Finite sample performance of information criteria on the number of dynamic shocks in high dimensional factor models

Diploma Thesis

submitted for the acquisition of the engineer degree Dipl.-Ing. at the
Institute for Mathematics in Economics
Technical University Vienna

under supervision of
Em.O.Univ.Prof. Manfred Deistler

by

Nico Didcock
Brentenmaisstraße 9
A-3012 Wolfsgraben

May 13, 2013

Deutsche Kurzfassung

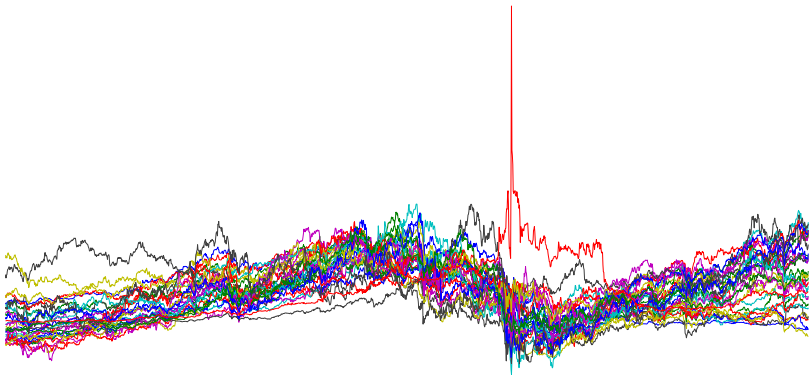
Rationale Übertragungsfunktionen treten bei Autoregressiven oder State Space Systemen auf. Diese Arbeit beschäftigt sich mit daraus resultierenden, rationalen, spektralen Dichten. Es wird unterstellt, dass das System weniger Eingänge als Ausgänge aufweist. Diese Annahme wird durch Faktorenmodelle motiviert, bei welchen eine kleine Anzahl an *Faktoren* eine Vielzahl an Variablen beschreibt. Der Aktienmarkt und makroökonomische Daten sind Anwendungsbeispiele für solche Faktorenmodelle. Der Vorteil dieser Modelle liegt darin, dass die Systemdynamik der Faktoren durch beträchtlich weniger Parameter beschrieben werden kann. Dadurch kann das Problem der Überanpassung vermieden werden. Faktoren eignen sich außerdem, große Datensätze zu verstehen, Zusammenhänge zwischen Variablen zu erkennen und hochdimensionale Systemdynamiken zu prognostizieren.

Das klassische Faktorenmodell ist für Modelle, bei der die Anzahl der Faktoren eine gewisse Schranke nicht überschreitet, generisch identifiziert. Das bedeutet, dass auf einer offenen und dichten Teilmenge des Parameterraumes eindeutige Parameter existieren, die dieses System als Faktorenmodell beschreiben. Das klassische Modell hat den Nachteil, dass die Realisationen im Stichprobemraum nirgends dicht, also dünn, sind. Die Annahme, dass die Fehler unkorreliert sind, ist eng mit der Identifizierbarkeit verknüpft. Eine Schwächung dieser Annahme erfordert eine unbeschränkte Anzahl an Variablen. Damit kann die Annahme gelockert werden und das Modell ist *asymptotisch* identifizierbar. Die Extraktion der Faktoren aus den Daten erfolgt mit Hilfe dynamischer Filtersequenzen, welche aus der Spektraldarstellung der spektralen Dichte geschätzt werden.

Die Schätzung der Anzahl der Faktoren beruht auf asymtotischen Resultaten, wonach gewisse, dynamische Eigenwerte divergieren, andere konvergieren. Diese Arbeit beschäftigt sich mit der Leistung der Schätzmethoden für endliche Stichproben. Eine Simulationsstudie vergleicht die asymtotische Schätzer mit gewissen, *optimalen* Schätzern.

Contents

1. Introduction



Factor models have received great attention in the analysis of macroeconomic and financial data. By simplification of complex data coherency they bring along inspiration for data interpretation while avoiding the usual mistake of overfitting in large data modelling. Originally, these models tried to explain the test scores of participants in intelligence tests by constructing the latent variable *intelligence* that would explain the correlation of the scores. The key word here is dimension reduction. While Charles Spearman has introduced the classic factor model at the beginning of the twentieth century it was only in the most recent decades that generalised versions of this model came into fashion because they were able to deal with large panel data. Indeed, these models could help an information society dealing with large amounts of data that are being produced in many fields, e.g. on the net. In a way, they could help produce information out of simple facts, useless on their own but valuable en masse.

In this work we will be facing the analysis of large panel data, that is, a set of observations $\mathcal{D} = \{y_{it} | i = 1, \dots, N; t = 1, \dots, T\}$ with large N and T. The mathematical model for the data will be a doubly indexed sequence $\{y_{it} | i \in \mathbb{N}; t \in \mathbb{Z}\}$ where we allow for negative integers in time to allow for infinite history and since we will analyse large dimensional processes we model an infinite dimensional process $y_t = (y_{it})_{i \in \mathbb{N}}$. The topics will include dimension reduction as well as prediction of the process. In a sense, there is a connection between these two tasks. The ARMA model of a large dimensional process will lead to far too many parameters to be estimated due to the so called curse of dimensionality. To overcome this problem one can estimate a low dimensional process, the factors, driving the original process and model the dynamics

of the system in the factors. Dimension reduction is a useful tool for itself, because we may be able to gain insights in the structure of the process under investigation.

Notation

Prime $\cdot'$ is for transpose, superscript H for adjoint and superindex N primarily stands for the trunctuated process $y_t^N = v\{y_{it} | i = 1, \dots, N; t \in \mathbb{Z}\}$. For stationary processes we can define the autocovariance at lag s, $\Gamma_s^x = \mathbb{E}[x_{t+s} x_t']$, and $\Gamma_s^{x,N} = \mathbb{E}[x_{t+s}^N x_t^{N\prime}]$. For the covariance I will drop the subscript zero and just write Γ^x. $\mathcal{I}$ is the identity of appropriate dimension and 0 can stand for zero as well as a matrix or vector of appropriate dimension containing zeros since this makes it easier to read and should not cause too much confusion. z denotes a complex variable as well as the lag operator $z y_t = y_{t-1}$. For a filter $a(z)$, which shall be analytic in the lag operator the transfer function is given as $a(e^{-i\lambda})$ as a function of $\lambda \in [-\pi, \pi]$. Since the filters in this work will only have z or $e^{-i\lambda}$ as inputs I will also write sloppily $a(\lambda)$ for the transfer function $a(e^{-i\lambda})$. $\hat{\cdot}$ denotes some estimate and $|\cdot|$ with a matrix input stands for determinant.

Assumtions

The underlying process of investigation y_t is always assumed to fulfil

A1 $\{y_{it} \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})\, i \in \mathbb{N}, t \in \mathbb{Z}\}$ and all finite vector processes y_t^N are weakly stationary zero mean processes with finite second order moments.

A2 For all $N \in \mathbb{N}$, y_t^N is a linearly regular process with absolutely summable covariances, such that the spectral density

$$\Sigma^{y,N}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} \Gamma_k^{y,N} \tag{1.1}$$

exists and is the Fourier transform of its covariance function.

In the following we will be exclusively dealing with decompositions of the process of the form

$$y_t = \chi_t + u_t \tag{1.2}$$

where we will always assume mutual orthogonality of the components, that is

$$\mathbb{E}[\chi_t u_s'] = 0 \quad \forall s, t \in \mathbb{Z}$$

such that

$$\Gamma_s^{y,N} = \Gamma_s^{\chi,N} + \Gamma_s^{u,N} \quad \forall s \in \mathbb{Z}, N \in \mathbb{N} \tag{1.3}$$

and, using obvious notations for $\Sigma^{\chi,N}$ and $\Sigma^{u,N}$,

$$\Sigma^{y,N} = \Sigma^{\chi,N} + \Sigma^{u,N} \quad \forall N \in \mathbb{N} \tag{1.4}$$

We will call χ_t the *common* and u_t the *idiosyncratic component* of the process. The common component is the part of the process that shall be explained by a small number of factors. These factors have a large influence on our data, and the influence needs to be common, i.e. each factor has influence on many variables. The factors are latent, although a primary object of investigation, hidden, or unmeasurable and need to be reconstructed or re-estimated from data. It is the main interest of this work to model and estimate the factors, to be more precise, the dimension of the factors which in general is assumed to be *small*. The common component is assumed to fulfil

$$\chi_t = \Lambda(z) z_t$$

with some factor process f_t of order q and some $N \times q$ filter $d(z)$. Tools that extract the common component $P(z)y_t = \chi_t$ have been widely discussed [Lippi et. al 2000, Forni & Lippi 2001, Deistler et. al. 2010, Stock & Watson 2002]. The most important techniques will be restated and used in the latter.

The Research Question

In this work I will concentrate on ways to determine the number of dynamic factors, that is, the minimum number of variables to explain the co-movements in the data. Several approaches have been discussed and here I want to compare the tools that have been established.

Outline

Section 2 briefly presents the classic factor model. Section 3 presents the structural approach in determining the number of factors. Properties of large transfer functions imply generic results for the structure of factor dynamics. In our setting the innovation process can be shown to be simple, i.e. of low rank and the number of factors is asymptotically identified as the rank of the innovation covariance matrix. Next, we will consider PCA models in Section 4. The nature of these models is rather statistical, that is, dimension reduction without further assumptions on the underlying process but they can be shown to consistently estimate the common components of large factor models. In Section 5 I give an overview on the generalisations of factor models. These generalisations have a drawback of asymptotic identifiability. Therefore, estimation procedures rely on the number of variables and observations growing unlimitedly. Estimation procedures are presented in Section 6. In the simulation study presented in Section 7 I want to determine the finite sample properties of these estimation procedures. Section 8 presents results.

2. The Classic Factor Model

Given an N dimensional stochastic process y_t, we say that the process suffices a classic factor model if there exist an integer $r < N$, an r dimensional process z_t, a constant $N \times r$ matrix Λ and an N dimensional process u_t, such that

$$y_t = \Lambda z_t + u_t \tag{2.1}$$

z_t are called the factors, the vectors forming Λ are called the factor loadings, u_t is the error of the model or the idiosyncratic component. The common component here takes the form Λz_t. The model assumptions formalise as:

CF1 $\mathbb{E}[z_t u_s'] = 0 \quad \forall t, s \in \mathbb{Z}$

CF2 $\Gamma^z = \mathcal{I}$

CF3 Γ^u is a diagonal matrix

The first assumption is a standard assumption on model errors and allows us to decompose the covariance of y_t as

$$\Gamma^y = \Lambda \Gamma^z \Lambda' + \Gamma^u \tag{2.2}$$

and we can always force CF2 by post-multiplying Λ by $(\Gamma^z)^{\frac{1}{2}}$ and pre-multiplying z_t by $(\Gamma^z)^{-\frac{1}{2}}$. The aim of the classic factor model was to explain the *correlation* of the variables so the most important assumption is CF3 which essentially means that the whole correlation in y_t is explained by the factors and the loadings. The essential idea is to find a *sparse* matrix Γ^u such that $\Gamma^y - \Gamma^u$ is simple, i.e. reduced rank and can be represented as $\Lambda\Lambda'$. Estimation of the model is concerned with extracting $\Lambda\Lambda'$ and Γ^u from (2.2) and Λ from $\Lambda\Lambda'$.

Existence

For a given stochastic process or vector, the existence of the structure is a question of the number of factors. Obviously, for $N = r$ we find a trivial factor model $y_t = z_t$. For less factors, the covariance needs to suffice the characteristic factor model structure, $\operatorname{rank}(\Gamma^y - \Gamma^u) = r$. (2.2) is a system of $N(N+1)/2$ unique equations to determine r^2 elements in $\Lambda\Lambda'$ and N parameters for the residual covariance. We can expect the structure to exist if the number of factors is large enough, i.e.

$$r \geq \sqrt{N(N-1)/2}$$

Identification

Conversely, a large number of factors inhibits the identification of the parameters. Λ is uniquely determined from $\Lambda\Lambda'$ up to right multiplication with orthogonal matrices

which comprises $r(r-1)/2$ restrictions for (2.2). Sometimes the loadings are forced to suffice a simple structure, where a large amount of entries is close to zero. This helps to obtain interpretable factors as the variables can be grouped according to the influence of the factors. The idea is to find an arbitrary matrix $\tilde{\Lambda}$ in the equivalence class $\{\Lambda|\Gamma^y = \Lambda\Lambda' + \Gamma^u\}$ and then rotate this matrix $\Lambda = \tilde{\Lambda}H$ to obtain the simple structure. A useful restriction to find representative loadings is to force $\Lambda'\Gamma^{u\,-1}\Lambda$ to be a diagonal matrix. If the number of parameters does not exceed the number of equations plus the number of restrictions we can expect identification. The model is generically identified below the so called *Ledermann-bound*.

$$r \leq (2N+1)/2 - \sqrt{((2N+1)/2^2) - (N^2 - N)}$$

Maximum Likelihood Estimation

Maximum likelihood methods have proved convenient for the estimation of parameters if the structure of the model is known, that is, a class of distributions $f(\cdot, \theta)$ of a stochastic variable is given and we are only interested in certain parameters θ of the distribution. The maximum likelihood method calculates the parameters that maximise the joint conditional probability of a given data set $\mathcal{D}$. The so called *likelihood function* is defined as

$$L_T(\theta|\mathcal{D}) = \prod_{t=1}^{T} f(y_t, \theta)$$

The aim is to maximise this function over the parameters. Here, $\theta = (\Lambda, \Gamma^u)$. Assume that y_t is normally distributed. Then the log likelihood can be differentiated and set to zero to obtain necessary conditions for the maximum likelihood estimator Λ_{ml}: (see the Appendix B for the differentiation of the functionals)

$$\log(L_T)(\Lambda, \Gamma^u|\hat{\Gamma}^y) = -\frac{t}{2}|\Gamma| - \frac{1}{2}\sum_{t=1}^{T} y_t'\Gamma_y^{-1}y_t = -\frac{t}{2}|\Gamma| - \frac{t}{2}\operatorname{tr}(\hat{\Gamma}^y(\Gamma^y)^{-1})$$

$$\begin{aligned} \operatorname{diag}(\Gamma^y)^{-1} &= \operatorname{diag}\hat{\Gamma}^y(\Gamma^y)^{-2} \quad |\cdot(\Gamma^y)^2(\Gamma^u)^{-1}\Lambda \\ \Lambda(\mathcal{I} + \Lambda'\Gamma^u\Lambda) &= (\hat{\Gamma}^y(\Gamma^u)^{-1})\Lambda \end{aligned}$$

Therefore, the maximum likelihood estimator for the loadings are eigenvectors of the sample covariance $\hat{\Gamma}^y$ relative to Γ^u, see the definition (4.3). Since the idiosyncratic variance is in general unknown, iterative algorithms are used to estimate Λ_{ml}.

Remarks

- Factor analysis is a useful tool in overcoming the curse of dimensionality for large datasets. Instead of modelling an $N \times N$ AR system in the variables we can model a $r \times r$ system in the factors and predict the future output via $\hat{y}_{t+1} = \Lambda \hat{z}_{t+1}$ according to a future estimate of the factors $\hat{z}_t$ which is less likely to be biased from of overfitting.

- The structure in (2.1) looks like a standard regression model. The fundamental difference between factor models and regression models is the fact that the regressors z_t are not known but have to be estimated from the data. We can see factor models as an approach to find regressors of low dimension.

- The classic factor model does not incorporate dynamics, either in the factors or the errors. However factor models are often used for prediction and therefore incorporating model dynamics is essential for the further purposes. The models in the later chapters will incorporate dynamics.

- PCA, which is discussed in Section 4, gives an approximation to the exact factor model. It is far easier to calculate, especially for large data sets the maximum likelihood method becomes computationally expensive. When the number of variables and observations tends towards infinity the two estimation procedures become equivalent, see Section 5.1.

So now we know what the classic factor model looks like I want to give an overview on how important extensions, such as dynamics in the common component, are usually modeled in the literature.

3. Representations of the Common Component

Some models discussed in this work, the exact factor model, the PCA model as well as the model by Chamberlain and Rothschild suggest modelling panel data in the form (2.1) with various assumptions about the covariance matrix of the errors. Now this type of representation is generally called *static*, since the common component at time t is determined by the *static factor* at time t only,

$$\chi_t = \Lambda z_t \tag{3.1}$$

It is clear that the dimension of the static factors cannot deceed the rank of the covariance matrix of the common component. With equality the static factors are called *minimal*. Let the common covariance have rank r then minimal static factors which satisfy $\Gamma^z = \mathcal{I}$ can be obtained via the factorisation of the covariance matrix

$$\Gamma^\chi = \Lambda \Lambda' \tag{3.2}$$

$$z_t = (\Lambda' \Lambda)^{-1} \Lambda' \chi_t \tag{3.3}$$

which is unique up to post multiplication by an orthogonal matrix $\tilde{\Lambda} = \Lambda H$ and $\tilde{z}_t = H' z_t$. Dynamics are inherently important since most applications aim for future prediction of the variables. We can allow the static factors to admit dynamics, such as

$$a(z) z_t = \nu_t \tag{3.4}$$

for some filter $a(z) = \sum_{k=0}^{\infty} a_k z^k$, which is usually assumed stable and one sided to ensure causality. The white noise process ν_t influences present and future variables, but still, the information content of the present factor cannot be improved by knowledge of past factors, i.e. we model a Markov chain $\mathbb{E}[y_t|z_t] = \mathbb{E}[y_t|z_t, z_{t-1}, \ldots]$. Prediction is calculated via $\mathbb{E}[y_t|\hat{z}_t] = \Lambda \hat{z}_t$ and $\hat{z}_t$ is estimated from (3.4).

One generalisation of the classic factor model is taking the dynamics of the model directly into account, that is, instead of static factors influencing the present only, a q dimensional orthonormal white noise process ϵ_t, the *dynamic shocks*, are modelled driving present as well as future variables. Such models have been discussed in [Lippi et. al 2000, Forni & Lippi 2001, Stock & Watson 2002, Deistler et. al. 2010]. Let's commence from the *wold decomposition* of the stationary time series

$$\chi_t = \omega(z) \epsilon_t \tag{3.5}$$

where $\omega(z)$ is the $N \times q$ transfer function. From (3.5) it follows that Σ^{χ} will have rank q *a.e.*

3.1. Rational Transfer Functions

Conversely, every rational transfer function of constant rank q can be factorised as

$$\Sigma^{\chi}(\lambda) = \frac{1}{2\pi} \omega(e^{-i\lambda}) \omega^H(e^{-i\lambda}) \tag{3.6}$$

where $\omega(z)$ is a real rational matrix with full column rank which has no poles and zeros for $|z| \leq 1$ and is unique up to post multiplication by constant orthogonal matrices [Deistler et. al. 2010]. For the latter discussion it makes sense to put the transfer function into its truncated *Smith McMillan form*,

$$w(z) = u d^{-1} n v = u n d^{-1} v = u(z) \begin{bmatrix} \frac{n_1}{d_1}(z) & & \\ & \cdot & \\ & & \frac{n_q}{d_q}(z) \end{bmatrix} v(z) \tag{3.7}$$

where u, v are unimodular, i.e. polynomial and have a determinant constant in z and unequal to zero and d_i and n_i are monic and coprime polynomials in z. The zeros

of n correspond to the finite zeros of ω and the poles of ω are the zeros of d Moreover, d_{i+1} divides d_i, n_i divides n_{i+1}. , see e.g. [Deistler et. al. 2010]. The Smith McMillan form exists for every rational transfer function matrix. We can define a stable, causal left inverse to (3.7) by $\omega^- = v^{-1}(d'd)^{-1}d'u^{-1}$ and reconstruct the shocks from the common component via

$$\epsilon_t = \omega^- \chi_t$$

A simple modification of (3.7) gives insight into the ARMA representation of the transfer function

$$ud^{-1}nv = \tilde{u}\tilde{d}^{-1}\tilde{n}v$$

such that $\tilde{u}$ is $N \times N$ and unimodular but the first q columns are equal to u, $\tilde{d}$ is the $N \times N$ diagonal matrix with first q diagonal elements as in d and the rest of the diagonal is filled up with ones. $\tilde{n}$ is the $N \times q$ matrix where the top square matrix is equal to n and the bottom $N-q$ columns are equal to zero. Now $\tilde{u}\tilde{d}^{-1}$ can be inverted and we result in the ARMA representation

$$\underbrace{a(z)}_{\tilde{d}\tilde{u}^{-1}} \chi_t = \underbrace{b(z)}_{\tilde{n}v} \epsilon_t \tag{3.8}$$

It is important to see that the static factors suffice the same dynamics as the common component. To see this, put (3.5) into (3.3) and get

$$z_t = (\Lambda'\Lambda)^{-1}\Lambda'\omega(z)\epsilon_t = k(z)\epsilon_t \tag{3.9}$$

The representation $y_t = \Lambda k(z)\epsilon_t$, where Λ is a tall constant zero lag matrix is called a *structural factor model* and has been discussed in [Lippi et. al 2009]. In general we will always assume the system to be stable, i.e.

$$\det(a(z)) \neq 0 \quad |z| \leq 1. \tag{3.10}$$

If we commence from an ARMA representation, e.g. (3.8) and the *stability condition* (3.10) holds then $a(z)$ has a power series expansion on an annulus containing the unit circle and is invertible for this region. Moreover, the process defined by $a^{-1}(z)b(z)\epsilon_t$ will be stationary and this solution will be causal. Summarising, the stability assumption guarantees the existence of a stationary process spanned by present and past values of the shocks and sufficing (3.8) for all t.

3.2. Zeroless Transfer Functions

In our setting ($N > q$) a number of interesting results have been derived. See **??** for a discussion on *large* transfer functions when the number of outputs exceeds the number of inputs. The paper discusses transfer function matrices resulting from state space systems of the form

$$x_{t+1} = Ax_t + B\epsilon_t \tag{3.11}$$
$$\chi_t = Cx_t + D\epsilon_t \tag{3.12}$$

It has been shown that for generic and minimal state space systems, i.e. for generic values of A, B, C, D such that the system (3.11)-(3.12) is minimal, the transfer function has no zeros, i.e. the numerator polynomials of n in (3.7) are all equal to one. In this case the transfer function is said to be *zeroless*. Since minimal state space systems are open and dense in the set of all rational transfer functions the result also holds for generic rational transfer functions. The zeroless case is particularly interesting because in this setting the common component as well as the static factors suffice AR systems. $k(z)$ is zeroless if and only if $\omega(z)$ is zeroless. $k(z)$ has a finite polynomial inverse k^- and z_t is the stationary solution of the stable AR system

$$\underbrace{z_t - a_1 z_{t-1} + \cdots + a_p z_{t-p}}_{a(z)z_t} = \nu_t = b\epsilon_t \tag{3.13}$$

To be more precise, the factors are determined by a finite number of past shock values and vice versa. The innovation process ν_t in (3.13) has rank q. These AR systems differ from the usual assumption that the innovation process has the same rank as the variables. Here, especially in the case $q < r$, the innovation process will have lower rank. These systems are called *singular AR systems* and will be discussed in more detail in Section 6.2.

Summarising, for the generic case we can put the dynamics of the static factors into the following state space form:

$$\begin{bmatrix} z_t \\ \cdot \\ z_{t-p+1} \end{bmatrix} = \begin{bmatrix} a1 & a_2 & \cdot & a_{p-1} & a_p \\ \mathcal{I} & 0 & \cdot & \cdot & 0 \\ 0 & \mathcal{I} & 0 & & \cdot \\ \cdot & & \ddots & & 0 \\ 0 & \cdot & 0 & \mathcal{I} & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ \cdot \\ z_{t-p} \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix} \epsilon_t \tag{3.14}$$

$$\chi_t = \begin{bmatrix} \Lambda & 0 \end{bmatrix} \begin{bmatrix} z_t \\ \cdot \\ z_{t-p} \end{bmatrix} \tag{3.15}$$

For state space systems we have the possibility of kalman filtering which will be discussed in Section 6.

Two Approaches

The latent variables can be represented by a linear static transformation of the static factors or by a linear dynamic transformation of the dynamic shocks. Clearly every static factor is a dynamic factor with a zero lag filter and therefore it should be clear that we can assume $q \leq r$. It is important to note that the factors z_t and therefore the common component each have spectral density of rank q. The goal of this dissertation is to examine the methods to determine q. The first approach commences from (3.13). If we know the static factors then q is determined as the rank of the covariance matrix of the innovations. The second approach estimates the rank of the spectral density of the common component. However, both approaches require some pre-estimation because in general we will not know the common component or the static factors. The next sections will discuss how to filter out the common component from the variables and how to estimate the static factors. Factor model estimation can be seen as the infinite dimensional analogue to PCA analysis and therefore it is natural that estimation is strongly connected to PCA estimation techniques. Principal Component Models will be discussed in the next section.

4. Principal Component Models

The nature of principal component models is purely statistical although there are several approaches towards principal components. One approach is to find a small number of linear combinations of the variables that can reconstruct the variables in a best possible way. Dynamic principal components are filtered variables that explain the variables best possible via non static filters. Another way of looking at the problem is to find the reduced rank covariance or spectral density that fits the population moments best in some norm, see e.g. [Deistler & Anderson 1989, Brillinger 2001].

Standard Principal Component Analysis (PCA) was invented by Karl Pearson at the beginning of the twentieth century and established as a fundamental basis for dimension reduction applications. The idea was to reduce the complexity of a large data set by finding the normalised linear combinations of stochastic variables with maximum variance. These linear combinations are then called the principal components and have the feature to explain the most variance of the original variables via linear combinations. It turns out that the linear combinations in both cases are equal and can be represented by the eigenvectors of the population covariance matrix of the original variables. A wide range of principal component models are available, an excellent discussion can be found in [Jolliffe 2002].

For this section, let y_t denote a finite, stochastic vector process. The PCA problem can be formalised as to find a reduced rank filter $P(z)$ such that $P(z)y_t$ is closest to the original process. The mathematical formalisation is to minimise the residual variance:

$$\min_{P(z)} \mathbb{E}[y_t - P(z)y_t]'[y_t - P(z)y_t] \tag{4.1}$$

over all filters with $\operatorname{rk} P(z) = q < N$. To solve this problem we recall the concept of eigenanalysis.

4.1. Classic PCA

Classic PCA deals with the problem formulated in (4.1) but restricts the filter to be a constant no lag matrix, $P(z) = P$ with $\operatorname{rk} P = r$. With this restriction the problem is solved by the spectral decomposition of the covariance matrix Γ^y. For now, I will drop the time subscript and will write y instead of y_t.Given a square matrix $A \in \mathbb{C}^{N\times N}$ the solutions $\mu_1, \cdot, \mu_N \in \mathbb{C}$ and $\gamma_1, \cdot, \gamma_N \neq 0 \in \mathbb{C}^N$ to the equations

$$A\gamma_j = \mu_j \gamma_j \quad j = 1, \cdot, N \tag{4.2}$$

are called the *eigenvalues* and the corresponding *eigenvectors* of A. Given in addition an invertible matrix B of the same dimension as A we call the solutions to

$$A\gamma_j = \mu_j B\gamma_j \quad j = 1, \cdot, N \tag{4.3}$$

the *eigenvalues* and the corresponding *eigenvectors* of A *relative to* B. For a stochastic vector y or stationary process y_t, μ_j^y shall denote the jth eigenvalue of Γ^y. We will consider normalised eigenvectors only, that is, $||\gamma_j|| = 1$, which can be constructed immediately since eigenvectors corresponding to the same eigenvalue form a linear subspace. If A is hermitian the eigenvalues are real and eigenvectors corresponding to different eigenvalues are orthogonal since

$$\mu_i^H \gamma_i^H \gamma_j = (\mu_i \gamma_i)^H \gamma_j = (A\gamma_i)^H \gamma_j = \gamma_i^H A^H \gamma_j = \gamma_i^H A \gamma_j = \mu_j \gamma_i^H \gamma_j$$

implies $\gamma_i^H \gamma_j = 0$ or $\mu_i^H = \mu_j$. It makes sense to order the eigenvalues in size, w.l.o.g. $\mu_1 \geq \cdot \geq \mu_N$. A very useful property of eigenvalues is the following:

$$\mu_k = \inf_{\dim H = k-1} \sup_{\alpha \in H^\perp} \frac{\alpha' A \alpha}{\alpha' B \alpha} \tag{4.4}$$

Every hermitian matrix $A \in \mathbb{C}^{N\times N}$ can be decomposed in the form

$$A = \sum_{k=1}^{N} \mu_k \gamma_k \gamma_k^H \tag{4.5}$$

with its eigenvalues μ_k and the corresponding orthonormal eigenvectors γ_k. If A is real valued then the eigenvalues as well as the eigenvectors are real valued. If A is in addition semi definite (e.g. a covariance matrix) then $\mu_1 \geq \cdot \geq \mu_N \geq 0$.

Now let us return to our stochastic vector y and let μ_j be the eigenvalues with corresponding eigenvectors γ_j of the covariance matrix Γ^y and define $\Gamma^{(r)} = [\gamma_1, \cdot, \gamma_r]$. The components in $z = {\Gamma^{(N)}}' y$ are called the *principal components* of y. Their covariance

matrix is given by

$$\mathbb{E}[zz'] = \begin{bmatrix} \gamma_1^H \\ \cdot \\ \gamma_N^H \end{bmatrix} \sum_{j=1}^{N} \mu_j \gamma_j \gamma_j^H [\gamma_1, \cdot, \gamma_N] = \begin{bmatrix} \mu_1 & & 0 \\ & \cdot & \\ 0 & & \mu_N \end{bmatrix} \tag{4.6}$$

which follows from the orthogonality of the eigenvectors and from (4.5). These variables are a fundamental set of independent variables, that is, the original variables can be reconstructed by linear combination

$$y = \Gamma^{(N)} z = \Gamma^{(N)} \Gamma^{(N)'} y$$

and they are mutually orthogonal. Lets return to the minimisation problem

$$\min_{\mathrm{rk}\, P = r} \mathbb{E}[y - Py]'[y - Py] \tag{4.7}$$

Equivalently, since every rank r matrix can be decomposed as $P = CB$ with an $N \times r$ matrix C and an $r \times N$ matrix B the problem can be formalised as

$$\min_{B,C} \mathbb{E}[y - CBy]'[y - CBy] \tag{4.8}$$

Then problem (4.8) is solved by $C = \Gamma^{(r)}$ and $B = C'$. $P = \Gamma^{(r)}\Gamma^{(r)\prime}$ corresponds to the orthogonal projection onto the r dimensional subspace spanned by the first r principal components. The resulting minimum in (4.8) is equal to

$$\sum_{j>r} \mu_j \tag{4.9}$$

Remarks

- Note that the classic factor model in the previous section was to explain the correlation among a number of stochastic variables, the aim in PCA models is to explain as much variance as possible by a lower dimensional process.

- There are several approaches that lead to principal component analysis. They correspond to the N normalised linear combinations $o_j' y$ with maximum variance while being mutually orthogonal. This can be seen by solving with Lagrange's method:

$$\max_{o \in \mathbb{R}^N} \mathrm{Var}[o'y] = \mathbb{E}[o'yy'o] = o'\Gamma^y o \quad \text{s.t.} \quad o'o = 1$$

$$
\begin{aligned}
L(o) &= o'\Gamma^y o - \mu(o'o - 1) \\
\frac{\partial L}{\partial o}(o) &= 2\Gamma^y o - 2\mu o = 0 \\
\Gamma^y o &= \mu o \quad \text{and} \quad o'o = 1
\end{aligned}
$$

So we see the connection to (4.2) and that the solution must correspond to an eigenvector of Γ^y. Since $\mathrm{Var}[\gamma_k' y] = \gamma_k' \Gamma^y \gamma_k = \mu_k ||\gamma_k|| = \mu_k$ the wanted linear combination is the eigenvector to the largest eigenvalue and since

$$
\mathbb{E}\gamma_i' y (\gamma_j' y)^H = \gamma_i^H \Sigma_y \gamma_j = \mu_i \gamma_i^H \gamma_j = 0
$$

we see that the orthogonality of the eigenvectors already implies the required zero correlation of the linear combinations.

- We can think of a given $T \times N$ set of data $\mathcal{D}$ as a point cloud of T elements embedded in an N dimensional vector space. Consider any orthonormal transformation $z = O'y$ with $O \in \mathbb{C}^{N \times r}$. It corresponds to the projection of the data onto a r dimensional space. Then $O = \Gamma^{(r)}$ maximises $\mathrm{tr}\,\Sigma_z$ and $\det \Sigma_z$. Both functionals can be seen as measures of the volume of the projected point cloud. This means that principal components maximises the individual variances of the projections onto one dimensional spaces as well as the *generalised volume*, $\det \Sigma_z$.

So, facing the task of dimension reduction we already know that standard principal components are a pretty good first guess how to solve our problem (4.1) by simply maximising the variance explained by linear combination of a small amount of latent variables. Here, the factors correspond to the first r principal components. However, this method does not take into account that we are not dealing with simple iid random variables because we are facing a time series. Since the whole information for principal components is contained in the covariance matrix we would get the same result for every permutation in time of the observed process. There are two possibilities how to bring the dynamical behaviour into account. Either we simply assume that the dynamics of the system are due to dynamics in the principal components, that is, we estimate a VAR model of the principal component scores, or we extend the concept of principal components to the socalled *dynamic principal components*.

4.2. Dynamic PCA

Now suppose we are not analysing some data vector y but a stochastic process, or a signal y_t. Let $q \leq N$ channels be available for the transmission of the signal so in general we will not be able to transmit the whole signal. Dynamic PCA projects the original signal onto a q dimensional signal such that the original variables can be re-constructed best possible from a limited amount of channels, see [Brillinger 2001]. Dynamic PCA

is for stochastic processes what classic PCA is for random variables. While the static problem in (4.7) has solutions related to the eigenvectors of the covariance matrix of Γ^y the unrestricted problem (4.1) turns out to have solutions related to eigenvectors related with the spectral density Σ^y generally called the *dynamic eigenvectors*. To construct the optimal unrestricted q ranked filter $P(z)$ we denote the projection error $u_t = [\mathcal{I} - P(z)]y_t$ and rewrite (4.1) as

$$\begin{aligned}\min \mathbb{E}[y_t - P(z)y_t]^H[y_t - P(z)y_t] &= \min \mathbb{E}\,\mathrm{tr}[y_t - P(z)y_t][y_t - P(z)y_t]^H \\ &= \min \mathrm{tr}\,\Gamma^u\end{aligned}$$

Since the covariance matrix is the inverse Fourier transform of the spectral density

$$\Gamma^u = \frac{1}{2\pi}\int_0^{2\pi} \Sigma^u(\lambda)d\lambda$$

and the spectral density of the error can be written as

$$\Sigma^u(\lambda) = [\mathcal{I} - P(e^{-i\lambda})]\Sigma^y(\lambda)[\mathcal{I} - P(e^{-i\mu})]^H$$

we result in the minimisation problem

$$\min \mathrm{tr}\int_0^{2\pi} [\mathcal{I} - P(e^{-i\lambda})]\Sigma^y(\lambda)[\mathcal{I} - P(e^{-i\lambda})]^H d\lambda \tag{4.10}$$

which we can solve for every $\lambda \in [-\pi, \pi]$ with classic PCA.

Dynamic Eigenvalues, Dynamic Eigenvectors and Principal Component Series

The vector valued function $\tilde{\gamma}_j : [-\pi, \pi] \to \mathbb{R}^N$ such that $\tilde{\gamma}_j(\lambda)$ is the eigenvector corresponding to the jth eigenvalue $\mu_j(\lambda)$ of $\Sigma^y(\lambda)$ for all $\lambda \in [-\pi, \pi]$ is called the jth *dynamic eigenvector* and the function $\lambda \mapsto \mu_j(\lambda)$ is called the jth *dynamic eigenvalue* of y_t. The dynamic eigenvectors are measurable in λ and can be normed to unity for every λ, see [Lippi et. al 2000, Brillinger 2001]. As a consequence, we find a Fourier series of the complex vector valued function as

$$\tilde{\gamma}_j(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi} \tilde{\gamma}_j(\lambda)e^{ik\lambda}d\lambda\right] e^{-ik\lambda}$$

and we find a filter $\gamma_j : \mathbb{Z} \to \mathbb{R}^N$ such that $\tilde{\gamma}_j = \gamma_j \circ \exp(-i\cdot)$. The scalar process $\gamma_j(z)^H y_t$ has spectral density $\mu_j(\lambda)$ and is called the jth *principal component series* of y_t. From (4.5) we can factorise the spectral density matrix as

$$\Sigma^y(\lambda) = \sum_{k=1}^{N} \mu_k(\lambda)\gamma_k(e^{-i\lambda})\gamma_k(e^{-i\lambda})^H \tag{4.11}$$

From (4.10) it can be seen that the unrestricted problem (4.1) is solved by the filter who's transfer function is given by

$$P(e^{-i\lambda}) = \sum_{k=1}^{q} \gamma_k(e^{-i\lambda})\gamma_k(e^{-i\lambda})^H \tag{4.12}$$

The filter $P(z)$ projects the process onto the space spanned by the first q dynamic principal components of $\Sigma^{y,N}$. The resulting minimum in (4.1), compared to (4.9), is equal to

$$\int_0^{2\pi} \sum_{j>q} \mu_j(\lambda) d\lambda$$

The *filter coefficients* in $B(z), C(z)$, such that

$$z_t = B(z)y_t = \sum_u b_u y_{t-u} \tag{4.13}$$

$$\chi_t = P(z)y_t = C(z)z_t = \sum_u c_u z_{t-u} \tag{4.14}$$

can be expressed as

$$b_u = \frac{1}{2\pi} \int_0^{2\pi} B(\lambda) e^{-iu\lambda} d\lambda \tag{4.15}$$

$$c_u = \frac{1}{2\pi} \int_0^{2\pi} C(\lambda) e^{-iu\lambda} d\lambda \tag{4.16}$$

where $C(\lambda) = [\tilde{\gamma}_1(\lambda), \cdot, \tilde{\gamma}_q(\lambda)]$ and $B(\lambda) = C(\lambda)^H$. Again, I refer to [Brillinger 2001] for proofs and an excellent discussion.

5. Generalised Factor Models

The goal of this section is to present what is called the *Generalised Dynamic Factor Model*. *Dynamic* means that the dynamics are now captured directly within the model. The term *generalised* refers to a relaxation of the crucial restriction CF3. First, I discuss the assumption in large panel data models that the number of variables is assumed not only to be large, but even unbounded.

Taking the limit $N \to \infty$

The identification of the classic factor model from Section 2 depends heavily on the assumption of diagonal noise. The assumption is quite unrealistic because the set of covariance matrices sufficing a classic factor structure is thin within the set of covariance

matrices. The idea is to generalise the model and let the noise be *weakly correlated*. The problem is how to define the term *weakly*. If we dismiss the orthogonality assumption then the model ceases to be identified. Reasonable assumptions to maintain identifiability are hard to find [Lippi et. al 2000]. An *approximate factor structure* has been introduced in [Chamberlain 1983, Chamberlain & Rothschild 1983], the model is presented in Section 5.1. Here, the model structure can be identified asymptotically if the number of variables increases without bounds. Obviously, this is a purely theoretical assumption. Despite the fact that we are never faced by an infinite number of variables we also loose the most important tool we have worked with so far, we have no principal component analysis for an infinite dimensional covariance matrix. Note that the PCA method is not nested, in the sense that estimation before and after an additional variable is taken into account will in general produce different results, i.e.: the space spanned by the factors (the first r principal components) will change and the projection of each single variable onto this space will as well. So the question we have to ask is if there is a method in this setting that can be seen as the limit of PCA estimation, i.e. if the projections will converge.

Averaging out the Common Component

Recall the PCA estimation method, where the factors are calculated as a linear combination of the form $\chi_t = Py_t$ with $P = [\gamma_1, \ldots, \gamma_r][\gamma_1, \ldots, \gamma_r]' = \sum_{k=1}^{r} \gamma_k \gamma_k'$. The idea behind this method is that considering

$$y_t = \chi_t + u_t = \sum_{k=1}^{r} \gamma_k \gamma_k' y_t + \sum_{l=r+1}^{N} \gamma_l \gamma_l' y_t \tag{5.1}$$

$$\Gamma^{y,N} = \underbrace{\sum_{k=1}^{r} \mu_k \gamma_k \gamma_k'}_{\Gamma^{\chi,N}} + \underbrace{\sum_{l=r+1}^{N} \mu_l \gamma_l \gamma_l'}_{\Gamma^{u,N}}$$

constructing linear combinations as possibly weighted eigenvectors of the r largest eigenvalues will cancel out u_t in (5.1) due to orthogonality and the principal components spanning the common component are maintained, $Py_t = \chi_t$, $\text{var}[Py_t] = \Gamma^{\chi,N}$ and $\text{var}[\gamma_k' u_t] = \gamma_k' \Gamma^{u,N} \gamma_k = 0$ such that the common component is restored and the idiosycratic component is cancelled out.

For the infinite setting we want to find infinite linear combinations that cancel out everything but the space spanned by the factors, but we cannot work with orthogonality any more and it is not trivial even finding a factor space as the finite dimensional principal components. We need to find a *sequence* of finite linear combinations $a^{N\ 1}$, such that

[1] To avoid confusion I need to point out that I have worked with a similar notation previously, namely $y^N = \{y_1, \ldots, y_N\}$. Here the sequence is definitely not nested, i.e. $a_j^N, \neq a_j^M$ for $j \leq N < M$.

$$a^{N\prime}\Gamma^{u,N}a^N \to 0 \tag{5.2}$$

Now assume that the eigenvalues of $\Gamma^{u,N}$ are bounded in N by c. From (4.4) we have

$$\frac{a^{N\prime}\Gamma^{u,N}a^N}{a^{N\prime}a^N} \leq \lambda_{r+1}^N < c \quad \forall N \in \mathbb{N}$$

and we achieve (5.2) for any sequence with $a^{N\prime}a^N = \|a^N\|_2^2 = \sum |a_i|^2 \to 0$. We call a sequence of real valued weights $\left(a^N\right)_{N\in\mathbb{N}} = \left(a_j^N\right)_{j\leq N, N\in\mathbb{N}}$ an *averaging sequence* if $\|a^N\|_2 \to 0$. The method is to define the factor space as the set of limits of linear combinations of averaging sequences, that is,

$$\mathcal{Z} = \{\lim_{N\to\infty} a^{N\prime}y_t^N |\ \|a^N\| \to 0\}$$

Now we want to apply this method we need to make sure that the common component is not cancelled out as well, so we will also assume that the first r eigenvalues will tend to infinity. It were Gary Chamberlain and Michael Rothschild who proved that these two conditions were almost enough to find a real infinite dimensional generalisation of the PCA method. In fact [Chamberlain 1983] and [Chamberlain & Rothschild 1983] proposed quite a few interesting results and will be summarised in the next section.

Define $\mathcal{H} = \bigcup_{N=1}^{\infty} \text{span}\{y_{1t}, \ldots, y_{Nt}\}$ then $\overline{\mathcal{H}}$ is a Hilbert space under the mean square inner product

$$\langle z_1, z_2\rangle = \mathbb{E}[z_1' z_2] \quad z_1, z_2 \in \overline{\mathcal{H}}$$

for $z_j \in \mathcal{H}$. We want to be sure the linear combination makes sense and assume $\text{Var}[z] < \infty$. This means that we restrict the linear combinations to

$$\langle z, z\rangle = a^{N\prime}\Gamma^{y,N}a^N < \infty$$

If r eigenvalues of the covariance matrix diverge and the rest are bounded then the subspace $\mathcal{Z}$ of $\overline{\mathcal{H}}$, corresponding to averaging sequences is r dimensional. The factor space can be seen as the space where each element correlates with an infinite amount of variables. Factors that influence a finite number of variables will be cancelled out by the averaging sequences. This is why this framework corresponds to a generalisation of the classic model, uncorrelated errors influence one variable only. The formal framework for this discussion is taken from [Chamberlain 1983, Chamberlain & Rothschild 1983] and repeated here. It is shown that there exists a seminorm on $\overline{\mathcal{H}}$ which is consistent with the finite dimensional version. the factor space is the space associated with averaging sequences and it is r dimensional. the projection onto the first principal components converges to the projection onto the factor space.

Lemma from [Chamberlain & Rothschild 1983]. *Suppose that* $\sup_{N\in\mathbb{N}} \mu_r^N = \infty$, $\sup_{N\in\mathbb{N}} \mu_{r+1}^N < c$ *and* $\mu_\infty \equiv \inf_{N\in\mathbb{N}} \mu_N^N > 0$. *Then the following results hold:*

(i) *There is a seminorm* $||\cdot||_2$ *defined on* $\overline{\mathcal{H}}$ [2] *such that* $a^{N\prime} y_t^N \to p \Rightarrow \|a^N\|_2 \to \|p\|_2$.

(ii) $\Xi \equiv \{p \in \overline{\mathcal{H}} : \|p\|_2 = 0\}$ *is an* r *dimensional linear subspace and therefore closed*

(iii) $\mathrm{Var}[q] \leq c\|q\|_2^2 \quad$ *for* $q \in \Xi^{\perp}$

(iv) *Let* P_N *be the orthogonal projection onto the first* r *(standardised) principal components and* P *the orthogonal projection* $\overline{\mathcal{H}} \to \Xi$ *then*

$$\lim_{N\to\infty} P_N p = Pp \tag{5.3}$$

Now we can define the factors z_{jt} as any orthonormal basis of Ξ and set $\lambda_{ij} \equiv \mathrm{Cov}[y_{it}, z_{jt}]$. Then

$$y_{it} = \sum_{j=1}^{r} \lambda_{ij} z_{jt} + u_{it} \tag{5.4}$$

and $u_t \perp z_t$ for all t.

5.1. The Approximative Factor Structure by Chamberlain and Rothschild

The first step generalising the factor model was done in [Chamberlain & Rothschild 1983]. This paper dealt with the analysis of large asset markets but the concept works for any infinite dimensional process. The idea was to propose a sequence of models as the number of variables tends towards infinity. Again, the model divides the variables into a common and an idiosyncratic component. Contrary to the classic factor structure they proposed what they call an *approximate r-factor structure*, that is, the idiosyncratic covariance is not assumed diagonal but contains in a sense bounded covariance. For every $N \in \mathbb{N}$ the covariance matrix of y_t^N shall be decomposed as

$$\Gamma^{y,N} = \Lambda^N \Lambda^{N\prime} + \Gamma^{u,N} \tag{5.5}$$

where Λ^N is $N \times r$, the sequence $\{\Lambda^N\}_{N\in\mathbb{N}}$ is nested and the sequence of the largest eigenvalues of $\Gamma^{u,N}$, $\{\mu_1^{u,N} | N \in \mathbb{N}\}$ is uniformly bounded in N. Compared to (2.2) the bounded eigenvalue assumption generalises CF3 since the eigenvalues of the diagonal matrix are simply the diagonal entries. Compared to the PCA model in Section 4 however, the sequence of the loadings are nested. One of the key findings of the paper was that (non nested) principal components provide consistent estimators for the common variance $\Lambda^N \Lambda^{N\prime}$ as $N \to \infty$. Therefore, a natural infinite dimensional generalisation of principal components was found and it makes perfect sense to estimate PCA for a finite dimensional samples of our infinite dimensional setting. The [Chamberlain & Rothschild 1983] prove the following characterisation theorem:

[2] Note that we use the same notation for the norm on l^2 as well as the seminorm on $\overline{\mathcal{H}}$

Theorem (**Characterisation of Approximate Factor Structure**). *Suppose*

CR1 $\sup_{N\in\mathbb{N}} \mu_r^N = \infty$

CR2 μ_{r+1}^N *is uniformly bounded in* N *and*

CR3 $\mu_\infty > 0$

hold. Define $\Lambda^{N,M}$ *as the* $N \times r$ *matrix whose* j*th column contains the first* N *elements of the vector* $(\mu_j^M)^{\frac{1}{2}}\gamma_j^M$. *Then*

(i) $\{\Gamma^{y,N}\}_{N\in\mathbb{N}}$ *has an approximate* r*-factor structure. Conversely, if the approximate* r*-factor structure exists then CR1 and CR2 hold.*

(ii) For any N,

$$\lim_{M\to\infty} \Lambda^{N,M}\Lambda^{N,M\prime} = \Lambda^N\Lambda^{N\prime}$$

(iii) The r*-factor structure is unique and 5.5 is identified.*

Remarks

- Every exact factor model suffices the conditions of an approximate factor structure. Therefore, the proposed model generalises the classic factor model.
- As it has been stated in Section 2, the maximum likelihood estimator of Λ in (2.2) can be obtained as the first r eigenvectors of the sample covariance matrix $\hat{\Gamma}^y$ relative to $\hat{\Gamma}^u$ given the maximum likelihood estimator Γ^u.
- In addition, it is shown that the PCA estimations relative to arbitrary positive definite weighting matrices are asymptotically equivalent for $N \to \infty$. Therefore, pca estimation, classic factor estimation and approximative r-factor structure estimation are asymptotically equivalent as the number of variables tends towards infinity.

The paper gave a crucial insight to the structure of r factor models. If the number of variables increase without bounds then the first r eigenvalues will diverge and the last $N - r$ will stay bounded. Given a set of data, the number of static factors is determined by this boundary value. This fact will be used in the following section where the estimation of the number of static factors is addressed.

5.2. The Generalised Dynamic Factor Model by Forni and Lippi

This section reviews the *Generalised Dynamic Factor Model*, which further generalises the previous models. Here, the factors are called *dynamic factors* because the loadings are no constant matrices any more but polynomial in the lag operator. the model has been widely discussed in [Lippi et. al 2000, Forni & Lippi 2001]. Before the model can

be presented we need to extend the concept of averaging sequences to the dynamic case.

Dynamic averaging Sequences

Recall the concept of averaging sequences which have been used in the approximate factor model by Chamberlain and Rothschild. By cancelling out the idiosyncratic component averaging over the cross section the factors are obtained via linear and static averaging sequences. For the idiosyncratic term we have (5.2). Sequences of linear combinations whose norm asymptotically converges to zero define an r dimensional factor space whose elements, the factors, span the common component. Here, the averaging sequences are in fact filters but the rest of the concept remains the same. We call a filter *sequence* $(a^N(z))_{N\in\mathbb{N}}$ a *Dynamic Averaging Sequence* (DAS) if $\lim_N \mathrm{Var}[a^N(z)y_t^N] < \infty$ and $\int_{-\pi}^{\pi} a^N(\lambda)^H a^N(\lambda)d\lambda = 0$. This means that the DAS converges to zero in an appropriate norm. The definition might need some discussion. The transfer functions $a^N(\lambda) = a^N(e^{-i\lambda})$ are defined on the space $\Theta = [-\pi, \pi]$. The complex vector space $\mathcal{L}_2^N(\Theta, \Sigma(\lambda))$ of all vectors of this form which suffice the condition $\int_{-\pi}^{\pi} a^N(\lambda)\Sigma(\lambda)a^N(\lambda)^H d\lambda < \infty$ form a Hilbert space with the inner product

$$\left\langle a^N, b^N \right\rangle \equiv \int_{-\pi}^{\pi} a^N(\lambda)\Sigma(\lambda)b^N(\lambda)^H d\lambda \tag{5.6}$$

The norm induced by (5.6) shall be denoted $\|\cdot\|_\Sigma$. Now any DAS suffices $a^N \in \mathcal{L}_2^N(\Theta, \mathcal{I}) \cap \mathcal{L}_2^N(\Theta, \Sigma^{y,N})$, $N \in \mathbb{N}$ and $\lim_{N\to\infty} \|a^N\|_\mathcal{I} = 0$. The dynamic factors define a q dimensional subspace which span the common component via linear dynamic relations. Dynamic averaging sequences are the key tool in separating the common from the idiosyncratic component, respectively. We call a stationary process u_t *weakly dependent* if $\lim a^N(z)u_t^N = 0$ in mean square sense for any DAS. We call $v_t \in \mathcal{H}_y$ an *aggregate* if there exists a DAS such that $\lim a^N(z)y_t^N = v_t$. The space spanned by the set of all aggregates of y_t is called the *Aggregation Space*.

The Model Framework

I now present the model framework for the generalised dynamic factor model as it has been presented in [Lippi et. al 2000, Forni & Lippi 2001]. A stochastic process y_t has a representation as a *Generalised Dynamic Factor Model* (GDFM) if there exist a q dimensional white noise process ϵ_t, a weakly dependent zero mean and stationary process u_t, mutually orthogonal to ϵ_t, whos covariances are absolutely summable and the spectral density exists, and square summable, one sided filters ω_{ik} such that

$$y_{it} = \sum_{k=1}^{q} \omega_{ik}(z)\epsilon_{kt} + u_{it} \tag{5.7}$$

$$\tag{5.8}$$

and the following restrictions hold:

GDFM1 $\mu_1^{u,N}(\lambda)$ is uniformly bounded in N and λ

GDFM2 $\mu_q^{\chi,N}(\lambda) \to \infty$ a.e. in $[-\pi, \pi]$

One of the most important results in the literature on factor models was a representation theorem given in [Forni & Lippi 2001]. It stated that an infinite dimensional stochastic process y_t whose spectral densities $\Sigma^{y,N}(\lambda)$ exist for every N has a representation as a generalised dynamic factor model if and only if

FL1 $\mu_q^{y,N}(\lambda) \to \infty$ a.e. in $[-\pi, \pi]$ and

FL2 $\mu_{q+1}^{y,N}(\lambda)$ is essentially bounded in N.

One direction is clear because it follows directly from Weyl's inequality. The other direction can be found in [Forni & Lippi 2001].

Identification

The common and idiosyncratic component are asymptotically identified. The first proof is in [Forni & Lippi 2001]. The asymptotic is with respect to $N, T \to \infty$. In [Filler 2010] the it is shown that the common component is the projection of the variables onto the aggregation space which gives a geometric insight on the identification property.

Estimation

Recall principal components are consistent estimators for the factors. In the dynamic setting the space spanned by the first q dynamic principal components converges to the space spanned by the dynamic shocks. This result has also been stated in [Lippi et. al 2000]. The spectral densities Σ^{χ} and Σ^{u} are consistently estimated as the first q and last $N - q$ terms in (4.11), respectively and let $P(z)$ denote the filter corresponding to (4.12) Then

$$\hat{\chi}_t = P(z)y_t \tag{5.9}$$

gives a N, T consistent estimate of the common component. This result is subject to the filter $P(z)$ to be known from the true spectral density Σ^{y}. The result holds if

the spectral density is estimated consistently, see [Lippi et. al 2000].

Remarks

- One sided filters are not a necessary for the identification of the model. All results are valid without this assumption. It mainly makes the interpretation of the model reasonable, since the main contribution of the literature was in macro economy, where causal and one way relationships are most plausible to assume.

6. Estimation

In the previous chapters we discussed the identification of the common component that is identified as the projection onto the aggregation space, i.e. the factors are identified via averaging sequences and the common component is the projection of the variables onto the space spanned by these factors. This works in theory because the idiosyncratic term has bounded variance and is cancelled out via the averaging sequences asymptotically for $N \to \infty$.

In this section we review some estimation procedures for determining the factors, the loadings as well as the integers r and q. Factor estimation in general requires the factor dimension to be given. Let's assume that the integers r and q are known, for now.

6.1. Filtering out the Common Component

As it has been stated, this work does not primarily deal with the estimation of the factor space or the latent variables. However, some techniques are important because the information criteria for determining the number of shocks depend on estimation performance of the factors. Therefore, the most important estimators, the PCA estimator and modified, generalised PCA estimators are presented briefly in this section.

6.1.1. The PCA estimator

In Section 4 we noted that the PCA estimator gives a good first guess to estimate the common component in factor models. Here, the factors are linear combinations of the variables with maximum variance. We estimate the factors as

$$\hat{z}_t = \begin{bmatrix} \hat{\gamma}_1' \\ \cdot \\ \hat{\gamma}_r' \end{bmatrix} y_t \quad \text{and} \quad \hat{\Lambda} = \begin{bmatrix} \hat{\gamma}_1 & \cdot & \hat{\gamma}_r \end{bmatrix} \tag{6.1}$$

In some cases it makes sense to scale the factors and the loadings

$$\hat{z}_t = \begin{bmatrix} \hat{\mu}_1^{-\frac{1}{2}} & & \\ & \cdot & \\ & & \hat{\mu}_r^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \hat{\gamma}_1' \\ \cdot \\ \hat{\gamma}_r' \end{bmatrix} y_t \quad \text{and} \quad \hat{\Lambda} = \begin{bmatrix} \hat{\gamma}_1 & \cdot & \hat{\gamma}_r \end{bmatrix} \begin{bmatrix} \hat{\mu}_1^{\frac{1}{2}} & & \\ & \cdot & \\ & & \hat{\mu}_r^{\frac{1}{2}} \end{bmatrix} \tag{6.2}$$

such that $\hat{\Gamma}^z = \mathcal{I}$ and $\hat{\Gamma}^\chi = \hat{\Lambda}\hat{\Lambda}'$. The PCA estimator is presumably the most common estimator for large data set problems. It was shown in [Chamberlain & Rothschild 1983] that the PCA estimator relative to *any positive definite weighting matrix* provides consistent estimates for the loadings of approximate factor structures. The h-step prediction performance of the pca estimator has been studied in [Stock & Watson 2002]. Again, under suitable conditions, the PCA estimator provides N, T consistent estimates for factors and the h-step factor prediction. In their setting, consistency results hold under weak cross sectionally dependent errors. Moreover, the estimator is robust with respect to small and idiosyncratic shifts in the loadings. [Doz et. al. 2011] use Kalman filter techniques to improve the factor estimates in the context of *nowcasting* in the presence of quarterly available data. They use a similar approximate factor structure for large panel data. In a first step the factors are estimated via the PC estimator. Next, the VAR parameter coefficients in (3.13) are estimated treating the estimations as the real factors via OLS methods. The second step of their estimation procedure updates the PCA factors by one run of the Kalman smoother. They prove consistency of the estimation performance and demonstrate via a simulation study that the estimator outperforms standard PCA techniques. The performance criterion was the mean square error of the projection of the true factors onto the space spanned by the factors obtained before and after the Kalman smoother updates. The simulation study presented here will use similar updating methods. Note however that we cannot estimate a standard OLS regression model for the VAR coefficients because in our setting we model a singular AR model where standard techniques are not feasible. After appropriate VAR coefficient estimates presented in Section 6.2 we do perform similar updating methods for the factors however. We introduce the Kalman filter, the Kalman smoother as well as an EM algorithm in Section 6.2.2. Here, the performance of the updating methods is measured with respect to performance regarding *consistency of q, not the factors*! We also consider a different *first step* estimator in the two-step-procedure that was proposed in [Lippi et. al 2005] and will be presented in the next section.

6.1.2. The One-Sided estimator by Forni and Lippi

For the dynamic case, [Forni & Lippi 2001] provided the basis for estimation of the *generalised dynamic factor model*. They show that the projection of y_{it} on all leads and lags of the q first dynamic principal components of y_t converges to χ_{it} in the mean squared sense. Dynamic principal components of the estimated spectral density then provide N, T consistent estimators for the factors. The problem with this procedure is that it is in general useless for prediction since the filters corresponding to

dynamic principal components tend to be two sided. The solution was then a compromise between relative PCA estimation and dynamic principal component methods. The method was proposed in [Lippi et. al 2005]. Instead of estimating the factors directly the method separates the sample covariance matrix into the common and the idiosyncratic part first. The covariance matrices as the inverse Laplace transform of the spectral density estimates

$$\hat{\Gamma}_k^{\chi} = \int_{-\pi}^{\pi} e^{ik\theta} \hat{\Sigma}^{\chi}(\theta) d\theta \quad \text{and} \quad \hat{\Gamma}_k^{u} = \int_{-\pi}^{\pi} e^{ik\theta} \hat{\Sigma}^{u}(\theta) d\theta \tag{6.3}$$

Second, the static factors are estimated as the principal components of $\hat{\Gamma}_k^{\chi}$ relative to $\hat{\Gamma}_k^{u}$. The paper [Lippi et. al 2000] shows consistency of the estimator for $N, T \to \infty$.

We now have two different estimates for the static factors. In general the estimator by Forni and Lippi provides better results for consistent estimation of the factors. Here I want to examine the performance when we try to estimate q. The first approach suggests to measure q as the rank of the innovation matrix of a VAR model of the factors. As it has been already stated we have to assume the AR model to be singular and this fact calls for individual treatment.

6.2. Singular Autoregressive Models

We commence from an $AR(p)$ system of the factors

$$a(z) z_t = \zeta_t \tag{6.4}$$

where $a(z) = \mathcal{I} - a_1 z - \cdots - a_p z^p, \quad a_j \in \mathbb{R}^{r \times r}$. Generally one assumes ζ_t to be a zero mean stationary white noise process. It is called the *innovation process*. If the innovation process is orthogonal to the past factors, i.e. $\mathbb{E} z_t \zeta_s = 0$ for $t < s$, and the system variables as well as the factors are known, then the innovation process is the one step prediction error. We will always assume the AR system to be stable as in (3.10).

6.2.1. Singularity in the generic case

In this section we need to distinguish between two kinds of solutions to AR systems. The first, henceforth referred to as the *process solution* is a stationary process z_t satisfying (6.4) for given $a(z)$ and ζ_t. The term *system solution* denotes AR coefficients $a(z)$ that satisfy the Yule Walker equations

$$\begin{bmatrix} a_1, & \cdot, & a_p \end{bmatrix} \underbrace{\begin{bmatrix} \Gamma^z & \Gamma_1^z & \cdot & \Gamma_{p-1}^z \\ \Gamma_1^z & \cdot & & \cdot \\ & & \cdot & \\ \Gamma_{p-1}^z & \cdot & \Gamma_1^z & \Gamma^z \end{bmatrix}}_{\Gamma^p} = \begin{bmatrix} \Gamma_1^z, & \cdot, & \Gamma_p^z \end{bmatrix} \tag{6.5}$$

for given Γ_p^z. A regular AR system $a(z)$ has a stationary process solution iff the system is stable. In this case the solution is unique and linearly regular, for a discussion see [Deistler et. al. 2011]. System solutions can be obtained via the Yule Walker equations (6.5) with estimates for Γ_j^z, they correspond to the Gauss Markov coefficients corresponding to OLS regression of the vector z_t onto its past $z_{t-1}, \cdot, z_{t-p}$ and are given by

$$\begin{bmatrix} \hat{a}_1, & \cdot, & \hat{a}_p \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_1^z, & \cdot, & \hat{\Gamma}_p^z \end{bmatrix} [\hat{\Gamma}^p]^{-1} \tag{6.6}$$

The system solution is unique iff the matrix $\hat{\Gamma}^p$ is non singular. It may be singular if the innovation process is singular, i.e. a transformation of a lower dimensional process of dynamic shocks

$$\zeta_t = b\epsilon_t \quad b \in \mathbb{R}^{r \times q}$$

An AR system with a singular innovation process is called a *singular AR system*. A singular AR system has a causal stationary solution iff the transfer function has a stable left coprime representation [Deistler et. al. 2011]. Most importantly, the Yule Walker equations yield no unique solution for the parameters. The corresponding transfer function is unique though, in the sense that, for every two solutions $a(z)$, $\tilde{a}(z)$ the transfer functions $a^{-1}(z)b$, $\tilde{a}^{-1}(z)b$ coincide [Deistler et. al. 2011]. A representative solution can be obtained by the Moore Penrose Pseudoinverse, that is,

$$\begin{bmatrix} \hat{a}_1, & \cdot, & \hat{a}_p \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_1^z, & \cdot, & \hat{\Gamma}_p^z \end{bmatrix} [\hat{\Gamma}^p]^{\#} \tag{6.7}$$

where $[\hat{\Gamma}^p]^{\#}$ denotes the Pseudoinverse. If the factor process $\hat{z}_t$ has a stable AR representation the minimum norm solution 6.7 yields a stable AR polynomial $[\hat{a}_1, \ldots, \hat{a}_p]$ [Filler 2010]. Another canonical representative has been proposed in [Deistler et. al. 2011]. Here, the r equations (6.5) are reduced to the system of equations corresponding to the first linearly independent rows of Γ^p. This corresponds to selecting a basis $\hat{z}_{t-j}^{(i)}$ spanned by the vector $[\hat{z}_{t-1}', \ldots, \hat{z}_{t-p}']$ in the Hilbert space of square integrable one dimensional random variables. Note that the dimension of this basis in not necessarily equal to q. The resulting equations yield unique solutions for the rows corresponding to the basis rows in Γ^p and the rest can be set to zero. The resulting system $\overline{a}(z)$ is stable and left coprime iff the Yule Walker equations contain a stable system solution.

Now we have some factor estimates and a stable AR polynomial we can apply Kalman filter techniques to improve factor estimation. A key question of this work is if the Kalman filter may improve estimation procedures that estimate the number of factors, r, and the number of shocks, q.

6.2.2. The Kalman Filter

The Kalman filter is one of the most popular state of the art procedures to improve the measurement of a process if the underlying system is known. We have the state space for the static factor model (suppressing the hat superscript for estimates now):

$$\underbrace{\begin{bmatrix} z_t \\ \cdot \\ z_{t-p+1} \end{bmatrix}}_{x_t} = \underbrace{\begin{bmatrix} a_1 & a_2 & \cdot & a_{p-1} & a_p \\ \mathcal{I} & & \cdot & \cdot & 0 \\ & \mathcal{I} & & & \cdot \\ \cdot & & \ddots & & \\ 0 & \cdot & 0 & \mathcal{I} & 0 \end{bmatrix}}_{F} \begin{bmatrix} z_{t-1} \\ \cdot \\ z_{t-p} \end{bmatrix} + \underbrace{\begin{bmatrix} b0 \end{bmatrix}}_{G} \epsilon_t \tag{6.8}$$

$$\chi_t = \begin{bmatrix} \Lambda & 0 \end{bmatrix} \begin{bmatrix} z_t \\ \cdot \\ z_{t-p} \end{bmatrix} \tag{6.9}$$

$$y_t = \underbrace{\begin{bmatrix} \Lambda & 0 \end{bmatrix}}_{H} \begin{bmatrix} z_t \\ \cdot \\ z_{t-p} \end{bmatrix} + u_t \tag{6.10}$$

The Kalman filter calculates the projection of the factors onto the observed variables up to time t given the state dynamics and some initial estimate for the factors. The Kalman smoother projects onto observed variables at every time point.

The Kalman Filter. *Let $\hat{x}_{t|s}$ denote the least squares approximation of x_t given the observed information $\{y_j | j \leq s\}$ and $e_t = y_t - H\hat{x}_{t|t-1}$ the resulting prediction error in y_t. Denote the one step error covariance in x_t by $P(t) = \mathbb{E}[x_t - \hat{x}_{t|t-1}][x_t - \hat{x}_{t|t-1}]'$, the one step error covariance in y_t by $\Sigma_e(t) = \mathbb{E}[e_t][e_t]'$ and define the Kalman filter $K(t) = FP(t)H'\Sigma_e(t)^{-1}$. Then*

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P(t)H'\Sigma_e(t)^{-1}e_t \tag{6.11}$$

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t} = F\hat{x}_{t|t-1} + K(t)e_t \tag{6.12}$$

$$\hat{x}_{t+n|t} = F\hat{x}_{t+n-1|t} \quad n > 1 \tag{6.13}$$

$$\Sigma_e(t) = HP(t)H' + \Sigma_\eta \tag{6.14}$$

$$P(t+1) = FP(t)F' + \underbrace{\Sigma_\zeta}_{GG'} - K(t)\Sigma_e(t)K(t)' \tag{6.15}$$

a proof can be found e.g. in [Deistler & Hannan 1988]. The simulation study used a Kalman filter Matlab Toolbox from
http://www.cs.ubc.ca/ murphyk/Software/Kalman/kalman.html that provides the filter and smoother as well as an EM Algorithm that calculates the maximum likelihood estimators for the loadings under normal noise assumptions, see the web page for details.

6.3. Estimation of r

As the model of Chamberlain and Rothschild has shown, in the approximate r factor setting the number of eigenvalues of Γ^y diverging equals the number of static factors. Therefore it is obvious to estimate r as the number of eigenvalues of $\hat{\Gamma}^y$ which exceed some boundary value. In Figure 1 the distribution of the first eigenvalues for a simulated factor model, which will be presented in the next section, are plotted for different values of N and T. Every box represents specific values of the dimension of the data and shows box plots of the distribution of the first eigenvalues. Obviously the number of static factors for this model equals two. In comparison Figure 2 shows the same plot of eigenvalues for a different model. The top right corner represents the case when $N = 100$ and $T = 250$ and gives a similar picture than the latter. Here, one can guess the number of static factors correctly to equal four. However, if the eigenvalues are calculated for decreasing $T = (25, 100, 250)$ (move left) or decreasing $N = (25, 100)$ (move down) we find cases where it is not obvious at all how many eigenvalues diverge and how many stay bounded. Moreover, this picture shows the necessity of N and T increasing simultaneously as large N or T values (top left and bottom right corners, respectively) do not exhibit the desired asymptotic behaviour compared to to moving up and right, corresponding to increasing both, N and T.

Weyl's inequality gives a first insight to the boundary value of the non-diverging eigenvalues. Given the structure (1.2) we have

$$\begin{aligned} \mu_l^\chi + \mu_N^u \leq \mu_l^y \leq \mu_l^\chi + \mu_1^u \qquad & l = 1 \dots r \\ \mu_N^u \leq \mu_l^y \leq \mu_1^u \qquad & l = r+1 \dots N \end{aligned}$$

We see that the $r + 1^{\text{st}}$ eigenvalue of y_t is bounded by the first eigenvalue of the idiosyncratic covariance. Existing literature dealt with the asymptotic behaviour of the eigenvalues of the noise covariance [Geman 1980, Yin et. al. 1988] by analysing the *spectral distribution*

$$F^A(x) = \frac{1}{N} \#\{l | \quad \mu_l^A \leq x\}$$

[Geman 1980] showed that for zero mean iid random variables u with variance σ^2 and finite moments the first eigenvalue of the noise covariance converges to $\mu_1^u \to (1 + \sqrt{c})^2 \cdot \sigma^2$ if

$$N, T \to \infty \quad \text{and} \quad \frac{N}{T} \to c$$

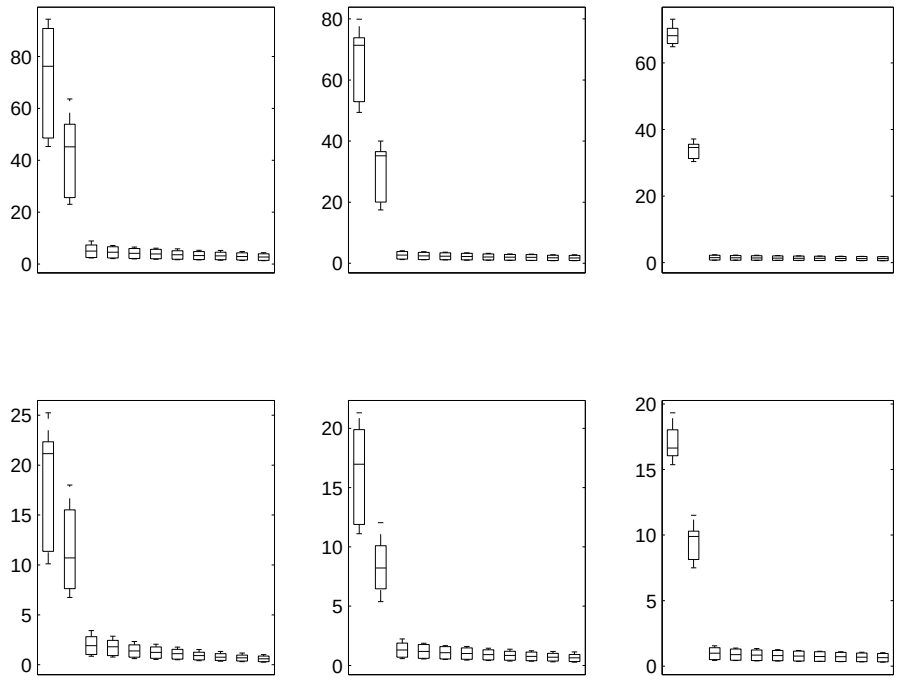


Figure 1: Eigenvalues for model 1 for increasing T horizontally $(25, 100, 250)$ and increasing N vertically $(25, 100)$

[Yin et. al. 1988] proved this statement by replacing the finite moment restriction by the weaker condition $\mathbb{E}u^4 < \infty$. [Onatski 2010] shows that F^{Γ^y} and F^{Γ^u} both weakly converge to the same spectral distribution with bounded support a.e. Moreover, for any fixed integer $i < r$, μ_i^y converges to the upper boundary of the support of this spectral distribution, μ_1^u. This finding motivates the estimator

$$\hat{r} = \#\{l| \quad \mu_l^y > (1+\delta)\hat{\mu}_{r_{\max}+1}^y\}$$

after choosing a suitable $r_{\max}$.

Information Criteria Approach

AIC and BIC criteria are inappropriate since the penalties in general depend on the number of sample, T, only. In this setting this is not feasible any more. The convergence of the eigenvalues depend on both, N and T. For large N, c and therefore μ_1^u will become large and standard information criteria are not consistent. Probably the most influential information criteria approach in determining the number of static factors was dome in [Bai & Ng 2002]. The trade off between goodness of fit criteria and a low dimensional factor space had to be generalised to the framework of large panel

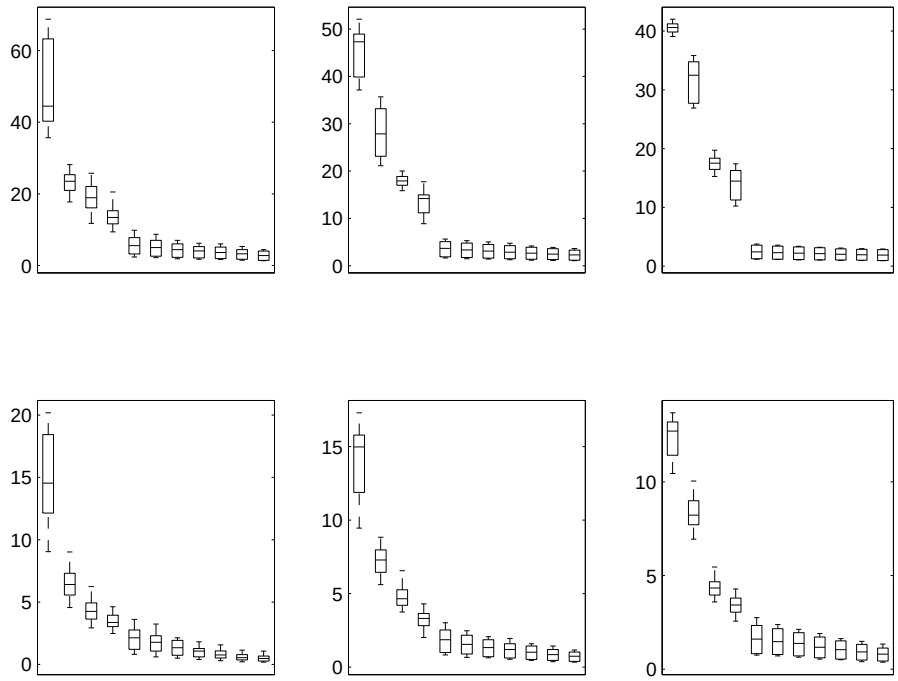


Figure 2: Eigenvalues for model 4 for increasing T horizontally $(25, 100, 250)$ and increasing N vertically $(25, 100)$

data. Denote σ^2 the average squared idiosyncratic variance over the cross section, $C = \min(N, T)$ and

$$V(k, \hat{z}_t) = \min_{\lambda_i \in \mathbb{R}^k} \frac{1}{NT} \sum_i \sum_t (y_{it} - \lambda_i{}' \hat{z}_t^k)$$

the average goodness of fit value over the cross dimension if k factors are chosen for a given $r_{\max}$ dimensional estimate $\hat{z}_t$. The authors of [Bai & Ng 2002] proposed the criteria

$$P(k) = V(k, \hat{z}_t) + \sigma^2 k g(N, T) \tag{6.16}$$

$$I(k) = \log(V(k, \hat{z}_t)) + k g(N, T) \tag{6.17}$$

where the penalty term $g(N, T)$ takes the forms $((N + T)/NT) \log (NT/(N + T))$, $((N + T)/NT) \log C^2$ or $\log(C^2)/C^2$. In align with AIC/BIC criteria, the estimator chooses the number of factors as the integer that minimises (6.16) or (6.17). If too many factors are chosen then the penalty term should surmount the additional decrease of the mean error variance. Conversely, too few factors will lead the errors to be large in the first place. In both cases the information criteria should be at its lowest for the

true number of factors. By contrast with the AIC and the BIC criteria the penalty function now depends on N as well. Note that the penalty criteria are symmetric in N and T. The paper proves consistent estimation of the number of static factors via the information criteria approach even under weak heteroscedasticity and weak time dependence of the idiosyncratic component.

BN1 $\mathbb{E}||z_t||^4 < \infty$, and $\hat{\Gamma}_0^z \to \Gamma_0^z > 0$ for $T \to \infty$

BN2 $||\lambda_i|| \leq M \leq \infty$, and $||\Lambda'\Lambda/N - D|| \to 0$ as $N \to \infty$ and $D > 0$

Theorem from [Bai & Ng 2002]. *Suppose assumptions BN1-2 hold and let $\hat{z}_t$ denote a consistent estimator of z_t. Denote*

$$\hat{r} = \arg\min_{k \leq r_{\max}} P(k) \tag{6.18}$$

Suppose (i) $g(N,T) \to 0$ and (ii) $C^2 g(N,T) \to \infty$ as $N,T \to \infty$, where $C = \min\{\sqrt{N}, \sqrt{T}\}$. Then

$$\lim_{N,T \to \infty} \mathbb{P}[\hat{r} = r] = 1$$

The full proof is given in [Bai & Ng 2002]. Let $\hat{z}_t^r$ denote the r factor estimator by the method of standard principal components, then $V(r, \hat{z}_t) = \frac{1}{N} \sum_{k=1}^{r} \hat{\mu}_k^y$. I will present a scetch for the proof if we consider the PCA estimator.

Proof. (Sketch) It is shown that

$$\lim_{N,T \to \infty} \mathbb{P}[P(k) < P(r)] = 0 \text{ for all } \; k \neq r \text{ and } k < r_{\max}$$

where

$$P(k) < P(r) \Leftrightarrow V(k) - V(r) < (r-k) g(N,T)$$

First, consider the PCA estimator for the factors. Then $V(k)$ equals $\frac{1}{N} \sum_{j>k} \hat{\mu}_j^N$

- $k < r$: $V(k) - V(r) = \frac{1}{N} \sum_{k<j \leq r} \hat{\lambda}_j^N$ has a positive limit and since $g(N,T) \to 0$ the result holds.
- $k > r$: $V(k) - V(r) = -\frac{1}{N} \sum_{r<j \leq k} \hat{\lambda}_j^N$ which converges to zero at rate $O_p(C^2)$, which is faster than the rate of convergence of $g(N,T)$.

□

Remarks

- (6.17) is sometimes preferred over (6.16) because it is independent of the unknown noise variance. In the simulation study this independence result will be tested on some data.

- If the penalties in (6.16), (6.17) are specified correctly then $r = \hat{r}$ if we define $\hat{r}$ as in (6.18). Conversely, if we know the correct model order, for every set of error measures $\{V(k)|k \leq r_{\max}\}$ that are decreasing in k, we can find lower and upper boundaries, $\underline{g}(N,T)$ and $\overline{g}(N,T)$, respectively, such that $r = \hat{r}$. In the case of the PCA estimator we find

$$\begin{aligned} P(k) < P(k+1) \quad &\Rightarrow \quad P(k) < P(k+j) \quad j \geq 1 \\ P(k) < P(k-1) \quad &\Rightarrow \quad P(k) < P(k-j) \quad j \geq 1 \end{aligned}$$

This follows immediately from

$$V(k) - V(k+j) = \frac{1}{N} \sum_{l=1}^{j} \hat{\mu}_{k+l}$$

$$V(k) - V(k-j) = -\frac{1}{N} \sum_{l=1}^{j} \hat{\mu}_{k-l}$$

and

$$\sum_{l=1}^{j} \hat{\mu}_{k+l} < \frac{j}{N} \hat{\mu}_{k+1} < j g(N,T)$$

Since the eigenvalues $\hat{\mu}_j$ are decreasing as a function of j the error measure $V(j)$ is convex in j. This allows us to determine feasible boundaries for the penalties by considering $V(r-1)$, $V(r)$ and $V(r+1)$ only. We have

$$\overline{g}(N,T) = \frac{1}{N} \hat{\mu}_{k-1} \tag{6.19}$$

$$\underline{g}(N,T) = \frac{1}{N} \hat{\mu}_{k+1} \tag{6.20}$$

In the simulation study we save computing $V(k)$ for all other values of $1 \leq k \leq r_{\max}$ which saves a huge amount of time for estimates calculated by the Kalman Filter and the EM algorithm.

6.4. Estimation of q

There are several ways in determining q, the number of primitive shocks, in the generalised dynamic factor model. It has already been stated that we focus on two approaches. First, we can commence from the static representation (3.3) from where we know that the static factors admit a singular AR- decomposition where the innovations are rank q. Second, we estimate q as the rank of some spectral density estimator.

The Estimator by Bai and Ng

One approach is to estimate the static factors z_t, estimate a (singular) AR model and estimate q as the estimated rank of the covariance matrix of the residuals $\hat{\Sigma}_\zeta$. This approach was discussed in [Bai & Ng 2007]. This paper used the spectral decomposition of the innovation covariance

$$\Gamma^\zeta = \sum_{j=1}^{r} \mu_j \gamma_j \gamma_j' = \sum_{j=1}^{q} \mu_j \gamma_j \gamma_j'$$

Inference on the number of shocks is derived from the fact that the last $r - q$ eigenvalues are exactly zero. Similarly, the eigenvalues of the estimated covariance will be close to zero. We define $\Gamma^\zeta(k) = \sum_{j=1}^{k} c_j \gamma_j \gamma_j'$ and, similarly, $\hat{\Gamma}^\zeta(k) = \sum_{j=1}^{k} \hat{\mu}_j \hat{\gamma}_j \hat{\gamma}_j'$ then

$$\|\Gamma^\zeta - \Gamma^\zeta(k)\| = \| \sum_{j=k+1}^{r} \mu_j \gamma_j \gamma_j' \| = 0 \quad \forall k \geq q \tag{6.21}$$

for any norm $\|\cdot\|$, e.g. the Frobenius norm $\|\cdot\|_F$. In practice the estimated eigenvalues and therefore (6.21) will be small, but not exactly zero. So in practice we will choose the number of primitive shocks as the minimal k such that $\hat{\Gamma}_\zeta(k)$ comes *close enough* to $\hat{\Gamma}_\zeta$ that is, we choose

$$\hat{q} = \underset{1 \leq k \leq q_{max}}{\arg\min} \{k : \|\hat{\Gamma}^\zeta - \hat{\Gamma}^\zeta(k)\|_F > C\} \tag{6.22}$$

subject to a boundary value $C > 0$. The performance of the estimator depends heavily on the rate of convergence of $\hat{\Sigma}^\zeta$ to Γ^ζ, or, more precisely to $H\Gamma^\zeta H'$ for some rotation matrix. [Bai & Ng 2007] show that $C = 0.5/\min(T^{2/5}, N^{2/5})$ is a good boundary value for factor models of large panel data.

Remarks

- If we choose the norm $\|\cdot\| = \mathrm{tr}\cdot = \sum \mu_j$ then

$$\|\hat{\Gamma}^\zeta - \hat{\Gamma}^\zeta(k)\| = \sum_{j=k+1}^{r} \hat{\mu}_j$$

  and therefore

$$\overline{C}(N,T) = \sum_{j=q}^{r} \hat{\mu}_j \tag{6.23}$$

$$\underline{C}(N,T) = \sum_{j=q+2}^{r} \hat{\mu}_j \tag{6.24}$$

The Information Criterion Approach by Hallin and Liska

A second approach was discussed in [Hallin & Liška 2007]. We know from [Forni & Lippi 2001] and FL1 that the first q dynamic eigenvalues of the spectral density matrix $\Sigma^{y,N}$ will diverge to infinity as $N \to \infty$, and the rest will be essentially bounded. If the spectral density is known this essential bound is subject to N only. If we define

$$\hat{q}^N = \arg\min P(k) \quad \text{where} \quad P(k) = \frac{1}{N} \sum_{j>k} \int_0^{2\pi} \mu_j^{y,N}(\lambda) d\lambda + kp(N)$$

for some penalty $p(N)$ such that

$$\lim p(N) = 0 \quad \text{and} \quad \lim Np(N) = \infty$$

holds then $\lim_N \mathbb{P}[\hat{q}^N = q] = 1$. The proof is the analogue to the static version presented in Section 6.3. The estimator can be shown consistent for the generalised dynamic factor model in Section 5.2 under some technical assumptions. These assumptions include *linear divergence* of the dynamic eigenvalues in N, that is, there exist constants c_i^-, c_i^+ such that

$$c_i^- \leq \liminf \mu_i^N(\lambda)/N \leq \limsup \mu_i^N(\lambda)/N \leq c_i^+ \quad a.e., i \leq q$$

and $c_i^+ < c_{i-1}^-$ and the non diverging eigenvalues are *bounded away from zero*, i.e. $\mu_j^{N,y}(\lambda) > c_\mu$ *a.e.* for $j > q$. A sketch of the proof is very similar to the static version:

$$\lim_N \mathbb{P}[P(k) < P(q)] = 0 \text{ for all } k \neq q \text{ and } k < q_{\max}$$

For $k > q$ we have

$$N[P(k) - P(q)] = -\sum_{j=q+1}^{k} \int_0^{2\pi} \mu_j^{y,N}(\lambda) d\lambda + \underbrace{N[k-q]p(N)}_{\to\infty} > 0 \text{ for } N \text{ large}$$

since the last term diverges and the first one is bounded by assumption. Conversely, for $k < q$

$$P(k) - P(q) = \sum_{j=k+1}^{q} \int_0^{2\pi} \mu_j^{y,N}(\lambda) d\lambda + \underbrace{[q-k]p(N)}_{\to 0} > 0 \text{ for } N \text{ large}$$

holds since by assumption the fist term is bounded away from zero. On the sample level, after spectral density estimation, the estimator is still consistent but the penalty calls for some modification. Periodogram smoothing techniques are summarised in the Appendix A. Assumptions on the weighting kernels of the density estimator assure the consistency result

$$\sup_N \max_{i,j} \sup_\lambda \left[\mathbb{E}|\Sigma^N(\lambda) - \hat{\Sigma}^{N,T}(\lambda)|_{ij}^2 \right] \leq L_1 M_T T^{-1} + L_2 M_T^{-4}$$

On the sample level the information criterion becomes

$$P(k) = \frac{1}{N} \sum_{i=k+1}^{N} \frac{1}{T-1} \sum_{l=1}^{T-1} \hat{\mu}_j^N(\lambda_l) + kp(N,T) \tag{6.25}$$

$$I(k) = \log \left[\frac{1}{N} \sum_{i=k+1}^{N} \frac{1}{T-1} \sum_{l=1}^{T-1} \hat{\mu}_j^N(\lambda_l) \right] + kp(N,T) \tag{6.26}$$

where $\lambda_l = 2\pi l/T/$. The sample estimators $\hat{q} = \arg\min P(k)$ and $\hat{q} = \arg\min I(k)$ can be shown consistent for q if the penalty in (6.25) and (6.25) suffices

$$p(N,T) \to 0 \quad \text{and} \quad \min \left[N, B_T^{-2}, B_T^{\frac{1}{2}} T^{\frac{1}{2}} \right] p(N,T) \to \infty$$

The paper [Hallin & Liška 2007] proposed the penalty functions

$$p(N,T) = [M_T^{-2} + M_T^{1/2} T^{-1/2} + N^{-1}] \log \min(N, M_T^2, M_T^{-1/2} T^{1/2}) \tag{6.27}$$

$$p(N,T) = \min(N, M_T^2, M_T^{-1/2} T^{1/2})^{-1/2} \tag{6.28}$$

$$p(N,T) = [M_T^{-2} + M_T^{1/2} T^{-1/2} + N^{-1}]^{-1} \log \min(N, M_T^2, M_T^{-1/2} T^{1/2}) \tag{6.29}$$

which were used in the simulation study.

Remarks

- The criterion works for every penalty function $cp(N,T)$ since the proof relies on asymptotic properties which are fulfilled for every pre-multiplication by positive constants. To choose the right penalty function on the sample level the paper proposed to estimate $\hat{q}$ for every information criterion corresponding to some penalty $cp(N,T)$ for $c > 0$. For $c = 0$ the estimator obviously chooses $\hat{q} = q_{\max}$. If c increases, $\hat{q}$ will decrease. The actual estimate is chosen as the second estimate to be chosen over a range of values of c, that is, the second *stability region* of the estimator. In the simulation study we want to compare the second stability region with the optimal penalties.

- From the Section 5.2 we know that the projection onto the first q dynamic principal components converges to the common component. Moreover, the projection onto more dynamic principal components will converge to the previous projection in m.s. [Lippi et. al 2000]. Therefore in practice it might be more important to determine lower boundaries on q.

The Approach by Stock and Amengual

Watson and Amengual [Watson & Amengual] proposed a method to estimate q via a modified version of the estimator in Section 6.3. They commence from the singular AR representation

$$y_t = \Lambda z_t + u_t \tag{6.30}$$

$$a(z)z_t = G\eta_t \quad \text{or} \quad z_t = \sum_{i=1}^{p} a_i z_{t-i} + G\eta_t \tag{6.31}$$

Combine (6.30) and (6.31)

$$y_t = \Lambda(\sum_{i=1}^{p} a_i z_{t-i} + G\eta_t) + u_t \tag{6.32}$$

and construct a new process

$$x_t = y_t - \Lambda \sum_{i=1}^{p} a_i z_{t-i} = \Lambda G\eta_t + u_t \tag{6.33}$$

which corresponds to a static factor model with q static factors such that the estimator for the number of static factors in Section 6.3 can be applied. Although this structure motivates this estimator note that it is only feasible if $a(z)$ and Λ are unknown.The paper proves consistency if Λa_i and z_{t-i} are estimated well enough. Denote $\hat{x}_t$ the new variables when Λa_i is replaced by estimates in (6.33). The estimator remains consistent if

$$\sum_i \sum_t (x_{it} - \hat{x}_{it})^2 = \mathcal{O}_p(\max(N, T))$$

The factors and the loadings are estimated via the PCA estimator for a consistent estimator of r. The AR coefficients are estimated via OLS regression onto the past of the estimated factors, or, as in our setting, via a singular AR estimator proposed in Section 6.2. Alternatively, the AR coefficients and the loadings can be estimated simultaneously via regression from y_t onto the past of z_t.

7. Simulation Study

In the following simulation study estimators for the number of dynamic factors are compared. All estimators rely on penalty values and the asymptomatic behaviour of various (dynamic) eigenvalues for large N and T. The Bai-Ng criterion depends

on the penalty $C(N,T)$ which arises from uncertainties in the innovation covariance and therefore depends on the pre-estimation of static factors and the estimation of AR coefficients. Therefore it needs pre-estimation of the number of static factors, r, before it can end up with an estimate for q. We will consider estimation performance if r is known, chosen by the IC approach in Section 6.3 or simply chosen as a *large* value, call it $r_{\max}$. We will consider two estimators for the static factors, the PCA method and the dynamic PCA method. Note that for the dynamic method we even need an initial estimate for q to estimate the static factors, to estimate r, to estimate q. An estimate can be obtained by choosing a large initial $\hat{q} = q_{\max}$. For both estimation procedures we consider updating the factor estimates via Kalman filter techniques. The same holds for the estimator from Stock and Amengual and the corresponding penalties $g(N,T)$. The Information criteria by Hallin and Liška $p(N,T)$ does not require a pre-estimation of static factors. It depends solemnly on some spectral density estimator from which the dynamic eigenvalues can be calculated. We sampled 35 generic and stable state space systems (A,B,C,D) of state dimension $r = 4$ and input order $q = 2$,

$$x_{t+1} = Ax_t + B\epsilon_t \tag{7.1}$$
$$z_t = Cx_t + D\epsilon_t \tag{7.2}$$
$$y_t = \Lambda z_t + \sqrt{\theta}u_t \tag{7.3}$$
$$(\mathcal{I} - dz)\, u_t = v_t \tag{7.4}$$

where the entries of the loadings Λ were sampled from a $\mathcal{N}(0,1)$ distribution. Following [Doz et. al. 2011], the weak cross correlation in u_t is modeled via v_t which is taken randomly from a normal distribution with covariance $\mathcal{T}$ such that $\mathcal{T}_{ij} = \sqrt{\alpha_j \alpha_i}\tau^{|i-j|}$, $\alpha_j = \beta_j/(1-\beta_j)\sum_{l=1}^{r}\lambda_{jl}^2$ and β_j is randomly generated from the uniform distribution $\mathcal{U}[.3,.7]$ and regulates the variance of the jth idiosyncratic component. $\mathcal{T}$ is a Toeplitz matrix and the cross correlation can be chosen by the parameter τ. The other model parameters are the noise parameter θ and the time dependence parameter d. Every model constellation is excited 100 times through a white noise input sampled from the $\mathcal{N}(0,1)$ distribution and a realisation of v_t. Performance for consistent estimation of q has been investigated for $N \in \{25,50,75,100\}$, $T \in \{25,50,75,100\}$, $d \in \{0,0.3\}$, $\tau \in \{0,0.5,0.9\}$ and $\theta \in \{0.5,1,1.5\}$.

8. Results

In the following some tables will illustrate some results from the simulation study. The optimal penalties have been calculated over all 3000 runs and averaged over both models and simulation runs. Standard deviations are given in brackets. In general, the tables are to be read from the bottom left box where the average penalties are shown for $N = T = 25$ and some variation in either estimation procedure or model parameters. The boxes towards the right correspond to higher T and moving to the top corresponds to increasing N. The top left of each box shows the corresponding penalties that have been proposed in Section 6.

8.1. The Bai Ng Criteria

First, I will analyse the results for the Bai-Ng criterion.

PCA vs FHLR

Table 1 compares the shows the theoretical boundary values $C = 1/\min(N,T)^{1/2}$ in the top left corners with the optimal values for $\theta = \tau = .5$ and $d = 0$. As a first observation we find that the optimal boundary values are not symmetric in N and T. As N increases, the bounds decrease at a faster rate than for increasing T. The PCA estimator performs well here since the theoretical penalty lies betewen the average optimal. The dynamic estimator (FHLR) has trouble for $N > T$ due to singularities in the estimated covariance $\hat{\Gamma}^u$. The penalties lie below the theoretical penalties in general which can be interpreted as a better fit however. In both cases the number of static factors have been chosen high, $\hat{r} = r_{\max} = 9$.

	VAR	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$		0.200			0.141			0.115			0.100		
	PCA		0.079	0.263		0.065	0.247		0.060	0.244		0.057	0.241
			(0.002)	(0.005)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	$FHLR$		NaN	0.154		NaN	0.193		0.000	0.251		0.071	0.198
			(NaN)	(0.009)		(NaN)	(0.006)		(0.000)	(0.006)		(0.004)	(0.007)
$N = 75$		0.200			0.141			0.115			0.115		
	PCA		0.082	0.270		0.068	0.253		0.064	0.250		0.061	0.248
			(0.002)	(0.005)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	$FHLR$		NaN	0.163		NaN	0.233		0.071	0.212		0.010	0.136
			(NaN)	(0.008)		(NaN)	(0.006)		(0.004)	(0.007)		(0.001)	(0.004)
$N = 50$		0.200			0.141			0.141			0.141		
	PCA		0.080	0.269		0.070	0.253		0.066	0.247		0.064	0.244
			(0.002)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	$FHLR$		NaN	0.168		0.059	0.187		0.010	0.156		0.010	0.182
			(NaN)	(0.007)		(0.003)	(0.007)		(0.000)	(0.004)		(0.000)	(0.003)
$N = 25$		0.200			0.200			0.200			0.200		
	PCA		0.090	0.313		0.085	0.295		0.084	0.293		0.083	0.291
			(0.002)	(0.005)		(0.001)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)
	$FHLR$		0.077	0.244		0.026	0.173		0.019	0.182		0.019	0.187
			(0.003)	(0.008)		(0.001)	(0.004)		(0.001)	(0.003)		(0.000)	(0.003)
			T=25			T=50			T=75			T=100	

Table 1: Average extreme penalty values for BN72 with a variation in estimation procedure: PCA vs FHLR $q = q_{HL}, r = r_{max}$.

The PCA Estimator

Table 2 shows first results for the PCA estimator only. We see again that the estimation is better in the lower triangular. The table also shows what happens if the number of static factors are either chosen too large or via the IC criterion. In both cases minimum as well as maximum values increase slightly without further consequences, the estimation method still seems consistent.

	VAR	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$		0.200			0.141			0.115			0.100		
	r		0.020	0.221		0.017	0.215		0.016	0.216		0.015	0.214
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{max}		0.079	0.263		0.065	0.247		0.060	0.244		0.057	0.241
			(0.002)	(0.005)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{IC}		0.021	0.220		0.017	0.216		0.022	0.222		0.028	0.226
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
$N = 75$		0.200			0.141			0.115			0.115		
	r		0.022	0.226		0.019	0.218		0.018	0.216		0.018	0.213
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{max}		0.082	0.270		0.068	0.253		0.064	0.250		0.061	0.248
			(0.002)	(0.005)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{IC}		0.048	0.247		0.034	0.230		0.039	0.231		0.043	0.236
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
$N = 50$		0.200			0.141			0.141			0.141		
	r		0.024	0.224		0.021	0.215		0.020	0.212		0.019	0.209
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{max}		0.080	0.269		0.070	0.253		0.066	0.247		0.064	0.244
			(0.002)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{IC}		0.075	0.265		0.060	0.245		0.060	0.242		0.060	0.242
			(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
$N = 25$		0.200			0.200			0.200			0.200		
	r		0.034	0.266		0.031	0.257		0.030	0.254		0.030	0.252
			(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	r_{max}		0.090	0.313		0.085	0.295		0.084	0.293		0.083	0.291
			(0.002)	(0.005)		(0.001)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)
	r_{IC}		0.090	0.313		0.085	0.295		0.084	0.293		0.083	0.291
			(0.002)	(0.005)		(0.001)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)
			T=25			T=50			T=75			T=100	

Table 2: Average extreme penalty values for BN72 with a variation in the PCA estimator via changes in the number of initial static factors: 1) r 2) r_{max} 3) r_{IC}.

Table 3 shows how the update methods influence estimation of q. Again, the top left corners show the true value, but now we see the distance between the minimum and maximum value for the initial PCA estimator, the Kalman filter, the Kalman smoother and the EM algorithm updates. A wider gap means that the probability of correct estimation increases. As can be seen the update methods increase the estimation performance in most cases, note however that the EM algorithm had trouble in the cases $N > T$.

	VAR	PE	max−min	PE	max−min	PE	max−min	PE	max−min	PE	max−min
$N = 100$		0.200		0.141		0.115		0.100		0.100	
	PCA		0.184		0.182		0.184		0.184		0.187
	KA		0.187		0.186		0.189		0.189		0.191
	SM		0.190		0.187		0.188		0.186		0.184
	EM		NaN		NaN		NaN		NaN		NaN
$N = 75$		0.200		0.141		0.115		0.115		0.115	
	PCA		0.189		0.185		0.186		0.186		0.189
	KA		0.192		0.190		0.193		0.193		0.194
	SM		0.195		0.191		0.193		0.191		0.191
	EM		NaN		NaN		NaN		NaN		NaN
$N = 50$		0.200		0.141		0.141		0.141		0.141	
	PCA		0.189		0.183		0.181		0.181		0.180
	KA		0.193		0.190		0.188		0.187		0.187
	SM		0.197		0.194		0.191		0.191		0.190
	EM		NaN		NaN		NaN		NaN		NaN
$N = 25$		0.200		0.200		0.200		0.200		0.200	
	PCA		0.223		0.210		0.209		0.209		0.205
	KA		0.226		0.213		0.212		0.211		0.208
	SM		0.229		0.214		0.213		0.213		0.208
	EM		NaN		NaN		NaN		NaN		NaN
		T=25		T=50		T=75		T=100		T=250	

Table 3: Average extreme penalty values for BN72 with a variation in the updates. PCA estimator and max number of initial static factors.

The FHLR Estimator

On the contrary, and most interestingly, Table 4 shows that the update methods work for the PCA estimator only. Here the FHLR method chooses the HL estimator for an initial guess for the number of dynamic factors. The number of static factors is assumed known. A run of the Kalman filter, smoother or the EM algorithm will in general shorten the gap between minimum and maximum value, except for very large values for the case $T >> N$.

	VAR	PE	max−min	PE	max−min	PE	max−min	PE	max−min	PE	max−min
$N = 100$		0.200		0.141		0.115		0.100		0.100	
	$FHLR$		NaN		NaN		0.250		0.139		0.165
	KA		NaN		NaN		0.066		0.104		0.102
	SM		NaN		NaN		0.066		0.108		0.136
	EM		NaN		NaN		NaN		NaN		NaN
$N = 75$		0.200		0.141		0.115		0.115		0.115	
	$FHLR$		NaN		0.231		0.155		0.125		0.174
	KA		NaN		0.064		0.114		0.058		0.128
	SM		NaN		0.062		0.120		0.061		0.161
	EM		NaN		NaN		NaN		NaN		NaN
$N = 50$		0.200		0.141		0.141		0.141		0.141	
	$FHLR$		NaN		0.143		0.140		0.177		0.185
	KA		NaN		0.098		0.048		0.060		0.125
	SM		NaN		0.107		0.057		0.091		0.160
	EM		NaN		NaN		NaN		NaN		NaN
$N = 25$		0.200		0.200		0.200		0.200		0.200	
	$FHLR$		0.177		0.158		0.172		0.179		0.180
	KA		0.118		0.074		0.103		0.118		0.148
	SM		0.142		0.098		0.129		0.140		0.183
	EM		NaN		NaN		NaN		NaN		NaN
		T=25		T=50		T=75		T=100		T=250	

Table 4: Average extreme penalty values for BN72 with a variation in the updates. FHLR with $q = q_{HL}$ estimator and true number of initial static factors.

Model Parameters

Next we see the impact of model parameter variation on the Bai-Ng criterion. Table 5 shows a general increase of the boundary values for a higher correlation in the cross section. Table 6 shows the results for an increase in θ where the boundaries increase enormously.

	τ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$		0.200			0.141			0.115			0.100		
	0		0.174	0.335		0.143	0.278		0.128	0.261		0.120	0.254
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	.5		0.176	0.335		0.151	0.279		0.138	0.264		0.132	0.257
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	0.9		0.186	0.393		0.191	0.351		0.192	0.339		0.192	0.334
			(0.003)	(0.005)		(0.003)	(0.006)		(0.003)	(0.006)		(0.004)	(0.006)
$N = 75$		0.200			0.141			0.115			0.115		
	0		0.177	0.343		0.151	0.292		0.138	0.277		0.131	0.269
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	.5		0.179	0.346		0.161	0.297		0.150	0.282		0.145	0.276
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	0.9		0.179	0.414		0.189	0.379		0.191	0.369		0.192	0.364
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
$N = 50$		0.200			0.141			0.141			0.141		
	0		0.177	0.354		0.159	0.305		0.149	0.290		0.143	0.283
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	.5		0.177	0.355		0.168	0.313		0.161	0.300		0.157	0.294
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	0.9		0.157	0.434		0.168	0.412		0.169	0.404		0.168	0.398
			(0.002)	(0.005)		(0.002)	(0.004)		(0.002)	(0.004)		(0.002)	(0.004)
$N = 25$		0.200			0.200			0.200			0.200		
	0		0.196	0.421		0.193	0.383		0.189	0.371		0.185	0.364
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	.5		0.192	0.429		0.196	0.396		0.198	0.389		0.197	0.384
			(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	0.9		0.133	0.489		0.145	0.510		0.147	0.516		0.148	0.514
			(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.003)		(0.001)	(0.003)
			T=25			T=50			T=75			T=100	

Table 5: Average extreme penalty values for BN72 with a variation in tau.

	θ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$		0.200			0.141			0.115			0.100		
	0.5		0.079	0.263		0.065	0.247		0.060	0.244		0.057	0.241
			(0.002)	(0.005)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	1		0.176	0.335		0.151	0.279		0.138	0.264		0.132	0.257
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	1.5		0.266	0.445		0.250	0.381		0.232	0.352		0.223	0.338
			(0.003)	(0.005)		(0.004)	(0.006)		(0.004)	(0.006)		(0.004)	(0.006)
$N = 75$		0.200			0.141			0.115			0.115		
	0.5		0.082	0.270		0.068	0.253		0.064	0.250		0.061	0.248
			(0.002)	(0.005)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	1		0.179	0.346		0.161	0.297		0.150	0.282		0.145	0.276
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	1.5		0.268	0.461		0.263	0.407		0.251	0.383		0.244	0.370
			(0.003)	(0.005)		(0.004)	(0.005)		(0.004)	(0.006)		(0.004)	(0.006)
$N = 50$		0.200			0.141			0.141			0.141		
	0.5		0.080	0.269		0.070	0.253		0.066	0.247		0.064	0.244
			(0.002)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	1		0.177	0.355		0.168	0.313		0.161	0.300		0.157	0.294
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	1.5		0.260	0.474		0.270	0.436		0.262	0.414		0.258	0.404
			(0.003)	(0.005)		(0.003)	(0.005)		(0.004)	(0.005)		(0.004)	(0.005)
$N = 25$		0.200			0.200			0.200			0.200		
	0.5		0.090	0.313		0.085	0.295		0.084	0.293		0.083	0.291
			(0.002)	(0.005)		(0.001)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)
	1		0.192	0.429		0.196	0.396		0.198	0.389		0.197	0.384
			(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	1.5		0.259	0.532		0.293	0.527		0.296	0.517		0.297	0.512
			(0.002)	(0.005)		(0.003)	(0.004)		(0.003)	(0.004)		(0.003)	(0.004)
			T=25			T=50			T=75			T=100	

Table 6: Average extreme penalty values for BN72 with a variation in theta.

On the other hand, consider an increase in d, that is, the introduction of time dependent errors. in Table 7 the lower boundaries increase but the top bounds decrease hence making consistent estimation less likely.

	d	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$		0.200			0.141			0.115			0.100		
	0		0.079	0.263		0.065	0.247		0.060	0.244		0.057	0.241
			(0.002)	(0.005)		(0.001)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	0.3		0.153	0.316		0.132	0.267		0.123	0.253		0.119	0.247
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
$N = 75$		0.200			0.141			0.115			0.115		
	0		0.082	0.270		0.068	0.253		0.064	0.250		0.061	0.248
			(0.002)	(0.005)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	0.3		0.161	0.333		0.143	0.283		0.135	0.269		0.132	0.264
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
$N = 50$		0.200			0.141			0.141			0.141		
	0		0.080	0.269		0.070	0.253		0.066	0.247		0.064	0.244
			(0.002)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)		(0.001)	(0.004)
	0.3		0.163	0.342		0.153	0.299		0.147	0.285		0.145	0.281
			(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
$N = 25$		0.200			0.200			0.200			0.200		
	0		0.090	0.313		0.085	0.295		0.084	0.293		0.083	0.291
			(0.002)	(0.005)		(0.001)	(0.004)		(0.002)	(0.004)		(0.001)	(0.004)
	0.3		0.178	0.409		0.184	0.380		0.182	0.368		0.181	0.362
			(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
			T=25			T=50			T=75			T=100	

Table 7: Average extreme penalty values for BN72 with a variation in d.

8.2. The Hallin Liška Criterion

Second, we want to analyse the Hallin-Liška criterion. Note the criterion is independent from a factor estimator so we won't compare estimation and update procedure techniques. The following tables show three boundary values calculated from [Hallin & Liška 2007] in the top left corners. Table 8 shows the minimum and maximum mean values for the boundaries (PE) as well as the mean boundary values for the second stability region (SR). It can be seen that the stability region tends to coincide with the optimal bounds. The penalty values themselves should be taken with care. Although the second criterion sometimes lies within the correct region the penalty bounds tend to be too small.

	VAR	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	PE		0.404	0.612		0.313	0.570		0.276	0.550		0.257	0.549
			(0.002)	(0.005)		(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	SR		0.512	0.940		0.361	0.543		0.296	0.494		0.280	0.522
			(0.002)	(0.017)		(0.001)	(0.007)		(0.001)	(0.008)		(0.002)	(0.009)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	PE		0.419	0.638		0.326	0.596		0.288	0.577		0.270	0.577
			(0.002)	(0.005)		(0.002)	(0.004)		(0.003)	(0.004)		(0.003)	(0.005)
	SR		0.537	0.954		0.381	0.576		0.319	0.534		0.292	0.542
			(0.003)	(0.017)		(0.001)	(0.007)		(0.001)	(0.007)		(0.001)	(0.008)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	PE		0.442	0.673		0.347	0.632		0.307	0.612		0.292	0.611
			(0.002)	(0.005)		(0.002)	(0.004)		(0.002)	(0.004)		(0.003)	(0.005)
	SR		0.578	1.000		0.421	0.639		0.356	0.564		0.328	0.556
			(0.003)	(0.016)		(0.002)	(0.008)		(0.001)	(0.005)		(0.001)	(0.005)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	PE		0.528	0.758		0.432	0.715		0.388	0.693		0.375	0.691
			(0.002)	(0.005)		(0.003)	(0.004)		(0.003)	(0.004)		(0.003)	(0.004)
	SR		0.731	1.267		0.558	0.854		0.488	0.705		0.453	0.666
			(0.004)	(0.016)		(0.003)	(0.013)		(0.002)	(0.007)		(0.002)	(0.005)
			T=25			T=50			T=75			T=100	

Table 8: Average extreme penalty values for HL with a variation in no variations but comparison with the second stability region SR.

Model Parameters

Again, we want to analyse the influence the model parameters on the estimation. τ increases in Table 9 and it can be seen that the the boundary values in general increases as well. However, an increase in θ reduces them which can be seen in Table 10. Again, error dynamics reduces the bounds and makes the gap shorter which can be observed in Table 11.

	τ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0		0.343	0.375		0.240	0.301		0.192	0.270		0.183	0.264
			(0.000)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	.5		0.353	0.386		0.252	0.311		0.206	0.278		0.196	0.271
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	0.9		0.450	0.488		0.358	0.405		0.315	0.365		0.306	0.359
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0		0.355	0.390		0.253	0.317		0.206	0.286		0.197	0.281
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	.5		0.368	0.406		0.269	0.331		0.223	0.297		0.214	0.291
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	0.9		0.488	0.537		0.401	0.454		0.360	0.415		0.352	0.409
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0		0.380	0.420		0.280	0.345		0.234	0.313		0.224	0.307
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)
	.5		0.399	0.443		0.301	0.366		0.256	0.330		0.247	0.323
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)
	0.9		0.554	0.624		0.475	0.544		0.436	0.507		0.429	0.502
			(0.002)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0		0.448	0.498		0.351	0.417		0.304	0.381		0.294	0.375
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	.5		0.477	0.529		0.382	0.450		0.338	0.412		0.328	0.404
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	0.9		0.676	0.810		0.600	0.744		0.561	0.717		0.553	0.713
			(0.002)	(0.004)		(0.002)	(0.003)		(0.002)	(0.003)		(0.002)	(0.002)
		T=25			T=50			T=75			T=100		

Table 9: Average extreme penalty values for HL with a variation in tau.

	θ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0.5		0.404	0.612		0.313	0.570		0.276	0.550		0.257	0.549
			(0.002)	(0.005)		(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	1		0.353	0.386		0.252	0.311		0.206	0.278		0.196	0.271
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	1.5		0.347	0.353		0.246	0.261		0.198	0.221		0.188	0.214
			(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0.5		0.419	0.638		0.326	0.596		0.288	0.577		0.270	0.577
			(0.002)	(0.005)		(0.002)	(0.004)		(0.003)	(0.004)		(0.003)	(0.005)
	1		0.368	0.406		0.269	0.331		0.223	0.297		0.214	0.291
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)		(0.001)	(0.003)
	1.5		0.362	0.371		0.262	0.281		0.215	0.241		0.205	0.233
			(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0.5		0.442	0.673		0.347	0.632		0.307	0.612		0.292	0.611
			(0.002)	(0.005)		(0.002)	(0.004)		(0.002)	(0.004)		(0.003)	(0.005)
	1		0.399	0.443		0.301	0.366		0.256	0.330		0.247	0.323
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)
	1.5		0.392	0.407		0.293	0.317		0.246	0.277		0.237	0.269
			(0.001)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0.5		0.528	0.758		0.432	0.715		0.388	0.693		0.375	0.691
			(0.002)	(0.005)		(0.003)	(0.004)		(0.003)	(0.004)		(0.003)	(0.004)
	1		0.477	0.529		0.382	0.450		0.338	0.412		0.328	0.404
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	1.5		0.467	0.493		0.371	0.402		0.325	0.363		0.315	0.355
			(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
			T=25			T=50			T=75			T=100	

Table 10: Average extreme penalty values for HL with a variation in theta.

	d	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0		0.404	0.612		0.313	0.570		0.276	0.550		0.257	0.549
			(0.002)	(0.005)		(0.002)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
	0.3		0.358	0.395		0.253	0.316		0.206	0.282		0.196	0.275
			(0.001)	(0.002)		(0.000)	(0.002)		(0.001)	(0.002)		(0.001)	(0.003)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0		0.419	0.638		0.326	0.596		0.288	0.577		0.270	0.577
			(0.002)	(0.005)		(0.002)	(0.004)		(0.003)	(0.004)		(0.003)	(0.005)
	0.3		0.373	0.414		0.270	0.336		0.223	0.302		0.213	0.295
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0		0.442	0.673		0.347	0.632		0.307	0.612		0.292	0.611
			(0.002)	(0.005)		(0.002)	(0.004)		(0.002)	(0.004)		(0.003)	(0.005)
	0.3		0.403	0.451		0.302	0.371		0.255	0.335		0.246	0.328
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0		0.528	0.758		0.432	0.715		0.388	0.693		0.375	0.691
			(0.002)	(0.005)		(0.003)	(0.004)		(0.003)	(0.004)		(0.003)	(0.004)
	0.3		0.481	0.540		0.383	0.458		0.336	0.419		0.326	0.412
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
			T=25			T=50			T=75			T=100	

Table 11: Average extreme penalty values for HL with a variation in d.

8.3. The Method by Stock and Amengual

Third we can analyse the finite sample properties for the estimator presented by Stock and Amengual. The last tables show the three IC criteria in the top left corners.

PCA vs FHLR

Table 12 presents results to compare the dynamic estimation method with standard PCA. The dynamic estimator chooses the initial number of dynamic factors from the Hallin-Liška method. The number of static factors is chosen to be high again, $\hat{r} = r_{\max} = 9$. The penalties are again quite distinct. Here, the optimal penalties are higher for the dynamic method.

	VAR	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	PCA		0.191	0.433		0.148	0.421		0.137	0.423		0.134	0.423
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	$FHLR$		0.355	0.541		0.359	0.543		0.367	0.554		0.378	0.546
			(0.004)	(0.007)		(0.004)	(0.007)		(0.004)	(0.008)		(0.004)	(0.008)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	PCA		0.206	0.461		0.159	0.454		0.147	0.457		0.142	0.456
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	$FHLR$		0.364	0.561		0.366	0.568		0.388	0.554		0.387	0.533
			(0.004)	(0.007)		(0.004)	(0.008)		(0.004)	(0.007)		(0.003)	(0.005)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	PCA		0.239	0.506		0.188	0.500		0.173	0.503		0.168	0.504
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	$FHLR$		0.367	0.593		0.377	0.580		0.372	0.537		0.340	0.526
			(0.004)	(0.006)		(0.004)	(0.006)		(0.003)	(0.004)		(0.003)	(0.004)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	PCA		0.335	0.596		0.272	0.580		0.254	0.582		0.248	0.582
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	$FHLR$		0.433	0.646		0.373	0.576		0.355	0.619		0.348	0.650
			(0.004)	(0.006)		(0.003)	(0.005)		(0.003)	(0.005)		(0.003)	(0.005)
		T=25			T=50			T=75			T=100		

Table 12: Average extreme penalty values for SWIC with a variation in estimation procedure: PCA vs FHLR $q = q_{HL}, r = r_{max}$.

The PCA Estimator

Table 13 shows the optimal bounds for the PCA estimator when the number of static factors is misspecified. The second criterion seems to perform well, even if the number of factors is overestimated.Table 14 shows the performance decrease after Kalman filter techniques.

	VAR	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N=100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	r		0.187	0.473		0.152	0.462		0.142	0.460		0.137	0.458
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	r_{max}		0.191	0.433		0.148	0.421		0.137	0.423		0.134	0.423
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	r_{IC}		0.187	0.470		0.152	0.461		0.141	0.454		0.136	0.446
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
$N=75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	r		0.200	0.501		0.162	0.496		0.150	0.497		0.145	0.496
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	r_{max}		0.206	0.461		0.159	0.454		0.147	0.457		0.142	0.456
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	r_{IC}		0.201	0.485		0.160	0.483		0.148	0.478		0.143	0.472
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
$N=50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	r		0.229	0.532		0.189	0.531		0.176	0.532		0.171	0.531
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	r_{max}		0.239	0.506		0.188	0.500		0.173	0.503		0.168	0.504
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	r_{IC}		0.237	0.509		0.188	0.508		0.174	0.507		0.169	0.507
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
$N=25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	r		0.310	0.616		0.267	0.618		0.253	0.619		0.248	0.620
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.004)		(0.002)	(0.004)
	r_{max}		0.335	0.596		0.272	0.580		0.254	0.582		0.248	0.582
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	r_{IC}		0.335	0.596		0.272	0.580		0.254	0.582		0.248	0.582
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
			T=25			T=50			T=75			T=100	

Table 13: Average extreme penalty values for SWIC with a variation in the PCA estimator via changes in the number of initial static factors: 1) r 2) r_{max} 3) r_{IC}.

	VAR	PE	max−min	PE	max−min	PE	max−min	PE	max−min	PE	max−min
$N = 100$	(1)	0.160		0.117		0.101		0.092		0.077	
	(2)	0.322		0.235		0.201		0.184		0.129	
	(3)	0.010		0.003		0.002		0.001		0.001	
	PC		0.242		0.274		0.286		0.289		0.295
	KA		0.235		0.261		0.272		0.272		0.274
	SM		0.224		0.242		0.250		0.248		0.245
	EM		NaN		NaN		NaN		NaN		NaN
$N = 75$	(1)	0.171		0.130		0.115		0.107		0.095	
	(2)	0.343		0.261		0.230		0.201		0.150	
	(3)	0.010		0.003		0.002		0.002		0.002	
	PC		0.256		0.295		0.310		0.314		0.325
	KA		0.249		0.285		0.300		0.302		0.310
	SM		0.236		0.265		0.276		0.276		0.280
	EM		NaN		NaN		NaN		NaN		NaN
$N = 50$	(1)	0.192		0.156		0.143		0.138		0.132	
	(2)	0.386		0.313		0.261		0.235		0.188	
	(3)	0.010		0.003		0.003		0.003		0.003	
	PC		0.266		0.312		0.330		0.336		0.348
	KA		0.263		0.306		0.324		0.330		0.341
	SM		0.250		0.289		0.305		0.311		0.321
	EM		NaN		NaN		NaN		NaN		NaN
$N = 25$	(1)	0.254		0.232		0.228		0.228		0.241	
	(2)	0.515		0.386		0.343		0.322		0.283	
	(3)	0.010		0.010		0.010		0.010		0.010	
	PC		0.261		0.308		0.328		0.333		0.345
	KA		0.261		0.307		0.327		0.333		0.345
	SM		0.248		0.289		0.308		0.314		0.325
	EM		NaN		NaN		NaN		NaN		NaN
		T=25		T=50		T=75		T=100		T=250	

Table 14: Average extreme penalty values for SWIC with a variation in the updates. PCA estimator and max number of initial static factors.

The FHLR Estimator

The same holds for the FHLR estimator which can be seen in table 15 where Kalman filter techniques reduce the boundary gap.

	VAR	PE	max−min	PE	max−min	PE	max−min	PE	max−min	PE	max−min
$N = 100$	(1)	0.160		0.117		0.101		0.092		0.077	
	(2)	0.322		0.235		0.201		0.184		0.129	
	(3)	0.010		0.003		0.002		0.001		0.001	
	$FHLR$		0.188		0.184		0.181		0.166		0.297
	KA		0.187		0.184		0.183		0.187		0.272
	SM		0.188		0.184		0.183		0.148		0.182
	EM		NaN		NaN		NaN		NaN		NaN
$N = 75$	(1)	0.171		0.130		0.115		0.107		0.095	
	(2)	0.343		0.261		0.230		0.201		0.150	
	(3)	0.010		0.003		0.002		0.002		0.002	
	$FHLR$		0.199		0.207		0.157		0.208		0.325
	KA		0.200		0.208		0.204		0.242		0.298
	SM		0.201		0.208		0.138		0.166		0.178
	EM		NaN		NaN		NaN		NaN		NaN
$N = 50$	(1)	0.192		0.156		0.143		0.138		0.132	
	(2)	0.386		0.313		0.261		0.235		0.188	
	(3)	0.010		0.003		0.003		0.003		0.003	
	$FHLR$		0.222		0.187		0.197		0.228		0.291
	KA		0.221		0.221		0.269		0.288		0.310
	SM		0.220		0.112		0.160		0.185		0.173
	EM		NaN		NaN		NaN		NaN		NaN
$N = 25$	(1)	0.254		0.232		0.228		0.228		0.241	
	(2)	0.515		0.386		0.343		0.322		0.283	
	(3)	0.010		0.010		0.010		0.010		0.010	
	$FHLR$		0.153		0.205		0.262		0.313		0.463
	KA		0.213		0.291		0.312		0.314		0.315
	SM		0.062		0.170		0.184		0.175		0.155
	EM		NaN		NaN		NaN		NaN		NaN
		T=25		T=50		T=75		T=100		T=250	

Table 15: Average extreme penalty values for SWIC with a variation in the updates. FHLR with $q = q_{max}$ estimator and $r = r_{max}$.

Model Parameters

Tables 16, 17 and 18 show the results for the parameter variation. unsurprisingly, an increase in τ increases the boundary values. However, an increase in θ *decreases* them which is most surprising. Again, time dependence in the errors worsens the estimation and the boundary values tend to decrease.

	τ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0		0.141	0.207		0.092	0.179		0.080	0.176		0.075	0.175
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	.5		0.155	0.215		0.105	0.183		0.092	0.179		0.086	0.178
			(0.001)	(0.002)		(0.000)	(0.002)		(0.000)	(0.002)		(0.000)	(0.002)
	0.9		0.278	0.321		0.224	0.260		0.211	0.246		0.206	0.239
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0		0.155	0.221		0.106	0.193		0.094	0.189		0.088	0.188
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	.5		0.173	0.234		0.123	0.198		0.110	0.195		0.104	0.193
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	0.9		0.325	0.374		0.275	0.311		0.264	0.297		0.259	0.288
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.001)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0		0.184	0.253		0.132	0.217		0.120	0.213		0.114	0.211
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	.5		0.209	0.270		0.155	0.229		0.143	0.223		0.137	0.221
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	0.9		0.402	0.473		0.357	0.413		0.351	0.400		0.348	0.392
			(0.002)	(0.003)		(0.002)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0		0.260	0.331		0.204	0.283		0.191	0.275		0.185	0.272
			(0.001)	(0.003)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	.5		0.298	0.364		0.239	0.311		0.226	0.301		0.220	0.296
			(0.001)	(0.003)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	0.9		0.531	0.662		0.480	0.629		0.473	0.628		0.466	0.625
			(0.003)	(0.004)		(0.002)	(0.003)		(0.002)	(0.003)		(0.002)	(0.003)
			T=25			T=50			T=75			T=100	

Table 16: Average extreme penalty values for SWIC with a variation in tau.

	θ	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0.5		0.191	0.433		0.148	0.421		0.137	0.423		0.134	0.423
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	1		0.155	0.215		0.105	0.183		0.092	0.179		0.086	0.178
			(0.001)	(0.002)		(0.000)	(0.002)		(0.000)	(0.002)		(0.000)	(0.002)
	1.5		0.147	0.167		0.098	0.122		0.085	0.112		0.079	0.108
			(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0.5		0.206	0.461		0.159	0.454		0.147	0.457		0.142	0.456
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	1		0.173	0.234		0.123	0.198		0.110	0.195		0.104	0.193
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	1.5		0.164	0.186		0.115	0.138		0.102	0.127		0.096	0.123
			(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)		(0.000)	(0.001)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0.5		0.239	0.506		0.188	0.500		0.173	0.503		0.168	0.504
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	1		0.209	0.270		0.155	0.229		0.143	0.223		0.137	0.221
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	1.5		0.199	0.227		0.147	0.173		0.134	0.161		0.128	0.156
			(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)		(0.000)	(0.001)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0.5		0.335	0.596		0.272	0.580		0.254	0.582		0.248	0.582
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	1		0.298	0.364		0.239	0.311		0.226	0.301		0.220	0.296
			(0.001)	(0.003)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
	1.5		0.282	0.322		0.227	0.259		0.214	0.245		0.208	0.239
			(0.001)	(0.002)		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
			T=25			T=50			T=75			T=100	

Table 17: Average extreme penalty values for SWIC with a variation in theta.

	d	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)	PE	min (sd)	max (sd)
$N = 100$	(1)	0.160			0.117			0.101			0.092		
	(2)	0.322			0.235			0.201			0.184		
	(3)	0.010			0.003			0.002			0.001		
	0		0.191	0.433		0.148	0.421		0.137	0.423		0.134	0.423
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	0.3		0.167	0.234		0.107	0.197		0.092	0.188		0.085	0.186
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 75$	(1)	0.171			0.130			0.115			0.107		
	(2)	0.343			0.261			0.230			0.201		
	(3)	0.010			0.003			0.002			0.002		
	0		0.206	0.461		0.159	0.454		0.147	0.457		0.142	0.456
			(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	0.3		0.184	0.252		0.125	0.212		0.110	0.204		0.103	0.201
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 50$	(1)	0.192			0.156			0.143			0.138		
	(2)	0.386			0.313			0.261			0.235		
	(3)	0.010			0.003			0.003			0.003		
	0		0.239	0.506		0.188	0.500		0.173	0.503		0.168	0.504
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.005)
	0.3		0.218	0.286		0.158	0.240		0.143	0.231		0.136	0.228
			(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
$N = 25$	(1)	0.254			0.232			0.228			0.228		
	(2)	0.515			0.386			0.343			0.322		
	(3)	0.010			0.010			0.010			0.010		
	0		0.335	0.596		0.272	0.580		0.254	0.582		0.248	0.582
			(0.002)	(0.006)		(0.002)	(0.005)		(0.002)	(0.005)		(0.002)	(0.004)
	0.3		0.306	0.374		0.243	0.318		0.226	0.307		0.218	0.303
			(0.001)	(0.003)		(0.001)	(0.002)		(0.001)	(0.002)		(0.001)	(0.002)
			T=25			T=50			T=75			T=100	

Table 18: Average extreme penalty values for SWIC with a variation in d.

9. Conclusion

Generalised Dynamic Factor Models have been discussed and emphasis has been on the identification of the number of dynamic shocks in large models. Two main approaches have been presented. The first one, the structural approach, motivates the modelling of an autoregressive system of the static factors and the number of dynamic shocks is identified as the rank of the innovation covariance. The second approach identifies the number of dynamic shocks as the number of dynamic eigenvalues that increase to infinity in the large panel data setting. The estimation methods rely heavily on the asymptotics $N, T \to \infty$. Possible problems with finite samples have been discussed. A simulation study presented finite sample analysis of the state of the art estimation procedures, the comparison of the estimation techniques as well as the impacts of model parameters such as noise, cross and serial error correlation. In general the theoretical penalties perform well, however, the optimal boundary values are dependent on the model parameters and the penalties could be adjusted. More importantly, the optimal penalties seem to be unsymmetrical in cross and time section dimension. The dynamic estimation method performed badly for $N > T$ due to singularities. Another drawback of the method could be that the dynamic estimator needs an initial number of dynamic factors to calculate the various criteria. Misspecification of the number of static factors hardly worsens the estimation performance. Kalman filter updates seem to increase the performance of the Bai-Ng estimator for the PCA estimation method. Kalman filter updates in combination with the dynamic method yielded bad results in all cases.

A. Periodogram Smoothing

The *periodogram* is defined as

$$I(\theta_j)^{N,T} \equiv \frac{1}{2\pi T}|\sum_{t=1}^{T} y_t e^{-it\theta_j}|^2$$

for the Fourier frequencies $\theta_j \equiv \frac{2\pi j}{T}, \quad j = 1, \ldots, T$ or $\theta_j \equiv \frac{2\pi j}{T}, \quad j = -M_T, \ldots, M_t$, and the alternative representation

$$I(\theta_j)^{N,T} = \frac{1}{2\pi} \sum_{k=-(T-1)}^{k=T-1} \hat{\Gamma}_k^{N,T} e^{-it\theta_j}$$

bur the periodogram is not a consistent estimator of the spectral density. However, we have the following lemma:

Lemma 1. *Let $y_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j}$ be an N dimensional stochastic process where z_t is a linear independent white noise process with finite fourth moments and the coefficients c_j fulfil $\sum_{j=-\infty}^{\infty} |c_j(k,i)||j|^2 < \infty \quad \forall i, j = 1, \ldots, N$. Let $\hat{\Gamma}_k^{N,T}$ be consistent estimators of $\Gamma_k^{N,T} \forall k \in \mathbb{Z}$ and let there be a sequence $\{M_T | T > 0\}$ and a weighting function ω fulfilling*

- *$M_T > 0 \quad \forall T$ and $M_T \to \infty, \quad \frac{M_T}{T} \to 0$ for $T \to \infty$*
- *ω is a piecewise differentiable function whose first three derivatives are bounded and*
- *$\omega(0) = 1$*
- *$\omega(x) = \omega(-x)$*
- *$|\omega(x)| \leq 1$*
- *$\omega(x) = 0 \quad \forall |x| > 1$*

then

$$\Sigma^{\hat{N},T}(\theta) \equiv \frac{1}{2\pi} \sum_{k=-M_T}^{M_T} \omega(\frac{k}{M_T}) \hat{\Gamma}_k^{N,T} e^{-ik\theta} \tag{A.1}$$

is a consistent estimator of $\Sigma^{N,T}$.

B. Matrix Operations

Let A be a square, invertible Matrix, then we can express the differentials of the functionals tr and $\log|\cdot|$ as

$$\frac{\partial \operatorname{tr} A}{\partial x} = \operatorname{tr}\left(\frac{\partial \mathrm{A}}{\partial \mathrm{x}}\right)$$
$$\frac{\partial \log |A|}{\partial x} = \operatorname{tr}\left(\mathrm{A}^{-1}\frac{\partial \mathrm{A}}{\partial \mathrm{x}}\right)$$

References

[Deistler & Hannan 1988] M. Deistler and E.J. Hannan, The Statistical Theory of Linear Systems, *Wiley*, New York, 1988

[Deistler et. al. 2010] Manfred Deistler, Brian D.O. Anderson, Ch. Zinner, W. Chen, Generalized Linear Dynamic Factor Models: An Approach via Singular Autoregressions, *European Journal of Control*, 2010, 3, 211-244

[Deistler & Anderson 1989] M. Deistler, B.D.O. Anderson, Linear dynamic errors-in-variables models: Some structure theory, *Journal of Econometrics*, Volume 41, Issue 1, May 1989, Pages 39-63

[Deistler et. al. 2011] M. Deistler, A. Filler and B. Funovits, *Communications in Information and Systems*, Volume 11, No 3, 2011, Pages 225-236

[Lippi et. al 2000] Mario Forni, Marc Hallin, Marco Lippi and Lucrezia Reichlin, The Generalized Dynamic-Factor Model: Identification and Estimation, *The Review of Economics and Statistics*, Vol. 82, No. 4, Nov. 2000, pp. 540-554

[Forni & Lippi 2001] Mario Forni and Marco Lippi, The Generalized Dynamic Factor Model: Representation Theory, *Econometric Theory*, Vol. 100, No. 471, Sep. 2005, pp. 830-840

[Lippi et. al 2005] Mario Forni, Marc Hallin, Marco Lippi and Lucrezia Reichlin, The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting, *Journal of the American Statistical Association*, Vol. 100, No. 471 (Sep., 2005), pp. 830-840

[Lippi et. al 2009] Mario Forni, Domenico Giannone, Marco Lippi and Lucrezia Reichlin, Opening the Black Box: Structural Factor Models with Large Cross Sections *Econometric Theory*, Vol. 25, Issue. 05, Oct. 2009, pp. 1319-1347

[Chamberlain 1983] Gary Chamberlain, Funds, Factors, and Diversification in Arbitrage Pricing Models *Econometrica*, Vol. 51, No. 5, Sep. 1983, pp. 1305-1323

[Chamberlain & Rothschild 1983] ,Gary Chamberlain and Michael Rothschild, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, Vol. 51, No. 5, Sep. 1983, pp. 1281-1304

[Bai & Ng 2002] Jushan Bai and Serena Ng, Determining the Number of Factors in Approximate Factor Models, *Econometrica*, Vol. 70 , No 1, Jan. 2002 , pp 191-221

[Bai & Ng 2007] ,Jushan Bai and Serena Ng, Determining the Number of Primitive Shocks in Factor Models *Journal of Business & Economic Statistics*, Vol. 25, No. 1, Jan. 2007, pp. 52-60

[Stock & Watson 2002] James H. Stock and Mark W. Watson, Forecasting Using Principal Components from a Large Number of Predictors *Journal of the American Statistical Association*, Vol. 97, No. 460, Dec. 2002, pp. 1167-1179

[Watson & Amengual] Mark W. Watson and Dante Amengual, Consistent Estimation of the Number of Dynamic Factors in a Large N And T Panel *Journal of Business and Economic Statistics*, Vol. 25, No. 1, Jan. 2007, pp. 91-96

[Hallin & Liška 2007] Marc Hallin and Roman Liška, Determining the Number of Factors in the General Dynamic Factor Model *Journal of the American Statistical Association*, Vol. 102, No. 478, Jun. 2007 , pp. 603-617

[Onatski 2010] Alexei Onatski, Determining the Number of Factors from Empirical Distribution of Eigenvalues, *The Review of Economics and Statistics*, Vol. 92, No. 4, Nov. 2010, Pages 1004-1016

[Brillinger 2001] David R. Brillinger, Time Series: Data Analysis and Theory *Classics in Applied Mathematics*, 36, Chapter 9, Sept. 2001, ISBN-10: 0898715016

[Jolliffe 2002] I.T. Jolliffe, Principal Component Analysis, *Springer Series in Statistics*, 2002, ISBN 978-0-387-95442-4

[Doz et. al. 2011] , Catherine Doz, Domenico Giannone, Lucrezia Reichlin, A two-step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics*, Volume 164, Issue 1, September 2011, Pages 188-205

[Filler 2010] Alexander Filler, Generalized Dynamic Factor Models: Structure Theory and Estimation for Single Frequency and Mixed Frequency Data, (Dissertation), Nov. 2010

[Geman 1980] Stuart Geman, A Limit Theorem for the Norm of Random Matrices, *Annals of Probability*, Volume 8, Number 22, 1980,pp. 252-261

[Yin et. al. 1988] , Y.Q. Yin, Z.D. Bai and P.R. Krishnaiah, On the Limit of the Largest Eigenvalue of the Large Dimensional Sample Covariance Matrix, *Probability Theory and Related Fields*, Vol. 78, Springer Verlag, 1988, pp. 509-521