative of the log-likelihood with respect to the factors, a useful result is the Jacobi formula that the derivative of the determinant is the adjoint. Mathematically, $\frac{\partial |M(q)|}{\partial q} = trace(adj(M(q))\frac{\partial M(q)}{\partial q})$ where M is a square matrix and q is a scalar (see e.g. Magnus and Neudecker (1999)). We have:

$$\frac{\partial |\Omega|}{\partial F_{tj}} = trace(\frac{\partial |\Omega|}{\partial \Omega}\frac{\partial \Omega}{\partial F_{tj}}) = \frac{1}{T}trace(C\frac{\partial (g'g + \eta'\eta)}{\partial F_{tj}}), \tag{20}$$

where C is the $N \times N$ cofactor matrix of Ω. The derivative on the right is, using the product rule:

$$\frac{\partial (g'g + \eta'\eta)}{\partial F_{tj}} = \frac{\partial g'}{\partial F_{tj}}g + g'\frac{\partial g}{\partial F_{tj}} + \frac{\partial \eta'}{\partial F_{tj}}\eta + \eta'\frac{\partial \eta}{\partial F_{tj}}, \tag{21}$$

so we need to calculate $\frac{\partial g}{\partial F_{tj}} = \frac{\partial}{\partial F_{tj}}(\varepsilon\Lambda')$ and $\frac{\partial \eta}{\partial F_{tj}}$.

For $\frac{\partial g}{\partial F_{tj}}$,

$$\frac{\partial g}{\partial F_{tj}} = \frac{\partial}{\partial F_{tj}}(\varepsilon\Lambda') = \frac{\partial \varepsilon}{\partial F_{tj}}\Lambda' + \varepsilon\frac{\partial \Lambda'}{\partial F_{tj}}, \tag{22}$$

so we also need to calculate $\frac{\partial \varepsilon}{\partial F_{tj}}$ and $\frac{\partial \Lambda}{\partial F_{tj}}$ where $\Lambda = X'F(\varepsilon'\varepsilon + F'F)^{-1}$ by the maximum likelihood formula in Section 3.3.

For $\frac{\partial \varepsilon}{\partial F_{tj}}$:

$$\frac{\partial \varepsilon}{\partial F_{tj}} = \frac{\partial}{\partial F_{tj}}(F - F^-\Phi') \tag{23}$$

$$= J_{tj} - \frac{\partial F^-}{\partial F_{tj}}\Phi' - F^-\frac{\partial \Phi'}{\partial F_{tj}} \tag{24}$$

$$= J_{tj} - J_{t+1,j}\Phi' - F^-\frac{\partial \Phi'}{\partial F_{tj}}, \tag{25}$$

where $\frac{\partial F}{\partial F_{tj}} = J_{tj}$ is a matrix which has zeros everywhere except a 1 in the t, j position and $\frac{\partial F^-}{\partial F_{tj}} = J_{t+1,j}$ is a matrix which has zeros everywhere except a 1 in the $t+1, j$ position. Also:

$$\Phi' = (F^{-\prime}F^-)^{-1}F^-F \tag{26}$$

$$(F^{-\prime}F^-)\Phi' = F^-F \tag{27}$$

$$\frac{\partial (F^{-\prime}F^-)}{\partial F_{tj}}\Phi' + (F^{-\prime}F^-)\frac{\partial \Phi'}{\partial F_{tj}} = \frac{\partial (F^{-\prime}F)}{\partial F_{tj}} \tag{28}$$

$$(F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi' + (F^{-\prime}F^-)\frac{\partial \Phi'}{\partial F_{tj}} = F^{-\prime}J_{tj} + J'_{t+1,j}F. \tag{29}$$

Hence:

$$\frac{\partial \Phi'}{\partial F_{tj}} = (F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - (F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi'), \tag{30}$$

therefore:

$$\frac{\partial \varepsilon}{\partial F_{tj}} = J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - (F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi'). \tag{31}$$

For $\frac{\partial \Lambda}{\partial F_{tj}}$ we have:

$$\Lambda' = (F'F + \varepsilon'\varepsilon)^{-1}F'X, \tag{32}$$

so:

$$(F'F + \varepsilon'\varepsilon)\Lambda' = F'X. \tag{33}$$

Differentiating:

$$\frac{\partial}{\partial F_{tj}}(F'F + \varepsilon'\varepsilon)\Lambda' + (F'F + \varepsilon'\varepsilon)\frac{\partial \Lambda'}{\partial F_{tj}} = \frac{\partial}{\partial F_{tj}}(F'X). \tag{34}$$

Now:

$$\frac{\partial}{\partial F_{tj}}(F'F + \varepsilon'\varepsilon) = F'\frac{\partial F}{\partial F_{tj}} + \frac{\partial F'}{\partial F_{tj}}F + \varepsilon'\frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon \tag{35}$$

$$= F'J_{tj} + J'_{tj}F + \varepsilon'\frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon. \tag{36}$$

Also needed is $\frac{\partial}{\partial F_{tj}}(F'X)$. This follows similarly:

$$\frac{\partial}{\partial F_{tj}}(F'X) = F'\frac{\partial X}{\partial F_{tj}} + \frac{\partial F'}{\partial F_{tj}}X = 0 + J'_{tj}X. \tag{37}$$

Substituting these back into Eq. (34), we have:

$$(F'J_{tj} + J'_{tj}F + \varepsilon'\frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon)\Lambda' + (F'F + \varepsilon'\varepsilon)\frac{\partial \Lambda'}{\partial F_{tj}} = J'_{tj}X. \tag{38}$$

Hence:

$$\frac{\partial \Lambda'}{\partial F_{tj}} = (F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F + \varepsilon'\frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon)\Lambda'). \tag{39}$$

$\frac{\partial \varepsilon}{\partial F_{tj}}$ has already been calculated above but is left as is for now so the notation is not too cumbersome.

Lastly for $\frac{\partial \eta}{\partial F_{tj}}$:

$$\frac{\partial \eta}{\partial F_{tj}} = \frac{\partial}{\partial F_{tj}}(X - F\Lambda') \tag{40}$$

$$= -\frac{\partial F}{\partial F_{tj}}\Lambda' - F\frac{\partial \Lambda'}{\partial F_{tj}} \tag{41}$$

$$= -J_{tj}\Lambda' - F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F + \varepsilon'\frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon)\Lambda'). \tag{42}$$

All the relevant parts have now been calculated but need to be put together. We have:

$$\frac{\partial |\Omega|}{\partial F_{tj}} = trace(C(\frac{\partial g}{\partial F_{tj}}g' + g'\frac{\partial g}{\partial F_{tj}} + \frac{\partial \eta}{\partial F_{tj}}\eta' + \eta'\frac{\partial \eta}{\partial F_{tj}})) \tag{43}$$

$$= 2trace(C(g'\frac{\partial g}{\partial F_{tj}} + \eta'\frac{\partial \eta}{\partial F_{tj}}))) \tag{44}$$

$$= 2trace(CH), \tag{45}$$

where:

$$H = g' \frac{\partial g}{\partial F_{tj}} + \eta' \frac{\partial \eta}{\partial F_{tj}} \tag{46}$$

$$= \Lambda\varepsilon' \frac{\partial(\varepsilon\Lambda')}{\partial F_{tj}} + \eta' \frac{\partial \eta}{\partial F_{tj}} \tag{47}$$

$$= \Lambda\varepsilon' (\frac{\partial \varepsilon}{\partial F_{tj}} \Lambda' + \varepsilon \frac{\partial \Lambda'}{\partial F_{tj}}) + \eta' \frac{\partial \eta}{\partial F_{tj}} \tag{48}$$

$$= \Lambda\varepsilon'(J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - (F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi'))\Lambda' \tag{49}$$

$$+\Lambda\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F\Lambda' + \varepsilon' \frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon)\Lambda') \tag{50}$$

$$-\eta'(J_{tj}\Lambda' - F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F + \varepsilon' \frac{\partial \varepsilon}{\partial F_{tj}} + \frac{\partial \varepsilon'}{\partial F_{tj}}\varepsilon)\Lambda')). \tag{51}$$

The first line of this long expression is $\Lambda\varepsilon' \frac{\partial \varepsilon}{\partial F_{tj}} \Lambda'$, the second line is $\Lambda\varepsilon'\varepsilon \frac{\partial \Lambda'}{\partial F_{tj}}$ and the third line is $\eta' \frac{\partial \eta}{\partial F_{tj}}$.

We can simplify this further using notation that $h(q) = q + q'$ for lines (50) and (51):

$$H = \Lambda\varepsilon'(J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - h(F^{-\prime}J_{t+1,j})\Phi'))\Lambda' \tag{52}$$

$$+\Lambda\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} + \varepsilon' \frac{\partial \varepsilon}{\partial F_{tj}})\Lambda') \tag{53}$$

$$-\eta'(J_{tj}\Lambda' + F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} + \varepsilon' \frac{\partial \varepsilon}{\partial F_{tj}})\Lambda')). \tag{54}$$

Finally, substituting for $\frac{\partial \varepsilon}{\partial F_{tj}}$ we get the full expression:

$$H = \Lambda\varepsilon'(J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - h(F^{-\prime}J_{t+1,j})\Phi'))\Lambda' \tag{55}$$

$$+\Lambda\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} \tag{56}$$

$$+\varepsilon'(J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - (F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi')))\Lambda') \tag{57}$$

$$-\eta'(J_{tj}\Lambda' + F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} \tag{58}$$

$$+\varepsilon'(J_{tj} - J_{t+1,j}\Phi' - F^-(F^{-\prime}F^-)^{-1}(F^{-\prime}J_{tj} + J'_{t+1,j}F - (F^{-\prime}J_{t+1,j} + J'_{t+1,j}F^-)\Phi')))\Lambda')). \tag{59}$$

$\frac{\partial|\Omega|}{\partial F_{0j}}$ and $\frac{\partial|\Omega|}{\partial F_{Tj}}$ are special cases; the above still applies but $\frac{\partial F}{\partial F_{0j}}$ and $\frac{\partial F^-}{\partial F_{Tj}}$ are matrices of zeros instead of the selection matrices such as J_{tj}.

It turns out that many of the terms are zero because the estimator of Φ coincides with the OLS estimator, hence $F^{-\prime}\varepsilon = 0$ by orthogonality, so the trace of the terms containing both F^- and ε are zero. So we get:

$$\tilde{H} = \Lambda\varepsilon'(J_{tj} - J_{t+1,j}\Phi')\Lambda' \tag{60}$$

$$+\Lambda\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} + \varepsilon'(J_{tj} - J_{t+1,j}\Phi'))\Lambda') \tag{61}$$

$$-\eta'(J_{tj}\Lambda' + F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - h(F'J_{tj} + \varepsilon'(J_{tj} - J_{t+1,j}\Phi'))\Lambda')). \tag{62}$$

$$= \Lambda\varepsilon'(J_{tj} - J_{t+1,j}\Phi')\Lambda' \tag{63}$$

$$+\Lambda\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F + \varepsilon'(J_{tj} - J_{t+1,j}\Phi') + (J'_{tj} - \Phi J'_{t+1})\varepsilon)\Lambda') \tag{64}$$

$$-\eta'(J_{tj}\Lambda' + F(F'F + \varepsilon'\varepsilon)^{-1}(J'_{tj}X - (F'J_{tj} + J'_{tj}F + \varepsilon'(J_{tj} - J_{t+1,j}\Phi') + (J'_{tj} - \Phi J'_{t+1})\varepsilon))\Lambda')), \tag{65}$$

where $h(q) = q + q'$ has been expanded for the second equality sign.

Looking at this as three terms, one on each line with the first having 2 parts, the second having 7 parts and the third having 8 parts, it turns out that the 2nd to 8th parts of the third term cancel in turn with the 1st to 7th parts (i.e. all) of the second term. This is because $\eta' F = \Lambda \varepsilon' \varepsilon$ (this is shown in Appendix B). For example the first part of the second term, $\Lambda \varepsilon' \varepsilon M J'_{tj} X$, equals the second part of the third term, $\eta' F M J'_{tj} X$. This leaves:

$$\tilde{H} = \Lambda \varepsilon' (J_{tj} - J_{t+1,j} \Phi') \Lambda' - \eta' J_{tj} \Lambda'. \tag{66}$$

This is very convenient as it removes the $(F^{-\prime} F^{-})^{-1}$ and $(F'F + \varepsilon'\varepsilon)^{-1}$ terms which would be difficult to factorise for F_{tj}.

Next, it is necessary to substitute $\varepsilon = F - F^{-\prime}\phi$ and $\eta = X - F\Lambda'$ to obtain a useful solution to the original problem:

$$\tilde{H} = \Lambda (F - F^{-\prime}\phi)' (J_{tj} - J_{t+1,j} \Phi') \Lambda' - (X - F\Lambda')' J_{tj} \Lambda' \tag{67}$$

$$= \Lambda F' J_{tj} \Lambda' - \Lambda \Phi F^{-\prime} J_{tj} \Lambda' + \Lambda F' J_{t+1,j} \Phi' \Lambda' + \Lambda \Phi F^{-^{'}} J_{t+1,j} \Phi' \Lambda' - X' J_{tj} \Lambda' + \Lambda F' J_{tj} \Lambda' \tag{68}$$

$$= 2\Lambda F' J_{tj} \Lambda' - \Lambda \Phi F^{-\prime} J_{tj} \Lambda' + \Lambda F' J_{t+1,j} \Phi' \Lambda' + \Lambda \Phi F' J_{t_j} \Phi' \Lambda' - X' J_{tj} \Lambda', \tag{69}$$

where the last line is obtained by noting that the first term in Eq. (68) equals the last term and that $F^{-^{'}} J_{t+1,j} = F' J_{tj}$.

3.4.2 Solving for the factors

We need to factorise this in order to analytically solve for the value of F_{tj} which sets $trace(C\tilde{H})$ to zero. It is slightly more convenient to factorise $trace(C\tilde{H}')$. The fifth term in Eq. (69) clearly doesn't contain F_{tj} so can be treated like a constant. The third term in Eq. (69) doesn't depend on F_{tj}. This can be seen as follows. The transpose of the third term is:

$$\Lambda \Phi J'_{t+1,j} F \Lambda'. \tag{70}$$

Looking at the $\Phi J'_{t+1,j} F$ part of this, we have, where the superscript indicates taking the column of the matrix:

$$\Phi J'_{t+1,j} F = \Phi^j F_{t+1} \tag{71}$$

where Φ^j is $K \times 1$ and $\Phi^j F_{t+1}$ is the outer product of the j^{th} column of Φ and the $t+1^{th}$ row of F. Note that F_t is $1 \times K$ and equals f'_t from before. The second term similarly doesn't depend on F_{tj}, since, after taking the transpose:

$$\Lambda J'_{tj} F^{-} = \Lambda^j F_{t-1} \tag{72}$$

since the t^{th} row of F^{-} is F_{t-1}.

So the only terms we need to factorise for F_{tj} are the first and fourth terms of Eq. (69). The transpose of the first term, after pre-multiplication by C is:

$$C\Lambda J'_{tj} F \Lambda' = C\Lambda^j F_t \Lambda'. \tag{73}$$

The s, i element of this is, where $s = 1, ..., N$ and $i = 1, ..., N$:

$$(C\Lambda^j F_t \Lambda')_{si} = C_s \Lambda^j F_t (\Lambda')^i = C_s \Lambda^j F_t \Lambda_i \tag{74}$$

$$= \sum_{n=1}^{N} \sum_{m=1}^{K} C_{sn} \Lambda_{nj} F_{tm} \Lambda_{im} \tag{75}$$

where each of the terms in the double sum is a scalar. So the trace of the fourth term of Eq. (69) can be written as follows, since the trace is the sum of the diagonal (i.e. i, i) elements:

$$\sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{m=1}^{K} C_{in} \Lambda_{nj} F_{tm} \Lambda_{im}. \tag{76}$$

This depends on F_{tj} where $m = j$ so this can be split up as follows:

$$\sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{m=1}^{K} C_{in} \Lambda_{nj} F_{tm} \Lambda_{im} = \sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{m=1, m \neq j}^{K} C_{in} \Lambda_{nj} F_{tm} \Lambda_{im} + \sum_{i=1}^{N} \sum_{n=1}^{N} C_{in} \Lambda_{nj} F_{tj} \Lambda_{ij} \tag{77}$$

where $\sum_{m=1, m \neq j}^{K}$ indicates summation over $m = 1, \ldots, K$ but not including j. The triple sum on the right of the equality sign no longer depends on F_{tj} and F_{tj} can be taken outside of the double sum (i.e. it equals $F_{tj} \sum_{i=1}^{N} \sum_{n=1}^{N} C_{sn} \Lambda_{nj} \Lambda_{ij}$). So this term has now been factorised into a constant term and a F_{tj} term times a coefficient.

The fourth term in Eq. (68) can be factorised in a very similar way, this is shown here for completeness. The transpose of the fourth term, after pre-multiplication by C is:

$$C\Lambda\Phi J'_{tj} F \Phi' \Lambda' = C\Lambda\Phi^j F_t \Phi' \Lambda'. \tag{78}$$

The s, i element of this is, where $s = 1, ..., N$ and $i = 1, ..., N$:

$$(C\Lambda\Phi^j F_t \Phi' \Lambda')_{si} = (C\Lambda)_s \Phi^j F_t (\Phi' \Lambda')^i = (C\Lambda)_s \Phi^j F_t (\Lambda\Phi)_i \tag{79}$$

$$\sum_{n=1}^{K} \sum_{m=1}^{K} (C\Lambda)_{sn} \Phi_{nj} F_{tm} (\Lambda\Phi)_{im} \tag{80}$$

where $(C\Lambda)_s$ is the s^{th} row of $C\Lambda$ and $(\Lambda\Phi)_i$ is the i^{th} row of $\Lambda\Phi$. The trace of this can be written as follows:

$$\sum_{i=1}^{N} \sum_{n=1}^{K} \sum_{m=1}^{K} (C\Lambda)_{in} \Phi_{nj} F_{tm} (\Lambda\Phi)_{im}. \tag{81}$$

This depends on F_{tj} where $m = j$ so we can split this up as follows:

$$\sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{m=1}^{K} (C\Lambda)_{in} \Phi_{nj} F_{tm} (\Lambda\Phi)_{im} \tag{82}$$

$$= \sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{m=1, m \neq j}^{K} (C\Lambda)_{in} \Phi_{nj} F_{tm} (\Lambda\Phi)_{im} + F_{tj} \sum_{i=1}^{N} \sum_{n=1}^{N} (C\Lambda)_{in} \Phi_{nj} (\Lambda\Phi)_{ij}. \tag{83}$$

By summing the constant terms and the F_{tj} coefficients over the various terms in Eq. (69), a linear equation can be formed and solved to find the value of F_{tj} which sets the derivative to zero. To be more explicit, $\frac{\partial|\Omega|}{\partial F_{tj}}$ can be calculated as:

$$\frac{\partial|\Omega|}{\partial F_{tj}} = k_1 + k_2 F_{tj} \tag{84}$$

where k_1 is the sum of the constant terms and constant coefficients and k_2 is the sum of the F_{tj} coefficients. Solving for this to be equal to zero gives a root for F_{tj} equal to $-k_1/k_2$. This is performed sequentially for each element of F, with coefficients being re-estimated after each element of F until convergence.

4 Macroeconomic application - test set up

4.1 FRED MD data set

The data set used is the FRED MD macroeconomic data set developed by McCracken & Ng (see McCracken and Ng (2015)). This data set is intended to supersede the usual Stock & Watson / Bai & Ng datasets as the standard macroeconomic data set used by academics. Indeed the data set has been used extensively since its creation (see e.g. Banerjee et al. (2017)). The data is monthly, and is updated continually on a monthly basis. The data for the present paper spans January 1959-January 2024. A few series are missing data for 1959 so the data set used here starts January 1960 (this is in line with McCracken and Ng (2015)). Two data points are lost at the start due to data transformations. This gives a time dimension of $T = 767$. There are 127 series organised into 8 groups. Group 1 is output and income, group 2 is labour market, group 3 is consumption and housing, group 4 is orders and inventories, group 5 is money and credit, group 6 is interest rate and exchange rates, group 7 is price and group 8 is stock market. The data has been transformed in order to be stationary and missing values have been filled by an EM algorithm in order to make the panel balanced. All of this is performed using Matlab code available on McCracken & Ng's website.[1] The only additional transformation made for the purpose of this paper is that each series has been standardised to have mean zero and standard deviation of one.

4.2 Number of factors

Opinion seems divided on the correct number of factors, K. Bai and Ng (2007) is a well known approach to estimate the number of number of dynamic factors which is based on examining the rank of the state equation errors. This approach indicates four factors when applied to a similar monthly macroeconomic dataset similar to the dataset used here. The Onatski (2005) approach is based on the eigenvalues of the cross sectional error correlation matrix, so allows for substantial cross sectional error correlation. This suggests two factors for a similar monthly macroeconomic data set. The latter test is for the number of static factors however this is still indicative for the number of dynamic factors, which should be less than or equal to the number of static factors.

It seems likely this difference in opinion may be because Onatski allows for substantial cross sectional error correlation but the Bai and Ng approach does not. Correlation within the data can either be modelled as coming from the factors or the cross sectional error term. The Bai and Ng approach effectively ignores any correlation which may come from the cross sectional error. Hence, intuitively, the Bai and Ng approach

[1] The data and code can be downloaded here: https://research.stlouisfed.org/econ/mccracken/fred-databases/

should add more factors in order to model the overall correlation in the data, and therefore overestimate the number of factors when there is substantial cross sectional error correlation. The Onatski approach on the other hand allows for a non-trivial amount of cross sectional correlation so should attribute more of the correlation within the data to the cross sectional error term, in turn reducing the number of factors required to model the data well.

There is a wealth of literature which deals with selecting the optimal number of factors in dynamic factor models, two examples are discussed above. This literature shows that it is a very difficult problem. As such, determining the number of factors is not the main topic of the present paper so the tests are run for $K = 1, ..., 5$ to broadly cover the most realistic cases, although it should be borne in mind that seems that either 1 or 5 factors is unlikely.

4.3 Approaches tested

The following five approaches are compared:

- PCs plus VAR in the factors (labelled as "PCVAR" in the charts) - factors are estimated by PCs, then a VAR in the PC factors is formed using OLS.
- Two step approach (labelled as "TwoStep" in the charts) - the Doz et al. (2011) two step approach (see Section 2.2)
- QML (labelled as "QML" in the charts) - the Doz et al. (2012) Quasi-Maximum Likelihood approach (see Section 2.2)
- S24 diagonal correlation approach ("NewDiagonal" in the results and discussion) - this is the new approach which assumes diagonal cross sectional covariance and is optimised numerically (see Section 3.2).
- S24 general correlation approach ("NewGeneral" in the results and discussion) - the new approach which allows general cross sectional error covariance and is optimised numerically (see Section 3.1).
- New analytical general correlation approach ("NewAnalytical" in the results and discussion) - the new approach introduced in the present paper which allows general cross sectional error covariance and is estimated analytically (see Section 3.3 and Appendix B).

Although the two numerical approaches introduced in S24 are not entirely novel, they are still new in the sense that they are very recent and have not been applied to a real world data set until now, hence they are labelled as "New" in the results.

In summary, there are six approaches tested in this paper, three existing diagonal approaches, two recent numerical approaches (one diagonal and one general) and one new general covariance approach introduced here which solves for the factors analytically. Note that for the two step and QML approaches, the code was downloaded from the authors' website[2]. There are two new general covariance approaches included, one is numerical introduced in S24 and the other is analytical introduced in the present paper.

[2]See https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZKNTUA

4.4 Reported metrics

The models are compared by various metrics. Most important is the log-likelihood, as this is the objective function of the new general approaches (NewGeneral and NewAnalytical). The RMSPE (see below) is based on the prediction errors. This is not the objective function of any of the models tested, so is unbiased in the sense that it does not relate to any model in particular. It is simply a diagnostic tool which has obvious practical relevance. Also included for discussion purposes is a measure of the off diagonal correlations. Lastly, cannonical correlations measure the closeness of the factors to NewAnalytical factors, this is included in order to show that NewGeneral factors are close to NewAnalytical factors. Further details of these metrics are below.

Log-likelihood: The log-likelihood is a metric which describes how well regressors and parameters describe observed data. The log-likelihood metric is as described in the overview of the general covariance method in Section 3.1. The input to the metric is just the data plus estimated factors and coefficients; the covariance matrices are obtained from the residuals for all approaches[3]. This is the underlying objective function maximised by the new general approaches. Note that ideally the evaluation criteria of a model should coincide with the objective function of the model so the most natural way to assess performance of the new general approach is the log-likelihood. In the same way for example, it would be more natural to judge PCA in terms of the mean squared cross sectional errors since this is the PCA optimisation objective function. The log-likelihood is not adjusted for the numbers of parameters as is done for model selection criteria such as the Bayesian Information Criteria since model selection in factor models is a complex topic so is not focussed on in the present paper.

Root mean square prediction error (RMSPE): The in sample root mean square prediction error (RMSPE) is explained fully in S24. A brief overview is as follows. Given the model estimated factors and state and measurement equation coefficients, the estimate of the factors in the current period is used to forecast the factors in the next period, using the state equation. Then, an estimate of the data in the next period is formed using the measurement equation and the forecasted factors. The final RMSPE is formed as the average over individuals of the square root of the time average of the squared deviations between the forecasted and actual data.

Measurement equation average correlation: This is the average absolute value of off-diagonal correlations of the $N \times N$ measurement error correlation matrix of $\hat{\eta}_t$. The input to the metric is just the data plus estimated factors and coefficients; the covariance matrices are obtained from the residuals for all approaches. This is less of a model performance statistic, but more for information and discussion purposes. I sometimes abbreviate "average absolute value of off-diagonal correlations of the measurement error correlation matrix" e.g. by "measurement equation correlations" but it should be clear from the context what is meant.

Canonical correlations: This is a measure of how close a set of estimated factors are to another set of estimated factors. Canonical correlations have been used e.g. in Bai and Li (2012). Canonical correlations transforms two datasets (in this case two sets of factors) in order that the correlations between corresponding columns of the two transformed datasets is maximised, and such that the columns of each of the transformed data sets are uncorrelated. The reader is referred to Härdle et al. (2003) for a full explanation. The function in matlab used here is the `cannoncorr` function. A benefit of using canonical correlations is that by construction they are independent of transformations of the factors, such as is used in normalisations. Displayed in the results tables is a matrix showing the closeness of all the various sets of estimated factors are to the new

[3]Note that QML estimates covariances through the EM algorithm, not via the resulting residuals. The likelihood metric here ignores this, as the aim here is to have a simple metric which is comparable across the various approaches tested.

analytical approach. This is in order to show the similarity between the S24 numerical approach with general covariance and the new analytical approach. Since there are usually more than one factor, the values in the matrices are the average of the canonical correlations. For example if $K = 3$, there are three factors so each displayed element in the $K = 3$ column is the average of the 3 canonical correlations.

5 Macroeconomic application - results and discussion

5.1 Overview of results tables

There is one table for each of the four metrics. The columns are the number of factors, and the rows are the six approaches tested as explained in Section 4.4. Each column is colour coded between yellow and green depending on value. Where appropriate, i.e. for Log-Likelihood and RMSPE, green is better, and for the remaining two metrics green is higher. It is reiterated that NewGeneral is the S24 numerical general covariance approach and NewAnalytical is the analytic general covariance approach introduced in the present paper.

Figure 1: Log-Likelihood for each approach

	K=1	K=2	K=3	K=4	K=5
PCVAR	-60610.8	-60606.9	-60029.2	-59788.9	-59180.3
TwoStep	-60623.5	-60605.3	-59608.9	-59277.1	-58360.4
QML	-60657.9	-60678.6	-59136.9	-59050.7	-58015.8
NewDiagonal	-60622.8	-60618.3	-59573.8	-59260.1	-58270.7
NewGeneral	-60142.4	-59605.9	-58953.1	-58499.7	-57481.9
NewAnalytical	-60142.9	-59612.6	-58952.4	-58498.1	-57481.5

Figure 2: RMSPE for each approach

	K=1	K=2	K=3	K=4	K=5
PCVAR	0.981	0.975	0.947	0.938	0.904
TwoStep	0.984	0.980	0.947	0.937	0.900
QML	0.991	0.988	0.942	0.935	0.902
NewDiagonal	0.985	0.982	0.946	0.936	0.899
NewGeneral	0.946	0.947	0.947	0.939	0.900
NewAnalytical	0.946	0.948	0.947	0.939	0.900

Figure 3: Average cross sectional correlation for each approach

	K=1	K=2	K=3	K=4	K=5
PCVAR	0.096	0.102	0.093	0.093	0.087
TwoStep	0.093	0.099	0.087	0.088	0.081
QML	0.090	0.082	0.076	0.073	0.072
NewDiagonal	0.091	0.096	0.084	0.085	0.078
NewGeneral	0.102	0.106	0.097	0.090	0.081
NewAnalytical	0.102	0.106	0.097	0.090	0.081

Figure 4: Canonical Correlation with NewAnalytical for each approach

	K=1	K=2	K=3	K=4	K=5
PCVAR	0.8454	0.9034	0.9668	0.9803	0.9816
TwoStep	0.8097	0.8781	0.9584	0.9700	0.9773
QML	0.7073	0.7244	0.8621	0.8939	0.9409
NewDiagonal	0.8011	0.8610	0.9543	0.9669	0.9799
NewGeneral	0.9997	0.9991	1.0000	1.0000	1.0000
NewAnalytical	1.0000	1.0000	1.0000	1.0000	1.0000

5.2 Discussion of results

From a theoretical point of view, note that the focus here should be on comparing NewDiagonal against NewGeneral, since these are like for like, the only difference being that the latter allows for general cross sectional covariance. Also comparing NewGeneral and NewAnalytical shows the differences between the numerical implementation of the general covariance model and the analytical implementation. The existing three diagonal approaches (PCVAR, TwoStep and QML) are only included for comparing these new approaches to existing diagonal approaches.

The most important metric for the general covariance approaches (NewGeneral and NewAnalytical) is the log-likelihood. It is best, i.e. least negative, for the general covariance approaches, with NewAnalytical being very close NewGeneral, the differences being around 0.01% or less. Both analytical approaches have much higher log-likelihood than PCVAR, which is used as initial factors to all approaches. The diagonal approaches get closer to the general approaches as the number of factors increases, most notably for $K = 3$ factors and above. This could be indicative of the extra factors shrinking the measurement errors so that they become less important. For example, for PCVAR, the first K eigenvalues represent the common component and the last $N - K$ represent the measurement errors, so the measurement errors shrink as K increases towards N. Note that shrinking the measurement errors decreases their covariance but not necessarily their correlation (this can be seen in the heatmaps which are discussed below and in Appendix A). Nevertheless, the reduced importance of the measurement errors may mean the general approaches are less able to utilise the remaining measurement error correlations as the number of factors is increased. The log-likelihood for NewDiagonal is very close to TwoStep, which shows that NewDiagonal is between PCVAR and QML.

For all values of K the RMSPE (a measure of prediction errors) is notably lower, i.e. better, for the general covariance approaches compared to all diagonal approaches for one or two factors, this shows that the fit is considerably more accurate for the new general approach. However, they are around the same as the diagonal models for three, four or five factors. For 3 or 4 factors, QML is a tiny bit better than the general approaches. Similarly to the log-likelihood case, it seems this is because as more factors are added they shrink the importance of the measurement errors so the general approaches are less able to utilise the remaining measurement error correlations.

The average of the estimated measurement equation error correlations are generally highest for the general covariance approaches. They are always higher than NewDiagonal, which is the most relevant diagonal model to compare the general approaches to. This likely shows that the general approaches are able to utilise more of the correlation of the measurement equation errors, which results in higher log-likelihood. PCVAR average correlation moves closer to the analytical approaches as the number of factors increases and is higher than

NewAnalytical for five factors. Again, it seems likely this is because as more factors are added they shrink the measurement errors. PCVAR has the highest average correlations out of all four diagonal models, and QML the lowest. It seems plausible that this is because PCVAR is just a simple linear transformation of the original data which ignores the correlations, whereas NewDiagonal explicitly assumes they are zero during optimisation. In other words NewDiagonal is inclined to push the correlations down compared to the initial PC input factors, at least for this particular data set. The general approaches on the other hand are inclined to push up the correlations. The 95% significance threshold for the value of the correlation coefficient when $T = 767$ is around 0.07[4] and all of the average correlations in the figures for all approaches are above this level, and much higher for the general covariance approaches. This seems to indicate that the diagonal assumption for this data set is invalid and it should therefore be better to use the general covariance approach for inference and forecasting. Obviously it would be better to do a joint test for all the correlations but formal testing of whether the diagonal assumption is valid is not the main focus of this paper.

The main novelty in this paper is in the estimation of the cross sectional error correlation matrix on a large macroeconomic data set. This should be of particular interest to macroeconomists as they show a clear structure to the correlation matrix, which is reasonably close to block diagonal, corresponding to the variable groupings. This is shown in Appendix A which contains heat map charts of the correlation matrix for the FRED-MD dataset. The heat map charts are simply just a graphical representation of the cross sectional correlation matrix but each element is represented by a colour on a scale depending on its value. Aside from the within-group block diagonal structure, there are a number of very high inter-group correlations, which are ignored by the diagonal approaches. The heat map charts for $K = 5$ are naturally less striking compared to $K = 1$ but there remains a large degree of cross sectional error correlation, despite the extra factors shrinking the cross sectional errors. Note that the cross sectional error correlation matrix could be estimated for the diagonal approaches such as PCVAR, by calculating the correlation matrix of the resulting measurement equation errors. The diagonal approaches however would not estimate these correlations optimally as is the case with the general approaches.

The main purpose of the cannonical correlations is to show that the NewAnalytical factors are close to NewGeneral. The last row contains ones because the figures show the cannonical correlation between the various approaches and NewAnalytical. NewGeneral has very high cannonical correlation with NewAnalytical, very close to 1 in all cases, which means the estimated factors are very similar and differences are negligible. The closeness of all approaches increases with the number of factors. This makes sense as the measurement errors shrink as more factors are added, so all approaches would equal each other in the limit as K approaches N. Given the similarity between NewAnalytical and NewGeneral in terms of cannonical correlations, as well as the very close similarity of the log-likelihood, RMSPE and average cross sectional correlation metrics, it seems reasonable to conclude overall that NewAnalytical comes very close to producing near identical results as NewGeneral. The differences are negligible and likely down to small numerical differences between the analytical method and Matlab's black box optimiser.

6 Conclusion

Spencer (2024) recently introduced a way to model the full general cross sectional covariance matrix in dynamic factor models. This is achieved by concentrating out the model parameters using maximum likelihood

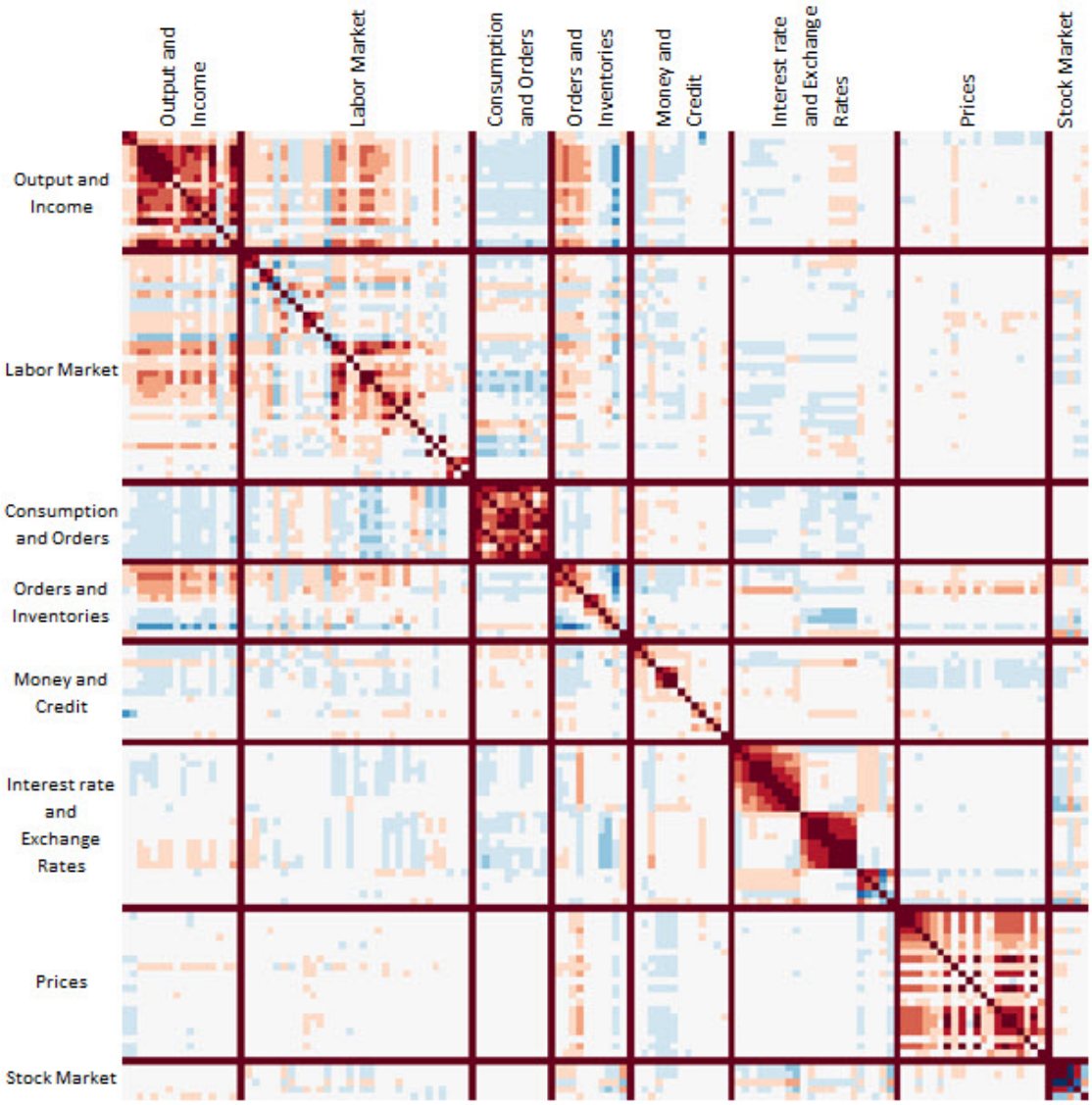
[4]This is using the correlation coefficient t statistic $\frac{r\sqrt{T-2}}{\sqrt{1+r^2}}$ where r is the correlation coefficient to be assessed for significance and T is the sample size, see e.g. Puth et al. (2014).

estimation. Importantly the cross sectional covariance matrix is concentrated out so does not require explicit estimation. The factors are obtained by numerical optimisation. The present paper applies the S24 general cross sectional covariance approach to a standard large macroeconomic data set. A new version of the S24 general covariance approach is also introduced, which estimates the factors analytically by solving the first order conditions with respect to the factors. Both the numerical and analytical general covariance approaches show promising results on the macroeconomic data set, in terms of log-likelihood and fitted errors compared to a variety of diagonal assumption approaches. The general covariance approaches estimate higher off-diagonal elements of the cross sectional error correlation matrix, resulting in better log-likelihoods and prediction errors. The correlation matrix shows a block diagonal structure according to the data groupings and there are a number of very high inter-group correlations, which are ignored by the diagonal approaches. This should be of particular interest to macroeconomists. Formal testing of the validity of the diagonal assumption is not the main focus of this paper, however the results seem to indicate that the diagonal assumption is not valid for the macroeconomic data set, and it should therefore be better to use a general covariance approach for inference and forecasting. The numerical and analytical general covariance approaches are nearly identical in terms of the four reported metrics so there is no significant difference between the two implementations.

7 Appendix A - Cross sectional correlation heat maps

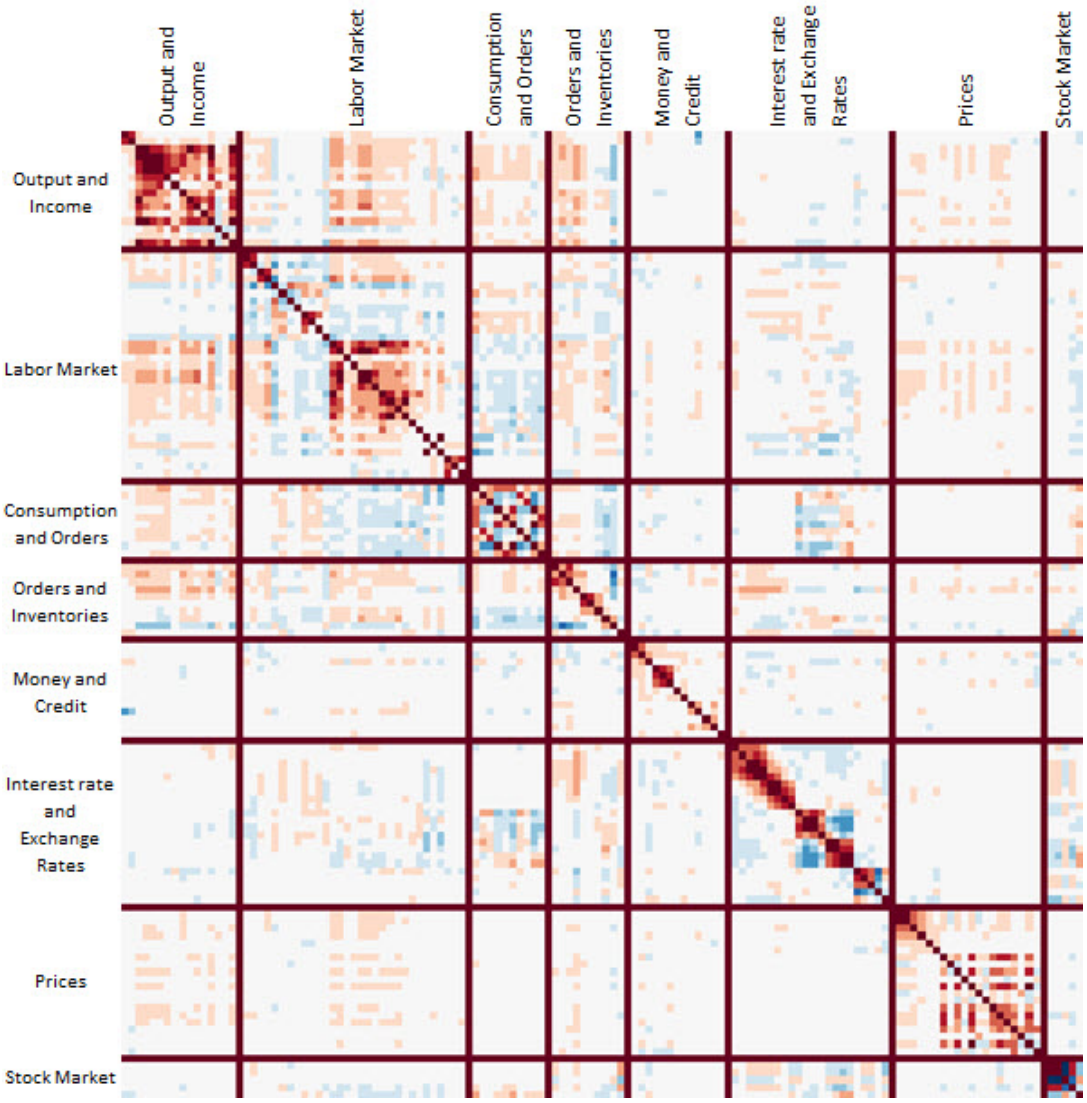
This appendix displays the heat map charts discussed in Section 5 for the full (i.e. $T = 767$) FRED-MD dataset. The heat map charts are simply just graphical representation of correlation matrices but each element is represented by a colour scale depending on its value. They are similar to surface plots but more of a two dimensional representation. The correlation matrices have values between -1 and 1. The closer the element of the correlation matrix is to 1, the darker red the colour of that element of the correlation matrix is represented and the closer to -1, the darker the blue. White is near zero correlation. The heatmaps are shown for the numerical general covariance model (NewGeneral) for $K = 1$ and $K = 5$ to show how the matrix changes as more factors are added, although the case $K = 1$ is almost certainly too low in reality. The factor loadings are omitted as they have been previously discussed in the literature (see e.g. McCracken and Ng (2015)).

Figure 5: Heat map of the correlation matrix of the measurement equation error for K=1



The above is a heat map chart of the 127×127 measurement error correlation matrix of the numerical general covariance approach for $K = 1$. This is the estimated correlation matrix of η but each element is represented by a colour scale depending on its value, between -1 and 1 as explained above. The average of the absolute values of the off-diagonal elements is around 0.10.

Figure 6: Heat map of the correlation matrix of the measurement equation error for K=5



The above heat map is the same as above but for $K = 5$. The average of the absolute values of the off-diagonal elements is around 0.08. Comparing this heat map with the $K = 1$ case (Figure 5) shows that although the extra factors shrink the measurement errors, the correlation remains quite high. Note that in both heat maps there is some block-within-block correlations, for example the first few series from the interest and exchange rates group are treasury rates, the next few are bond spreads and the remaining few are exchange rates.

8 Appendix B: Proof that $\eta' F = \Lambda \varepsilon' \varepsilon$.

Eq. (66) of Section 3.4 relies on the relation $\eta' F = \Lambda \varepsilon' \varepsilon$. This is shown as follows. Starting with the cross sectional equation we have:

$$X - \eta = F\Lambda' \tag{85}$$

$$F'X - F'\eta = F'F\Lambda'. \tag{86}$$

Substituting $\Lambda' = (F'F + \varepsilon'\varepsilon)^{-1}F'X$:

$$F'X - F'\eta = F'F(F'F + \varepsilon'\varepsilon)^{-1}F'X. \tag{87}$$

Adding and subtracting $\varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}F'X$:

$$F'X - F'\eta = (F'F + \varepsilon'\varepsilon)(F'F + \varepsilon'\varepsilon)^{-1}F'X - \varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}F'X \tag{88}$$

$$F'X - F'\eta = F'X - \varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}F'X, \tag{89}$$

so:

$$F'\eta = \varepsilon'\varepsilon(F'F + \varepsilon'\varepsilon)^{-1}F'X = \varepsilon'\varepsilon\Lambda'. \tag{90}$$

Taking the transpose of this shows finally that $\eta' F = \Lambda \varepsilon' \varepsilon$.

9 Acknowledgements

I am grateful for comments by Zacharias Psaradakis, Ron Smith, Walter Beckert and Stephen Wright which have substantially improved this paper.

References

Jushan Bai and Kunpeng Li. Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436 – 465, 2012. doi: 10.1214/11-AOS966. URL `https://doi.org/10.1214/11-AOS966`.

Jushan Bai and Kunpeng Li. Maximum likelihood estimation and inference for approximate factor models of high dimension. *The Review of Economics and Statistics*, 98(2):298–309, 2016. ISSN 00346535, 15309142. URL `http://www.jstor.org/stable/43830349`.

Jushan Bai and Serena Ng. Determining the number of primitive shocks in factor models. *Journal of Business Economic Statistics*, 25:52–60, 02 2007. doi: 10.1198/073500106000000413.

Anindya Banerjee, Massimiliano Marcellino, and Igor Masten. Structural fecm: Cointegration in large-scale structural favar models. *Journal of Applied Econometrics*, 32(6):pp. 1069–1086, 2017. ISSN 08837252, 10991255. URL `https://www.jstor.org/stable/26609783`.

Francis X. Diebold, Maximilian Göbel, Philippe Goulet Coulombe, Glenn D. Rudebusch, and Boyuan Zhang. Optimal combination of arctic sea ice extent measures: A dynamic factor modeling approach. *International Journal of Forecasting*, 37(4):1509–1519,

2021. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2020.10.006. URL https://www.sciencedirect.com/science/article/pii/S0169207020301606.

Catherine Doz, Domenico Giannone, and Lucrezia Reichlin. A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205, 2011. URL https://EconPapers.repec.org/RePEc:eee:econom:v:164:y:2011:i:1:p:188-205.

Catherine Doz, Domenico Giannone, and Lucrezia Reichlin. A quasi—maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):1014–1024, 2012. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/23355337.

James D. Hamilton. Time series analysis : James D. Hamilton, 1994, (Princeton University Press, Princeton, NJ), 799 pp., US 55.00, $ISBN0 - 691 - 04289 - 6.$ 11(3) : 494 − −495, $September1994.URL$https://ideas.repec.org/a/eee/intfor/v11y1995i3p494-495.html.

W. Härdle, L. Simar, and W. Hdrdle. *Applied Multivariate Statistical Analysis*. Springer, 2003. ISBN 9783540030799. URL https://books.google.co.uk/books?id=xt31OYzJxnoC.

Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, second edition, 1999. ISBN 0471986321 9780471986324 047198633X 9780471986331.

Michael McCracken and Serena Ng. Fred-md: a monthly database for macroeconomic research. *Journal of Business Economic Statistics*, 34, 09 2015. doi: 10.1080/07350015.2015.1086655.

Alexei Onatski. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92, 02 2005. doi: 10.1162/REST$_{a0}$0043.

Pilar Poncela, Esther Ruiz, and Karen Miranda. Factor extraction using kalman filter and smoothing: This is not just another survey. *International Journal of Forecasting*, 37, 03 2021. doi: 10.1016/j.ijforecast.2021.01.027.

Marie-Therese Puth, Markus Neuhäuser, and Graeme D. Ruxton. Effective use of pearson's product–moment correlation coefficient. *Animal Behaviour*, 93:183–189, 2014. ISSN 0003-3472. doi: https://doi.org/10.1016/j.anbehav.2014.05.003. URL https://www.sciencedirect.com/science/article/pii/S0003347214002127.

Spencer. Maximum likelihood estimation of dynamic factor models using general cross sectional covariance. *Working paper*, 2024. URL https://eprints.bbk.ac.uk/id/eprint/53612/.

Mark W. Watson and Robert F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23 (3):385–400, 1983. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(83)90066-0. URL https://www.sciencedirect.com/science/article/pii/0304407683900660.