DIPLOMARBEIT

Titel der Diplomarbeit

„Comparing three estimation methods in the context of generalized dynamic factor models"

Verfasser

Alexander Braumann

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien, im September 2012

Contents

Abstract

This thesis considers three different estimation methods for the unobserved static factors z_t in the context of generalized dynamic factor models. These models assume that a high-dimensional multivariate time-series x_t is driven by a very low dimensional factor process z_t. This unobserved factors partly explain the observations x_t through the equation $x_t = Lz_t + \xi_t$. The models are generalized, because they allow for weak cross-sectional dependence among the idiosyncratic components ξ_{it}. A drawback of this general assumption is that the common (Lz_t) and idiosyncratic component ξ_t are only asymptotically (for the number of different time-series going to infinity) identified. Therefore all three estimation methods for the static factors are based on restricted models. Despite these restrictions they consistently estimate the linear space of static factors under generalized assumptions. However, no further analytical properties for all three methods are available. Therefore a simulation study was conducted, in order to understand their behavior and compare their performances for finite dimensional panels. Also a data set from the US macroeconomy was analyzed in order to get authentic parameters for the simulation study.

The aim of this simulation study was to study the performance of the estimators for different models and parameters. The results suggest that all three estimators suffer from neglecting the generalized dependence structure of the idiosyncratic component. If the idiosyncratic process is white noise, a ranking of the estimators seems to be possible. However, high local dependencies could weaken or even erase relative advantages that could be observed in the white noise case. This effect seems to be stronger for the cases where the dimension of the dynamic factors is smaller than the dimension of the static factors. Different noise-to-signal ratios and especially a long-memory static factor process z_t clearly influence absolute and relative performance of the estimators.

1   Introduction

1.1   Motivation

Non-technical introduction   Statistical factor models were developed in the context of multivariate data. They assume that different observed entities linearly depend on unobserved so-called factors (see [39] for an introduction). Generalized dynamic factor models are a very general class of factor models for time-series data [26], [27], [14]. They don't face strict assumptions common to classical factor models that may conflict with reality. In the last decade they have become very popular in macroeconomics, because they are able to deal with large panels, i.e. a very large number of variables and/or a large number of observations. Today macroeconomists are now able to observe a huge amount of aggregated and disaggregated data and use this for example in order to forecast future economic variables [45] or to assess the impact of economic policy (structural analysis) [8].

Example 1.1.   Consider disaggregated and aggregated macro-economic time-series data displayed in Figure 1 that describe the US macro-economy[1]. This multivariate time-series consists of 109 different economic entities observed between the years 1959 and 2006 (see Section 6 for a complete description).

□

In order to work with high dimensional multivariate time-series data, many models suffer from the so-called "curse of dimensionality". This means that the number of parameters grows faster than the number of variables in the model. Take for example the vector auto-regressive model. The number of model parameters is proportional to the squared number of model variables. Generalized dynamic factor models try to overcome this curse by assuming that the observed data is driven by a much lower dimensional unobserved factor process.

As illustrated in Figure 1 the idea is to summarize high dimensional time-series data (in Example 1.1 the dimension is 109) by a few factors (in this case a 2 dimensional process).

The knowledge of the factor process is the basis for forecasting and structural analysis. This thesis analyses three different ways of extracting these factors from the data, i.e. it compares three different estimators procedures by conducting a simulation study.

Starting point   The object of analysis will be an $n \times T$ panel

$$X' = \begin{pmatrix} x_{11} & x_{12} & \cdots & \cdots & \cdots & \cdots & x_{1T} \\ \vdots & & & & & & \vdots \\ x_{n1} & x_{n2} & & & & & x_{nT} \end{pmatrix},$$

a finite realization of the stochastic process $\boldsymbol{x} = \{x_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ where $x_{it} \in L_2(\mathbb{P}, \mathbb{C})$.

Factor models are characterized by splitting the observation x_{it} into a *common component* (also called latent variable) and an *idiosyncratic component*

$$x_{it} = \chi_{it} + \xi_{it}, \quad i \in \mathbb{N}, \ t \in \mathbb{Z} \tag{1}$$

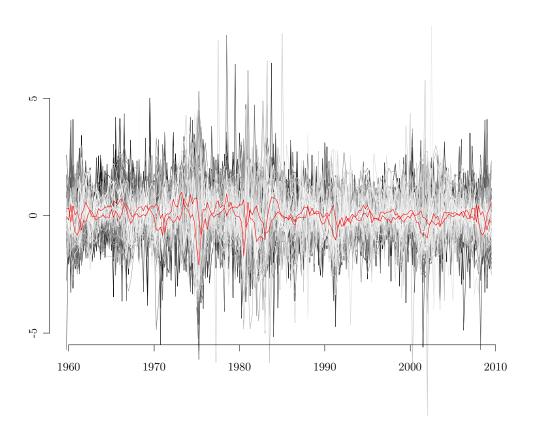[1]This data-set can be downloaded from http://www.princeton.edu/ mwatson/publi.html.

Figure 1: Disaggregated macroeconomic time-series data from Stock and Watson [50]. The 109 already transformed univariate time-series have been plotted in gray-scale whereas the two factors are represented by the red lines.

This very general setting of a double sequence (x_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$) is a characteristic of generalized dynamic factor models. Let the number of variables (the cross-section dimension) be finite, i.e. $i = 1, \ldots, n$. Then the equations

$$\underset{(n \times 1)}{x_t^n} = \underset{(n \times 1)}{\chi_t^n} + \underset{(n \times 1)}{\xi_t^n} = \underset{(n \times q)}{w^n(z)} \; \underset{(q \times 1)}{u_t} + \underset{(n \times 1)}{\xi_t^n} \;, \quad t \in \mathbb{Z} \qquad (2)$$

where the transfer function $w^n(z) = \sum_{j=-\infty}^{\infty} w_j z^j$, $w_j \in \mathbb{R}^{n \times q}$ is called the factor loading matrix, describe all different kinds of factor models in the context of multivariate data. It will be assumed that the process χ_t^n and ξ_t^n fulfill Assumption A and that the Laurent series expansion of $w^n(z)$ converges within and on the complex unit circle.

If $w^n(z) = L$ and the processes u_t and ξ_t^n are white noise, then Equation (2) describes the so-called *static factor model*. With $w^n(z) = L \in \mathbb{R}^{n \times q}$ constant, and $(u_t), (\xi_t^n)$ not necessarily white noise, the representation of the so-called *quasi-static factor model* [15] is described. Dynamic factor models are characterized through $w^n(z) = \sum_{j=-\infty}^{\infty} w_j z^j$, $w^n(z) = \sum_{j=0}^{\infty} w_j z^j$ and $w^n(z) = \sum_{j=0}^{s} w_j z^j$ respectively.

The distinction between *approximate factor models* [11], [12] and *strict factor models* (or factor models with idiosyncratic noise) arises from the way the process ξ_t^n is modeled. Strict factor models[2] are characterized by a diagonal spectral density matrix (or the covariance matrix in the static case) of ξ_t^n, whereas in the context of approximate factor models a weak dependence between the idiosyncratic variates is assumed (see Section 3). An approximate dynamic factor model is called *generalized dynamic factor model*. A formal definition will be given in Section 3.1.

The q dimensional process u_t drives the latent process χ_t^n and is called the *dynamic factor process*. In the case of a static factor model or if the order of the polynomial matrix $w^n(z)$ is finite, the process u_t can be stacked to an $r \geq q$ dimensional process z_t, the so-called *static factor process*. The unknown integer parameters q and r are of central interest. First, they indicate to what extend the dimension of the observations n can be reduced. Second, the number of dynamic factors has a special economic interpretation (see Section 3.5). In contrast to principal component models [28], where the number of principal components is the result of a tradeoff between approximation accuracy and dimensionality reduction, the number of static factors in factor models is intrinsic in the sense that it is a property of the data.

Two identification problems in the context of factor models are the identification of the latent and idiosyncratic component, given the observation x_{it} and the identification of the static factors given the latent component. As far as the latter problem is concerned, static factors are identifiable only up to post-multiplication by a non-singular[3] matrix. Concerning the former problem, approximate factor models are in contrast to factor models with strict idiosyncratic noise not identifiable for a finite number of variables n. The unique separation into common and idiosyncratic component can however be achieved as n goes infinity [11], [12], [26] (see Section 3.2).

The three estimation procedures for the static factors are based on simpler models that may be identifiable for finite n. These simpler models are all assuming that the idiosyncratic spectrum density matrix is diagonal or that ξ_t^n is white-noise with a diagonal covariance matrix. It will be shown that all three estimators are consistent in the sense that their estimates converge to a certain basis of the linear static factor space as $n, T \to \infty$ (see Section 4).

---

[2]The static factor model with idiosyncratic noise is the classical factor model, with a long history dating back, to the beginning of the twentieth century ([15], p. 156).

[3]Or an orthogonal matrix, if the static factors are assumed to be orthonormal.

Finite sample properties of the three estimation methods were studied by several simulation studies [49], [17], [18]. The added-value of this thesis is to extend the simulation setting of the three mentioned articles in several ways (see Section 5).

Section 2 discusses the theoretical concepts that will be needed later in Section 4 where the estimation procedures will be discussed.

Assumption A (*General assumption*): For any $n \in \mathbb{N}$

A1. $(x_t^n, \ t \in \mathbb{Z})$ is a centered, (wide sense) stationary[4] process

A2. The auto-covariances of (x_t^n) are absolutely summable, that is

$$\sum_{s=-\infty}^{\infty} ||\gamma_x(s)|| < \infty.$$

This implies that (x_t^n) has a spectral density $\Sigma_x^n(\cdot) : \Theta \to \mathbb{C}^n$ which is defined by

$$\Sigma_x^n(\theta) := \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{i\theta s} \gamma_x(s)$$

1.2 The research question

In this simulation study the goal is to assess the finite sample behavior of three different estimation procedures for the static factors in the context of a generalized dynamic factor model. Two questions govern the analysis. First, as the estimators are misspecified in the generalized context, the influence of this misspecification on the absolute and relative estimation performance will be of interest. Second, a finite sample comparison for different model specifications shall shed light on the relative performances of these three estimation procedures and therefore indicate points in the parameter space, where some method may outperform the other.

1.3 Framework and Notation

The double sequence $\{x_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ of random variables will be defined on an arbitrary probability space $\mathbb{P} := (\Omega, \mathfrak{A}, P)$. For each x_{it}, the first index always refers to the cross-section dimension, whereas the second index represents the time-dimension. Therefore x_{it} is the random value of the i-th cross-section at time t. The first n random values at time t will be written as $x_t^n = (x_{1t}, \ldots, x_{nt})'$. Often the superscript n will be suppressed and the single-indexed x_t will be written instead x_t^n. A double sequence $\boldsymbol{x} = \{x_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ will be written in bold letters and is always indexed on $\mathbb{N} \times \mathbb{Z}$. A cross-section sequence for a given $t \in \mathbb{Z}$ is written as $\boldsymbol{x}_t = \{x_{it}, \ i \in \mathbb{N}\}$. If the context is clear, $z_t \in \mathbb{R}$ means $z_t(\omega) \in \mathbb{R}$, $\omega \in \Omega$. I do not distinguish between sample and random variables as far as notation is concerned. In general X is a $T \times n$ data matrix.

The basis for abstract geometric argumentation are sub-spaces of the Hilbert-space $L_2(\mathbb{P})$, which is the space of all complex valued $\mathfrak{A}$-measurable square integrable functions. For the closed linear sub-spaces $\overline{\text{span}}\{y_{it}, \ i = 1, \ldots, n, \ t \in M, \ y_{it} \in L_2(\mathbb{P})\}$, $n \in \mathbb{N}$, $M \subseteq \mathbb{Z}$, abbreviations like $\mathcal{H}_y(t)$, $\mathcal{H}(y_s^n, \ 1 \leq s \leq t)$ or $\mathcal{H}_y$ will be used and explained in context. For example $\mathcal{H}_x(T)$

---

[4]Only wide sense stationary processes will be discussed. Therefore stationary is a short cut for wide sense stationary.

corresponds to $M = \{1, \ldots, T\}$. I say an n-dimensional vector x belongs to $\mathcal{H}_y$ if all components of x are in $\mathcal{H}_y$. Let z be the back-shift operator defined on $\mathcal{H}_x$ by $zx_{it} = x_{it-1}$ for every $i \in \mathbb{N}$ and $t \in \mathbb{Z}$. The variable z will be used as a complex variable as well as the back-shift operator.

For a given n, the h-lag covariance matrix of x_t^n, χ_t^n, ξ_t^n is denoted by $\gamma_\chi(h) = \mathbb{E}\chi_t^n \chi_{t-h}^n{}'$. The spectral densities of x_t^n, χ_t^n, ξ_t^n are denoted by $f_x^n(\theta), f_\chi^n(\theta)$ and $f_\xi^n(\theta)$ respectively. They are defined on the set $\Theta := [-\pi, \pi]$.

For a quadratic matrix $A \in \mathbb{R}^{k \times k}$, $A > 0$ means that A is positive definite, and $A \geq 0$ means that A is positive semi-definite. The transposed matrix of a general $m \times n$ matrix A will be written as A'. I denote A_i as its i-th column and $A_{(i)}$ or $A(i)$ as its i-th row. An element at position (i, j) is either denoted by A_{ij} or $A(i, j)$. The norm $||A||$ stands for any matrix-norm (e.g. the Frobenius norm, $||A|| = ||\text{vec}(A)||_2$ or the spectral matrix norm, $||A|| = \sqrt{\lambda_{\max}(A'A)}$), where $\text{vec}(A)$ is defined in the Appendix (see Definition C.9) and $||x||_2$ is the spectral norm. I will always use underlines for the vec-operator, i.e. $\underline{A} = \text{vec}(A)$. If not stated otherwise, all matrices as well as the processes are considered to be real-valued.

The closed complex unit ball of radius one is sometimes denoted as $\mathcal{U}_1 := \{z \in \mathbb{C} : |z| \leq 1\}$.

2 Theoretical Foundations

The main basic theorems which I will use later are stated in this section. If more elaborated concepts are used somewhere, I will refer to specialized literature.

2.1 The theory of stationary processes

Hilbert geometry: As for every x_{it}^n the variance is finite, the theoretical focus is on the space $L_2(\mathbb{P})$. This space equipped with the scalar product $(f, g) := \int_\Omega f(\omega)\overline{g(\omega)}\, d\lambda(\omega)$ is a Hilbert space[5]. The scalar product induces the L_2-norm $||x_{it}||^2 = \mathbb{E}|x_{it}|^2$ on L_2. For this space the theoretically valuable projection theorem can be used.

Theorem 2.1 (Projection theorem): Let the set H together with the scalar product $(\cdot, \cdot)$ be a Hilbert space and $M \subseteq H$ a closed linear sub-space. Then[6] $M \oplus M^\perp$, that is every $x \in H$ can be uniquely decomposed into $x = y + z$ with $y \in M$ and $z \in M^\perp$. The summand y is the orthogonal projection $\mathcal{P}_M x$ of x on M and minimizes the L_2-distance $||x - w||_2$, $w \in M$.

The structure of stationary processes *Theorem 2.2 (Wold theorem):* Every stationary n-dimensional process (x_t) can be represented in a unique way as $x_t = y_t + z_t$, where (y_t) is a regular, (z_t) a singular (deterministic) process, and where (y_t) and (z_t) are orthogonal, i.e. $\mathbb{E}y_t z_s' = 0$, $\forall s, t$ and that $y_t, z_t \in \overline{\text{span}}\{x_{is}, i \leq n, s \leq t\}$ holds.
  Wold representation: Every regular process (y_t) can be represented as

$$y_t = \sum_{j=0}^\infty k_j u_{t-j}, \quad \sum_{j=0}^\infty ||k_j||^2 < \infty \tag{3}$$

where (u_t) is white noise and $\overline{\text{span}}\{u_{is}, i \leq n, s \leq t\} = \overline{\text{span}}\{y_{is}, i \leq n, s \leq t\}$ holds.

Theorem 2.3 (Spectral factorization): Let x_t be an n-dimensional process having a rational spectral density matrix f_x of rank $q \leq n$. Then the $(n \times n)$ dimensional rational matrix can be factorized as

$$f_x(\theta) = \frac{1}{2\pi}\, w(e^{-i\theta})\, w^*(e^{-i\theta}) \tag{4}$$

where the $(n \times q)$ dimensional spectral factor $w(z)$, $z \in \mathbb{C}$ is a rational matrix which has no poles and zeros for $|z| \leq 1$. The transfer function $w(z)$ is therefore *stable* and *mini-phase*.
  This factorization is unique up to post-multiplication with constant orthogonal matrices.

---

[5]This statement is true, if L_2 are the equivalent classes for the relation $f \sim g \Leftrightarrow f = g$ a.e. on Ω.
[6]$M^\perp = \{z \in H, (z, m) = 0, \forall m \in M\}$ and $A \oplus B$ is the orthogonal direct sum of A and B.

2.2   Principal components

The classical statistical definition of principal components is based on finding a linear transformation $P = (p_1, \ldots, p_n)$ with orthonormal columns p_j of given variates $x = (x_1, \ldots, x_n)'$ such that the new variates $y = P'x$ have maximal variance.

Definition 2.4. For a random vector $x = (x_1, \ldots, x_n)$ with existing second moments, $\Sigma = cov(x)$, the first *loading* $p_1 \in \mathbb{R}^n$ defines the first principal component (*score*) $y_1 = p_1'x$ that maximizes $var(y_1) = p_1'\Sigma p_1$ subject to $p_1'p_1 = 1$. It is easy to see that p_1 solves $(\Sigma - \lambda_1 I_n)p_1 = 0$, which reveals λ_1 as the biggest eigenvalue with $\lambda_1 = var(y_1)$ and p_1 its corresponding eigenvector.

The k-th principal component p_k solves the maximization problem

$$\max_{p \in \mathbb{R}^n} p'\Sigma p - \lambda_k(p'p - 1) - \sum_{j<k} \phi_j(p'p_j)$$

The variance of the k-th score $var(y_k)$ is equal to λ_k and $(p_1, \ldots, p_n)$ form an orthonormal basis of $\mathbb{R}^n$.

In the context of time series, the characterization given in [38] is more appealing.

Suppose that we wish to predict x by a r-dimensional sub-space of $\mathrm{span}\{x_1, \ldots, x_n\}$ in the sense that we would like to minimize

$$\mathbb{E}(x - CB'x)'(x - CB'x), \quad B, C \in \mathbb{R}^{n \times r} \tag{5}$$

This is achieved by taking $B = C = (p_1, \ldots, p_r)$. This characterization of principal components is also the starting point for dynamic principal components (see [38], Ch. 11.2 or [10] Ch. 9).

Definition 2.5 (Static principal components of the process x_t): Let x_t be a stationary n dimensional process with covariance matrix $\gamma_x(0)$ and let $(p_1, \ldots, p_n)$ be the eigenvectors corresponding to the spectral decomposition of $\gamma_x(0)$ (see Theorem C.2). The variate $y_{jt} = p_j'x_t, \quad j = 1, \ldots, n$ is called the j-th principal component of the process $(x_t)_{t \in \mathbb{Z}}$.

Principal components of a stationary process x_t can also be given in a dynamic sense. They are defined as linear filters $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$, $b_j \in \mathbb{R}^{r \times n}$ such that x_t is approximated best in the least squares sense by a filtered process of itself. This filtered process has spectral density q. We therefore search for linear filters $B(z)$ and $C(z) = \sum_{j=-\infty}^{\infty} c_j z^j$, $c_j \in \mathbb{R}^{n \times r}$ such that

$$\mathrm{tr}\left[\mathbb{E}(x_t - C(z)B(z)x_t)(x_t - C(z)B(z)x_t)'\right]$$

is minimized for fixed rank q.

It can be shown that similar to the static case, the solution to this problem is given through the spectral decomposition of the spectral density of the process x_t (see [28]).

Definition 2.6 (Dynamic principal components of the process x_t):   Let x_t be a zero mean stationary n-dimensional process with spectral density f_x. Decompose

$$f_x(\theta) = O_1(e^{-i\theta})\Lambda_1(\theta)O_1(e^{-i\theta})^* + O_2(e^{-i\theta})\Lambda_2(\theta)O_2(e^{-i\theta})^*$$

where $O_1(e^{i\theta}), O_2(e^{i\theta})$ contain the eigenvectors and $\Lambda_1(\theta), \Lambda_2(\theta)$ the eigenvalues of the spectral density matrix $\Sigma_x(\cdot)$. The matrices $O_1(e^{i\theta}), O_2(e^{i\theta})$ are the corresponding transfer functions to the $n \times q$ and $n \times (n-q)$ dimensional linear filters $O_1(z), O_2(z)$.

The q dynamic principal components are then defined by $y_t = O_1(z)^* x_t$.

Remark 2.7.

1. Principal components can be seen as coordinates in the eigenbasis of $\gamma_x(0)$. This can be seen from Definition 2.4.

2. Definitions 2.5 and 2.6 are given at population level. If one replaces $\gamma_x(0)$ by the sample covariance matrix $S := \frac{1}{T-1} \sum_{t=1}^{T} x_t x_t' = \frac{1}{n} X'X$, then its eigenvectors $\hat{p}_1, \ldots, \hat{p}_n$ define the (sample) principal components $\hat{p}_j' x_t, \quad j = 1, \ldots, n$. The spectral density f_x can be estimated in many ways (see [29]). The sample dynamic principal components in this thesis are calculated from the smoothed periodogram estimates of the spectral density.

3. For any integer q, $1 \leq q \leq n$, consider the orthonormal transformation $y = B'x$, where $B \in \mathbb{R}^{n \times q}$, $\Sigma_y = B'\Sigma B$. Then by setting $B = (p_1, \ldots, p_q)$ the matrix B maximizes the trace of Σ_y. This result also holds in the sample case. Then $B = (\hat{p}_1, \ldots, \hat{p}_q)$ maximize $\text{tr}(B'X'XB)$ (see [38]).

$\square$

2.3 The Kalman Filter apparatus

Two of the three discussed estimation methods are using the Kalman filter apparatus in order to calculate the estimates. It is a very powerful iterative algorithm originally introduced as a method primarily for use in aerospace-related research, especially used in electrical engineering, but also in medicine, soil sciences and economics[7]. The procedure is based on so-called state space models.

Consider the following state-space model with given (H, F, G, Q, R). For every $t \in \mathbb{Z}$:

$$\underset{(n \times 1)}{x_t} = \underset{(n \times m)}{H} \underset{(m \times 1)}{s_t} + \underset{(n \times 1)}{\xi_t} \tag{6a}$$

$$\underset{(m \times 1)}{s_{t+1}} = \underset{(m \times m)}{F} \underset{(m \times 1)}{s_t} + \underset{(m \times q)}{G} \underset{(q \times 1)}{u_t} \tag{6b}$$

with $\xi_t \sim \text{WN}(0, R)$, $u_t \sim \text{WN}(0, Q)$, $\mathbb{E}\xi_t u_s' = 0$, $\forall s, t \in \mathbb{Z}$ and $\mathbb{E}s_0 u_t' = \mathbb{E}s_0 \xi_t' = 0$, $t > 0$. The Kalman filter is a linear filter which is the solution to the following optimization problem:

For the state space system in Equations (6), observations $X = (x_1, x_2, \ldots, x_T)$ and a known distribution of $s_1 \sim \mathcal{N}(\mu_1, V_1)$, find those linear estimates for s_t, $t = 1, \ldots, T$ that minimize component-wise the mean square distance to s_t. By the projection theorem this approximation is given by $s_{t|t-1} = \mathcal{P}(s_t | x_1, \ldots, x_{t-1})$ (see [2]).

---

[7]See [48] for more references.

Denote the best linear estimator (in mean square sense) of s_t given $x_1, \ldots, x_{t-1}$ by $s_{t|t-1}$ and the one-step ahead prediction error of x_t given $X_1^{t-1} := \{x_1, \ldots, x_{t-1}\}$ by $e_t := x_t - Hs_{t|t-1}$. The accuracy of the state prediction is given by the covariance matrix $P_{t|t-1} := \mathbb{E}(s_t - s_{t|t-1})(s_t - s_{t|t-1})'$. Then for $t = 1, \ldots, T$, the *prediction equations* are given by:

$$s_{t+1|t} = Fs_{t|t-1} + K_t e_t \tag{7a}$$

$$K_t = FP_{t|t-1}H'\Sigma_t^{-1} \tag{7b}$$

$$\Sigma_t = HP_{t|t-1}H' + R \tag{7c}$$

$$P_{t+1|t} = FP_{t|t-1}F' - K_t\Sigma_t K_t' + GQG' \tag{7d}$$

The best estimation for $t = 1$ is given by $s_{1|0} = 0$ with covariance matrix $P_{1|0} = \mathbb{E}s_1 s_1'$. The *updating equations* are given by:

$$s_{t|t} = s_{t|t-1} + P_{t|t-1}H'\Sigma_t^{-1}e_t \tag{8a}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'\Sigma_t^{-1}HP_{t|t-1} \tag{8b}$$

Substituting in (7a) and (7d) gives $s_{t+1|t} = Fs_{t|t}$ and $P_{t+1|t} = FP_{t|t}F' + GQG'$.

Remark 2.8 (Remarks on the Kalman Filter).

- The Kalman Filter equations can be derived based on Hilbert geometry (projections), as done in [32], based on multivariate normal distribution theory, as done in [19], or from a Bayesian perspective (see [42]). For example, the proof of Equations (7) and (8) using the orthogonal decomposition of $\mathcal{H}(x_1, \ldots, x_t)$ is not very difficult. The interested reader is referred to Theorem 3.2.1 in [32].

- Often it is assumed that $G = I_m$ and u_t is a m-dimensional random vector[8]. The general case of Equations (6) is discussed in [19] and [2], Chapter 3.

- Note that the matrices H, F, R, G, Q, P_1 and the vector a_1 are assumed to be known a-priori. The next section deals with the situation where most of these variables are unknown.

- Equations (7a) and (7d) relate the state approximations $s_{t|t-1}$ and its uncertainty $P_{t|t-1}$ to the approximation $s_{t+1|t}$ and uncertainty $P_{t+1|t}$ for the next period. The *prediction equations* are derived by assuming that data until $t-1$ is available for state prediction at time t. The higher the uncertainty about state prediction at $t-1$ and the more accurate the determination of the innovations e_t (see Equation (7b)), the higher the influence of a bad prediction of the observations on future state predictions. The *updating equations* are needed after a new observation x_t has arrived. The determined prediction error e_t changes the approximation of the unknown state. This again depends on the state prediction accuracy and the precision of the calculation of e_t. The uncertainty about the state approximation $P_{t|t}$ decreases, meaning that $P_{t|t} \leq P_{t|t-1}$, where $\leq$ is the Loewner order.

---

[8]This is done in the `R`-packages `dse`, `?` and the code of Shumway and Stoffer.

- If a Gaussian state space model is assumed, meaning that ξ_t, u_t jointly follow a normal distribution, then $s_{t|t-1} = \mathbb{E}(s_t|X_1^{t-1})$ and $P_{t|t-1} = \mathbb{E}(s_t - s_{t|t-1})(s_t - s_{t|t-1})'|X_1^{t-1})$. This means that $s_{t|t-1}$ provides the best estimators in the class of all $\sigma(X_1^{t-1})$ measurable and square integrable functions. The second equation holds, because $(s_t - s_{t|t-1})$ is orthogonal to $(x_1, \ldots, x_T)$.

- The calculation of the Likelihood is a byproduct of solving Equations (7). As the likelihood of the observations is the joint density of $X = (x_1, \ldots, x_T)$ depending on a parameter θ,

$$\mathcal{L}(X;\theta) = f_\theta(x_1) \prod_{t=2}^{T} f_\theta(x_t|x_1, \ldots, x_{t-1}),$$

where $f_\theta(x_t|x_1, \ldots, x_{t-1})$ is the density of a normal distribution with mean $Hs_{t|t-1}$ and covariance Σ_t. The *prediction error decomposition*[9] of the Log-Likelihood is given by

$$\ell(X;\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\left(\sum_{t=1}^{T}\ln(\det(\Sigma_t)) + e_t'\Sigma_t e_t\right), \tag{9}$$

where e_t is defined above. Note all the factors needed for calculating $\mathcal{L}(X;\theta)$ are given by Equations (7).

$\square$

Kalman smoother. The Kalman smoother is a fixed interval smoother[10]. For a given "interval" $t = 1, \ldots, T$, it computes $s_{t|T} = \mathcal{P}(s_t|X) = (\mathbb{E}s_t\underline{X})(\mathbb{E}\underline{X}\underline{X}')^{-1}$ and $P_{t|T} = \mathbb{E}(s_t - s_{t|T})(s_t - s_{t|T})'$, where $X = (x_1, \ldots, x_T)$ and $\underline{X} = \mathrm{vec}(X)$. The projections are computed for $t = T, T-1, \ldots, 1$ with the help of the Kalman Filter recursions. For $t < u \leq T$:

$$J_t = (P_{t|t-1} - P_{t|t-1}H'K_t')P_{t+1|t}^{-1} = P_{t|t}F'P_{t|t-1}^{-1} \tag{10a}$$

$$s_{t|u} = s_{t|t} + J_t(s_{t+1|u} - s_{t+1|t}) \tag{10b}$$

$$P_{t|u} = P_{t|t} + J_t(P_{t+1|u} - P_{t+1|t})J_t' \tag{10c}$$

The matrix J_t in Equation (10a) is called the *Kalman smoothing matrix* . These equations are derived by Hilbert projection arguments for example in [32], page 93–95. The second equality of (10a) can be easily proved by substituting (8b) into (10a).

Remark 2.9 (Remarks on the Kalman smoother).

1. The Kalman smoother is not the orthogonal projection (or conditional expectation, if the state space model is Gaussian) but the iterative procedure that computes the projections $s_{t|T}, t = 1, \ldots, T$.

---

[9]Terminology introduced by Harvey (1989).
[10]For a discussion about different types of smoothing see [2].

2. The matrix $P_{t+1|t}$ must not be invertible in general. Consider the identity $P_{t+1|t} = FP_{t|t}F' + GQG'$. The matrix GQG' must not have full column rank, as G is a real matrix of dimension $m \times q$. If $P_{t+1|t} \geq 0$ is not invertible, i.e. it has eigenvalues equal to zero, then the pseudo-inverse is taken instead. Let $P_{t+1|t} = U \operatorname{diag}(d_1, d_2, \ldots, d_k, 0, \ldots, 0) U'$, where $U^{-1} = U'$ is a unitary matrix. Then $P_{t+1|t}^{-1} = U' \operatorname{diag}(1/d_1, 1/d_2, \ldots, 1/d_k, 0, \ldots, 0) U$. The integer k is equal to the rank of $P_{t+1|t}$.

3. Following [19], Chapter 4.3, an illustrative relation to Equations (10) is given by Hilbert geometry. Let $\mathcal{H}_x$ be the Hilbert space spanned by all the components of $x_1, \ldots, x_T$. This space can be written as the direct orthogonal sum $\mathcal{H}_x = \bigoplus_{t=1}^{T} \mathcal{H}(e_t)$, where $\mathcal{H}(e_t)$ is the Hilbert space spanned by the components of the innovations $e_t = x_t - Hs_{t|t-1}$. The set $\{e_1, e_2, \ldots, e_T\}$ is a basis of $\mathcal{H}_x$. The projection $s_{t|T}$ can now be written as

$$s_{t|T} = \mathcal{P}(s_t|X) = \mathcal{P}(s_t|X_1^{t-1}) + \mathcal{P}(s_t|e_t, e_{t+1}, \ldots, e_T)$$

$$= s_{t|t-1} + \sum_{k=t}^{T} (s_t, e_k) \Sigma_k^{-1} e_k$$

where Σ_k has been defined in Equation (7c), and $(s_t, e_k) = \mathbb{E}(s_t e_k')$. A different system of iterative equations arises out of the need to calculate the covariances (s_t, e_k). For $t = T, T-1, \ldots, 1$:

$$s_{t|T} = s_{t|t-1} + P_{t|t-1} r_{t-1} \tag{11}$$

$$r_{t-1} = H\Sigma_t^{-1} e_t + \sum_{j=t+1}^{T} L_t' L_{t+1}' \cdots L_{j-1}' H' \Sigma_j^{-1} e_j, \quad L_t := (F - K_t H) \tag{12}$$

$$r_T = 0 \tag{13}$$

The vectors r_T can be calculated iteratively by

$$r_{t-1} = H'\Sigma_t^{-1} e_t + L_t' r_t$$
$$r_T = 0$$

It is shown in [19] that one can get to this sets of equations by starting from Equations (10). Note also that no inversion of $P_{t+1|t}$ is needed in these recursions[11].

$\square$

Equations (8a) and (10b) can be used to define an iterative system of equations for the *Lag-One Covariance Smoother*. It will be needed to obtain expected sufficient statistics for the EM algorithm.

---

[11]The `R`-package `KFAS` uses these equations to compute the Kalman smoother. Again it can be seen that different equations can be used to compute the smoothing projections.

Lemma 2.10 (Lag-One Covariance Smoother – version 1) Consider model (6) and define $P_{t_1,t_2|u} := \mathbb{E}(s_{t_1} - s_{t_1|u})(s_{t_2} - s_{t_2|u})'$. Then the lag-one covariance smoother $P_{t,t-1|T}$, $t = 2, 3, \ldots, T$ can be defined by:

$$P_{T,T-1|T} = (I_m - P_{T|T-1}H'\Sigma_T^{-1}H)FP_{T-1|T-1} \tag{14}$$

and for $t = T, T-1, \ldots, 3$:

$$P_{t-1,t-2|T} = P_{t-1|t-1}J'_{t-2} + J'_{t-1}(P_{t,t-1|T} - FP_{t-1|t-1})J'_{t-2} \tag{15}$$

where J_t is defined in Equation (10a).

Proof. The proof uses the classical recursive equations and can be found in [48]. □

Lemma 2.11 (Lag-One Covariance Smoother – version 2) The covariance matrix $P_{t,t-1|T} \in \mathbb{R}^{m \times m}$, $t = 2, \ldots, T$ defined in Lemma 2.10 can also be calculated by

$$P_{t,t-1|T} = L_{t-1}P_{t-1|t-2} - P_{t|t-1}N'_{t-1}L_{t-1}P_{t-1|t-2} \tag{16}$$

where $L_t = F - K_tH$ and $N_{t-1} = H'\Sigma_t^{-1}H + \sum_{j=t+1}^{T} L'_t L'_{t+1} \cdots L'_{j-1}H'\Sigma_j^{-1}HL_{j-1} \cdots L_{t+1}L_t$ are both $m \times m$ matrices.

Proof. A proof of these recursions can be found in [19]. □

2.4 Expectation Maximization (EM) Algorithm

The EM Algorithm [16] is an iterative procedure to compute the maximum likelihood estimates for a given probability model with parameters $\tau \in \mathcal{T}$ (where the parameter space is not specified explicitly) under the setting of incomplete data. One assumes that only part of the sample can be directly observed. There are two sample spaces $\mathcal{A}$ and $\mathcal{B}$, where $b \in \mathcal{B}$ (*incomplete data*) can be, whereas $a \in \mathcal{A}$ (*complete data*) can not be observed directly. In addition there exists a mapping $\iota : \mathcal{A} \to \mathcal{B}$, $\iota(a) = b$, meaning that a can be indirectly observed through b. All we know is that $a \in \iota^{-1}(\{b\}) \subseteq \mathcal{A}$.

If $f_a(a|\tau)$ and $f_b(b|\tau)$ denote the respective sampling densities, then $f_b(b|\tau) = \int_{\iota^{-1}(\{b\})} f_a(a|\tau)da$.

The aim of the EM Algorithm is to find $\hat{\tau} = \arg\max_{\tau \in \mathcal{T}} f_b(b|\tau)$ given observations b. The algorithm does this by making essential use of the sampling density of the complete data $f_a(a|\cdot)$.

Definition 2.12 (EM Algorithm):

Initialization Set the parameter $\tau^{(0)}$.

E-step Assume that $\tau^{(k)}$, $k \in \mathbb{N}$ is known already. Calculate

$$Q(\tau|\tau^{(k)}) = \mathbb{E}(\log f_a(a|\tau^{(k)})|b, \tau), \quad \tau \in \mathcal{T}$$

If the complete data sampling density comes from an exponential family, this means calculating the expected sufficient statistics $\mathbb{E}(t(a)|b, \tau^{(k)})$, where t is the sufficient statistic for the complete data density f_a.

M-step Recalculate the parameters. Choose $\tau^{(k+1)} = \arg\max_{\tau \in \mathcal{T}} Q(\tau|\tau^{(k)})$.

Repeat both steps until the relative change of the incomplete data likelihood is smaller than a tolerance level.

Remark 2.13 (*Remarks on the EM algorithm*).

- In the factor model setting, $a = (s_1, s_2, \ldots, s_T, x_1, x_2, \ldots, x_T)$ and $b = (x_1, \ldots, x_T)$. For an explanation of this notation and a detailed discussion on how to use the EM algorithm in the context of a state space model, see Section 4.4.

$\square$

3   The generalized dynamic factor model

This chapter follows the paper of Forni and Lippi [27] and starts with the most general definition of a generalized dynamic factor model. In order to meet the setting of the estimation procedures, which are introduced in the next section, several technical and practical assumptions are made.

3.1   General definition

Factor models are based on the assumption, that observations can be divided into a strongly and weakly dependent part (see Equation (1)). The concept of an *idiosyncratic process* or *weakly dependent process* is therefore central for establishing a factor structure for a time series.

An intuitive example is coming from finance. Suppose a portfolio manager can choose between infinitely many assets. He sets up a portfolio, a linear combination of those assets, where each one has a certain risk of return (variance). Now if you can distribute that risk by spreading it over your infinitely many different assets such that the risk of the portfolio will be zero, your assets don't share a common systemic risk, only idiosyncratic risk and you would say that your asset sequence is weakly dependent. In the end, you averaged the risk to zero.

In the context of a double-sequence $\boldsymbol{x}$, the motivating example was given for a fixed $t \in \mathbb{Z}$ and different $i \in \mathbb{N}$, but you can also take averages across cross section and time dimension. In the papers of [11] and [12] cross section averaging was performed on $\boldsymbol{x}_t$, that is $y_t = \lim_n \sum_{i=1}^n a_i^n x_{it}$, where the averaging sequence satisfies[12] $\lim_n \sum_{i=1}^n |a_i^n|^2 = 0$. The arithmetic mean $a_i^n = (1/n)$, $i = 1, \ldots, n$ is a simple example for such an averaging sequence. To extend the idea of averaging, suppose that for a fixed $n \in \mathbb{N}$ every a_i^n is a measurable complex function on Θ. Averaging is now done on $\boldsymbol{x}$ and it means building $s_n = \sum_{i=1}^n a_i^n(z) x_{it}$ where $a_i^n(z) x_{it} = \sum_{j=-\infty}^\infty a_{ij}^n x_{it-j}$.

This concept of averaging across the cross section and the time dimension is central for identification and estimation.

In order to properly define such a *dynamic averaging sequence*, a little amount of spectral theory is needed.

Definition 3.1 (*Dynamic averaging sequence*):   Let f_x^n be the spectral density of $(x_t^n)_{n \in \mathbb{N}}$ and denote $L_2^n(\Theta, f_x^n)$ the complex linear space of all n-dimensional vectors $a^n = (a_1^n, \ldots, a_n^n)^*$, where $a_i^n : \Theta \to \mathbb{C}$ and $(a^n, a^n)_{f_x^n} = \int_{-\pi}^\pi a^n(\theta)^* f_x^n(\theta) a^n(\theta) \, d\theta < \infty$. This scalar product makes $L_2^n(\Theta, f_x^n)$ a Hilbert space, also called the frequency domain of the process x_t^n.

Now let $a^n \in L_2^n(\Theta, I_n) \cap L_2^n(\Theta, f_x^n)$. Then $(a^n)_{n \in \mathbb{N}}$ is a *dynamic averaging sequence (DAS)* if $\lim_n ||a^n||_{I_n} = \lim_n \int_{-\pi}^\pi a^n(\theta)^* I_n a^n(\theta) \, d\theta = 0$.

This definition makes perfect sense, if the following definition of a weakly dependent process is considered. It also reveals in which sense the finite sums s_n are converging.

Definition 3.2 (*Idiosyncratic process*):   A double sequence $\boldsymbol{x}$ on $\mathbb{N} \times \mathbb{Z}$, where $(x_t^n)_{t \in \mathbb{Z}}$ fulfills Assumption A for all $n \in \mathbb{N}$, is called *idiosyncratic* or *weakly dependent* if for any DAS $(a^n)_{n \in \mathbb{N}}$ the following holds

$$\lim_n \mathbb{V}(a^n(z) x_t^n) = \lim_n ||a^n(z) x_t^n||_{I_n} = \lim_n \int_{-\pi}^\pi a^n(\theta)^* f_x^n(\theta) a^n(\theta) \, d\theta = \lim_n ||a^n||_{f_x^n} = 0.$$

---

[12]Note that $(x_{it}, \ i \in \mathbb{N})$ is sequence of square integrable stochastic variables and that y_t is the L_2-limit of the finite sums $\sum_{i=1}^n a_i^n x_{it}$.

Remark 3.3.

1. An idiosyncratic process can be characterized via its eigenvalue function $\lambda_1^x : \theta \mapsto \sup_{n \in \mathbb{N}} \lambda_{1n}^x(\theta)$. Here, $\lambda_{1n}^x(\theta)$ is the largest eigenvalue of the complex matrix $f_x^n(\theta)$ for a fixed $\theta \in \Theta$. The process $\boldsymbol{x}$ is idiosyncratic if and only if λ_1^x is essentially bounded on Θ (see for example [21], Theorem 2.2.4).

2. Although the last definition is very formal, the intuition behind remains. Note that the means square convergence of $y_t^n = \sum_{i=1}^n \sum_{h=-\infty}^{\infty} a_{ih} x_{j,t-h}$ to zero means that the variance of y_t^n vanishes when $n \to \infty$. A double sequence is idiosyncratic if the components are independent at all leads and lags with constant variances of its components, but an idiosyncratic process may also have some kind of weak dependencies over its cross-section and time.

3. The sequence $(a_1(\theta), a_2(\theta), \ldots)$ used above is often called *dynamic averaging sequence.*

$\square$

Definition 3.4 (*Generalized dynamic factor model*):   The data-generating process of $\boldsymbol{x}$ follows a generalized dynamic factor model or $\boldsymbol{x}$ can be represented as a q-dynamic factor sequence if

there exists a q-dimensional orthonormal white noise process $u_t = (u_{1t}, \ldots, u_{qt})'$,

a double sequence $\boldsymbol{\xi} = \{\xi_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ fulfilling Assumption A.

and square-summable filters $b_{ij}(z) = \sum_{k=-\infty}^{\infty} \beta_k^{(ij)} z^k, \quad i \in \mathbb{N}, \ j = 1, \ldots, q$

such that for any $i \in \mathbb{N}$ and $t \in \mathbb{Z}$

$$
\begin{aligned}
x_{it} &= \chi_{it} + \xi_{it} \\
&= b_{i1}(z)u_{1t} + b_{i2}(z)u_{2t} + \ldots + b_{iq}(z)u_{qt} + \xi_{it} = b_i(z)u_t + \xi_{it}
\end{aligned}
\tag{17}
$$

and the following holds

(gdfm a) $\xi_{it} \perp u_{j,t-k}$ for any $i \in \mathbb{N}$, $j = 1, 2, \ldots, q$ and $k \in \mathbb{Z}$

(gdfm b) $\boldsymbol{\xi}$ is weakly dependent or idiosyncratic (see Definition 3.2)

(gdfm c) $\boldsymbol{\chi}$ is strongly dependent, that is $\lambda_{qn}^\chi(\theta) \to \infty$ as $n \to \infty$ almost everywhere on Θ

Remark 3.5.

- Note that q does not depend on n.

- As in the static case, we have $f_x(\theta) = f_\chi(\theta) + f_\xi(\theta)$.

- The orthogonality assumption of common and idiosyncratic component has an economic interpretation: Sargent, T.J. (1989). Two models of measurements and the investment accelerator. Journal of Political, Economy 97, 251–287, (see [25], p. 1323)

- As shown in [27], Theorem 4, p. 117, the given assumptions of Definition 3.4 imply that the representation of Equation (17) is unique. This means that the q-dynamic factor sequence is identifiable (see Section 3.2).

$\square$

3.2 Identifiability

In the following identifiability for the (quasi-)static and dynamic case will be briefly summarized.

The (quasi-) static case. In this case the covariance matrix of the observations $x_t^n = L^n z_t + \xi_t^n$ is given by

$$\gamma_x^n(0) = \gamma_\chi^n(0) + \gamma_\xi^n(0) = L\gamma_z^n(0)L' + \gamma_\xi^n(0), \quad n \in \mathbb{N}. \tag{18}$$

In this context two problems of identifiability arise. The first deals with the unique decomposition of a given $\gamma_x^n(0)$ into $\gamma_\chi(0) \in \mathbb{R}^{n \times n}$, which is positive semi-definite, singular and symmetric, and $\gamma_\xi(0)$, which is positive semi-definite and diagonal, such that Equation (18) is satisfied.

The second problem deals with finding all possible L and $\gamma_z(0)$ such that $\gamma_\chi(0) = L\gamma_z(0)L'$. In fact if z_t is a static factor then all static factors can be determined by choosing some non-singular, quadratic matrix R and defining $\tilde{z}_t := R^{-1}z_t$ and $\tilde{L} := LR$[13]. If $\mathbb{E}z_t z_t' = I_r$ is assumed, then L is unique up to right-multiplication by orthogonal matrices. This is easily seen by $I_r = \mathbb{E}\tilde{z}_t\tilde{z}_t' = R^{-1}I_r(R^{-1})'$, which implies that R^{-1} must be orthogonal. An immediate implication of this is

> The static factors can only be identified up to a non-singular linear transformation and a rotation respectively. Therefore extracting common static factors means estimating the r-dimensional linear space of the static factors (see Section 3.3).

The answer to the first problem is, that $\gamma_\chi(0)$ and $\gamma_\xi(0)$ are generically unique if the dimension of static factors $r = \mathrm{rk}\gamma_\chi(0)$ is smaller than or equal to the so-called *Lederman bound*, i.e.

$$r \leq \frac{2n+1}{2} - \sqrt{\frac{(2n+1)^2}{4} - n^2 + n} \tag{19}$$

This inequality results from the comparison of the number of parameters on both sides of Equation (18). The number of parameters on the left side is $\frac{1}{2}n(n+1)$. The number of free parameters of the right side of this equation is $nr + n$. In order to guarantee identifiability the condition that $L'L$ or $L'\gamma_\xi(0)^{-1}L$ is a diagonal matrix is imposed (see [39]). This implies $\frac{1}{2}r(r-1)$ restrictions. By building the difference between the known parameters and the unknown parameters $\frac{1}{2}n(n+1) - nr - n + \frac{1}{2}r(r-1) \geq 0$ and solving for r gives the inequality from above.

If strict inequality in Equation (19) holds, then one could say that the factor model gives a simpler interpretation of the data in the sense that the number of parameters needed to describe the second moments is smaller than the free parameters of the covariance matrix of the observations.

Equation (19) is of importance for my simulation study, as the number of static factors and the dimension of the cross-section are not independent from each other. For example, generating data with 5 variables which are driven by 3 static factors is not feasible.

The dynamic case where f_ξ is diagonal, has been studied by [46] and they showed that for a given spectral density of the observations f_x, the spectral density matrices f_χ and f_ξ are generically unique for $q \leq n - \sqrt{n}$, where $q = \mathrm{rk}f_\chi$, and satisfy

---

[13]See [14]

$$f_x(\theta) = w(e^{-i\theta})f_u(\theta)w(e^{-i\theta})^* + f_\xi(\theta), \quad \theta \in \Theta.$$

The case of an approximate factor structure. It can be easily seen that in case the number of free parameters of $\gamma_x(0)$ is equal to the number of free parameters of $\gamma_\xi(0)$, no identifiability is possible for finite cross-section dimension n. Therefore in the approximate setting one looks at sequences of factor models

$$x_t^n = L^n z_t + \xi_t^n, \quad n \in \mathbb{N} \tag{20}$$

and at sequences of their moment (covariances in the static- and quasi-static case and spectral densities in the dynamic case) relations.

In the case of a (quasi-) static approximate factor model, the problem of uniquely determining $\gamma_\chi^n(0)$ and $\gamma_\xi^n(0)$ from $\gamma_x^n(0)$ can be solved for $n \to \infty$ ([11], [12]). Key conditions for asymptotic identifiability are

1. The first r eigenvalues of $\gamma_x^n(0)$ diverge to ∞ as $n \to \infty$

2. The $r+1, \ldots, n$-th eigenvalue of $\gamma_x^n(0)$ is uniformly bounded for all $n \in \mathbb{N}$

For *generalized dynamic factor models* a similar characterization has been given by [27]. The necessary and sufficient key conditions in order to represent the double sequence (x_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$) by a sequence of generalized factor models 20 are that

1. the first q eigenvalues of the spectral density matrix $f_x^n(\theta)$ diverge almost everywhere in Θ as $n \to \infty$

2. the $(q+1)$-th eigenvalue of $f_x^n(\theta)$ is uniformly bounded for almost all $\theta \in \Theta$ and for all $n \in \mathbb{N}$.

They also showed that Assumptions (gdfm b) and (gdfm c) imply asymptotic identifiability.

Remark 3.6. In fact for finite n and T we are working with an unidentified model.   □

3.3   Static factors

Consider the generalized dynamic factor sequence of Definition 3.4, which can also be written as $(x_t^n, \ t \in \mathbb{Z})_{n \in \mathbb{N}}$. Static factors will be defined for a fixed $n \in \mathbb{N}$.

Definition 3.7 (Static factors):   Let $(\chi_t^n)_{t \in \mathbb{Z}}$ be an n-dimensional stationary process fulfilling Assumptions A. The r dimensional vector process (z_t) is called *static factor process* of dimension r, if $\chi_t = L^n z_t$ with $L^n \in \mathbb{R}^{n \times r}$ for all $t \in \mathbb{Z}$.

A *minimal static factor process* is a static factor process with minimal dimension $r \in \mathbb{N}$ such that $\chi_t^n = L^n z_t$.

Lemma 3.10 in Section 3.4 gives a structural interpretation of static factors. Static factors can also be motivated from a Hilbert-space geometric point of view [11], [12], which is related

to Section 3.1. Consider the sequence $\boldsymbol{x}_t$ for a fixed $t \in \mathbb{Z}$. Now average across the x_{it} by talking the weights $\alpha \in \ell_2(\mathbb{N})$, i.e. $\alpha = (\alpha_i)_{i \in \mathbb{N}}$ and $\sum_{j=1}^{\infty} |\alpha_i|^2 < \infty$, and define the so-called *static aggregation set*[14] $\mathcal{A}^* := \{y \in L_2(\mathbb{P}), \ \exists y_n = \sum_{j=1}^{n} \alpha_i x_{it}, \ y_n \overset{L_2}{\to} y\}$. The components of $z_t = (z_{1t}, \ldots, z_{rt})$ are a basis of $\mathcal{A}^*$. This averaging is often called *de-noising* and also exists in the finite sample case (see Section 4.1). Sloppily said, the static factors are the essence that is left after de-noising the data.

Remark 3.8.

1. The terminology *minimal static factor* [14] comes from structural analysis, where the state of a state space realization of the latent process χ_t^n is also a static factor. Static factors in this thesis are always meant to be minimal static factors.

2. Note that in Definition 3.7 the number of static factors may vary for different n. Assumption B3. will fix this by assuming that the minimal static factor dimension will be constant for all but finite n. This assumption also implies that the space of static factors is of finite dimension. Without this assumption, there doesn't have to exist a finite amount of static factors.

<div align="right">□</div>

3.4   Structural analysis

Having introduced the generalized dynamic factor model and explained what static factors and dynamic factors are, this section discusses the modeling of the latent process χ_t^n. This results in a restricted version of the model introduced in Definition 3.4 that will be used in the simulation study. The following assumptions are closely related to those of [25] and are therefore labeled as the assumptions of a *structural factor model* (see [25], p. 1323).

Assumption B (Structural factor model):

B1. For every $n \geq n_0$, the latent process χ_t^n has a rational spectral density with rank q almost everywhere on Θ. The rank q does not depend on the cross-sectional dimension n.

B2. The rational spectral factor $w(z)$ can be written as $w(z) = A_n k(z)$, where $A_n \in \mathbb{R}^{n \times r}$ and $k(z)$ is a $r \times q$ polynomial matrix which does not depend on n.

B3. Let $\lambda_j^n(\Sigma_\chi^n)$ be the j-th largest eigenvalue of the covariance matrix of the latent process χ_t^n. Then

$$\lim_{n \to \infty} \inf \frac{1}{n} \lambda_r^n(\Sigma_\chi^n) > 0$$

B4. The models $x_t^n = \chi_t^n + \xi_t^n$ are nested in n, meaning that χ_{it}, ξ_{it} do not depend on bigger n for $i \leq n$.

---

[14]In [27] an aggregation set for the dynamic factors is defined. Related to that definition I call $\mathcal{A}^*$ the static aggregation set, which is also defined in [11] by $\mathcal{P}_1$.

Remark 3.9 (Remarks on Assumption B).

- Following Remark C.1, Assumption B1. implies that the latent process has a Wold representation $\chi_t^n = \sum_{j=0}^{\infty} w_j u_{t-j}$, where $u_t \sim \mathrm{WN}(0, I_q)$.

- Assumption B4. implies that the sequence of rational transfer functions $(w^n(z))_{n \in \mathbb{N}}$ is nested.

- Assumption B2. can be replaced by the following Assumption

B2.* The dimension, m say, of a minimal state space realization of a stable and mini-phase spectral factor of $f_\chi^n(\cdot)$ is independent of n, for $n \geq n_0$.

For a every $n \in \mathbb{N}$, we choose the following state space realization for $w^n(z)$

$$\chi_t = H^n s_t \tag{21a}$$
$$s_{t+1} = F^n s_t + G^n u_{t+1} \tag{21b}$$

with $w^n(z) = H^n(I - F^n z)^{-1} G^n$. Even more can be said about $(H^n, F^n, G^n)_{n \in \mathbb{N}}$.

Theorem 3.10 ([21], page 33): Let $(w^n(z))_{n \in \mathbb{N}}$ be a nested sequence of rational transfer functions of dimensions $n \times q$. Every $w^n(z)$ has rank q and zeros only outside the unit circle. Let $n \geq n_0$ such that B2.* holds. Then there exist minimal stable state space realizations (21) of $w^n(z) = H^n(I - F^n z)^{-1} G^n$ such that F and G are independent of n and the sequence $(H^n)_{n \in \mathbb{N}}$ is nested.

The dimensions of the dynamic and static factors as well as the minimal state are related by $q \leq r \leq m$. The number of static factors is determined by Lemma 3.11.

Lemma 3.11 [see[14], page 215] Let χ_t be a stationary vector process of dimension n. Then the dimension of a minimal static factor is the rank, call it r, of the zero-lag variance matrix $\gamma_\chi(0) = \mathbb{E}\chi_t \chi_t'$.

Proof. Due to Equations (21), there exists a ρ-dimensional z_t, such that $\chi_t = L z_t$, $L \in \mathbb{R}^{n \times \rho}$, $q \leq \rho \leq m$. The equation $\gamma_\chi(0) = L\gamma_z(0)L'$ implies $\rho \geq r$. In order to show $\rho = r$, consider the factorization

$$\gamma_\chi(0) = (U_1, U_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} (U_1, U_2)', \quad U_i' = U_i^{-1}, \, i = 1, 2$$

where $\Lambda_1 \in \mathbb{R}^{r \times r}$ and $\Lambda_2 = 0$. The matrix $M := U_1 \Lambda_1^{1/2}$ has rank r and $\Sigma_\chi = MM'$. The equation

$$z_t = (M'M)^{-1} M' \chi_t$$

defines a r dimensional static factor with covariance matrix I. This definition is not unique. It is trivial to verify, by considering $\mathbb{E}(\chi_t - M z_t)(\chi_t - M z_t)'$, that $\chi_t = M z_t$. $\square$

Note that the dimension r is also independent of n for $n \geq n_0$, because $(H^n)_{n \in \mathbb{N}}$ is nested. Now

$$z_t = (M'M)^{-1}M'w(z)u_t = k(z)u_t \tag{22a}$$
$$\chi_t = Mk(z)u_t = w(z)u_t \tag{22b}$$

where $M \in \mathbb{R}^{n \times r}$ clearly depends on n. Thus Assumption B2. follows from Assumption B2.*.

- Assumption B3. implies that the number of static factors is unique and that $\mathrm{rk}A_n = r$ for $n \geq n_0$.

$\square$

Assumption B2. let's us define the static factors by $z_t := k(z)u_t$, where $k(z) = \sum_{j=0}^\infty k_j z^j$, $k_j \in \mathbb{R}^{r \times q}$. Both $w(z)$ and $k(z)$ are tall matrices, because their number of rows is larger than their number of columns. Each polynomial matrix can be represented by its Smith-McMillan Form $w(z) = u(z)d(z)v(z)$ (see [32]).

Definition 3.12 (Zeroless transfer function, [14], p. 216): An $n \times q$ transfer function $w(z)$ is called zeroless if the numerator polynomials of the diagonal matrix in its Smith-McMillan form are all equal to one.

We already know that $(w^n(z))_{n \in \mathbb{N}}$ can be parameterized by $(H^n, F, G)_{n \in \mathbb{N}}$, where $(H^n)_{n \in \mathbb{N}}$ is nested. Clearly $\mathrm{vec}(H^n, F, G) \in \mathbb{R}^{nm+m^2+mq}$. The following Theorem 3.13 states that for values of (H^n, F, G) which form an open and dense subset of $\mathbb{R}^{nm+m^2+mq}$ the transfer function $w(z)$ is zeroless.

Theorem 3.13 ([14], page 216): Consider an $n \times q$ rational transfer function $w^n(z)$ with a minimal state space realization (H^n, F, G) with state dimension m. If $n > q$ holds, then for given m, the transfer function $w^n(z)$ is zeroless for generic values of (H^n, F, G).

The property of being zeroless is important as the following Theorem 3.14 shows.

Theorem 3.14 ([14], page 216): Let the latent process χ_t satisfy Assumptions B1, B2*, B4 and let z_t be an associated minimal static factor of dimension r. Then the following statements for the process z_t are equivalent:

1. The spectral factors $k(z)$ of the spectral density f_z are zeroless

2. There exists a polynomial left inverse k^- corresponding to the Smith-McMillan form of $k(z)$ and thus the unobserved input process u_t is determined from a finite number of outputs $z_1, \ldots, z_\ell$ for some $\ell \in \mathbb{N}$

3. The process z_t is a stationary solution of a stable AR system

$$z_t = -e_1 z_{t-1} - \cdots - e_p z_{t-p} + \nu_t, \; e_j \in \mathbb{R}^{r \times r} \tag{23}$$

where $e(z) = I_r + e_1 z + \cdots + e_p z^p$ and $\det(e(z)) \neq 0$, $|z| \leq 1$ and ν_t is a zero mean white noise process with $\mathrm{rk}\Sigma_\nu = q$.

The following definition shall briefly summarize the arguments from above and introduce the model which will be used for the simulation study.

Definition 3.15 (Structural factor model):
    For any $t \in \mathbb{Z}$ and $n \in \mathbb{N}$:

$$x_t \;=\; \chi_t \;+\; \xi_t \;=\; L \quad z_t \;+\; \xi_t \tag{24a}$$
$$\scriptstyle (n \times 1) \qquad (n \times 1) \quad (n \times 1) \qquad (n \times r)\,(r \times 1) \quad (n \times 1)$$

$$e(z)\, z_t \;=\; b \quad u_t, \tag{24b}$$
$$\scriptstyle (r \times r) \qquad\quad (r \times q)$$

where $u_t \sim \mathrm{WN}(0, I_q)$ and $\det e(z) \neq 0, \ \forall |z| \leq 1$. For convenience we assume $e(z) = I_r + e_1 z + \ldots + e_p z^p$. Further

1. The processes x_t, z_t and ξ_t fulfill Assumption A

2. ξ_t is weakly dependent

3. χ_t is strongly dependent

4. $\mathbb{E}\xi_t \chi_s' = 0$ for all $s, t \in \mathbb{Z}$

5. The largest r eigenvalues of Σ_χ^n diverge to ∞ as $n \to \infty$ (see B3.)

6. The observation matrix L depends on the cross-section dim. n and will be nested.

3.5   Economic interpretation of dynamic factors

Although static factors do not have a direct economic interpretation and for forecasting purposes, little is to be gained from a clear distinction between the static factors and the dynamic factors ([6], p. 54), the dynamic factors indeed do have a structural economic interpretation.

    Structural analysis for multivariate time-series, like structural vector autoregressive models (SVAR), tries to answer the following macroeconomic research questions

> What and how many are the shocks that drive economic fluctuations? What is the relative importance of supply and demand disturbances? What are the effects of macroeconomic policies? ([24])

    The dynamic factors are the primitive or common shocks of the economy. The dimension q is the number of such shocks that drive the latent process. The identification of such shocks (supply shocks, demand shocks, fiscal or monetary policy shocks) and their impact on macro variables like GDP output or unemployment is topic of ongoing research ([24], [23]).

    The interpretation of u_t as innovations for the latent process χ_t^n is equivalent to the fact that $\chi_t = w(z)u_t$ is one wold-representation of the latent process χ_t. Assumptions B imply that $w(z)$ and $w^-(z)$ are rational and causal transfer functions, therefore $\mathcal{H}_\chi = \mathcal{H}_u$. Further $||w_j|| \to 0$ geometrically as the power series expansion of $w(z)$ is continuous and converges for all $|z| < r$ with $r > 1$.

4 Estimation of static factors

Section 3.2 already discussed the asymptotic identification of the latent and idiosyncratic component in a generalized dynamic factor model as well as the non-uniqueness of the static factors. The following estimation procedures are developed under models that are restricted versions of the structural generalized dynamic factor model (see Definition 3.15). For these models, the latent and common component can be identified in the finite sample case. A consequence of the imposed restrictions is that all three estimators are misspecified in the generalized context. Nevertheless they consistently estimate the space of the static factors.

4.1 Cross-section averaging

Starting with equation (24a), the idea is to find a weighting matrix $W \in \mathbb{R}^{n \times r}$ with $W'W = nI_r$ such that

$$n^{-1}W'x_t^n \overset{p}{\to} Oz_t, \qquad O \in \mathbb{R}^{r \times r}, \ \det(O) \neq 0. \tag{25}$$

To see how this can be achieved decompose the weighted cross-section

$$n^{-1}W'x_t = \underbrace{n^{-1}W'Lz_t}_{(a)} + \underbrace{n^{-1}W'\xi_t}_{(b)}$$

Assumption (gdfm b) implies that the eigenvalues of Σ_{ξ_t} are bounded for all $n \in \mathbb{N}$ (see Lemma C.6) and by using Theorem C.3, summand (b) converges component-wise in mean square to 0. It follows that (b) vanishes in probability. Part (a) converges to Oz_t if the condition

$$n^{-1}W'L \overset{p}{\to} O \tag{26}$$

holds.

Note that the weighting n-tuple $n^{-1}W_j$ fulfills $\left(n^{-1}W_j\right)'\left(n^{-1}W_j\right) = n^{-2}\sum_{i=1}^{n}w_{ij}^2 = \frac{1}{n}$ which converges to 0 as the cross-section dimension goes to ∞. Therefore for every $j = 1, \ldots, r$, the sequence $(n^{-1}W_j)$ is a static averaging sequence.

4.1.1 Method of asymptotic principal components

The method of principal components weights by $W := \hat{P} = (\hat{p}_1, \ldots, \hat{p}_r) \in \mathbb{R}^{n \times r}$, where $\hat{p}_j$ is the eigenvector corresponding to the j-th largest eigenvalue of the sample covariance matrix $S = (1/T)X'X$.

The estimator is derived by solving the following non-parametric[15] optimization problem

$$\min_{z_1, \ldots, z_T, L} V(Z, L) = \frac{1}{nT} \sum_{t=1}^{T} (x_{it} - L_{(i)}z_t)'(x_{it} - L_{(i)}z_t) \tag{27}$$

$$= (nT)^{-1} \operatorname{tr}\left[(X - ZL')(X - ZL')'\right] \tag{28}$$

---

[15]Non-parametric in the sense that the likelihood of $X = (x_1, \ldots, x_T)$ can not be described by a finite number of parameters.

$$\text{subject to} \quad L'L/n = I_r \tag{29a}$$
$$\text{or subject to} \quad Z'Z/T = I_r \tag{29b}$$

where $Z \in \mathbb{R}^{T \times r}$ is the static factor panel, that is $Z = (z_1, \ldots, z_T)' \in \mathbb{R}^{T \times r}$ and $X = (x_1, \ldots, x_T)' \in \mathbb{R}^{T \times n}$ is the observation panel.

Note that the solution is not unique (as noted in [5]). One solution can be obtained by minimizing over Z given L, which gives (after building the derivative with respect to Z^{16} and setting it equal to zero) together with the side condition in (29a) the matrix $\tilde{Z} = XL/n$. Then

$$V(\tilde{Z}, L) = (nT)^{-1}\text{tr}(XX') - (n^2 T)^{-1}\text{tr}(L'X'XL)$$

and this is equivalent to maximizing $\text{tr}(L'X'XL)$ subject to $L'L/n = I_r$. Following Remark 2.7 the solution is given by $\hat{L} = \sqrt{n}\hat{P}$.

Principal component estimator: For a given $T \times n$ panel $X = (x_1, \ldots, x_T)$, where $x_t = \chi_t + \xi_t$, the common component χ_t is estimated by setting

$$\hat{z}_t = n^{-1/2}\hat{P}'x_t \quad \text{and} \quad \hat{L} = \sqrt{n}\,\hat{P}, \qquad t = 1, \ldots, T \tag{30}$$

where $\hat{P}$ are the eigenvectors corresponding to the r largest eigenvalues of the matrix $X'X \in \mathbb{R}^{n \times n}$.

Remark 4.1.

- Equation (29a) is just a normalization condition. We could also use (29b) and minimize over L given Z. The optimization problem is then identical to maximizing $\text{tr}(Z'(XX')Z)$. This is solved by setting $\hat{Z} = \sqrt{T}\hat{V}$, where $\hat{V}$ are the eigenvectors corresponding to the r largest eigenvalues of the matrix $XX' \in \mathbb{R}^{T \times T}$ (see [5]).

  This solution would be preferable, if $T > n$.

$\square$

4.1.2   Asymptotic behavior of the PC-estimator

The most important asymptotic characteristics of an estimator are its consistency and its asymptotic distribution. The former has been studied in [49] and [5] as well as in [18]. The latter has been discussed in [4]. I follow the assumptions of Stock and Watson [49]. Suppose the n-dimensional process x_t has the representation $x_t = Lz_t + \xi_t$, where $L \in \mathbb{R}^{n \times r}$ depends on n.

Assumption C (Assumptions of Stock and Watson):

C1. $(L'L/n) \to I_r$ for $n \to \infty$

C2. $\mathbb{E}z_t z_t' = diag(\sigma_1, \ldots, \sigma_r)$ with $\sigma_1 > \sigma_2 > \cdots > \sigma_r > 0$.

---

[16] The definition of the derivative $\frac{\partial V}{\partial Z} = \left(\frac{\partial V}{\partial Z_{ij}}\right)_{i=1,\ldots,T, j=1,\ldots,r}$ can be found in [40], page 664.

C3. $|L_{ij}| < \overline{h}$ for all $i, j \in \mathbb{N}$

C4. $\frac{1}{T} \sum_{t=1}^{T} z_t z_t' \xrightarrow{p} \mathbb{E} z_t z_t'$

C5. $\lim_{n \to \infty} \sup_{t \in \mathbb{Z}} \sum_{s=-\infty}^{\infty} |\mathbb{E}(\xi_t' \xi_{t+s}/n)| < \infty$

C6. $\lim_{n \to \infty} \sup_{t \in \mathbb{Z}} \sum_{i=1}^{n} \sum_{j=1}^{n} |(\gamma_\xi^n(0))_{ij}/n| < \infty$

C7. $\lim_{n \to \infty} \sup_{s,t \in \mathbb{Z}} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} |cov(\xi_{is}\xi_{it}, \xi_{js}\xi_{jt})| < \infty$

Discussion of Assumptions 4.2. The Assumptions of Stock and Watson are closely related to the Assumptions made in Defintion 3.15 and to the Assumptions E of Doz, Giannone and Reichlin.

- Lemma C.5 implies that $\lambda_k(L'L/n) \to 1$, $k = 1, \ldots, r$ as $n \to \infty$, i.e. $\lambda_k(L'L) = O(n)$. Because the first r eigenvalues of LL' are equal to the eigenvalues of $L'L$, Assumption C1. implies that the first r eigenvalues of $\gamma_\chi^n(0)$ and therefore of $\gamma_x^n(0)$ diverge to ∞.

  A maybe more intuitive way to interpret this assumption is to say that each factor contributes to the average variance of x_{it} for every n and T. Let $L_{(i)}$ be the i-th row of $L \in \mathbb{R}^{n \times r}$. Then for all $t \in \mathbb{Z}$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{V} x_{it} \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(z_t' L_{(i)}' L_{(i)} z_t) = \mathbb{E} \left(\sum_{i=1}^{n} \left(z_t'(L_{(i)}' L_{(i)}/n) z_t \right) \right)$$
$$= \mathbb{E} \left(z_t'(L'L/n) z_t \right) \to \sum_{j=1}^{r} \mathbb{V} z_{jt}$$

- Assumptions C1. and C2. also imply that the static factors are identified up to a sign change. Let $\chi_t = L z_t = L R R^{-1} z_t$ where $R \in \mathbb{R}^{r \times r}$ is a non-singular matrix. The factorizations (L, z_t) and $(\tilde{L}, \tilde{z}_t) = (LR, R^{-1} z_t)$ are observationally equivalent and fulfill both Assumptions C. The first assumption implies that

$$\tilde{L}'\tilde{L}/n \to I_r, \quad \tilde{L}'\tilde{L}/n = R'(L'L/n)R \to R'R$$

  and this implies that R is (asymptotically) an orthogonal matrix. If some $\tilde{z}_t$ fulfills the second assumption with $\tilde{z}_t = R z_t$, then $diag(\sigma_1, \ldots, \sigma_r) = R\mathbb{E}(z_t z_t')R'$. It follows that $R^{-1} diag(\sigma_1, \ldots, \sigma_r)R = \mathbb{E}(z_t z_t')$ and therefore $\mathbb{E}(z_t z_t')$ must converge to a diagonal matrix. This asymptotic identification up to a sign change is no restriction, as the static factors in an approximate factor model are only asymptotically identified.

- Assumptions C3. and C4. are of technical nature. The former is related to Assumption E1.

- Assumption C5. allows for idiosyncratic serial correlations, whereas C6. allows for cross-correlation between the idiosyncratic variates. Both assumptions are very central to the generalized dynamic factor model class. The last Assumption C7. is again technical. Note that all three Assumptions do not demand a stationary process (ξ_t^n). This generality is needed to deal with time-varying factor loading matrices, in particular factor loadings with

a stochastic drift. This will become important in the Empirical Section 6 as macroeconomic multivariate time series seem to suffer from structural breaks.

Assumptions C5., C6., C7. are related to the assumption of bounded idiosyncratic covariances, i.e. $\exists M$ such that $||\gamma_\xi^n(0)|| \leq M < \infty$ for all $n \in \mathbb{N}$.

Note that Assumption C6. implies

$$\sum_{i=1}^n \sum_{j=1}^n |(\gamma_\xi^n(0))_{ij}/n| = 1'|(\Sigma_\xi^n)|1/n = 1'/\sqrt{n}|\gamma_\xi^n(0)|1/\sqrt{n} < \infty$$

where $1 = (1,\ldots,1)' \in \mathbb{R}^n$. Again using inequalities from Theorem C.2, this does not imply that the eigenvalues of $\gamma_\xi^n(0)$ are bounded for all $n \in \mathbb{N}$. In the paper of Stock and Watson the boundedness of the eigenvalues of the estimates idiosyncratic covariance matrix $\hat{\gamma}_\xi^n(0)$ is deducted from Assumption C6. by using Assumptions C5. and C7.:

Lemma 4.3 (*see [49], Appendix*) Define the set $M = \{a \in \mathbb{R}^n : a'a/n = 1\}$. Then

$$\sup_{a \in M} n^{-2} a'\hat{\gamma}_\xi^n(0)a \xrightarrow{p} 0, \quad n, T \to \infty$$

$\square$

The consistency proof is based on the idea that the largest eigenvalue of $T^{-1}X'X$ converges in probability to the largest eigenvalue of the unobservable sample covariance matrix of the latent component $T^{-1}LZ'ZL'$. This is the statement of the next Lemma:

Lemma 4.4 Define $R(a) := (n^2T)^{-1}a' \sum_{t=1}^T x_t x_t' a$ and $R^*(a) := (n^2T)^{-1}a' \sum_{t=1}^T Lz_t z_t' L' a$. Then

$$|\sup_{a \in M} R(a) - \sup_{a \in M} R^*(a)| \xrightarrow{p} 0, \quad n, T \to \infty$$

where $M = \{a \in \mathbb{R}^n : a'a/n = 1\}$ as in Lemma 4.3.

The following Lemma follows from Lemmata 4.3 and 4.4 and refers to Equation (26).

Lemma 4.5 (*see [49], Appendix*) Let $\hat{L} \in \mathbb{R}^{n \times r}$ the matrix of eigenvectors corresponding to the r largest eigenvalues of $T^{-1}X'X$ scaled such that $\hat{L}'\hat{L}/n = I_r$. Define $S = diag(sgn(\hat{L}'\hat{L}))$. Then

$$S\hat{L}'L/n \xrightarrow{p} I_r$$

Theorem 4.6 (*Consistency of principal component estimator*): Let S_i denote a variable with value of ± 1, let $n, T \to \infty$ and suppose that Assumptions C hold. Then

$$S_i \hat{z}_{jt} - z_{jt} \xrightarrow{p} 0, \quad j = 1,\ldots,r \tag{31}$$

where $\hat{z}_t = n^{-1/2}\hat{P}x_t$ has been defined in Equation (30).

Proof. See [49].   □

Remark 4.7. Again note that the z_{jt} are just one basis of the latent space $\mathcal{H}(\chi_{it}, \; i \in \mathbb{N})$. The statement of the last theorem is that the principal component estimator is a consistent estimator of the linear space of static factors.   □

4.2   Identification of a singular AR system

In Section 3.4 assumptions where given such that the static factors follow a singular AR process

$$z_t = -e_1 z_{t-1} - \cdots - e_p z_{t-p} + \nu_t. \tag{32}$$

The Yule-Walker equations are a standard method to identify vector auto-regessive systems,

$$(-e_1, \ldots, -e_p)\Gamma_p' = (\gamma_1, \ldots, \gamma_p) \tag{33a}$$

$$\Sigma_\nu = \gamma_0 + (e_1, \ldots, e_p)(\gamma_1, \ldots, \gamma_p)' \tag{33b}$$

In the case $q < r$ the matrix Γ_{p+1} will be singular and the Block-Toeplitz matrix $\Gamma_p \in \mathbb{R}^{rp \times rp}$ may be singular. It therefore may be the case that no unique solution of the Yule-Walker equations exists

Lemma 4.8 Let z_t be our r dimensional static factor process that fulfills Assumption A. Suppose $e(z)z_t = bu_t$, where $\det e(z) \neq 0$, $\forall |z| \leq 1$ and u_t is q dimensional white noise. The polynomial order of $e(z)$ is p. Then (i) if $q = r$ and $\mathrm{rk}(b) = r$ the auto-covariance matrix $\Gamma_{Z,r} > 0$ for all $r \in \mathbb{N}$. If (ii) $q < r$ and $\mathrm{rk}(b) = q$, then $\det \Gamma_{Z,p+1} = 0$.

Proof. Statement (i) is standard in the theory of stationary processes and will therefore not be proven here. For statement (ii) choose a $b_\perp \in \mathbb{R}^{r \times (r-q)}$ such that $b_\perp' b = 0$. Then the matrix $b_\perp(I_r, e_1, e_2, \ldots, e_p)$ is in the left kernel of $\Gamma_{Z,p+1}$, because

$$(I_r, e_1, e_2, \ldots, e_p) \; \mathbb{E} \begin{pmatrix} z_t z_t' \\ \vdots \\ z_{t-p} z_t' \end{pmatrix} = \mathbb{E}b(u_t z_t') = bb'$$

$$(I_r, e_1, e_2, \ldots, e_p) \; \mathbb{E} \begin{pmatrix} z_t z_{t-k}' \\ \vdots \\ z_{t-p} z_{t-k}' \end{pmatrix} = \mathbb{E}b(u_t z_{t-k}') = 0.$$

It follows that $b_\perp'(I_r, e_1, e_2, \ldots, e_p)\Gamma_{Z,p+1} = 0$.

                                                                                                        □

We skip the discussion of finding a solution at the population level and refer to [14] or [21]. One solution of the (singular) Yule-Walker equations is the so-called *minimum norm solution* which corresponds to taking the Moore-Penrose pseudo-inverse of Γ_p in order to determine $e(z)$.

If the static factors z_t were known, the matrix Γ_p would be estimated by the sample covariances

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T-j} z_{t+j} z_t' \;=\; \frac{1}{T} \sum_{s=1+j}^{T} z_s z_{s-j}'$$

$$\hat{\gamma}_{-j} = (\hat{\gamma}_j)' = \frac{1}{T} \sum_{t=1}^{T-j} z_t z_{t+j}'$$

The blocks of the sample covariance matrix $\hat{\Gamma}_p$ are $\{\hat{\Gamma}_p\}_{i,j} = \hat{\gamma}_{i-j}$, $i,j = 1, \ldots, p$. Define

$$\mathtt{Gp} := \begin{bmatrix} z_1 & z_2 & \cdots & z_p & z_{p+1} & \cdots & z_T & 0 & \cdots & 0 \\ 0 & z_1 & \ddots & & & & & z_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & & & \ddots & 0 \\ 0 & \cdots & 0 & z_1 & z_2 & \cdots & z_{T-p+1} & z_{T-p+2} & \cdots & z_T \end{bmatrix} \in \mathbb{R}^{pr \times T+p-1}$$

Then $\hat{\Gamma}_p' = \frac{1}{T} \mathtt{Gp}\, \mathtt{Gp}'$. The i-th row of the matrix $\mathtt{Gp}$ is

$$\mathtt{Gp}[i,] = (\; \underbrace{0, \ldots, 0}_{i-1 \text{ elements}}, \; z_1, z_2, \ldots, z_T, \; \underbrace{0, \ldots, 0}_{p-i \text{ elements}} \;), \quad i = 1, \ldots, p$$

The determination of the rank of $\Gamma_p = \mathbb{E}(z_t', \ldots, z_{t-p}')'(z_t', \ldots, z_{t-p}')$ is not trivial. It can be done by looking at

$$\tilde{\Gamma}_p' = \frac{1}{T} \mathtt{Hp} \cdot \mathtt{Hp}', \quad \mathtt{Hp} = \begin{bmatrix} z_p & z_{p+1} & \cdots & z_T \\ z_{p-1} & & & z_{T-1} \\ \vdots & & & \vdots \\ z_1 & z_2 & \cdots & z_{T-p+1} \end{bmatrix} \in \mathbb{R}^{pr \times T-p+1}$$

The matrix $\mathtt{Hp}$ is simly a column-truncated version of $\mathtt{Gp}$. The j-th column of $\mathtt{Hp}$ is the $j + p - 1$-th column of $\mathtt{Gp}$.

Note that Γ_p and $\tilde{\Gamma}_p$ have the same kernel a.e., as $\det \Gamma_p = 0$ implies the existence of $a \in \mathbb{R}^{rp}$ with $\Gamma_p a = 0$. This implies that $(z_p', \ldots, z_1')a = 0$ a.e. and therefore $\mathtt{Hp}'a = 0$.

Estimation of $\mathrm{rk}\Gamma_p$ can be done heuristically by looking at the singular values of $\mathtt{Hp}$ and determining $s = \mathrm{rk}\Gamma_p$ by imposing a certain threshold on the minimal value of those eigenvalues.

Note that $\hat{\gamma}_j, \mathtt{Gp}, \mathtt{Hp}$ are unknown, because the static factors z_t are unobserved. Therefore z_t have to be replaced by their estimates $\hat{z}_t$ which are determined by principal component analysis or the Kalman smoothing algorithm (see next Sections).

Suppose we had successfully estimated the unknown integer $\mathrm{rk}\Gamma_p = s \le pr$. We write

$$\hat{\Gamma} = (O_1^p, O_2^p) \begin{pmatrix} \Lambda_1^p & 0 \\ 0 & \Lambda_2^p \end{pmatrix} \begin{pmatrix} O_1^{p\prime} \\ O_2^{p\prime} \end{pmatrix} \tag{34}$$

where Λ_i^p, $i = 1, 2$ are diagonal matrices containing the first s and last $pr - s$ eigenvalues of $\hat{\Gamma}_p$. Note that although $\text{rk}\Gamma_p = s \leq pr$, the last $pr - s$ eigenvalues are in general not zero. Also looking at the eigenvalues of $\hat{\Gamma}_p$ is not recommended, because the kernel of $\hat{\Gamma}_p$ is different from the kernel of Γ_p.

Theorem 4.9 ([21], page 68): Let the process z_t be generated according to (32). If $\text{rk}\Gamma_p = s < pr$ and if all nonzero eigenvalues of Γ_p are distinct, then

$$-(\hat{e}_1, \ldots, \hat{e}_p) = (\hat{\hat{\gamma}}_1, \ldots, \hat{\hat{\gamma}}_p) O_1^p (\Lambda_1^p)^{-1} O_1^{p\prime}, \tag{35}$$

where $\hat{\hat{\gamma}}_j = \frac{1}{T} \sum_{t=1}^{T-j} \hat{z}_{t+j} \hat{z}_t'$, is a consistent estimator of the minimum norm solution of the Yule-Walker equations. We call the estimator in (35) the *Yule-Walker estimator of the minimum norm solution* .

In the standard case where $\text{rk}\Sigma_\nu = r$ and the static factors z_t were known, the Yule-Walker estimator always yields a stable autoregression. The next Theorem 4.10 deals with the case where $s \leq pr$ and where the factors have to be estimated.

Theorem 4.10 ([14], page 220):

1. If $\text{rk}\Gamma_p = pr$ holds, then the Yule-Walker estimator corresponding to (33), i.e. when the γ_j are replaced by $\hat{\hat{\gamma}}_j$ yields a stable autoregression

2. For $\text{rk}\Gamma_p = s < pr$, the solution (35) corresponds to a stable autoregression

4.3 Two step approach by Doz, Giannone, Reichlin

In the paper of Doz, Giannone and Reichlin (DGR) [18] the estimators of the factors z_t minimize the least-squares distance between z_{jt} and a suitable linear combination of x_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$, for all $j = 1, \ldots, r$.

Assume that the process $(x_t^n)_{t \in \mathbb{Z}}$ fulfills Assumption A and has a static factor representation

$$\underset{(n \times 1)}{x_t} = \underset{(n \times r)}{L} \underset{(r \times 1)}{z_t} + \underset{(n \times 1)}{\xi_t} \tag{36a}$$

$$z_t = -e_1 z_{t-1} - e_2 z_{t-2} - \ldots - e_p z_{t-p} + \nu_t \tag{36b}$$

and assume that the assumptions of Definition 3.15 are fulfilled.

The two-stage estimator: At the population level, the estimator is the component-wise orthogonal projection of z_t on the linear sub-space $\mathcal{H}_x := \mathcal{H}_x(T)$. Let z_t be our r dimensional static factor process, with $z_{it} \in L_2(\mathbb{P})$. Theorem 2.1 tells us that each z_t can be uniquely decomposed into $z_t = z_{t|T} + (z_t - z_{t|T})$, where $z_{t|T} = \mathcal{P}(z_t|X_1^{nT})$, $z_{j,t|T} \in \mathcal{H}_x$ and $(z_{jt} - z_{j,t|T}) \in \mathcal{H}_x^\perp$, $j = 1, \ldots, r$. For given matrices H, F, G, the Kalman smoother calculates $z_{t|T} = \sum_{j=1}^T a_{jt} x_j$, $a_{jt} \in \mathbb{R}^{r \times n}$, $j = 1, \ldots, n$, $t = 1, \ldots, T$. As $\mathcal{P}$ is an orthogonal projection,

$\mathbb{E}z_t x_s' = \mathbb{E}z_{t|T}x_s' = \sum_{j=1}^T a_{jt}\mathbb{E}x_j x_s'$ for $s = 1, \dots, T$. For $\underline{X} = \mathrm{vec}(x_1, \dots, x_T) \in \mathbb{R}^{nT \times 1}$ the last equations can be combined to

$$\mathbb{E}z_t \underline{X}' = (a_{1t}, \dots, a_{Tt})\, \Gamma_X \tag{37}$$

where $\Gamma_X = \mathbb{E}(\underline{X}\,\underline{X}')$. Equation (37) has a unique solution as Γ_X is the sum of a positive semi-definite and a positive definite matrix (due to Assumption E8.) and has therefore full rank. The least squares estimation can therefore be written as

$$z_{t|T} := \mathcal{P}(z_t | X_1^{nT}) = \underset{(r \times rT)}{S_t'} \quad \underset{(rT \times rT)}{\Gamma_Z} \quad \underset{(rT \times nT)}{(I_T \otimes L')} \quad \underset{(nT \times nT)}{\Gamma_X^{-1}} \quad \underset{(nT \times 1)}{\underline{X}} \tag{38}$$

where $S_t' = (0, 0, \dots, 0, I_r, 0, \dots, 0)$ is a selection matrix having the identity matrix at the t-th block, $\Gamma_Z = \mathbb{E}(\underline{Z}\,\underline{Z}')$, and $\underline{X} = \mathrm{vec}(X_1^{nT})$, $X_1^{nT} = (x_1, \dots, x_T) \in \mathbb{R}^{n \times T}$. Model (36) can also be written as $\underline{X} = (I_T \otimes L)\underline{Z} + \underline{\xi}$, where $\underline{Z} = \mathrm{vec}(z_1, z_2, \dots, z_T)$ and $\underline{\xi} = \mathrm{vec}(\xi_1, \xi_2, \dots, \xi_T)$. Then

$$\Gamma_X = (I_T \otimes L)\Gamma_Z(I_T \otimes L)' + I_T \otimes \gamma_\xi(0) \tag{39}$$

where $\Gamma_Z = (\mathbb{E}\underline{Z}\,\underline{Z}')$ is the $r(T+1) \times r(T+1)$ auto-covariance matrix of the process z_t.

Remark 4.11. Depending on which assumptions are made about the static factor process, Equation (39) can be further simplified. Suppose that z_t and ξ_t are white noise processes with covariance matrices I_r and I_n respectively. Then the observations x_t are uncorrelated and

$$z_{t|T} = (\mathbb{E}z_t x_t')(\mathbb{E}x_t x_t')^{-1}x_t = L'(LL' + I_n)^{-1}x_t,$$

as $\mathbb{E}z_t \underline{X}' = (0, \dots, 0, \mathbb{E}(z_t x_t'), 0, \dots, 0)$.
Using Lemma C.7 with $A = C = I_r, U = L, V = L'$ this expression becomes

$$L'(I_r - L(L'L + I_r)^{-1}L')x_t = (L'L + I_r - L'L)(L'L + I_r)^{-1}L'x_t = (L'L + I_r)^{-1}L'x_t.$$

An estimator for $z_{t|T}$ in this case is $\hat{z}_{t|T} = (\hat{L}'\hat{L} + I_r)^{-1}\hat{L}'x_t$, where the columns of $\hat{L}$ are the eigenvectors corresponding to the largest r eigenvalues of $X'X$. As the eigenvectors are orthonormal, this estimator is proportional to the consistent PC-estimator of Section 4.1.1. □

In order to derive an estimator for $z_{t|T}$ one further assumes

Assumption D.

D1. The idiosyncratic process is white-noise with diagonal covariance matrix $\gamma_\xi^n(0)$. The restricted covariance matrix will be denoted by $\gamma_{\xi,R}(0)^n$ in the following[17].

---

[17]The superscript n for the cross-section dimension will be abandoned if the cross-section is fixed.

and rewrites model (36) in state-space form. For $t \in \mathbb{Z}$

$$
x_t = \underbrace{(L, 0, \ldots, 0)}_{H} \overbrace{\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix}}^{s_t} + \xi_t = H s_t + \xi_t \tag{40a}
$$

$$
s_{t+1} = \underbrace{\begin{pmatrix} -e_1 & -e_2 & \cdots & -e_p \\ I_r & 0 & 0 & 0 \\ 0 & \cdots & \ddots & \vdots \\ 0 & \cdots & I_r & 0 \end{pmatrix}}_{F} s_t + (b', 0, \ldots, 0)' u_{t+1} = F s_t + G u_{t+1} \tag{40b}
$$

where $s_t \in \mathbb{R}^{rp \times 1}$ (random), $H \in \mathbb{R}^{n \times rp}$, $F \in \mathbb{R}^{rp \times rp}$ and $G \in \mathbb{R}^{rp \times q}$.

Equation (47b) is the *companion form* of Equation (46b) and the state is equal to the stacked factors.

Two-stage estimation of static factors 4.12.   The matrices $H, F, G, \gamma_\xi(0)$ are unknown and observations are available for $t = 1, \ldots, T$. Assuming, that we already know $r, q, p \in \mathbb{N}$, the static factors are estimated in two-stages

1. *Estimate by parameters of the state space model.*

    - $\hat{L} \in \mathbb{R}^{n \times r}$ is equal to the eigenvectors corresponding to the largest r eigenvalues of $X'X$, where $X = (x_1, \ldots, x_T) \in \mathbb{R}^{n \times T}$
    - $\hat{\gamma}_\xi(0) = diag(\Psi_{11}, \ldots, \Psi_{nn})$, with $\Psi = \frac{1}{T} X'X - \hat{L}\hat{L}'$.
    - The parameters $e_1, \ldots, e_p$ are estimated by solving (singular) Yule-Walker equations (see Section 4.2).
    - $\hat{b} \in \mathbb{R}^{n \times q}$ is equal to the first q eigenvectors of the spectral decomposition of $\hat{\Sigma}_\nu$ (see Equations (33)).

2. *Treat the parameters as known and estimate $s_{t|T}$ through the Kalman Filter apparatus.*

    - The best approximation of s_1 given no further information, i.e. $s_{1|0}$ (see Section 2.3 for notation), is set to the principal component estimates, that is $\hat{s}_{1|0} := (\hat{z}_{\hat{p}}^{PC}, \ldots, \hat{z}_1^{PC})$, where p is the order of polynomial $e(z)$.
      The variance of s_1 is determined by the Lyapunov[18] Equation. As the static factors are assumed to be stationary, the state process is also stationary. Therefore by considering equations (40)

$$
\Gamma_p := \mathbb{E} s_{t+1} s_{t+1}' = \mathbb{E} \left(F \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p+1} \end{pmatrix} + \begin{pmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \left((z_{t-1}', z_{t-2}', \ldots, z_{t-p}') F' + (\nu_t', 0, \ldots, 0) \right)
$$

$$
= F \Gamma_p F' + G G'
$$

---

[18] Aleksandr Mikhailovich Lyapunov, *June 6 1857, †November 3 1918.

Applying the vec operator (see Definition C.9) yields

$$\text{vec}(\Gamma_p) = (F \otimes F)\,\text{vec}(\Gamma_p) + \text{vec}(GG')$$
$$(I_{r^2 p} - F \otimes F)\text{vec}(\Gamma_p) = \text{vec}(GG')$$
$$\text{vec}(\Gamma_p) = \underset{(r^2 p \times r^2 p)}{(I_{r^2 p} - F \otimes F)^{-1}}\ \underset{(r^2 p \times 1)}{\text{vec}(GG')} \qquad (41)$$
$$\underset{(r^2 p \times 1)}{}$$

The initial covariance matrix $\mathbb{V}s_1$ is based on $(I_{r^2 p} - \hat{F} \otimes \hat{F})^{-1}\hat{G}\hat{G}'$.

- Make use of the recursive equations in Section 2.3 in order to calculate $\hat{s}_{t|T}$ for $t = 1, \ldots, T$.

Two-stage estimator for static factors: The estimator of z_t is then equal to the first r components of the estimated smoothed states, i.e. $\hat{z}_{t|T} := (\hat{s}_{1,t|T}, \ldots, \hat{s}_{r,t|T})$, for $t = 1, \ldots, T$.

$\square$

Consistency of the two-stage estimator. The linear orthogonal projection of the components of the static factor on the Hilbert space spanned by the components of the observations is consistent for large cross-section and time dimension.

However this statement is only proved for $r = q$, whereas the case $r > q$ has not been addressed yet. The reason why this case needs to be addressed separately is given in Facts 4.15.

Assumption E (Assumptions of Doz, Giannone and Reichlin):

E1. Uniformly bounded variance, $\exists M > 0$ s.t. $\mathbb{E}(x_{it} x_{it}) \le M$ for all $i \in \mathbb{N}$.

E2. $z_t = \sum_{j=0}^\infty c_j \varepsilon_{t-j}$, with $\sum_{j=0}^\infty ||c_j|| < \infty$ and (ε_t) is stationary of order 4.

E3. $\xi_t = \sum_{j=0}^\infty d_j \zeta_{t-j}$, with $\sum_{j=0}^\infty ||d_j|| < \infty$ and (ζ_t) is white noise with uniformly bounded fourth moments, $\exists M \in \mathbb{N}$ s.t. $\mathbb{E}\zeta_{it}^4 \le M,\ \forall i \in \mathbb{N}, t \in \mathbb{Z}$.

E4. $\mathbb{E}z_t z_t' = I_r$ *(Normalization)*

E5. $\limsup_{n \to \infty} \sum_{s=-\infty}^\infty ||\gamma_\xi(s)|| < \infty$ *(Weak dependence)*

E6. $\liminf_{n \to \infty} \frac{1}{n}\lambda_r(\Sigma_\chi^n) > 0$ *(Persistence of static factors)*

E7. $\limsup_{n \to \infty} \frac{1}{n}\lambda_1(\Sigma_\chi^n) < \infty$

E8. $\inf_{n \in \mathbb{N}} \lambda_n(\Sigma_\xi^n) > 0$

E9. For every $n \in \mathbb{N}: \lambda_1^\chi > \lambda_2^\chi > \ldots > \lambda_r^\chi, \quad \lambda_k^\chi = \lambda_k(\Sigma_\chi^n)$

Assumption F (Further assumptions):

F1. The noise ν_t can be written as $\nu_t = bu_t$, where $b \in \mathbb{R}^{r \times q}$ and u_t is a q dimensional white noise.

Remark 4.13 (Discussion of Assumptions).

- Assumption E1. is a technical assumption and related to Assumption C3. of Stock and Watson. It is needed in the proof of Theorem 4.16.

- Assumptions E2. and E3. define the processes χ_t and ξ_t a.e. and guarantee the existence of their spectral densities. The conditions on the moments of ε_t and ζ_t are needed for proving the consistency of the sample second moments and therefore of the AR coefficients.

- Although I assumed ξ_t to be weakly dependent, Assumption E5. is the original corresponding assumption in [18]. It also implies weak dependence as

$$||\Sigma_\xi^n(\theta)|| = ||(1/2\pi) \sum_{s=-\infty}^{\infty} \gamma_\xi(s)e^{i\theta s}|| \leq (1/2\pi) \sum_{s=-\infty}^{\infty} ||\gamma_\xi(s)|| < \infty, \quad \forall n \in \mathbb{N}.$$

- The assumption $\mathbb{E}z_t z_t' = I_r$ implies that the static factors are identified up to orthogonal rotation. To see this choose a non-singular quadratic matrix R and define $\tilde{z}_t := Rz_t$. Then $\mathbb{E}\tilde{z}_t\tilde{z}_t' = RR' = I_r$. In contrast to Stock and Watson [49], who assume $L'L/n = I_r$, a specific basis of $\mathcal{H}(\chi_{it}, i \in \mathbb{N})$ is chosen. Define $A := L'L$ and look at its spectral decomposition $A = UDU'$, where $U' = U^{-1}$. Define $L^* := LU$ and $z_t^* := U'z_t$. It can be easily seen, that this specific selection does not change model except for the case that the loading matrices are not nested any more. In the following the static factors z_t will be that specific basis and are therefore identified up to a sign change.

- Assumptions E6. and E7. guarantee that r static factors have a non-negligible influence on all the observations and that this influence is somehow "stationary" for all observations. Both state that the first r eigenvalues of the latent covariance matrix diverge of linear order to infinity.

  These assumptions also imply that the rank of the factor loading matrix L^n must be equal to r for $n \geq n_0 \in \mathbb{N}$ as $\Sigma_\chi^n = L^n(\mathbb{E}z_t z_t')L^{n'}$ and $\text{rk}\mathbb{E}z_t z_t' = r$.

- Assumption D1. is necessary to write the model in state-space form and use the Kalman Filter in order to estimate the static factors. It is a restriction and implies that this approximation model is a quasi-static strict factor model. It also means that only the restricted covariance structure of ξ_t can be used for the estimation of z_t.

- Assumption F1. is not made in the paper [18] and it challenges their proofs as the consistency results of their estimator depend on the smallest eigenvalue of $\mathbb{E}\nu_t\nu_t'$. In general $T > p$ and this implies that Γ_Z is not invertible, if the static factor process is generated by a singular AR system (see Facts 4.15).

$\square$

The consistency of the projection of Equation (38) will be shown in two steps, first at the population level, when the second moments and the parameter matrices of the system are known, and then at the sample level, when only the observations X_1^{nT} are known. The full proofs are given in [18], I will only sketch the main points.

Theorem 4.14. Consider model (36) where the dimension of the dynamic factors is equal to the dimension of the static factors. Under Assumptions E, $z_{t|T} := \mathcal{P}(z_t|X_1^{nT})$ is a consistent estimator for z_t when $n, T \to \infty$, i.e.

$$z_{t|T} \xrightarrow{p} z_t, \quad n, T \to \infty \tag{42}$$

In order to proof Theorem 4.14 the following facts will be used:

4.15. Facts

1. $\lambda_{\max}(\Gamma_Z^{-1}) = O(1)$ as shown in [18], Lemma 1.

2. By using $\underline{X} = (I \otimes L)\underline{Z} + \underline{\xi}$, $\Gamma_{\xi,R} = I_T \otimes \Sigma_{\xi,R}$ and Lemma C.7 the inverse of Γ_X (see (39)) is given by

$$\Gamma_{X,R}^{-1} = I_T \otimes \Sigma_{\xi,R}^{-1} - (I_T \otimes \Sigma_{\xi,R}^{-1}L)(\Gamma_Z^{-1} + I_T \otimes L'\Sigma_{\xi,R}^{-1}L)^{-1}(I_T \otimes L'\Sigma_{\xi,R}^{-1}) \tag{43}$$

3. For two matrices of suitable dimensions A, H and $\det A \neq 0$ the following holds

$$(A + H)^{-1} = A^{-1} - (A + H)^{-1}HA^{-1} \tag{44}$$

The first two facts assume that Γ_Z is non-singular. Lemma 4.8 shows that for $q < r$ and $T > p$ this matrix can not be inverted. The proof for $q < r$ must therefore rely on another decomposition of $\Gamma_{X,R}^{-1}$.

$\square$

Proof (Theorem 4.14). Use Lemma C.10 and Equation (43) to write

$$
\begin{aligned}
z_{t|T} &= S_t'\Gamma_Z(I_t \otimes L')\Gamma_X^{-1}\underline{X} \\
&= S_t'\Gamma_Z\left[I_T \otimes L'\Sigma_{\xi,R}^{-1} - (I_T \otimes L'\Sigma_{\xi,R}^{-1}L)(\Gamma_Z^{-1} + I_T \otimes L'\Sigma_{\xi,R}^{-1}L)^{-1}(I_T \otimes L'\Sigma_{\xi,R}^{-1})\right]\underline{X} \\
&= S_t'(\Gamma_Z^{-1} + I_T \otimes L'\Sigma_{\xi,R}^{-1}L)^{-1}(I_T \otimes L'\Sigma_{\xi,R}^{-1})\underline{X}
\end{aligned}
$$

Define $M := L'\Sigma_{\xi,R}^{-1}L$. Use relation (44) to expand $(\Gamma_Z^{-1} + I_T \otimes M)^{-1}$. This gives

$$z_{t|T} = \underbrace{S_t'(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})\underline{X})}_{=:z_{t|T}^{(1)}} - \underbrace{S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}(I_T \otimes M)^{-1}(I_T \otimes L'\Sigma_{\xi,R}^{-1})\underline{X}}_{=:z_{t|T}^{(2)}}$$

The first summand is equal to z_t plus something that vanishes asymptotically

$$s_{t|T}^{(1)} = S_t'(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})(Lz_t + \xi_t) = z_t + M^{-1}L'\Sigma_{\xi,R}^{-1}\xi_t$$

If the second term converges in mean squares sense to zero, then it will also converge in probability to zero. As $\lambda_{\max}(\Sigma_{\xi,R}^{-1}) = 1/\lambda_{\min}(\Sigma_{\xi,R})$, it follows that

$$||M^{-1}L'\Sigma_{\xi,R}^{-1}\xi_t||_2^2 = \text{tr}(M^{-1}L'\Sigma_{\xi,R}^{-1}\gamma_\xi(0)\Sigma_{\xi,R}^{-1}LM^{-1}) \leq \frac{\lambda_{\max}(\gamma_\xi(0))}{\lambda_{\min}(\Sigma_{\xi,R})}\text{tr}(M^{-1}) \to 0$$

which is guaranteed by Assumption E6..

The second term will $z_{t|T}^{(2)}$ be split up again by using $\underline{X} = (I_T \otimes L)\underline{Z} + \underline{\xi}$

$$z_{t|T}^{(2.1)} = S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}\underline{Z}$$

and

$$z_{t|T}^{(2.2)} = S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}(I_T \otimes M)^{-1}(I_T \otimes L'\Sigma_{\xi,R}^{-1})\underline{\xi}.$$

We look again at their L_2 convergence to show that both terms vanish asymptotically

$$||z_{t|T}^{(2.1)}||_2^2 = \text{tr}(S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}(\Gamma_Z^{-1} + I_T \otimes M)^{-1}S_t)$$

$$\leq \lambda_{\max}(\Gamma_Z^{-1})\text{tr}(M^{-2}) = O\left(\frac{1}{n^2}\right)$$

The following inequalities use the sub-multiplicity of the spectral norm.

$$||z_{t|T}^{(2.2)}||_2^2 = \text{tr}(S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})\Gamma_\xi \times$$

$$\times (I_T \otimes \Sigma_{\xi,R}^{-1}LM^{-1})\Gamma_Z^{-1}(\Gamma_Z^{-1} + I_T \otimes M)^{-1}S_t)$$

$$\leq ||\Gamma_Z^{-1}(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})\Gamma_\xi(I_T \otimes \Sigma_{\xi,R}^{-1}LM^{-1})\Gamma_Z^{-1}||\,\text{tr}(S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-2}S_t)$$

$$\leq ||\Gamma_Z^{-2}||\,||I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1}||\,||\Gamma_\xi||\,||I_T \otimes \Sigma_{\xi,R}^{-1}LM^{-1}||\,\text{tr}(S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-2}S_t)$$

Note that $||\Gamma_Z||^2 = O(1)$, $||\Gamma_\xi|| = O(1)$ and

$$\text{tr}(S_t'(\Gamma_Z^{-1} + I_T \otimes M)^{-2}S_t) \leq \text{tr}(S_t'(I_T \otimes M)^{-2}S_t) = \text{tr}(M^{-2}) = O\left(\frac{1}{n^2}\right),$$

$$||I_T \otimes \Sigma_{\xi,R}^{-1}LM^{-1}|| = ||M^{-1}L'\Sigma_{\xi,R}^{-1}|| = O\left(\frac{1}{\sqrt{n}}\right),$$

$$||I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1}|| = ||\Sigma_{\xi,R}^{-1}LM^{-1}|| = O\left(\frac{1}{\sqrt{n}}\right)$$

Thus we have $z_{t|T} = z_t + O_p(1/\sqrt{n}) + O_p(1/n) + O_p(1/n\sqrt{n})$.

$\square$

The next Theorem shows that also the sample counterpart $\hat{z}_{t|T}$ of $z_{t|T}$ is consistent. Again a sketch of the proof will be given.

Theorem 4.16. Let $\hat{z}_t$ be the principal components estimator of z_t. Consider the assumptions of Theorem 4.14 and assume further that $\limsup \frac{T}{n^3} = O(1)$. Then

$$\hat{z}_{t|T} = \widehat{\mathcal{P}}(z_t|X_1^{nT}) = \underset{(r \times rT)}{S'_t} \quad \underset{(rT \times rT)}{\hat{\Gamma}_Z} \quad \underset{(rT \times nT)}{(I_t \otimes \hat{L}')} \quad \underset{(nT \times nT)}{\hat{\Gamma}_{X,T}^{-1}} \quad \underset{(nT \times 1)}{\underline{X}} \tag{45}$$

where $\hat{L}$ is equal to the eigenvectors corresponding to the r largest eigenvalues of the sample covariance matrix of X_1^{nT}, $\hat{\Gamma}_{X,T}$ is the sample-counterpart of Γ_X, and

$$\hat{\Gamma}_Z = \begin{pmatrix} \hat{\gamma}_Z(0) & \hat{\gamma}_Z(-1) & \cdots & \hat{\gamma}_Z(-T+1) \\ \hat{\gamma}_Z(1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \hat{\gamma}_Z(T-1) & \cdots & \cdots & \hat{\gamma}_Z(0) \end{pmatrix}$$

with

$$\hat{\gamma}_Z(h) = \frac{1}{T-h} \sum_{t=h+1}^{T} \hat{z}_t \hat{z}'_{t-h}$$

is a consistent estimator of z_t for $n, T \to \infty$.

The idea of the proof is the same as before at population level. The estimate $\hat{z}_{t|T}$ will be split up into three parts and for each convergence to its population counterpart will be shown. The proof is a sketch, because it will use the following unproven facts.

4.17. Facts

1. The PC-estimates for L and $\Sigma_{\xi,R}$ are consistent. The former is equal to $\hat{L} = (\hat{p}_1, \ldots, \hat{p}_r)$, where $\hat{p}_j$ is the eigenvector of $X'X$ corresponding to the j-th largest eigenvalue. The latter will be a diagonal matrix where the diagonal entries come from $\frac{1}{T-1}X'X - \hat{L}\hat{L}'$. Proposition 2 in [18] gives the formal statements:

   - For any $i,j:\ \hat{L}_{ij} - L_{ij} = O_p(1/\sqrt{n}) + O_p(1/\sqrt{T})$
   - If $\hat{\Psi} = \frac{1}{T-1}X'X - \hat{L}\hat{L}'$, then for any $(i,j):\ \hat{\psi}_{ij} - (\gamma_\xi(0))_{ij} = O_p(1/\sqrt{n}) + O_p(1/\sqrt{T})$

2. The matrix $\hat{\Gamma}_Z$ is a consistent estimate of Γ_Z. This is proven in Proposition 4 in [18] and it is not a standard argument, because the static factors z_t have to be estimated.

   - $||\hat{\Gamma}_Z - \Gamma_Z|| = O_p(1/n) + O_p(1/\sqrt{T})$
   - $||\hat{\Gamma}_Z|| = O_p(1),\ ||\hat{\Gamma}_Z^{-1}|| = O_p(1)$
   - $||\hat{\Gamma}_Z^{-1} - \Gamma_Z^{-1}|| = O_p(1/n) + O_p(1/\sqrt{T})$

$\square$

Proof (of Theorem 4.16).   Let $\hat{z}_{t|T}^{(1)}, \hat{z}_{t|T}^{(2.1)}, \hat{z}_{t|T}^{(2.2)}$ be the sample counterparts of $z_{t|T}^{(1)}, z_{t|T}^{(2.1)}, z_{t|T}^{(2.2)}$ (Theorem 4.14) defined by

$$\hat{z}_{t|T}^{(1)} = \hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1}x_t$$

$$\hat{z}_{t|T}^{(2.1)} = S_t'(\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1}\hat{\Gamma}_Z^{-1}(I_T \otimes \hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1})(I_T \otimes L)\underline{Z}$$

$$\hat{z}_{t|T}^{(2.2)} = S_t'(\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1}\hat{\Gamma}_Z^{-1}(I_T \otimes \hat{M})^{-1}(I_T \otimes \hat{L}'\hat{\Sigma}_{\xi,R}^{-1})\underline{\xi}$$

where $\hat{M} = \hat{L}'\hat{\Sigma}_{\xi,R}^{-1}\hat{L}$.

Consider the L_2 norm[19] of $\hat{z}_{t|T}^{(1)} - z_{t|T}^{(1)} = (\hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1} - M^{-1}L'\Sigma_{\xi,R}^{-1})x_t$

$$||\hat{z}_{t|T}^{(1)} - z_{t|T}^{(1)}||_2 \le ||\hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1} - M^{-1}L'\Sigma_{\xi,R}^{-1}||\,||x_t||_2$$

By Assumption E1. $\exists c > 0$ such that $||x_t||_2 \le c < \infty$ for all $n \in \mathbb{N}$. Lemma 5 (vi) in [18] uses the consistency of the principal components in order to show

$$\frac{1}{n}||\hat{M}'\hat{\Sigma}_{\xi,R}^{-1}\hat{M} - M'\Sigma_{\xi,R}^{-1}M|| = O_p(1/n) + O_p(1/\sqrt{T}).$$

Therefore $\hat{z}_{t|T}^{(1)} = z_{t|T}^{(1)} + O_p(1/n) + O_p(1/\sqrt{T})$ and $\hat{z}_{t|T}^{(1)} = z_t + O_p(1/\sqrt{n}) + O_p(1/\sqrt{T})$. For the second and third term, define

$$\Delta := (\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1}\hat{\Gamma}_Z^{-1}(I_T \otimes \hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1}) - (\Gamma_Z^{-1} + I_T \otimes M)^{-1}\Gamma_Z^{-1}(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})$$

Then $\hat{z}_{t|T}^{(2.1)} - z_{t|T}^{(2.1)} = S_t'\Delta(I_T \otimes L)\underline{Z}$ and $\hat{z}_{t|T}^{(2.2)} - z_{t|T}^{(2.2)} = S_t'\Delta\underline{\xi}$.
Split up $\Delta = \Delta_1 + \Delta_2 + \Delta_3$ where

$$\Delta_1 = (\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1}\hat{\Gamma}_Z^{-1}(I_t \otimes (\hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1} - M^{-1}L'\Sigma_{\xi,R}^{-1}))$$

$$\Delta_2 = (\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1}(\hat{\Gamma}_Z^{-1} - \Gamma_Z^{-1})(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})$$

$$\Delta_3 = \left((\hat{\Gamma}_Z^{-1} + I_T \otimes \hat{M})^{-1} - (\Gamma_Z^{-1} + I_T \otimes M)^{-1}\right)\Gamma_Z^{-1}(I_T \otimes M^{-1}L'\Sigma_{\xi,R}^{-1})$$

and use the Facts 4.17 and Lemma 5 of [18] in order to show $||\Delta_i||_2 = O_p(1/n^2\sqrt{n}) + O_p(1/n\sqrt{nT})$, $i = 1, 2, 3$.

Note that $||\underline{Z}||^2 = \mathbb{E}(\underline{Z}'\underline{Z}) = \sum_{t=1}^T \text{tr}(\Sigma_z) = rT$. Also $||\underline{\xi}||^2 = \sum_{t=1}^T ||\xi_t||^2 = \sum_{t=1}^T \text{tr}(\Sigma_{\xi,R})$. Therefore $||\underline{Z}|| = O_p(\sqrt{T})$ and $||\underline{\xi}|| = O(nT)$,

Finally the inequalities

$$||S_t'\Delta(I_T \otimes L)\underline{Z}||_2 \le ||S_t'||\,||\Delta||\,||I_T \otimes L||\,||\underline{Z}|| = O_p(\sqrt{T}/n^2) + O_p(1/n)$$

$$||S_t'\Delta\underline{\xi}||_2 \le ||S_t'||\,||\Delta||\,||\underline{\xi}|| = O_p(\sqrt{T}/n^2) + O_p(1/n)$$

together with the Assumption $\limsup \frac{T}{n^3} = O(1)$ give $\hat{z}_{t|T} = z_t + O_p(1/\sqrt{n}) + O_p(1/\sqrt{T})$.
□

---

[19]The matrix $A := (\hat{M}^{-1}\hat{L}'\hat{\Sigma}_{\xi,R}^{-1} - M^{-1}L'\Sigma_{\xi,R}^{-1})$ is a linear function between the rT dimensional product space of $L_2(\mathbb{P})$ and the nT dimensional product space of $L_2(\mathbb{P})$ and each space is equipped with the product norm $||x||_2^2 = \mathbb{E}x'x$.

4.4   The QML estimation approach

The quasi-maximum likelihood estimation approach is based on a Gaussian state space framework. The estimated static factors $\check{z}_{t|T}$ are equal to the estimated expected values of z_t given all observations x_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$ under parameters that maximize the likelihood of the sample $X := (x_1, \ldots, x_T) \in \mathbb{R}^{n \times T}$. The discussion of this approach is based on the paper of Doz, Giannone and Reichlin (DGR) [17].

The model is still an approximate dynamic factor model, i.e.

$$x_t \;=\; \underset{(n \times r)}{L} \;\; \underset{(r \times 1)}{z_t} \;+\; \underset{(n \times 1)}{\xi_t} \tag{46a}$$
$$\underset{(n \times 1)}{}$$

$$z_t = -e_1 z_{t-1} - e_2 z_{t-2} - \ldots - e_p z_{t-p} + b u_t \tag{46b}$$

where all assumptions of Definition 3.15 are fullfilled. Its companion form is

$$x_t = \underbrace{(L, 0, \ldots, 0)}_{H} \overbrace{\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix}}^{s_t} + \xi_t = H s_t + \xi_t \tag{47a}$$

$$s_{t+1} = \underbrace{\begin{pmatrix} -e_1 & -e_2 & \cdots & -e_p \\ I_r & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & I_r & 0 \end{pmatrix}}_{F} s_t + (b', 0, \ldots, 0)' u_{t+1} = F s_t + G u_{t+1} \tag{47b}$$

where $s_t \in \mathbb{R}^{rp \times 1}$ (random), $H \in \mathbb{R}^{n \times rp}$, $F \in \mathbb{R}^{rp \times rp}$ and $G \in \mathbb{R}^{rp \times q}$. Additionally one assumes

Assumption G (Quasi-maximum likelihood assumptions):

G1.

$$\begin{pmatrix} \xi_t \\ u_t \end{pmatrix} \sim \mathcal{N} \left(0, \begin{pmatrix} R & 0 \\ 0 & I_q \end{pmatrix} \right), \quad \text{i.i.d.} \tag{48}$$

G2. R is a diagonal matrix, where $R_{ii} = \vartheta$, $i = 1, \ldots, n$ (spherical noise) or where $R_{ii} = \vartheta_i$, $i = 1, \ldots, n$ (heteroscedastic noise)

Remark 4.18.

- Due to Assumption G1., Model (47) is a state-space model.

- The approximate factor model assumptions of DGR [17] are

$$\lim_{n \to \infty} \inf \frac{1}{n} \lambda_{\min}(L'L) > 0$$

$$\lim_{n \to \infty} \sup \frac{1}{n} \lambda_{\max}(L'L) < \infty$$

$$0 < \lim_{n \to \infty} \inf \lambda_{\min}(\gamma_\xi(0)) \le \lim_{n \to \infty} \sup \lambda_{\max}(\gamma_\xi(0)) < \infty$$

The weak dependence of the process ξ_t implies the right inequality of the third line as Assumption (gdfm b) implies the existence of $0 < M < \infty$ such that $||f_\xi(\theta)|| < \infty$ a.e. on Θ for all $n \in \mathbb{N}$. This implies that $||\gamma_\xi(0)|| = || \int_{-\pi}^{\pi} f_\xi(\theta)\, d\theta|| \le \int_{-\pi}^{\pi} ||f_\xi(\theta)||\, d\theta \le 2\pi M < \infty$ for all $n \in \mathbb{N}$.

The first line corresponds to the strong dependence of the latent process χ_t. As in DGR [18] it is assumed that the first r eigenvalues of $\gamma_\chi(0)$ diverge of linear order to infinity. The first inequality of the third line is a technical assumption and implies that the idiosyncratic stochastic variables won't be degenerated in the limit.

- In this parametric model the static factors z_t are estimated by taking the conditional expectation of z_t given all observations X, i.e. $\check{z}_t = \mathbb{E}_{\hat{\tau}}(z_t | \mathcal{H}_x)$. If the maximum likelihood estimates $\hat{\tau}$ where known, this could be calculated by the Kalman smoother recursions (see Section 2.3). The aim of this section is therefore to illustrate a way in order to calculate the parameters $\hat{\tau}$.

- Note that because of $s_{t+1} = Fs_t + Gu_{t+1}$, the vector s_t is not a real state in the sense that it depends only on past inputs.

- The parameters $\tau = (H, F, G, Q, R, \mu_1, V_1)$ determine the distribution of $\underline{X} = \text{vec}(x_1, \ldots, x_T)$. By specifying matrix F, the dynamics of the static factors are modeled. Setting F such that $e_1 = \cdots = e_p = 0$, means that the static factors are modeled as white noise. Setting R means specifying the amount of mis-specification as far as the dynamics of the idiosyncratic component is concerned.

  In the papers [18], [17] the parameters are given by $(e(z), H, \gamma_\xi^*(0))$ and $(I_r, H, \gamma_\xi^*(0))$ respectively, which is just another notation for the equivalent description given above. The matrices R and F correspond to $\gamma_\xi^*(0)$ and $e(z)$ respectively.

$\square$

According to the Bayes formula, the joint density of $\underline{X} = \text{vec}(x_1, \ldots, x_T) \in \mathbb{R}^{nT \times 1}$ can be written as the quotient of conditional and marginal densities

$$f_{\underline{X}}(x|\tau) = \frac{f_{\underline{X}|\underline{Z}}(x|z, \tau)\, f_{\underline{Z}}(z|\tau)}{f_{\underline{Z}|\underline{X}}(z|x, \tau)} \tag{49}$$

where $\underline{Z} = \text{vec}(z_1, \ldots, z_T) \in \mathbb{R}^{rT \times 1}$.

For $\underline{X} = (I_T \otimes L)\underline{Z} + \xi$, Assumptions G imply that $(\underline{X}', \underline{Z}')' \sim \mathcal{N}(\mu, \Sigma)$, $\mu = (\mu_x', \mu_z')'$, $\mu_x = (I_T \otimes L)\mu_z$, $\Sigma_{xz} = (I_T \otimes L)\Sigma_{zz}$, $\Sigma_{zx} = \Sigma_{xz}'$, $\Sigma_{xx} = (I_T \otimes L)\Sigma_{zz}(I_T \otimes L)' + \Gamma_{\xi,R}$. The covariance matrix Σ is suitably partitioned into $\Sigma_{xx}, \Sigma xz, \Sigma_{zx}$, and Σ_{zz}. By applying Theorem C.14

$$\underline{X}|\underline{Z} \sim \mathcal{N}((I_T \otimes L)\underline{Z}, \Gamma_{\xi,R}),$$

where $\Gamma_{\xi,R} = I_T \otimes \Sigma_{\xi,R}$ due to Assumption G2.

Consider $\underline{Z} \sim \mathcal{N}(\mu_z, \Gamma_Z)$, where $\Gamma_Z \in \mathbb{R}^{rT \times rT}$. If $q = r$, then Γ_Z is always invertible and the density of $\underline{Z}$ exists. For the case $q < r$ the density does not exist if $T > p$. In the following the case $q < r$ is not analyzed although simulation studies have been conducted with this assumption (see Section 6).

Based on Equation (49), the Log-Likelihood can be written as

$$\ell(\underline{X}|\tau) = \log f_{\underline{X}|\underline{Z}}(x|z, \tau) + \log f_{\underline{Z}}(z|\tau) - \log f_{\underline{Z}|\underline{X}}(z|x, \tau) =$$

$$-\frac{nT}{2}\log(2\pi) - \frac{1}{2}\log\det(\Gamma_{\xi,R}) - \frac{1}{2}(\underline{X} - (I_T \otimes L)\underline{Z})'\Gamma_{\xi,R}^{-1}(\underline{X} - (I_T \otimes L)\underline{Z})$$

$$-\frac{rT}{2}\log(2\pi) - \frac{1}{2}\log\det(\Gamma_Z) - \frac{1}{2}\underline{Z}'\Gamma_Z^{-1}\underline{Z}$$

$$+\frac{rT}{2}\log(2\pi) + \frac{1}{2}\log\det(\Omega_\tau) + \frac{1}{2}(\underline{Z} - \hat{\underline{Z}}_\tau)'\Omega_\tau^{-1}(\underline{Z} - \hat{\underline{Z}}_\tau)$$

where $\Gamma_{Z,T} = \mathbb{E}\underline{Z}\underline{Z}'$, $\hat{\underline{Z}}_\tau = \mathbb{E}_\tau(\underline{Z}|\underline{X})$, and $\Omega_\tau = \mathbb{E}(\underline{Z} - \hat{\underline{Z}}_\tau)(\underline{Z} - \hat{\underline{Z}}_\tau)'$.

Evaluated at $\underline{Z} = \hat{\underline{Z}}_\tau$ and by using Lemma C.10 this becomes

$$\ell(\underline{X}|\tau) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log\det(\Sigma_{\xi,R}) - \frac{1}{2}\mathrm{tr}\left[(X - \hat{Z}_\tau L')\Sigma_{\xi,R}^{-1}(X - \hat{Z}_\tau L')'\right]$$

$$-\frac{1}{2}\log\det(\Gamma_{\hat{Z},T}) - \frac{1}{2}\hat{\underline{Z}}'\Gamma_{\hat{Z},T}^{-1}\hat{\underline{Z}} + \frac{1}{2}\log\det(\Omega_\tau) \qquad (50)$$

where $X = (x_1, \ldots, x_T) \in \mathbb{R}^{T \times n}$.

In the following two statements will be discussed in more detail:

- In the paper of [17] it has been shown, that the estimator $\mathbb{E}_{\hat{\tau}}(\underline{Z}|\underline{X})$ for $\underline{Z}$ is consistent for large n and T, i.e. $n, T \to \infty$.

- The ML-estimates $\hat{\tau}$ and therefore the estimator of the static factors $\mathbb{E}_{\hat{\tau}}(\underline{Z}|\underline{X})$ are not directly calculated. Instead the iterative EM-algorithm will be used.

Consistency of QML-estimator. The conditional expectation of the static factors given all observations, taken under the ML-parameters is consistent for large cross-section and time dimensions.

Denote τ^c as the case where the parameter matrices $(H, F, G, Q, R, \mu_1, V_1)$ are known and $e(z) = I_r$.

Assumption H (Further assumptions for consistency):

H1. There exists a positive constant M such that for all $i, j \in \mathbb{N}$ and for all $T \in \mathbb{Z}$

  (a) $\mathbb{E}\left(\sqrt{T}(\frac{1}{T}\sum_{t=1}^T \xi_{it}\xi_{jt} - \psi_{ij})\right)^2 < M$, where ψ_{ij} is the (i, j) element of $\gamma_\xi(0)$.

  (b) $\mathbb{E}||\frac{1}{\sqrt{T}}\sum_{t=1}^T z_t\xi_{jt}||^2 < M$

(c) $\mathbb{E}\|\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t' - I_r\right)\|^2 < M$

where $\|\cdot\|$ is a matrix norm.

H2. There exists $\delta > 0$ such that $\underline{c} \leq \psi_{ii} - \delta \leq \psi_{ii} + \delta \leq \overline{c}$ for all $i \in \mathbb{N}$, where $\underline{c}$ and $\overline{c}$ are the constant terms defining the constrained maximization of the likelihood, i.e. $0 < \underline{c} \leq \psi_{ii} \leq \overline{c} < \infty$ for all $i \in \mathbb{N}$.

H3. The first r eigenvalues of Σ_χ^n are of order $n \mapsto n$ (see Remark 4.18) and

$$0 < \lim_{n \to \infty} \inf \lambda_{\min}(\gamma_\xi(0))$$

The result of the following Theorem 4.19 is that the time average of the squared deviations between the factors that lie in the true factor space and the estimated factors vanishes as n, T go to ∞.

Theorem 4.19 (see [17], Proposition 1): Consider the model of Definition 3.15 with $r = q$ and $b = I_r$. The observation equation can be written as $\underline{X} = (I_T \otimes L)\underline{Z} + \underline{\xi}$. Under Assumptions G and H the following holds

$$\frac{1}{T}(\underline{Z} - (I_T \otimes \hat{\beta})\hat{\underline{Z}}_{\hat{\tau}})'(\underline{Z} - (I_T \otimes \hat{\beta})\hat{\underline{Z}}_{\hat{\tau}}) = O_p\left(\frac{1}{\Delta_{nT}}\right) \tag{51}$$

where $\hat{\beta} \in \mathbb{R}^{r \times r}$ is the multivariate OLS coefficient-matrix of regressing Z on $\hat{Z}_{\hat{\tau}} = \mathbb{E}_{\hat{\tau}}(\underline{Z}|X_1^{nT})$, i.e. $\hat{\beta} = (\hat{Z}_{\hat{\tau}}'\hat{Z}_{\hat{\tau}})^{-1}\hat{Z}_{\hat{\tau}}'Z \in \mathbb{R}^{r \times r}$, $Z = (z_1, \dots, z_T) \in \mathbb{R}^{T \times r}$, and $\Delta_{nT} = \min\{\sqrt{T}, \frac{n}{\log n}\}$.

In order to prove this theorem the following facts (which are proven in [17]) will be used.

Facts 4.20.

1. Consider $\tau = \tau^c$, the parameter where $e(z) = I_r$ and the covariance matrix of ξ_t is a diagonal matrix $R = \Sigma_{\xi, R}$. Then under Assumptions H1. and the assumptions of Remark 4.18, the Log-Likelihood under this parameter can be written as

$$\frac{1}{nT}\ell(X|\tau^c) = -\frac{1}{2}\log(2\pi) - \frac{1}{2n}\log\det(\Sigma_{\xi, R}) - \frac{1}{2} + O_p\left(\frac{\log n}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \tag{52}$$

as $n, T \to \infty$.

2. Lemma 2 in [17] states that under Assumptions H1. and the assumptions of Remark 4.18 the following inequality holds

$$\frac{1}{nT}\operatorname{tr}(X - \hat{Z}_{\hat{\tau}})\Sigma_{\xi, R}^{-1}(X - \hat{Z}_{\hat{\tau}})' \geq \frac{1}{nT}\operatorname{tr}\left[(L'\hat{\Sigma}_{\xi, R}^{-1}L)'(Z - \hat{Z}_{\hat{\tau}}\hat{\beta})'(Z - \hat{Z}_{\hat{\tau}}\hat{\beta})\right]$$

$$- 2\left(\frac{1}{T}(Z - \hat{Z}_{\hat{\tau}}\hat{\beta})'(Z - \hat{Z}_{\hat{\tau}}\hat{\beta})\right)^{\frac{1}{2}}\left(O_p(1/\sqrt{T}) + O_p(1/n)\right)^{\frac{1}{2}}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\frac{\psi_{ii}}{\hat{\psi}_{ii}} + O_p(1/\sqrt{T}) + O_p(1/n),$$

where ψ_{ii} and $\hat{\psi}_{ii}$ are the diagonal entries of $\Sigma_{\xi,R}$ and $\hat{\Sigma}_{\xi,R}$ respectively.

This inequality is basically obtained by regressing the observations X_1^{nT} on the QML-estimates $\underline{\hat{Z}}_{\hat{\tau}}$. The matrix $\hat{\beta}$ has been defined in Theorem 4.19.

$\square$

Proof (of Theorem 4.19).   Obviously

$$V_{nT} := \frac{1}{T}(\underline{Z} - (I_T \otimes \hat{\beta})\underline{\hat{Z}}_{\hat{\tau}})'(\underline{Z} - (I_T \otimes \hat{\beta})\underline{\hat{Z}}_{\hat{\tau}}) = \frac{1}{T}\mathrm{tr}\left((Z - \hat{Z}_{\hat{\tau}}\hat{\beta})'(Z - \hat{Z}_{\hat{\tau}}\hat{\beta})\right) \geq 0$$

The aim is therefore to show that $V_{nT} \leq 0$ somehow. The starting point is the difference of $\ell(X_1^{nT}|\tau^c)$ and $\ell(X_1^{nT}|\hat{\tau})$. Assumption H2. assures that τ^c is in the set of possible maximum-likelihood parameters. Therefore $0 \geq \frac{2}{nT}(\ell(X_1^{nT}|\tau^c) - \ell(X_1^{nT}|\hat{\tau}))$.

Substituting Equations (50) and (52) gives

$$0 \geq -\frac{1}{n}\log\det(\Sigma_{\xi,R}) - 1 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\log n}{n}\right) + \frac{1}{n}\log\det(\hat{\Sigma}_{\xi,R})$$

$$+\frac{1}{nT}\mathrm{tr}\left((X - \hat{Z}_{\hat{\tau}}\hat{L}')\hat{\Sigma}_{\xi,R}^{-1}(X - \hat{Z}_{\hat{\tau}}\hat{L}')'\right)$$

$$+\frac{1}{nT}\log\det(\Gamma_{\hat{Z},T}) - \frac{1}{nT}\log\det(\Omega_{\hat{\tau}}) + \frac{1}{nT}\underline{\hat{Z}}'\Gamma_{\hat{Z},T}^{-1}\underline{\hat{Z}}$$

Using the second fact of 4.20

$$0 \geq \left[-\frac{1}{n}\log\det(\Sigma_{\xi,R}) - 1 + \frac{1}{n}\log\det(\hat{\Sigma}_{\xi,R}) + \frac{1}{n}\sum_{i=1}^{n}\frac{\psi_{ii}}{\hat{\psi}_{ii}}\right]$$

$$+\left[\frac{1}{nT}\log\det(\Gamma_{\hat{Z},T}) - \frac{1}{nT}\log\det(\Omega_{\hat{\tau}}) + \frac{1}{nT}\underline{\hat{Z}}'\Gamma_{\hat{Z},T}^{-1}\underline{\hat{Z}}\right]$$

$$+O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\log n}{n}\right)$$

$$+\left[\frac{1}{nT}\mathrm{tr}\left((L'\hat{\Sigma}_{\xi,R}^{-1}L)(Z - \hat{Z}_{\hat{\tau}}\hat{L})'(Z - \hat{Z}_{\hat{\tau}}\hat{L})\right) - 2\left(\frac{1}{T}\mathrm{tr}((Z - \hat{Z}_{\hat{\tau}}\hat{L})'(Z - \hat{Z}_{\hat{\tau}}\hat{L}))\left(O_p(1/\sqrt{T}) + O_p(1/n)\right)\right)^{\frac{1}{2}}\right]$$

The first three summands are non-negative, which implies

$$0 \geq \frac{1}{nT}\mathrm{tr}\left((L'\hat{\Sigma}_{\xi,R}^{-1}L)(Z - \hat{Z}_{\hat{\tau}}\hat{L})'(Z - \hat{Z}_{\hat{\tau}}\hat{L})\right) - 2\left(V_{nT}\right)^{\frac{1}{2}}O_p\left(\frac{1}{\sqrt{\Delta_{nT}}}\right) + O_p\left(\frac{1}{\Delta_{nT}}\right)$$

$$\geq \lambda_{\min}\left(\frac{L'\hat{\Sigma}_{\xi,R}^{-1}L}{n}\right)V_{nT} - 2\sqrt{V_{nT}}O_p\left(\frac{1}{\sqrt{\Delta_{nT}}}\right) + O_p(1/\Delta_{nT})$$

By assumption $\liminf_n \lambda_{\min}(L'\hat{\Sigma}_{\xi,R}^{-1}L/n) > 0$ and therefore

$$0 \leq V_{nT} - \sqrt{V_{nT}}O_p(1/\sqrt{\Delta_{nT}}) + O_p(1/\Delta_{nT})$$

which is a quadratic inequality in $y = \sqrt{V_{nT}}$. Deriving conditions for y such that this inequality holds implies $V_{nT} = O_p(1/\Delta_{nT})$.

$\square$

The EM Algorithm for State Space models.   The determination of the parameter $\hat{\tau}$ that maximizes the likelihood of the observations $\underline{X} = \text{vec}(X)$ is quite involved. The EM Algorithm for state space models [52], [48] is one way to do that.

   We start with Equations (47). As the states $S = (s_1, \ldots, s_T)$ are unobserved we face an incomplete data setting (see Remark 2.13). The EM algorithm is derived by assuming that the complete data (S, X) was available. Then for a given parameter τ, the sufficient statistics for the expected conditional complete data likelihood are determined. Finally given the expected sufficient statistics for this likelihood, multivariate OLS regression (or equivalently by building the derivates with respect to the parameters) lead to a new τ that maximizes the expected conditional likelihood. The aim is to calculate $\hat{\tau} = \arg\max \ell(\underline{X}, \tau)$. An iterative application of the EM algorithm gives a sequence $\tau^{(n)}$ such that $\ell(\underline{X}, \tau^{(n)}) \geq \ell(\underline{X}, \tau^{(n-1)})$ for all $n \in \mathbb{N}$ (see [16]).

Assumption I.

   I1. $s_1 \sim \mathcal{N}(\mu_1, V_1)$ where $\mu_1 = \text{vec}(\hat{z}_p, \ldots, \hat{z}_1)$ are the principal component estimates and $V_1 = \hat{\Gamma}_p$, which is determined through Equation (41). The state s_1 is linear independent of $\{\xi_t, u_t, t = 1, \ldots, T\}$.

Lemma 4.21   Consider the Structural Dynamic Factor Model in companion form (see Equation (47)) and assume that Assumptions G hold. Further assume that $r = q$ and $p = 1$. Then $Q = GG'$ with $\det Q \neq 0$, and the complete data likelihood is given by

$$-2\ell(S, X | \tau) = c + \log \det(V_1) + (s_1 - \mu_1)' V_1^{-1}(s_1 - \mu_1) +$$

$$+ T \log \det(Q) + \sum_{t=2}^{T}(s_t - Fs_{t-1})' Q^{-1}(s_t - Fs_{t-1}) +$$

$$+ T \log \det(R) + \sum_{t=1}^{T}(x_t - Hs_t)' R^{-1}(x_t - Hs_t)$$

Proof.   Let $f(S, X)$ be the joint density of states and observations for $t = 1, \ldots, T$[20]. By iteratively applying the well known relation $f(x|y) = f(x, y)/f(y)$, we get

$$f(S, X) = f(x_T | x_{T-1}, \ldots, x_1, s_1, \ldots, s_T) f(x_{T-1}, \ldots, x_1, S) =$$
$$= f(x_T | x_{T-1}, \ldots, x_1, s_1, \ldots, s_T) f(x_{T-1} | x_{T-2}, \ldots, x_1, s_1, \ldots, s_T) f(x_{T-2}, \ldots, x_1, S) =$$
$$= f(x_T | s_T) f(x_{T-1} | s_{T-1}) f(x_{T-2}, \ldots, x_1, S) = \ldots$$
$$\ldots = f(s_1, s_2, \ldots, s_T) \prod_{t=1}^{T} f(x_t | s_t)$$

   Therefore $f_\tau(S, X) = f(s_1) \prod_{t=2}^{T} f(s_t | s_{t-1}) \prod_{t=1}^{T} f(x_t | s_t)$. Using Lemmata C.11 and C.12 together with the assumptions made, one can show that $s_t | s_{t-1} \sim \mathcal{N}(s_t - Fs_{t-1}, Q)$, $t = 2, \ldots, T$ and $x_t | s_t \sim \mathcal{N}(x_t - Hs_t, R)$. The variable c is a constant.

---

[20]Throughout the proof the densities all depend on τ.

□

Lemma 4.22 Let $\tau^{(k)}$ be the parameter of step $k \in \mathbb{N}$. Under the assumptions of Lemma 4.21, the expected conditional likelihood $Q(\tau|\tau^{(k)}) = \mathbb{E}(\log f_\tau(s_1, \ldots, s_T, x_1, \ldots, x_T)|X_1^T, \tau^{(k)})$ can be written as

$$
\begin{aligned}
Q(\tau|\tau^{(k)}) = {} & \ln \det(P_{1|0}) + \mathrm{tr}\left[V_1^{-1}(P_{1|T} + (s_{1|T} - \mu_1)(s_{1|T} - \mu_1)')\right] \\
& + (T-1)\ln \det(Q) + \mathrm{tr}\left[Q^{-1}(S_{11} - S_{10}F' - FS_{10}' + FS_{00}F')\right] \\
& + T \ln \det(R) + \mathrm{tr}\left[R^{-1}\sum_{t=1}^n ((x_t - Hs_{t|T})(x_t - Hs_{t|T})' + HP_{t|T}H')\right] \quad (53)
\end{aligned}
$$

where

$$
S_{11} = \sum_{t=2}^T (s_{t|T}s_{t|T}' + P_{t|T})
$$

$$
S_{10} = \sum_{t=2}^T (s_{t|T}s_{t|T-1}' + P_{t,t-1|T})
$$

$$
S_{00} = \sum_{t=1}^{T-1} (s_{t|T}s_{t|T}' + P_{t|T})
$$

The matrix $P_{t,t-1|T} = \mathbb{E}(s_t - s_{t|T})(s_{t-1} - s_{t-1|T})'$ has been defined in Lemma 2.10. For the definitions of $s_{t|T}$ and $P_{t|T}$ see Section 2.3.

Proof. The idea is to exploit the independence between $(s_t - s_{t|T})$ and $\underline{X} = \mathrm{vec}(X_1^{nT})$ and the $\sigma(\underline{X})$-measurability[21] of $s_{t|T}$ for $t = 1, \ldots, T$. Let $s_1 - \mu_1 = (s_1 - s_{1|T}) + (s_{1|T} - \mu_1)$ in order to write

$$
\begin{aligned}
\mathbb{E}[(s_1 - \mu_1)'V_1^{-1}(s_1 - \mu_1)|X_1^{nT}] &= \mathbb{E}[(s_1 - s_{1|T})'V_1^{-1}(s_1 - s_{1|T})] + (s_{1|T} - \mu_1)'V_1^{-1}(s_{1|T} - \mu_1) = \\
&= \mathbb{E}[\mathrm{tr}((s_1 - s_{1|T})V_1^{-1}(s_1 - s_{1|T})')] + \mathrm{tr}((s_{1|T} - \mu_1)V_1^{-1}(s_{1|T} - \mu_1)') = \\
&= tr(V_1^{-1}(P_{1|T} + (s_{1|T} - \mu_1)(s_{1|T} - \mu_1)')
\end{aligned}
$$

Expand $s_t - Fs_{t-1} = (s_t - s_{t|T}) + s_{t|T} - Fs_{t-1|T} - F(s_{t-1} - s_{t-1|T})$ and write for every $t = 2, \ldots, T$

$$
\begin{aligned}
\mathbb{E}[(s_t - Fs_{t-1})'Q^{-1}(s_t - Fs_{t-1})|X_1^{nT}] = {} & \left[\mathbb{E}(s_t - s_{t|T})'Q^{-1}(s_t - s_{t|T}) + s_{t|T}'Q^{-1}s_{t|T}\right] - \\
& - \left[\mathbb{E}(s_t - s_{t|T})'Q^{-1}F(s_{t-1} - s_{t-1|T}) + s_{t|T}'Q^{-1}Fs_{t-1|T}\right] \\
& - \left[\mathbb{E}(s_{t-1} - s_{t-1|T})'F'Q^{-1}(s_t - s_{t|T}) + s_{t-1|T}'F'Q^{-1}s_{t|T}\right] \\
& + \left[\mathbb{E}(s_{t-1} - s_{t-1|T})'F'Q^{-1}F(s_{t-1} - s_{t-1|T}) + s_{t-1|T}'F'Q^{-1}Fs_{t-1|T}\right]
\end{aligned}
$$

---

[21]The sigma-algebra $\sigma(\underline{X})$ is the smallest sigma-algebra such that all x_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$ are measurable.

Note that all other summands are zero due to the independence between $(s_t - s_{t|T})$ and $s_{u|T}$. Due to $\mathbb{E}(x'Mx) = \text{tr}(\mathbb{E}xMx')$ for $M = M'$, the summands correspond to $\text{tr}(Q^{-1}S_{11})$, $\text{tr}(Q^{-1}S_{10}F')$, $\text{tr}(Q^{-1}FS'_{10})$ and $\text{tr}(Q^{-1}FS_{00}F')$ respectively.

Write $x_t - Hs_t = (x_t - Hs_{t|T}) - H(s_t - s_{t|T})$ and note that $(x_t - Hs_{t|T})$ belongs to $\mathcal{H}(X_1^{nT})$ whereas $H(s_t - s_{t|T})$ belongs to $\mathcal{H}(X_1^{nT})^\perp$. This implies

$$\mathbb{E}[(x_t - Hs_t)'R^{-1}(x_t - Hs_t)|X_1^{nT}] = (x_t - Hs_{t|T})'R^{-1}(x_t - Hs_{t|T})$$
$$+ H\mathbb{E}[(s_t - s_{t|T})(s_t - s_{t|T})']H'$$

which gives the last summand of Equation (53) after summing over $t = 1, \ldots, T$.

$\square$

Equation (53) is the basis for determining the new parameters (H, F, Q, R, μ_1, V_1). Following [47] (p. 121, Equation 7), one could derive F by building the derivative $\frac{\partial Q(\cdot|\tau^{(k)})}{\partial F}(\tau) = 0$. If we assume to be in step k, the transition matrix $F^{(k)} = S_{10}S_{00}^{-1}$. The same applied for Q, R, μ_1, V_1 yields

Maximization step 4.23.

$$F^{(k)} = S_{10}S_{00}^{-1}$$
$$H^{(k)} = \left(\sum_{t=1}^T z_{t|T}X_{(t)}\right)(s_{1|T}s_{1|T}' + P_{1|T} + S_{11})^{-1}$$
$$Q^{(k)} = (T-1)^{-1}(S_{11} - S_{10}S_{00}^{-1}S'_{10})$$
$$R^{(k)} = T^{-1}\sum_{t=1}^T (x_t - H^{(k)}s_{t|T})(x_t - H^{(k)}s_{t|T})' + H^{(k)}P_{t|T}H^{(k)'}$$
$$\mu_1^{(k)} = s_{1|T}$$
$$V_1^{(k)} = P_{1|T}$$

$\square$

Calculate $H^{(k)}$ by means of SUR[22]-equations: For $H^{(k)}$ one could apply the same rule or argue by multivariate regression, which also gives the maximum likelihood estimator if the errors where normally distributed. Assume again that (S, X) where known and start with $x_t = Hs_t + \xi_t$, $t = p, \ldots, T$ (see Remark 4.25) where $s_t = (z_t, \ldots, z_{t-p+1})$. Because we assumed that $\xi_t \sim \text{WN}(0, R)$ normally distributed, the errors are independent across time. The data-matrix of our observations is now structured by columns, i.e. $X := (X_1, X_2, \ldots, X_n)$, $X_k \in \mathbb{R}^{T-p+1}$. Denote the state matrix by $S = (s_1, \ldots, s_T) \in \mathbb{R}^{(T-p+1)\times m}$ with $m = rp$. The errors are put in a matrix $\varepsilon = (\xi_1, \xi_2, \ldots, \xi_T)' = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

The model can therefore be written as

---

[22]SUR – seemingly unrelated regression.

$$X_k = S\,H_k' + \varepsilon_k, \quad k = 1, \ldots, n$$

$$\underset{(nT \times 1)}{\mathrm{vec}(X)} = \underset{(nT \times nrp)}{(I_n \otimes S)} \underset{(nrp \times 1)}{\mathrm{vec}(\beta)} + \underset{(nT \times 1)}{\mathrm{vec}(\varepsilon)}$$

where $\beta = (H_1', H_2', \ldots, H_n')$ and $H_k' = [H_{(k)}]'$. Note that because of $H = (L, 0, 0, \ldots, 0)$, the parameter vector $\underline{\beta} := \mathrm{vec}(\beta)$ has $n(p-1)r$ restrictions, which can be written as

$$\underset{(nrp \times 1)}{\underline{\beta}} = \underset{(nrp \times nr)}{W} \underset{(nr \times 1)}{\underline{\gamma}} \tag{54}$$

where W is a matrix of zeros having entries of one at positions $(i,j) = ((k-1)r+1, rp(k-1)+1), \ldots, (kr, rp(k-1)+r)$, $k = 1, \ldots, n$. By substituting (54) into $\underline{X} = (I_n \otimes S)\underline{\beta} + \underline{\varepsilon}$ one gets

$$\underline{X} = [(I_n \otimes S)W]\,\underline{\gamma} + \underline{\varepsilon}$$

where $(I_n \otimes S)W = (I_n \otimes S_r)$ with $S_r = (s_1, s_2, \ldots, s_r) \in \mathbb{R}^{(T-p+1) \times r}$. Therefore W selects elements of a block matrix that contain the first r components of s_t, which is z_t.

Now because $\mathbb{E}(\underline{\varepsilon}\underline{\varepsilon}'|X) = (R \otimes I_T)$ generalized least squares (GLS) would be the right thing to do. It can be easily shown (as done in [30]) that this is equivalent to OLS equation by equation if all equations have the regressors S_r. Therefore

$$\hat{\underline{\gamma}} = (S_r'S_r)^{-1}S_r'\underline{X},$$

which would determine $\hat{\beta}$ and therefore $H^{(k)}$. Now because S is not available in practice, expected sufficient statistics of $(S_r'S_r)$ and $S_r'X$ are needed. Note that $S_r' = Z := (z_1, \ldots, z_T)$. We denote the selection of the t-th row of a matrix M as $t \mapsto M(t)$ and take the conditional expected values

$$\mathbb{E}(ZX|X) = \sum_{t=1}^{T} \mathbb{E}(Z_{(t)}X_{(t)}|X) = \sum_{t=1}^{T} z_{t|T}X_{(t)} = \hat{Z}X$$

where $\hat{Z} = (z_{1|T}, z_{2|T}, \ldots, z_{T|T})'$. The second term is the inverse of

$$\mathbb{E}(ZZ'|X) = \sum_{t=1}^{T} \mathbb{E}(Z_{(t)}Z_{(t)}'|X) = \sum_{t=1}^{T} (z_{t|T}z_{t|T}' + P_{t|T})$$

where the second equality comes from $z_t = z_{t|T} + (z_t - z_{t|T})$ and the fact that $(z_t - z_{t|T})$ is orthogonal on X and $z_{t|T}$ is measurable with respect to $\sigma(X)$.

In order to summarize, a non-formal overview of the EM algorithm is given.

Algorithm 4.24 (EM algorithm for state space models).

Initialization by principal component estimation and estimation of the factor dynamics. This step is identical to the first step of 4.12 on page 34.

- $H^{(0)}, R^{(0)}, \mu_1^{(0)}, V_1^{(0)}$ are based on principal component estimation.
- $F^{(0)}, G^{(0)}, Q^{(0)}$ are estimated by means of solving (singular) Yule Walker equations.

For steps $k = 1, 2, \ldots, \text{iter.max}$:

E-step Calculate expected sufficient statistics by running the Kalman filter and -smoothing recursions for given $\tau^{(k)}$. They depend on $s_{t|T}, P_{t|T}, P_{t,t-1|T}, \ t = 1, \ldots, T$ and can be calculated by the Kalman Filter recursions and Lemma 2.10 (or Lemma2.11).

The Kalman filter also gives the value of the current log-likelihood $\ell(\underline{X}, \tau^{(k)})$.

M-step Follow 4.23 in order to calculate $\tau^{(k+1)}$ that maximizes $Q(\tau | \tau^{(k)})$.

Convergence Iterate between Expectation- and Maximization step until the relative increase of the log-likelihood is smaller a given tolerance level:

$$\frac{\ell(\underline{X}, \tau^{(k)}) - \ell(\underline{X}, \tau^{(k+1)})}{|\ell(\underline{X}, \tau^{(k)})|} < \text{tolerance}$$

$\square$

Remark 4.25 (What happens if $p > 1$). If $p > 1$, the state s_t must be stacked, i.e. $s_t = (z_t', z_{t-1}', \ldots, z_{t-p}')'$, which implies that t must start at $t = p$. Therefore the first $(p-1)$ observations do not influence the estimation procedure and estimates for z_t are available only for $t = p, \ldots, T$. $\square$

5   Simulation study

All three estimation methods consistently estimate the linear space of the static factors when n and T become very large, as argued in Section 4. To my knowledge there is only a paper by J. Bai [4] that analyzes the asymptotic distribution of the principal component estimator. Therefore a simulation study has been conducted in order to compare the estimators for different model specifications.

The starting point will be the structural factor model, which has already been discussed in Section 3.4. For any fixed $n, T \in \mathbb{N}$ and $t = 1, \ldots, T$

$$x_t \ \underset{(n \times 1)}{=} \ \underset{(n \times r)}{L} \quad \underset{(r \times 1)}{z_t} \ + \ \underset{(n \times 1)}{\xi_t} \tag{55a}$$

$$\underset{(r \times r)}{e(z)} z_t = \ \underset{(r \times q)}{b} \quad u_t, \tag{55b}$$

where $e(z)$ is stable, $u_t \sim WN(0, I_q)$ and ξ_t is either white-noise or weakly dependent. The loadings are always generated according to $L_{ij} \sim \mathcal{N}(0,1)$ i.i.d. and the rows of b are uniformly drawn from the q-dimensional unit sphere.

The *integer parameters* are the number of dynamic (q) and static (r) factors and the lag (p) of the autoregressive process of the static factors. *Real parameters* are the factor loadings L, the parameter matrices $(e_1, \ldots, e_p)$, the matrix b and the noise-to-signal ratio ρ_i. Estimation results have been evaluated by means of multivariate linear regression and canonical correlation analysis respectively.

One can also approach model (55) from a different angle. Assumptions B2. and B3. imply that $z_t = k(z)u_t$, where $k(z)$ has a power series expansion in and on the complex unit circle. The entries of the matrix $k(z)$ are rational functions in $z \in \mathbb{C}$. By looking at a right matrix-fraction-description (MFD) $k(z) = N(z)c^{-1}(z)$, where $N(z)$ is a $r \times q$ polynomial matrix of order s, one obtains the following representation:

$$x_t \ \underset{(n \times 1)}{=} \ \underset{(n \times q)}{L_0} \quad \underset{(q \times 1)}{f_t} \ + L_1 f_{t-1} + \ldots + \ \underset{(n \times q)}{L_s} \ \underset{(q \times 1)}{f_{t-s}} + \xi_t \tag{56a}$$

$$f_t \ = \ \underset{(q \times q)}{a_1} \quad \underset{(q \times 1)}{f_{t-1}} + \ldots + a_k f_{t-k} + u_t \tag{56b}$$

for some $s \geq 0$ and $k \geq 0$. Note that $0 < q \leq r \leq (s+1)q$ holds.

Figure 2 describes the simulation analysis in graphical terms. For a fixed panel size $T \times n$, parameters $r, q, p, \sigma_i, L, e(z), b$ and a specification of the process ξ_t, data matrices $Z \in \mathbb{R}^{T \times r}, \xi, X \in \mathbb{R}^{T \times n}$ are generated 25 times[23]. Given X, each estimation method calculates estimators for the static factors. Their different performances will be assessed in the *evaluation* step, whose outcome is a measure of performance for each estimation method. The *data generation, model estimation* and *model evaluation* steps are done for different factor loadings L, static factor dynamics $(e(z), b)$ and noise-to-signal ratios σ_i. The parameters $e(z), b$ can also be held fixed,

---

[23]It is true that this number if very low in order to make statistically reliable comparisons. The low number of repetitions is due to computational restrictions.
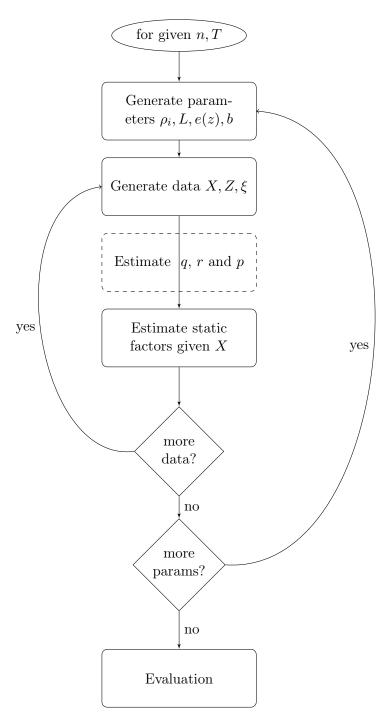
Figure 2: Simple illustration of the simulation setting.

for example if they have been estimated from a real data-set. The generation of parameters has also been repeated 25 times. The number of variables and observations vary, i.e. $(T, n) \in \{25, 50, 100, 150\} \times \{10, 25, 50, 100\}$. The maximum number has been motivated by current macro-economic data-sets[24]. A more detailed description of the different models and parameter settings is given in Section 5.1. Note that the model selection step (choosing q, r and p from the data) is optional and has been omitted in this thesis.

Note, that comparing results of simulations where $(e(z), b)$ have been chosen at random, means comparing the estimation performance for different models. If one estimation method outperforms the other, this could indicate a uniform dominance of that method on the preselected parameter set.

The goal   was to identify settings, where one estimation procedure performs better than the other. To structure this vague formulation, the following questions have been raised.

Questions 5.1.

(a) What influence do different types of noise-misspecifications have on the estimation methods?

(b) Does the assumption $q < r$ have an impact on the estimation results?

(c) What is the influence of different noise-to-signal ratios on the estimation accuracy?

(d) Are the estimation results invariant with respect to the location of the roots of $\det(e(z))$? A root near the complex unit circle would mean that the process had a long memory.

$\square$

Question (a) addresses a key point of estimation methods in generalized dynamic factor models. It asks for the effect of misspecification. Question (c) can be motivated by the theoretical observation that for the case $q < r$, the density of the observations (likelihood) does not exist[25] and therefore a formal description of the EM-algorithm is not straightforward (see Section 4.4). The comparison between models with $r = q$ and $q < r$ may therefore provide a hint for the theoretical analysis of the case $q < r$. Question (d) is a classical statistical question and one expects the noise-to-signal ratio to depend negatively on the estimation accuracy. Moreover it would be interesting to know, whether the relative estimation performances are invariant with respect to the noise-to-signal ratio. A longer memory of the static factor process will make the estimation of the observation covariance matrix more difficult and therefore lead to a decrease in absolute performance of the estimation methods. Question (e) also asks what happens to the relative performances in that case.

Simulation studies done in other papers.   Comparable simulation studies have been conducted in the papers of Stock and Watson [49] and Doz, Giannone and Reichlin [17], [18].

---

[24]The data-set which has been analyzed in Section 6 is of dimension 194×107.

[25]See Definition C.13 for the general definition of a normally distributed n-dimensional random vector x with $\mathrm{rk}\mathbb{E}(xx') = r < n$.

5.1 Simulation scenarios

A simulation scenario is determined by the integer parameters q, r, p, by the spectrum of ξ_t, by the parameters of the ARMA representation $(e(z), b)$ of z_t, and by the signal-to-noise ratios of the observations x_{it}.

Model parameters. The dimension of the static factor process r can be greater of equal to the dimension of the dynamic factor process q. The lag-parameter for the static factor dynamic p will be 1, 2 or 3, and the noise-to-signal ratio[26] $\rho_i = \mathbb{V}\xi_{it}/\mathbb{V}x_{it}$, $i = 1, \ldots, n$ is either fixed to the values $\{0.5, 0.75, 0.9\}$ or drawn uniformly from the interval $[0.1, 0.9]$.

Idiosyncratic component. There are four different specifications for the idiosyncratic component. The first one is $\xi_t \sim \mathcal{N}(0, I_q)$ iid, and denoted as the *strict* case. The second setting $\xi_t \sim \mathcal{N}(0, \Xi_0)$ iid, assumes a small amount of cross-sectional dependence among the idiosyncratic influences. The entries of Ξ_0 are defined by

$$\xi_{it} = \alpha_{it} + \frac{1}{2}(\alpha_{i-1t} + \alpha_{i+1t}), \quad \alpha_{it} \sim \mathcal{N}(0, 1), \ \mathbb{E}\alpha_{it}\alpha_{jt} = \delta_{ij}, \ i, j = 1, \ldots, n,$$

where $\alpha_{0t} = \alpha_{n+1t} = 0$. To obtain the correct noise-to-signal ratio, ξ_{it} needs to be weighted[27]. I denote this setting by *cross*.

The third noise setting adds the feature of time-dependence (it is therefore called *cross-time*)

$$\xi_{it} = \alpha_{it} + \frac{1}{2}(\alpha_{i-1t} + \alpha_{i+1t}) + \frac{1}{5}\alpha_{it-1}, \quad \alpha_{it} \sim \mathcal{N}(0, 1), \ \mathbb{E}\alpha_{it}\alpha_{jt} = \delta_{ij}, \ i, j = 1, \ldots, n,$$

where $\alpha_{0t} = \alpha_{n+1t} = 0$. Again, re-weighting of ξ_{it} needs to be done. These equations define Ξ_1.

The last most general noise setting (denoted as the *general* setting) is taken from the simulation study conducted by [18]. It assumes that the idiosyncratic component has a Wold representation defined by

$$\xi_{it} = d\xi_{it-1} + \varepsilon_{it}, \ d = \frac{1}{2}, \quad \varepsilon_t \sim \mathcal{N}(0, \Psi) \text{ iid},$$

where

$$\Psi_{ij} = \psi^{|i-j|}\left(1 - d^2\right)\sqrt{a_i}\sqrt{a_j}, \quad 0 < \psi < 1, \ i, j = 1, \ldots, n,$$

and $a_i = (\sigma_i/(1-\sigma_i))\mathbb{V}\chi_{it}$. Note that ψ controls for the amount of cross-sectional correlation. It is a key value in order to address Question (a). For the *general* noise-setting, $\psi = 0.5$. The cases where $\psi = 0.75, 0.9, 0.95$ are denoted by *g.psi075, g.psi09, g.psi095* respectively.

---

[26]Note that the signal here are the observations x_t (not χ_t) and the noise is ξ_t. One could also interpret the latent process χ_t as the unobserved noise-free signal.

[27]Define $\tilde{\xi}_{it} := \alpha\xi_{it}$ Then $\mathbb{V}\tilde{\xi}_{it}/\mathbb{V}\chi_{it} = \rho_i/(1 - \rho_i)$ and the squared correction factor becomes $\alpha^2 = (1/\mathbb{V}\xi_{it})(\rho_i/(1 - \rho_i))\mathbb{V}\chi_{it}$. In the first case $\mathbb{V}\xi_{it} = 3/2$, in the second $\mathbb{V}\xi_{it} = 77/50$.

Static factors. The r-dimensional static factors follow a singular $AR(p)$-process

$$z_{it} = e_1(i)z_{t-1} + e_2(i)z_{t-2} + \ldots + e_p(i)z_{t-p} + b(i)\, u_t, \ i = 1, \ldots, r$$

where $u_t \sim WN(0, I_q)$. Define $b = (b_{ij})_{i=1,\ldots,r,\, j=1,\ldots,q}$, as $b_{ij} := (\tilde{b}_{ij}/\|\tilde{b}(i)\|_2)$, with $\tilde{b}_{ij} \sim \mathcal{N}(0,1)$. Then $\nu_t = bu_t \sim WN(0, D)$, where $D_{ii} = 1, \ i = 1, \ldots, r$.

The matrix $e(z)$ has been modeled in three different ways. First, as diagonal, i.e. $e(z) = diag(e_1(z), \ldots, e_r(z))$, with random coefficients for the polynomials $e_i(z)$. Second, as determined by means of an empirical analysis (see Section 6). The third way is connected to an attempt to address the problem of how to uniformly draw from the stability region of $e(z)$. A standard fact is that this region is open.

Lemma 5.2 The parameter space for stable vector r-dimensional autoregressive processes of order p is an open subset of $\mathbb{R}^{2rp}$.

Proof. The process is described by $a(z)x_t = u_t$. Being stable means $\det a(z) \neq 0$ for all $|z| \leq 1$, which is equivalent to $|\lambda_{\max}(A)| < 1$, where A is the companion form of $a(z)$. Calculating the largest eigenvalue of a matrix and the modulus is a continuous operation[28] and the set $(-1, 1)$ is open.   □

For the univariate case $(r = 1)$, a sampling procedure for parameters of an $ARMA(p, q)$ system based on the Levinson-Durbin algorithm has been discussed [7]. For the multivariate case no similar procedure has been found. Often the parameters are generated by the "generate and test" methods. The following procedure is going in this direction.

Algorithm 5.3 (*Generate parameters of $e(z)$*). Let k be the number of restrictions on the matrix $(e_1, e_2, \ldots, e_p) \in \mathbb{R}^{r \times rp}$.

1. Draw $r^2 p - k$ times from standard normal distribution to set $e^{(0)} = (e_1^{(0)}, e_2^{(0)}, \ldots, e_p^{(0)})$.

2. $(m-1) \mapsto (m)$: As long as there exist $y \in \mathbb{C}$ such that $|y| < 1$ and $\det e(y) = 0$, multiply every non-restricted entry of $e^{(m-1)}$ by $0 < \vartheta < 1$ to obtain $e^{(m)}$.

                                                                                                        □

In general $e^{(0)}$ is unstable. If ϑ is set near to one, the determinant of $e(z)$ will have zeros near the unit circle. If one wishes to have a certain distance of the smallest root to the unit circle, a ϑ near zero will help. Tables 1 and 2 indicate the support of the empirical distribution of the modulus of the smallest root of randomly generated $e(z)$.

The number of iterations of Algorithm 5.3 increases with ϑ. Generating a system with a root very close to the unit circle therefore comes at a price, as can be seen in Table 3. The proposed method is just another method of choosing random coefficient matrices. It is not a method for uniformly selecting coefficients of polynomials $e(z)$ from their stability region. This argument is supported by Figure 3. In this figure also the effect of choosing ϑ near to one can be seen.

Factor loading matrix. The factors loadings $L \in \mathbb{R}^{n \times r}$ have been modeled in such a way that the factors have a "stationary" influence on the observations. This means that $L_{ij} \sim \mathcal{N}(0,1)$ i.i.d., $i = 1, \ldots, n, \ j = 1, \ldots, r$.

---

[28]See Theorem C.3 for the definition of the largest eigenvalue.
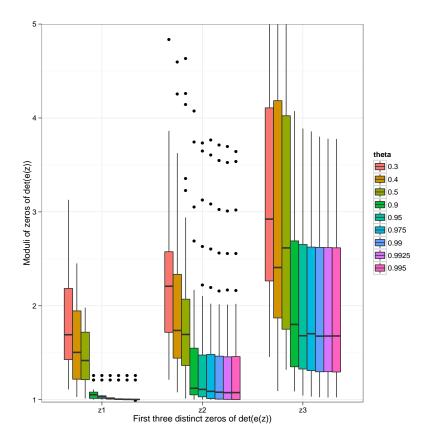
Figure 3: The empirical distribution of distinct zeros z_0 of $\det(e(z_0))$ for randomly generated stable 4×4 polynomial matrices $e(z)$ of order one.
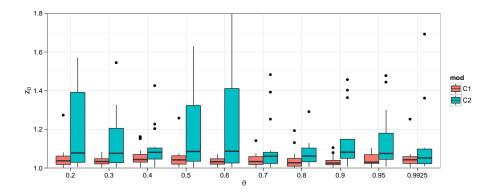
Figure 4: The empirical distribution of the smallest root z_0 of $\det(e(z))$ for models C1 and C2.

ϑ	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3	1.0280	1.4930	1.8590	1.9800	2.4350	3.1720
0.4	1.0130	1.1990	1.4720	1.5760	1.9050	2.3790
0.5	1.0090	1.1980	1.4620	1.4740	1.7160	1.9530
0.9	1.0020	1.0380	1.0650	1.0890	1.0960	1.7890
0.95	1.0010	1.0140	1.0280	1.0590	1.0440	1.7890
0.975	1.0000	1.0070	1.0130	1.0470	1.0220	1.7890
0.99	1.0000	1.0040	1.0070	1.0410	1.0090	1.7890
0.9925	1.0000	1.0020	1.0040	1.0390	1.0060	1.7890
0.995	1.0000	1.0010	1.0030	1.0380	1.0040	1.7890

Table 1: Summary statistics for the modulus of the smallest root of $\det(e(z))$ for different ϑ. The algorithm has been applied to polynomial matrices $e(z)$ of dimension 3×3 and order 1.

ϑ	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3	1.1580	1.3950	1.6760	1.7920	2.0320	3.2410
0.4	1.0200	1.1910	1.3890	1.5470	1.8270	2.4820
0.5	1.0010	1.1730	1.4830	1.4490	1.6670	1.9890
0.9	1.0010	1.0280	1.0470	1.0510	1.0770	1.1090
0.95	1.0000	1.0150	1.0260	1.0250	1.0350	1.0520
0.975	1.0010	1.0080	1.0140	1.0140	1.0200	1.0240
0.99	1.0000	1.0020	1.0050	1.0050	1.0080	1.0100
0.9925	1.0000	1.0020	1.0040	1.0040	1.0050	1.0070
0.995	0.9041	1.0010	1.0030	0.9994	1.0040	1.0050

Table 2: Summary statistics for the modulus of the smallest root of $\det(e(z))$ for different ϑ. The algorithm has been applied to polynomial matrices $e(z)$ of dimension 4×4 and order 1.

5.2   The models

The following data-generating processes had been the basis of analysis[29].

Models A1 ($q = r = 1$, $p = 1, 2, 3$)   This model has been used in [18] in order to demonstrate the finite sample performance of the TS-estimator (see Section 4.3). For $i = 1, \ldots, n$, $t = 1, \ldots, T$

$$x_{it} = L(i)z_t + \xi_{it}, \quad H(i), z_t \in \mathbb{R}$$
$$z_t = az_{t-1} + u_t, \quad |a| < 1, \ u_t \sim \mathrm{WN}(0, 1)$$

Normally the parameter a is uniformly drawn from $[0.6, 0.9]$. For generating processes with a longer memory, a is uniformly drawn from $[0.998, 0.9992]$.

---

[29]Note that the word model can be misleading here. A specific model is characterized by its fixed parameters $r, q, p, e(z), b$. The following categories A1, A2, A3, B1, B12, B2, C1, and C2 represent rather a collection of different models, because they involve the random generation of $e(z)$ and b.

ϑ	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.9	2	8	9	9	11	17
0.95	4	16	17	18	21	34
0.975	8	33	34	36	43	68
0.99	19	81	86	91	107	171
0.9925	26	108	114	121	143	228
0.995	39	162	172	176	214	251

Table 3: Number of iterations until the 3×3 polynomial matrix $e(z)$ of order 1 was stable for different ϑ.

Models A2 ($q = r = 3$, $p = 1, 2, 3$)   In [17] a simulation study has been conducted based on the following model. For $i = 1, \ldots, n$, $t = 1, \ldots, T$

$$x_{it} = L(i)z_t + \xi_{it}, \quad H(i)', z_t \in \mathbb{R}^{3 \times 1}$$
$$e(z)z_t = u_t, \quad u_t \sim \mathrm{WN}(0, I_3)$$

where $e(z) = \begin{pmatrix} 1 - a_1 z & 0 & 0 \\ 0 & 1 - a_2 z & 0 \\ 0 & 0 & 1 - a_3 z \end{pmatrix}$. For $p = 1$ the coefficients are uniformly drawn from different intervals, i.e. $a_1 \sim U_{[0.85, 0.9]}$, $a_2 \sim U_{[0.75, 0.85]}$, $a_3 \sim U_{[0.65, 0.75]}$. If the process shall have a long memory, then $a_1 = 0.999$, $a_2 \sim U_{[0.95, 0.9925]}$, $a_3 \sim U_{[0.85, 0.95]}$. Also the more general parameterization (see Section 5.1) for $e(z)$ has been used for $p = 1, 2, 3$.

Models A3 ($q = r = 4$, $p = 1, 2, 3$)   The specification corresponds to the specification of model A2 with the appropriate dimensions. For the case $e(z) = diag(e_1(z), \ldots, e_4(z))$ and $p = 1$, the coefficients were generated according to the rule $a_1 \sim U_{[0.85, 0.9]}$, $a_2 \sim U_{[0.75, 0.85]}$, $a_3 \sim U_{[0.7, 0.8]}$, $a_4 \sim U_{[0.65, 0.75]}$. For the long-memory case, $a_1 = 0.999$, $a_2 \sim U_{[0.95, 0.9925]}$, $a_3 \sim U_{[0.85, 0.95]}$, and $a_4 \sim U_{[0.8, 0.9]}$. Again random and stable non-diagonal $e(z)$ with higher orders $p = 1, 2, 3$ were also generated.

Models B1 ($q = 1, r = 3$)   For $i = 1, \ldots, n$, $t = 1, \ldots, T$

$$x_{it} = L(i)z_t + \xi_{it}, \quad H(i)', z_t \in \mathbb{R}^{3 \times 1}$$
$$e(z)z_t = bu_t, \quad u_t \sim \mathrm{WN}(0, 1)$$

where $e(z)$ is specified as in Model A2 and the rows of b are chosen randomly from the unit sphere in $\mathbb{R}^q$ (see Section 5.1).

Model B12 ($q = 2, r = 3$)   This model has been motivated by the data-set of Section 6. The parameters $p, b, e(z)$ and L have been estimated and are held fixed during the simulation analysis. For this model, $p = 4$.

Models B2 $(q = 2, r = 4)$ The specification is the same as in Model B1 and differs only in the dimensions of the variables. The parameters $p, b, e(z)$ and L are either generated randomly or set to estimated values which had been determined in the empirical study of Section 6.

The following models assume restrictions on the parameter matrices of $e(z)$. Assume that the DGP is described by model (56).

Model C1   Let

$$x_t = L_0 f_t + L_1 f_{t-1} + \ldots + L_s f_{t-s} + \xi_t$$
$$f_t = e_1 f_{t-1} + e_2 f_{t-2} + u_t$$

where u_t and f_t are q dimensional stationary processes. We set $s = 1$. Define $z_t = (f_t', f_{t-1}')'$, then

$$x_t = (L_0, L_1) z_t + \xi_t$$
$$z_t = F z_{t-1} + b u_t$$

with $F = \begin{pmatrix} e_1 & e_2 \\ I_q & 0 \end{pmatrix}$ and $b = (I_q, 0)' \in \mathbb{R}^{2q \times q}$.

Model C2   Let

$$x_t = L_0 f_t + L_1 f_{t-1} + \ldots + L_s f_{t-s} + \xi_t$$
$$f_t = e_1 f_{t-1} + u_t$$

where u_t and f_t are q dimensional stationary processes and $s = 1$. Define $z_t = (f_t', f_{t-1}')'$, then

$$x_t = (L_0, L_1) z_t + \xi_t$$
$$z_t = F z_{t-1} + b u_t$$

with $F = \begin{pmatrix} e_1 & 0 \\ I_q & 0 \end{pmatrix}$ and $b = (I_q, 0)' \in \mathbb{R}^{2q \times q}$.

Checking the assumptions.

- *Strong dependence:* The r-dimensional process z_t has been modeled as a non-orthonormal basis of the factor space. It is easy to see that $\gamma_z(0) > 0^{30}$. As $L_{ij} \sim \mathcal{N}(0, 1)$ it follows that $\frac{1}{n}(L'L)_{ij} = \frac{1}{n}(L_i)'L_j = \frac{1}{n}\sum_{k=1}^{n} L_{ki}L_{kj} \to \delta_{ij}$ as $n \to \infty$. Using Lemma C.5 it follows that the eigenvalues of $L'L/n$ converge to 1.

  Now $\gamma_\chi^n(0) = L^n \gamma_z(0) L^{n\prime} = (L^n M)(L^n M)' \in \mathbb{R}^{n \times n}$, where $\gamma_z(0) = MM'$ and $\mathrm{rk}M = r$. The first r eigenvalues of $\gamma_\chi(0)$ are equal to the first r eigenvalues of

---

[30]The components of z_t are linear independent a.e. for all $t \in \mathbb{Z}$ if and only if $\gamma_z(0) > 0$.

$$(L^n M)'(L^n M) = M'L^{n'}L^n M \geq \lambda_{\min}(L^{n'}L^n)M'M$$

which implies $\lambda_{\min}(M'L^{n'}L^n M) \geq \lambda_{\min}(L^{n'}L^n)\lambda_{\min}(M'M)$. As $\mathrm{rk}M = r$, its smallest eigenvalue is greater than zero and following the argument from above $\lambda_{\min}(L^{n'}L^n) \to \infty$ as $n \to \infty$. It follows that the first r eigenvalues of $\gamma_\chi(0)$ diverge to infinity for large n.

With similar arguments the divergence of the first q eigenvalues of the spectrum f_χ can be shown. As $\chi_t = Lz_t$ and $z_t = k(z)u_t$, where $k(z) = e^{-1}(z)b$ is a $r \times q$ dimensional rational transfer function without roots in and on the complex unit circle, it follows that $f_z(\theta) = \frac{1}{2\pi}e^{-1}(e^{-i\theta})bb'(e^{-1}(e^{-i\theta}))^*$. Because $\det(e^{-1}(z)) = (\det(e(z)))^{-1} \neq 0$ for all $|z| \leq 1$ and $\mathrm{rk}b = q$, the rank of the spectrum of z_t is equal to q for all $\theta \in \Theta$. In our setting $\mathrm{rk}L^n = r$ for all $n \in \mathbb{N}$ and therefore

$$\mathrm{rk}(f_\chi(\theta)) = \mathrm{rk}(Lf_z(\theta)L') = q, \quad \forall \theta \in \Theta.$$

As $\mathrm{rk}(k(z)) = q$ the smallest eigenvalue of $k^*(z)k(z)$ is greater than zero and

$$(L^n k(e^{-i\theta}))^*(L^h k(e^{-i\theta})) \geq \lambda_{\min}(L^{n'}L^n)k^*(z)k(z).$$

- *Weak dependence:* The covariances of all four noise classifications have the following structure:

$$\gamma_\xi^n(0) = \begin{pmatrix} c_0 b_{11} & c_1 b_{12} & \cdots & \cdots & c_{n-1} b_{1n} \\ c_1 b_{12} & c_0 b_{22} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \\ \vdots & & & \ddots & c_1 b_{n-1,n} \\ c_{n-1} b_{1n} & \cdots & \cdots & c_1 b_{n-1,n} & c_0 b_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

The b_{ij} correspond to the noise-to-signal ratio corrections, whereas $0 < c_j < 1$. In the general noise setting $b_{ij} = \sqrt{a_i}\sqrt{a_j}$, where $a_i = (\sigma_i/(1 - \sigma_i))\mathbb{V}\chi_{it}$. The variance of χ_{it} is bounded for all $i \in \mathbb{N}$ with a very high probability[31] or can be easily bounded by scaling the loadings $L(i)$.

This implies the existence of a $K \in \mathbb{N}$ independent of n such that $x'\gamma_\xi^n(0)x \leq K \sum_{j=0}^{n-1}|c_j|$. To see this consider the case $n = 3$:

$$\gamma_\xi(0) = c_0 \underbrace{\begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix}}_{=:M_0} + c_1 \underbrace{\begin{pmatrix} 0 & b_{12} & 0 \\ b_{12} & 0 & b_{23} \\ 0 & b_{23} & 0 \end{pmatrix}}_{=:M_1} + c_2 \underbrace{\begin{pmatrix} 0 & 0 & b_{13} \\ 0 & 0 & 0 \\ b_{13} & 0 & 0 \end{pmatrix}}_{=:M_2}$$

It is well known that for symmetric matrices M_i the norm inequality $||M_i||_2 \leq ||M_i||_1 = \max_{j=1,\ldots,n}\sum_{k=1}^n |M_i(k,j)|$ holds. Therefore $x'\gamma_\xi^n(0)x \leq \sum_{j=0}^{n-1} c_0\lambda_{\max}(M_j)$. If we model L such that $|L_{ij}| \leq \bar{\ell}$ for all $i \in \mathbb{N}, j = 1, \ldots, r$, then $K = 2\bar{\ell}$.

---

[31]Note, that $\mathbb{V}\chi_{it} \leq \lambda_{\max}(\gamma_z(0))L(i)L(i)'$, and that $L(i)L(i)' \sim \chi_r^2$.

In the first three specifications (strict, cross, cross-time) $\gamma_\xi^n(0)$ is a band matrix and its maximal eigenvalue is therefore bounded for all $n \in \mathbb{N}$. In the last specification the sequence $\left(\sum_{j=0}^{n-1} |c_j|\right)_{n \in \mathbb{N}}$ converges in $\mathbb{R}$ as n goes to infinity. It can be easily verified that also the spectrum of ξ_t is uniformly bounded for all $n \in \mathbb{N}$.

It can also easily be seen that $\inf_{n \in \mathbb{N}} \lambda_{\min}(\Sigma_\xi^n) > 0$ in all four settings.

5.3   Evaluation

Assumption 5.4. The model-parameters $r, q, p \in \mathbb{N}$ are assumed to be known and will not be estimated with model-selection procedures[32]. $\qquad\square$

Multivariate Linear Regression.   We assume the estimated factors $\breve{Z} \in \mathbb{R}^{T \times r}$ are a linear function of its simulated counterparts $Z \in \mathbb{R}^{T \times r}$ plus a stochastic error term. The model is a special case of the SUR (seemingly unrelated regression) model[33]. Let $Y = (Y_1, Y_2, \ldots, Y_r) \in \mathbb{R}^{T \times r}$ be the dependent variables and $X = (X_1, X_2, \ldots, X_m) \in \mathbb{R}^{T \times m}$ be the independent variables. The model is then

$$Y = X\beta + \varepsilon, \quad \beta \in \mathbb{R}^{m \times m} \tag{57}$$

$$\text{vec}(Y) = (I_m \otimes X)\,\text{vec}(\beta) + \text{vec}(\varepsilon), \tag{58}$$

where $\mathbb{E}(\varepsilon\varepsilon'|X) = \Sigma \otimes I_m$ with $\Sigma_{ij} = \mathbb{E}\varepsilon_{it}\varepsilon_{jt}$. It is therefore assumed that the errors may be linear dependent across equations but linear independent across time.

It can be easily shown that the OLS estimator $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_m)$ is determined by OLS equation by equation. This means that $\hat{\beta}_i = (X'X)^{-1}X'Y_i$, $i = 1, \ldots, m$.

Now suppose that $r = m$ and $X = Z$, $Y = \breve{Z}$, $Z, \breve{Z} \in \mathbb{R}^{T \times r}$, where $\breve{Z}$ is the result of one of the three estimates of Section 4. The measure for the estimation accuracy is the following trace statistic:

$$\mathcal{R}^2 := 1 - \frac{\text{tr}(e'e)}{\text{tr}(\breve{Z}'\breve{Z})} = \frac{\text{tr}(\breve{Z}'\breve{Z}) - \text{tr}(e'e)}{\text{tr}(\breve{Z}'\breve{Z})} = \frac{\text{tr}(\breve{Z}'I_T\breve{Z} - e'e)}{\text{tr}(\breve{Z}'\breve{Z})} = \frac{\breve{Z}'Z(Z'Z)^{-1}Z'\breve{Z}}{\text{tr}(\breve{Z}'\breve{Z})}$$

where $e = \breve{Z} - Z\hat{\beta} \in \mathbb{R}^{T \times r}$. The last equation holds because

$$e'e = \breve{Z}'(I_t - Z(Z'Z)^{-1}Z')(I_t - Z(Z'Z)^{-1}Z')\breve{Z} = \breve{Z}'I_T\breve{Z} - \breve{Z}'Z(Z'Z)^{-1}Z'\breve{Z}.$$

Note that one could also set $X = \breve{Z}$, $Y = Z$ and define $\mathcal{R}_*^2 = \frac{Z'\breve{Z}(\breve{Z}'\breve{Z})^{-1}\breve{Z}'Z}{\text{tr}(Z'Z)}$.

---

[32]The only exception is model B2, where results from the empirical analysis of Section 6 are used for the simulation study. This results however will not be reported, because the EM algorithm did not work in this setting.

[33]A brief introduction can be found for example in [30].

Canonical Correlation Analysis.   The idea of canonical correlation analysis is to assess the correlation between two random vectors of possibly different dimensions. Consider the n-dimensional random vector $x = (x_1, x_2) : \Omega \to \mathbb{R}^n$, where $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ with $n = n_1 + n_2$. Let $\mathbb{E}x = \mu = (\mu_1', \mu_2')' \in \mathbb{R}^n$ be the mean of x and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad \Sigma > 0$$

be the covariance matrix of x. If one assumes that $\Sigma_{11} > 0$ and $\Sigma_{22} > 0$ respectively, then one could consider the n_1- and n_2-dimensional spaces $\mathcal{H}_1 := \mathrm{span}\{x_{11}, \ldots, x_{1n_1}\} \subseteq L_2(\mathbb{P})$ and $\mathcal{H}_2 := \mathrm{span}\{x_{21}, \ldots, x_{2n_2}\} \subseteq L_2(\mathbb{P})$.

> The idea of canonical correlation analysis (CCA) is to find new coordinate systems in $\mathcal{H}_1$ and $\mathcal{H}_2$ which display unambiguously the system of correlation between x_1 and x_2 (see [3]).

As x_1 and x_2 are already a basis of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, this means finding a non-singular matrix which transforms the old bases x_1 and x_2 into new bases $y_1 \in \mathbb{R}^{n_1}$ and $y_2 \in \mathbb{R}^{n_2}$ respectively. This new bases will be the pairs of canonical correlations.

Definition 5.5 (CCA at population level, [3], p. 495):   The r-th pair of canonical variates is the pair of linear combinations $y_1^{(r)} = a^{(r)'}x_1$ and $y_2^{(r)} = b^{(r)'}x_2$, each of unit variance, uncorrelated with the first $r - 1$ pairs of canonical variates and having maximum correlation. The correlation betwen $y_1^{(r)}$ and $y_2^{(r)}$ is the r-th *canonical correlation*. The components of $y_1^{(r)}$, $y_2^{(r)}$ are called the *canonical components*. The vectors $a^{(r)}$, $b^{(r)}$ are called the *coefficients of the canonical components*.

Theorem 5.6 (Sample CCA, [3], p. 500):   Let $x^{(1)}, \ldots, x^{(N)}$ be N observations from $x \sim \mathcal{N}(\mu, \Sigma)$. The covariance matrices are estimated by their sample counterparts, $\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x^{(i)}})(x^{(i)} - \overline{x^{(i)}})'$. The maximum likelihood estimators of the canonical correlations are the $\min(n_1, n_2)$ roots of

$$\det \begin{pmatrix} -c\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -c\hat{\Sigma}_{22} \end{pmatrix} = 0 \tag{59}$$

The maximum likelihood estimators of the coefficients of the j-th canonical component satisfy

$$\begin{pmatrix} -c_j\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -c_j\hat{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \hat{a}^{(j)} \\ \hat{b}^{(j)} \end{pmatrix} = 0 \tag{60}$$

and

$$\hat{a}^{(j)'}\hat{\Sigma}_{11}\hat{a}^{(j)} = 1, \quad \hat{b}^{(j)'}\hat{\Sigma}_{22}\hat{b}^{(j)} = 1$$

for $j = 1, \ldots, \min(n_1, n_2)$.

The results of a sample CCA between the simulated and the estimated static factors are r canonical correlations $1 \geq c_1 \geq \cdots \geq c_r \geq 0$ which should be near 1. One could look at the different c_j, especially on the last one to assess the different performances of the estimation procedures.

Another possibility would be to summarize the r canonical correlations by one statistic, the group correlation coefficient [22], which is defined as

$$\rho_{x_1,x_2}^2 := 1 - \prod_{j=1}^{min(n_1,n_2)} (1 - c_j) \tag{61}$$

where c_j are the canonical correlation coefficients.

Evaluation 5.7. The following statistics are proposed for evaluation the "quality" of the estimation procedures.

Multivariate goodness of fit (R^2) $\mathcal{R}^2 = 1 - \frac{\sum_{i=1}^{r} \sum_{t=1}^{T} e_{it}^2}{||\breve{Z}||_F^2} = 1 - \frac{||e||_F^2}{||\breve{Z}||_F^2} = 1 - \frac{vec(e)'vec(e)}{vec(\breve{Z})'vec(\breve{Z})}$

Group correlation coefficient $\rho_{Z,\breve{Z}} = 1 - (1 - c_1) \cdots (1 - c_r)$

Last canonical correlations Compare the last $c_j^{(i)}$, $j = r, r-1, r-k$ for each $i = 1, 2, 3$ (estimation method)

$\square$

The multivariate R^2 has been selected for further analysis.

Notation for analysis. For each data-set of dimension $T \times n$ that comes from a model μ of type A1, A2, A3, B1, B2, C1 and C2, the performance of an estimation method $\nu \in \{PCA, TS, QML\}$ is denoted by $R^2(\mu, \nu, n, T)$. The variable μ is either used for a specific model of a certain type (fixed $e(z), b$) or for the type itself (random $e(z), b$). To assess differences in the relative performances the distribution of $\frac{R^2(\mu,\nu_1,n,T)}{R^2(\mu,\nu_2,n,T)}$ is considered. Define $\mathcal{R}_{\mu,\nu_1/\nu_2} := \left\{ \frac{R^2(\mu,\nu_1,n,T)}{R^2(\mu,\nu_2,n,T)} \right\}$ as the set of relative performances between methods ν_1 and ν_2 for a model (of type) μ and fixed n, T. The set $\mathcal{R}_{\mu,\nu_1/\nu_2}$ contains 625 items, if μ represents a model type[34], e.g. A1, and it contains 250 items if μ is a fixed model with estimated parameters from the Stock and Watson data-set.

---

[34]For each fixed model μ of a certain type 25 estimation results are available. The parameters were randomly drawn 25 times.

5.4 Results

The *strategy of analysis*: For the strict case $\xi_t \sim \mathrm{WN}(0, diag)$ one expects the TS- and QML-method to outperform the PC-method. Departing from this case, the impact of a non-diagonal spectrum of ξ_t on the absolute and relative estimation performances will be shown. Furthermore, the impact of modelling a long-memory for the static factor process and high noise-to-signal ratios respectively on these results is of interest.

Absolute performances for each model are given in Section A. The EM-algorithm did not work without problems. A more detailed report about this is given at the end of this section.

The case: ξ_t is white-noise with diagonal covariance matrix

When the idiosyncratic process ξ_t is white-noise with a diagonal covariance matrix, then the models (40) and (47) are state-space models. Therefore, we expect a performance advantage of the TS- and QML-method compared to the PC-method. Results are reported for simulations where the polynomials $e(z)$ were non-diagonal, stable and $\sigma_i \sim U_{[0.1, 0.9]}$, $i = 1, \ldots, n$.

n/T	A1	A2	A3	B1	B2	C1	C2
n=10							
T=25	1.024 (0.026)	1.024 (0.023)	1.018 (0.024)	1.038 (0.058)	1.007 (0.052)	1.061 (0.086)	1.039 (0.068)
T=50	1.033 (0.019)	1.037 (0.02)	1.034 (0.023)	1.059 (0.06)	1.025 (0.059)	1.082 (0.08)	1.06 (0.061)
T=100	1.035 (0.015)	1.046 (0.021)	1.042 (0.027)	1.064 (0.072)	1.031 (0.066)	1.084 (0.082)	1.066 (0.06)
T=150	1.035 (0.014)	1.05 (0.019)	1.046 (0.027)	1.066 (0.071)	1.028 (0.076)	1.085 (0.08)	1.064 (0.062)
n=25							
T=25	1.008 (0.009)	1.019 (0.011)	1.022 (0.015)	1.029 (0.024)	1.024 (0.024)	1.049 (0.045)	1.035 (0.038)
T=50	1.01 (0.006)	1.025 (0.01)	1.033 (0.017)	1.033 (0.02)	1.039 (0.021)	1.061 (0.04)	1.047 (0.029)
T=100	1.011 (0.004)	1.026 (0.009)	1.039 (0.018)	1.036 (0.019)	1.047 (0.021)	1.066 (0.03)	1.053 (0.024)
T=150	1.011 (0.004)	1.028 (0.008)	1.041 (0.017)	1.036 (0.018)	1.047 (0.02)	1.066 (0.022)	1.054 (0.02)
n=50							
T=25	1.004 (0.005)	1.011 (0.006)	1.015 (0.008)	1.016 (0.013)	1.017 (0.014)	1.028 (0.028)	1.022 (0.023)
T=50	1.005 (0.002)	1.013 (0.004)	1.019 (0.007)	1.018 (0.01)	1.023 (0.013)	1.039 (0.021)	1.03 (0.016)
T=100	1.005 (0.002)	1.013 (0.003)	1.02 (0.005)	1.019 (0.009)	1.025 (0.012)	1.037 (0.014)	1.031 (0.01)
T=150	1.005 (0.001)	1.014 (0.003)	1.02 (0.004)	1.02 (0.009)	1.026 (0.012)	1.036 (0.01)	1.031 (0.008)
n=100							
T=25	1.002 (0.002)	1.006 (0.003)	1.008 (0.004)	1.01 (0.008)	1.01 (0.007)	1.016 (0.019)	1.013 (0.015)
T=50	1.002 (0.001)	1.006 (0.002)	1.009 (0.003)	1.01 (0.005)	1.012 (0.006)	1.019 (0.011)	1.017 (0.009)
T=100	1.002 (0.001)	1.006 (0.002)	1.009 (0.002)	1.01 (0.004)	1.013 (0.006)	1.019 (0.006)	1.017 (0.005)
T=150	1.002 (0.001)	1.006 (0.001)	1.009 (0.002)	1.01 (0.004)	1.013 (0.006)	1.018 (0.004)	1.016 (0.004)

Table 4: Medians and mean absolute deviations of $\mathcal{R}_{\mu, ts/pc}$.

Tables 4–6 and Figures 5–7 confirm a performance advantage on average[35]. By looking at the spread and the distributions of the results, a small sample effect ($n = 10$) leads to outliers in the region, where the PC-estimator outperforms the TS- and QML-estimator respectively[36]. These outliers disappear for larger panels except for the models C1 and C2.

For small panels the spread of results seems to be higher for models where $q < r$ compared to models where $r = q$. Nevertheless, especially for a small amount of variables ($n = 10, 25$)

---

[35]Although the medians are reported, the results also hold for the mean. The median was taken as a robust statistic of the mean.

[36]This effect was strongest for models B2. An example where this happened is the specification $e(z) = diag(e_1(z), \ldots, e_4(z))$ of models A3. It disappeared, when $e_3(z) = 1 - 0.35z$ and $e_4(z) = 1 - 0.2z$.
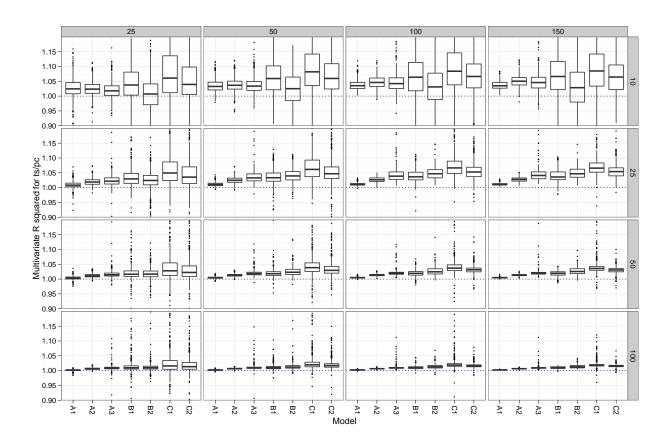
Figure 5: The distribution of $\mathcal{R}_{\mu,ts/pc}$ for different model specifications with $\xi_t \sim \mathrm{WN}(0, diag)$.

n/T	A1	A2	A3	B1	B2	C1	C2
n=10							
T=25	1.03 (0.038)	1.039 (0.055)	1.038 (0.067)	1.061 (0.071)	1.037 (0.062)	1.123 (0.114)	1.079 (0.078)
T=50	1.038 (0.025)	1.075 (0.048)	1.086 (0.078)	1.086 (0.073)	1.081 (0.066)	1.182 (0.113)	1.132 (0.072)
T=100	1.039 (0.02)	1.09 (0.044)	1.114 (0.089)	1.099 (0.072)	1.107 (0.069)	1.199 (0.111)	1.143 (0.071)
T=150	1.038 (0.017)	1.094 (0.042)	1.121 (0.086)	1.101 (0.072)	1.112 (0.068)	1.204 (0.097)	1.148 (0.059)
n=25							
T=25	1.009 (0.012)	1.026 (0.02)	1.039 (0.033)	1.038 (0.028)	1.034 (0.029)	1.081 (0.065)	1.052 (0.04)
T=50	1.011 (0.007)	1.033 (0.015)	1.056 (0.032)	1.042 (0.024)	1.051 (0.027)	1.099 (0.056)	1.065 (0.035)
T=100	1.011 (0.004)	1.035 (0.012)	1.061 (0.026)	1.043 (0.023)	1.055 (0.024)	1.092 (0.036)	1.067 (0.027)
T=150	1.011 (0.004)	1.036 (0.01)	1.061 (0.021)	1.042 (0.022)	1.055 (0.023)	1.088 (0.028)	1.066 (0.024)
n=50							
T=25	1.004 (0.006)	1.015 (0.01)	1.023 (0.015)	1.021 (0.015)	1.022 (0.018)	1.05 (0.039)	1.037 (0.027)
T=50	1.005 (0.003)	1.015 (0.006)	1.025 (0.01)	1.021 (0.011)	1.027 (0.015)	1.049 (0.024)	1.037 (0.016)
T=100	1.005 (0.002)	1.016 (0.005)	1.025 (0.007)	1.022 (0.009)	1.027 (0.013)	1.043 (0.014)	1.035 (0.011)
T=150	1.005 (0.002)	1.016 (0.004)	1.024 (0.005)	1.021 (0.009)	1.027 (0.013)	1.04 (0.01)	1.034 (0.009)
n=100							
T=25	1.002 (0.002)	1.007 (0.004)	1.011 (0.007)	1.012 (0.008)	1.013 (0.009)	1.026 (0.022)	1.022 (0.018)
T=50	1.002 (0.002)	1.007 (0.003)	1.011 (0.004)	1.011 (0.006)	1.013 (0.007)	1.024 (0.012)	1.02 (0.009)
T=100	1.002 (0.001)	1.007 (0.002)	1.01 (0.003)	1.011 (0.005)	1.013 (0.006)	1.02 (0.006)	1.017 (0.005)
T=150	1.002 (0.001)	1.007 (0.002)	1.01 (0.002)	1.011 (0.005)	1.013 (0.006)	1.019 (0.004)	1.017 (0.003)

Table 5: Medians and mean absolute deviations of $\mathcal{R}_{\mu,qml/pc}$.
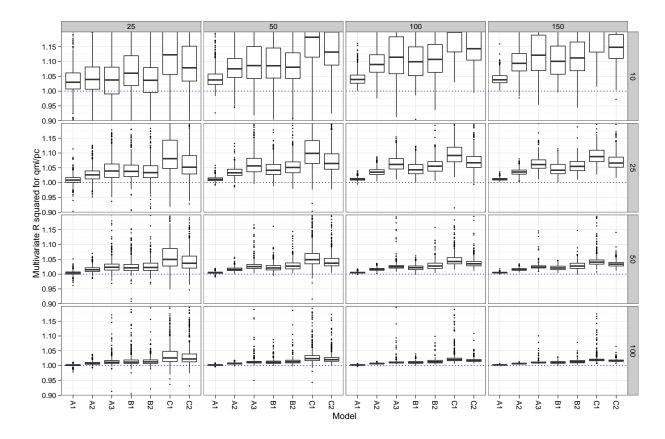
Figure 6: The distribution of $\mathcal{R}_{\mu,qml/pc}$ for different model specifications with $\xi_t \sim \text{WN}(0, diag)$.

n/T	A1	A2	A3	B1	B2	C1	C2
n=10							
T=25	1.005 (0.015)	1.015 (0.039)	1.016 (0.054)	1.017 (0.036)	1.028 (0.053)	1.047 (0.077)	1.033 (0.061)
T=50	1.004 (0.007)	1.034 (0.032)	1.049 (0.056)	1.023 (0.028)	1.056 (0.051)	1.081 (0.065)	1.059 (0.051)
T=100	1.004 (0.004)	1.041 (0.03)	1.061 (0.062)	1.024 (0.026)	1.072 (0.053)	1.096 (0.065)	1.072 (0.049)
T=150	1.003 (0.003)	1.039 (0.026)	1.064 (0.061)	1.027 (0.025)	1.082 (0.058)	1.097 (0.065)	1.074 (0.055)
n=25							
T=25	1 (0.003)	1.007 (0.012)	1.014 (0.022)	1.005 (0.01)	1.009 (0.016)	1.024 (0.032)	1.013 (0.02)
T=50	1 (0.001)	1.008 (0.008)	1.02 (0.016)	1.004 (0.005)	1.007 (0.008)	1.027 (0.023)	1.013 (0.014)
T=100	1 (0.001)	1.007 (0.006)	1.019 (0.011)	1.003 (0.003)	1.006 (0.005)	1.019 (0.014)	1.013 (0.011)
T=150	1 (0)	1.006 (0.005)	1.018 (0.008)	1.002 (0.002)	1.005 (0.004)	1.018 (0.012)	1.012 (0.011)
n=50							
T=25	1 (0.001)	1.003 (0.005)	1.007 (0.009)	1.003 (0.005)	1.004 (0.007)	1.013 (0.017)	1.01 (0.012)
T=50	1 (0)	1.002 (0.002)	1.006 (0.004)	1.001 (0.002)	1.002 (0.003)	1.008 (0.008)	1.006 (0.006)
T=100	1 (0)	1.002 (0.002)	1.005 (0.003)	1.001 (0.001)	1.001 (0.002)	1.004 (0.003)	1.003 (0.003)
T=150	1 (0)	1.001 (0.001)	1.004 (0.002)	1.001 (0.001)	1.001 (0.001)	1.003 (0.002)	1.002 (0.002)
n=100							
T=25	1 (0.001)	1.001 (0.002)	1.003 (0.003)	1.001 (0.002)	1.002 (0.003)	1.007 (0.009)	1.006 (0.007)
T=50	1 (0)	1 (0.001)	1.002 (0.002)	1.001 (0.001)	1.001 (0.001)	1.003 (0.003)	1.002 (0.003)
T=100	1 (0)	1 (0)	1.001 (0.001)	1 (0)	1 (0)	1.001 (0.001)	1.001 (0.001)
T=150	1 (0)	1 (0)	1.001 (0.001)	1 (0)	1 (0)	1.001 (0.001)	1.001 (0.001)

Table 6: Medians and mean absolute deviations of $\mathcal{R}_{\mu,qml/ts}$.
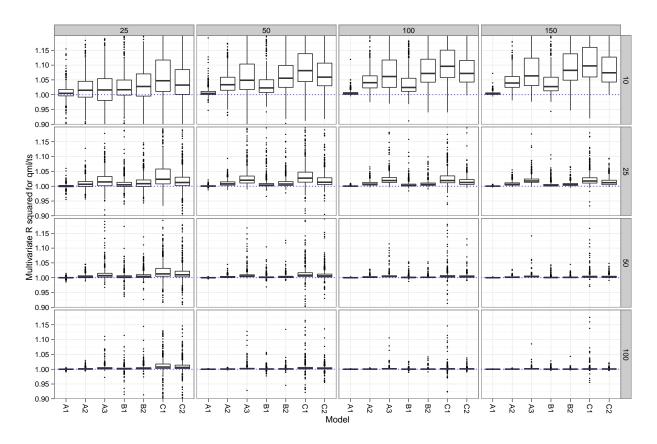
Figure 7: The distribution of $\mathcal{R}_{\mu,qml/ts}$ for different model specifications with $\xi_t \sim \mathrm{WN}(0, diag)$.

the QML-method clearly outperforms the PC-method and demonstrates superiority over the TS-estimator. For larger panels the performance advantage decreases, the advantage of QML- over TS- disappears very rapidly. The results are invariant with respect to the order of $e(z)$. I simulated models with $p = 1, 2, 3$ and in this case only the cases where $p = 1$ are reported.

Be aware, that a comparison of the relative performances between the different models is problematic, because the location of the roots of $e(z)$ has an influence on the relative performance gain (see more below). Thus from the results given above, one can not come to the conclusion that the performance advantage of the TS- and QML-estimator respectively is higher for models where $r > q$.

The effect of misspecification, i.e. a non-diagonal spectrum of ξ_t

Tables 7–9 and Figures 8–10 show the effects of local cross-sectional (weak) dependence of the ξ_{it} for the same set of models as above. Again the polynomials $e(z)$ were chosen randomly, non-diagonal and stable, and $\sigma_i \sim U_{[0.1,0.9]}$, $i = 1, \ldots, n$. Results for the noise-specifications *cross*, *cross-section* and *general*, and two more settings with very high local dependence, *general,* $\psi = 0.75$ and *general,* $\psi = 0.9$, are shown. A global strong local dependence ($\psi = 0.75, \psi = 0.9$) is a strong assumption, but nevertheless the results given below suggest a negative effect on the relative performance advantages of the TS- and QML-estimator over the PC-estimator. An overview of the results is given in Discussion 5.8.

n/noise	A1	A2	A3	B1	B2	C1	C2
n=10							
strict	1.035 (0.014)	1.05 (0.019)	1.046 (0.027)	1.066 (0.071)	1.028 (0.076)	1.085 (0.08)	1.064 (0.062)
cross	1.015 (0.018)	1.01 (0.039)	1.017 (0.034)	1.047 (0.093)	1.013 (0.081)	1.042 (0.093)	1.057 (0.074)
cross-time	1.023 (0.021)	1.032 (0.038)	1.02 (0.033)	1.022 (0.066)	1.02 (0.084)	1.044 (0.088)	1.017 (0.09)
g	1.012 (0.015)	1.027 (0.029)	1.013 (0.023)	1.044 (0.051)	1.012 (0.054)	1.055 (0.055)	1.029 (0.07)
g ($\psi = 0.75$)	0.999 (0.024)	1.012 (0.031)	1.009 (0.021)	0.984 (0.05)	0.936 (0.089)	1.043 (0.069)	0.997 (0.052)
g ($\psi = 0.9$)	0.993 (0.065)	1.001 (0.018)	1.005 (0.015)	0.813 (0.227)	0.926 (0.099)	0.998 (0.052)	0.983 (0.058)
n=25							
strict	1.011 (0.004)	1.028 (0.008)	1.041 (0.017)	1.036 (0.018)	1.047 (0.02)	1.066 (0.022)	1.054 (0.02)
cross	1.008 (0.006)	1.02 (0.012)	1.021 (0.014)	1.041 (0.032)	1.035 (0.023)	1.047 (0.034)	1.036 (0.032)
cross-time	1.01 (0.005)	1.029 (0.014)	1.031 (0.016)	1.037 (0.029)	1.039 (0.032)	1.04 (0.034)	1.026 (0.033)
g	1.006 (0.005)	1.027 (0.018)	1.024 (0.015)	1.031 (0.023)	1.028 (0.025)	1.044 (0.03)	1.034 (0.022)
g ($\psi = 0.75$)	1.001 (0.008)	1.012 (0.014)	1.013 (0.014)	1.016 (0.025)	0.994 (0.045)	1.006 (0.063)	1.003 (0.04)
g ($\psi = 0.9$)	0.984 (0.021)	1.003 (0.015)	1.003 (0.013)	0.984 (0.063)	0.929 (0.086)	0.976 (0.058)	0.94 (0.084)
n=50							
strict	1.005 (0.001)	1.014 (0.003)	1.02 (0.004)	1.02 (0.009)	1.026 (0.012)	1.036 (0.01)	1.031 (0.008)
cross	1.004 (0.003)	1.016 (0.007)	1.017 (0.008)	1.023 (0.015)	1.027 (0.012)	1.036 (0.019)	1.025 (0.017)
cross-time	1.005 (0.002)	1.014 (0.006)	1.021 (0.009)	1.022 (0.012)	1.028 (0.015)	1.029 (0.019)	1.028 (0.018)
g	1.003 (0.002)	1.014 (0.005)	1.017 (0.007)	1.019 (0.012)	1.019 (0.012)	1.027 (0.018)	1.025 (0.013)
g ($\psi = 0.75$)	1.002 (0.003)	1.011 (0.009)	1.012 (0.009)	1.014 (0.012)	1.018 (0.017)	1.013 (0.03)	1.003 (0.027)
g ($\psi = 0.9$)	0.995 (0.009)	1.003 (0.01)	1.006 (0.008)	1.002 (0.025)	0.972 (0.039)	0.984 (0.041)	0.95 (0.072)
n=100							
strict	1.002 (0.001)	1.006 (0.001)	1.009 (0.002)	1.01 (0.004)	1.013 (0.006)	1.018 (0.004)	1.016 (0.004)
cross	1.002 (0.001)	1.007 (0.002)	1.009 (0.002)	1.012 (0.005)	1.015 (0.005)	1.018 (0.007)	1.014 (0.007)
cross-time	1.002 (0.001)	1.007 (0.002)	1.01 (0.003)	1.011 (0.005)	1.016 (0.007)	1.018 (0.006)	1.016 (0.007)
g	1.002 (0.001)	1.006 (0.002)	1.009 (0.003)	1.01 (0.005)	1.012 (0.007)	1.016 (0.007)	1.014 (0.005)
g ($\psi = 0.75$)	1.001 (0.001)	1.008 (0.004)	1.01 (0.006)	1.01 (0.007)	1.013 (0.009)	1.014 (0.013)	1.01 (0.012)
g ($\psi = 0.9$)	1 (0.002)	1.005 (0.005)	1.007 (0.006)	1.007 (0.008)	1 (0.019)	0.99 (0.03)	0.973 (0.043)

Table 7: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,ts/pc}$ for $T = 150$.
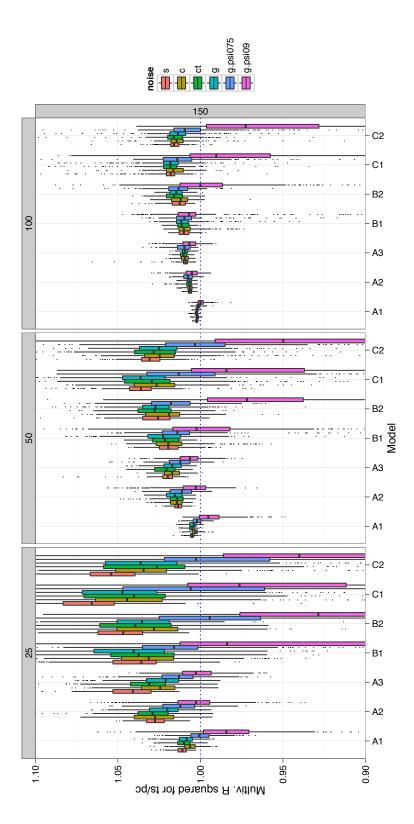
Figure 8: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,ts/pc}$. In this illustration, T has been fixed to 150 and n varies.

n/ψ	A1	A2	A3	B1	B2	C1	C2
n=10							
s	1.038 (0.017)	1.094 (0.042)	1.121 (0.086)	1.101 (0.072)	1.112 (0.068)	1.204 (0.097)	1.148 (0.059)
c	1.024 (0.031)	1.008 (0.087)	1.052 (0.075)	1.114 (0.107)	1.068 (0.103)	1.174 (0.133)	1.122 (0.094)
ct	1.03 (0.029)	1.055 (0.089)	1.049 (0.081)	1.088 (0.096)	1.079 (0.13)	1.131 (0.106)	1.106 (0.084)
g	1.011 (0.021)	1.061 (0.067)	1.036 (0.065)	1.07 (0.058)	1.048 (0.066)	1.136 (0.084)	1.092 (0.072)
g.psi075	0.99 (0.058)	1.013 (0.068)	1.014 (0.047)	0.973 (0.092)	0.964 (0.101)	1.095 (0.08)	1.029 (0.038)
g.psi09	0.973 (0.16)	0.996 (0.044)	1.003 (0.042)	0.853 (0.139)	0.957 (0.104)	1.017 (0.039)	1.013 (0.034)
n=25							
s	1.011 (0.004)	1.036 (0.01)	1.061 (0.021)	1.042 (0.022)	1.055 (0.023)	1.088 (0.028)	1.066 (0.024)
c	1.009 (0.007)	1.031 (0.023)	1.033 (0.029)	1.06 (0.044)	1.056 (0.03)	1.09 (0.051)	1.062 (0.038)
ct	1.01 (0.006)	1.044 (0.024)	1.059 (0.042)	1.048 (0.03)	1.065 (0.045)	1.085 (0.053)	1.06 (0.039)
g	1.005 (0.005)	1.046 (0.035)	1.041 (0.028)	1.045 (0.026)	1.047 (0.035)	1.071 (0.048)	1.064 (0.032)
g.psi075	0.997 (0.015)	1.016 (0.03)	1.021 (0.034)	1.034 (0.035)	1.018 (0.067)	1.051 (0.05)	1.033 (0.033)
g.psi09	0.943 (0.056)	0.999 (0.044)	1 (0.035)	0.954 (0.06)	0.964 (0.074)	1.016 (0.034)	0.995 (0.027)
n=50							
s	1.005 (0.002)	1.016 (0.004)	1.024 (0.005)	1.021 (0.009)	1.027 (0.013)	1.04 (0.01)	1.034 (0.009)
c	1.004 (0.003)	1.021 (0.009)	1.025 (0.012)	1.028 (0.016)	1.033 (0.013)	1.049 (0.023)	1.035 (0.019)
ct	1.005 (0.002)	1.021 (0.01)	1.032 (0.015)	1.025 (0.014)	1.034 (0.017)	1.047 (0.023)	1.04 (0.018)
g	1.003 (0.002)	1.019 (0.008)	1.023 (0.012)	1.024 (0.012)	1.026 (0.018)	1.042 (0.024)	1.039 (0.017)
g.psi075	1.001 (0.005)	1.017 (0.023)	1.02 (0.02)	1.021 (0.019)	1.034 (0.029)	1.043 (0.034)	1.028 (0.029)
g.psi09	0.984 (0.023)	1.002 (0.028)	1.01 (0.023)	1.007 (0.028)	0.985 (0.055)	1.015 (0.032)	1.003 (0.024)
n=100							
s	1.002 (0.001)	1.007 (0.002)	1.01 (0.002)	1.011 (0.005)	1.013 (0.006)	1.019 (0.004)	1.017 (0.003)
c	1.002 (0.001)	1.007 (0.002)	1.012 (0.003)	1.013 (0.004)	1.017 (0.005)	1.021 (0.007)	1.017 (0.007)
ct	1.002 (0.001)	1.008 (0.003)	1.012 (0.004)	1.012 (0.005)	1.018 (0.007)	1.022 (0.006)	1.019 (0.007)
g	1.002 (0.001)	1.008 (0.003)	1.012 (0.005)	1.012 (0.005)	1.016 (0.008)	1.02 (0.008)	1.018 (0.005)
g.psi075	1.001 (0.001)	1.01 (0.006)	1.016 (0.012)	1.014 (0.009)	1.021 (0.016)	1.032 (0.018)	1.021 (0.018)
g.psi09	0.999 (0.004)	1.007 (0.012)	1.01 (0.014)	1.012 (0.014)	1.015 (0.031)	1.019 (0.031)	1.008 (0.019)

Table 8: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,qml/pc}$ for $T = 150$.
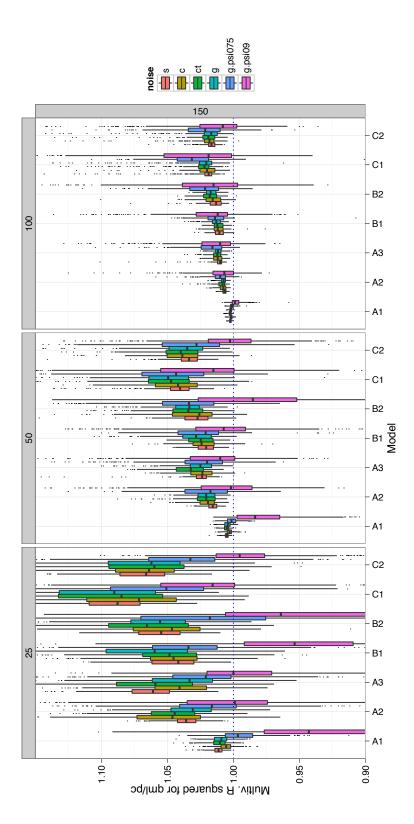
Figure 9: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,qml/pc}$. In this illustration, T has been fixed to 150 and n varies.

n/ψ	A1	A2	A3	B1	B2	C1	C2
n=10							
s	1.003 (0.003)	1.039 (0.026)	1.064 (0.061)	1.027 (0.025)	1.082 (0.058)	1.097 (0.065)	1.074 (0.055)
c	1.008 (0.013)	0.996 (0.049)	1.03 (0.053)	1.054 (0.067)	1.05 (0.063)	1.113 (0.104)	1.054 (0.048)
ct	1.006 (0.012)	1.027 (0.054)	1.031 (0.056)	1.054 (0.054)	1.069 (0.063)	1.087 (0.074)	1.08 (0.067)
g	0.999 (0.008)	1.027 (0.045)	1.021 (0.044)	1.025 (0.04)	1.032 (0.045)	1.072 (0.061)	1.052 (0.049)
g.psi075	0.996 (0.036)	1.001 (0.04)	1.004 (0.032)	0.993 (0.052)	1.026 (0.074)	1.057 (0.06)	1.029 (0.037)
g.psi09	0.995 (0.081)	0.993 (0.026)	0.999 (0.029)	1.059 (0.222)	1.019 (0.04)	1.023 (0.029)	1.032 (0.043)
n=25							
s	1 (0)	1.006 (0.005)	1.018 (0.008)	1.002 (0.002)	1.005 (0.004)	1.018 (0.012)	1.012 (0.011)
c	1.001 (0.001)	1.008 (0.011)	1.011 (0.015)	1.012 (0.012)	1.018 (0.015)	1.035 (0.029)	1.026 (0.023)
ct	1.001 (0.002)	1.014 (0.013)	1.025 (0.024)	1.009 (0.007)	1.021 (0.018)	1.034 (0.032)	1.028 (0.021)
g	0.999 (0.001)	1.016 (0.017)	1.014 (0.015)	1.009 (0.011)	1.017 (0.016)	1.024 (0.023)	1.028 (0.023)
g.psi075	0.997 (0.006)	1.004 (0.018)	1.008 (0.021)	1.012 (0.019)	1.022 (0.04)	1.046 (0.041)	1.042 (0.044)
g.psi09	0.964 (0.045)	0.995 (0.028)	0.997 (0.024)	0.987 (0.083)	1.028 (0.068)	1.055 (0.07)	1.053 (0.069)
n=50							
s	1 (0)	1.001 (0.001)	1.004 (0.002)	1.001 (0.001)	1.001 (0.001)	1.003 (0.002)	1.002 (0.002)
c	1 (0)	1.004 (0.003)	1.007 (0.006)	1.003 (0.003)	1.006 (0.006)	1.011 (0.009)	1.007 (0.007)
ct	1 (0)	1.006 (0.005)	1.01 (0.007)	1.002 (0.001)	1.005 (0.004)	1.013 (0.011)	1.009 (0.007)
g	1 (0)	1.004 (0.004)	1.006 (0.005)	1.002 (0.002)	1.006 (0.007)	1.012 (0.01)	1.011 (0.008)
g.psi075	1 (0.001)	1.005 (0.011)	1.007 (0.012)	1.005 (0.007)	1.01 (0.016)	1.03 (0.027)	1.022 (0.023)
g.psi09	0.989 (0.014)	0.999 (0.018)	1.003 (0.018)	1.005 (0.022)	1.02 (0.049)	1.046 (0.052)	1.064 (0.074)
n=100							
s	1 (0)	1 (0)	1.001 (0.001)	1 (0)	1 (0)	1.001 (0.001)	1.001 (0.001)
c	1 (0)	1.001 (0.001)	1.002 (0.001)	1.001 (0.001)	1.001 (0.001)	1.002 (0.002)	1.002 (0.002)
ct	1 (0)	1.001 (0.002)	1.002 (0.001)	1 (0)	1.001 (0.001)	1.003 (0.002)	1.002 (0.002)
g	1 (0)	1.001 (0.001)	1.002 (0.002)	1.001 (0.001)	1.002 (0.002)	1.003 (0.002)	1.003 (0.002)
g.psi075	1 (0)	1.002 (0.002)	1.005 (0.005)	1.002 (0.002)	1.006 (0.008)	1.013 (0.011)	1.012 (0.013)
g.psi09	0.999 (0.001)	1.001 (0.006)	1.003 (0.009)	1.005 (0.011)	1.015 (0.023)	1.036 (0.036)	1.045 (0.049)

Table 9: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,qml/ts}$ for $T = 150$.
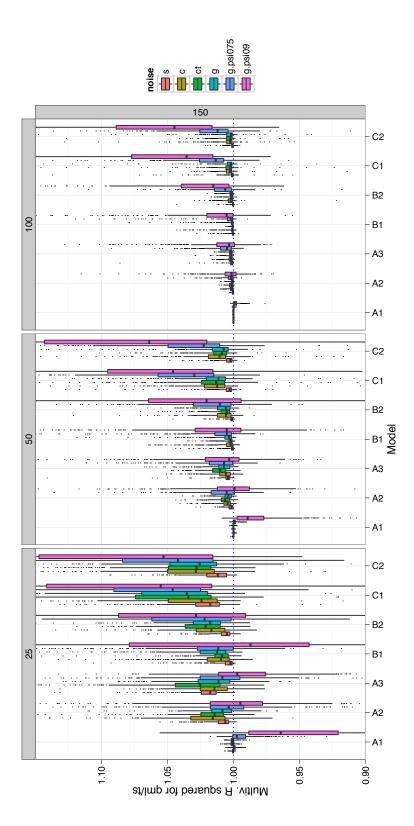
Figure 10: The effect of non-diagonal idiosyncratic spectra on $\mathcal{R}_{\mu,qml/ts}$. In this illustration, T has been fixed to 150 and n varies.
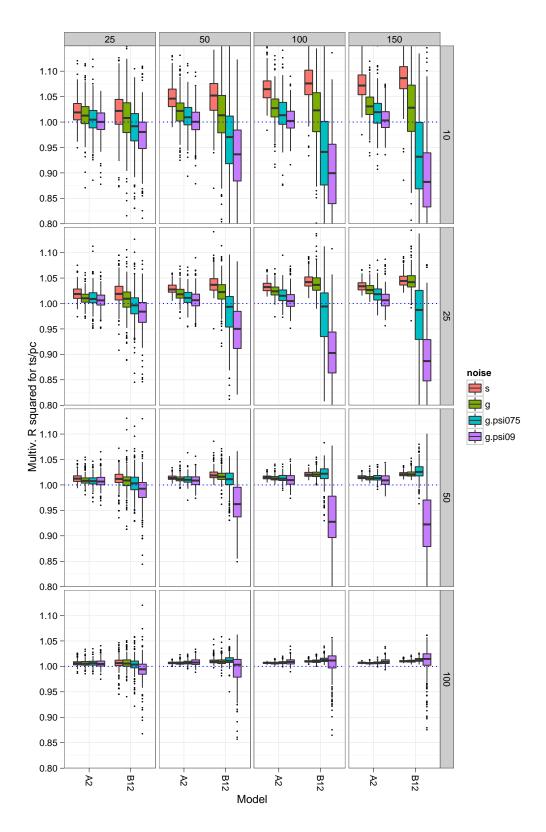
Figure 11: Simulation with parameters coming from the Stock and Watson data-set. The effect of strong local cross-dependencies for different n and T.

Discussion 5.8. (The effects of a non-diagonal spectrum of ξ_t)

- As in the *strict* case, absolute performances decrease with the number of factors also in the case of noise-misspecification. This can be seen by comparing the performances between models A1, A2 and A3, as well as between models B1 and B2 (see the corresponding tables in Section A). Figures 12 and 13 also suggest, that the relative performances of TS and QML over PC increase with the number of static factors. This corresponds to the results obtained in [17].

- In general, the absolute performances decrease with the degree of cross-sectional dependence for all three estimators. The only exception are cases where the number of variables is very small ($n = 10$). It is not clear which effects lead to an increase of absolute performance with an increasing ψ.

- The simulation results above suggest a relative advantage of the TS and QML estimator over the PC estimator if the level of cross-sectional dependence is moderate. This holds for most models of all types A, B, C. Furthermore it confirms the results of [17], who simulated models with $r = q = 1, 3$, $e(z) = diag(1 - 0.9z, \ldots, 1 - 0.9z)$ and $\psi = 0.5$[37]. Note, that there is also a small number of models, where the PC results are better than the TS and QML results respectively. They are the "outlier" cases in the figures shown above.

- If the degree of dependence (ψ) increases, then especially for models of type A1 ($r = q = 1$) and for the cases $r > q$, the performance advantage decreases, vanishes or even turns into a disadvantage. This is especially true for the TS estimator. A drop in its absolute performance can be observed. The picture partially changes for the QML estimator. It dominates (except for the case $r = q = 1$) the TS estimator if the local dependence is very high (but still weak).

- The results from above are confirmed for models where the parameters were estimated from the Stock and Watson data set (see Figure 11). They also indicate a negative effect of local dependence on the relative performance advantages of TS and QML over the PC estimator. Also here, this advantage can even turn into a disadvantage.

- The variation of results is very high for simulations with a small number of different time-series ($n = 10, 25$). This is also due to the fact, that the smallest roots of the non-diagonal and random polynomial $e(z)$ are nearly uniformly drawn from its support (see Tables 1 and 2)[38]. Especially in this case, no uniform statement about performance advantages can be made. The negative effect of high local dependencies decreases when the number of variables increases.

- Interestingly the QML method performs worse than the TS estimator if $r = q = 1$ and n is small.

---

[37]In their paper, the following statement could be read

"Because of the explicit modelling of the dynamics and the cross-sectional heteroscedasticity, the maximum likelihood estimates dominate the principal components. . ." ([17], page 13).

Unfortunately they do not report standard errors (or give reports about significance) for their results.

[38]If $e(z) = diag(e_1(z), \ldots, e_r(z))$ with $e_i(z) \sim U_{[a_i, b_i]}$, then the variance of results decreases with the length of the intervals $[a_i, b_i]$ and the smaller the number of models considered respectively. This can be seen in Figures 12 and 13.

$\square$

To statistically support the negative effect of a strong local dependence on the relative performance advantages, one could perform non-parametric (robust) permutation tests (independence tests). K-Sample independence tests use resampling techniques in order to approximate the distribution of a test statistic which tests the null hypothesis

$$H_0 : \mathcal{L}(y|x) = \mathcal{L}(y), \tag{62}$$

where the random variables x, y may be measured at arbitrary scales and may also be multivariate. $\mathcal{L}(y)$ specifies the random law of y. Here y are the relative performances coming from different "groups" y. The groups are then the different noise-specifications and different levels of cross-sectional dependence. These tests assume, that the group-association (e.g. the value of ψ) had been selected at random, which was not the case in this simulation. Classical examples of tests that fit in this setting are the Wilcoxon-Mann-Whitney test (U test, Wilcoxon rank sum test) [34]. For K different groups, testing whether a classification into groups has a significant influence on the relative performances can be done by a non-parametric version of the Kruskal-Wallis test [34], [37].

One could performs tests for every model of different types A, B, C and report the relative amount of cases, where the null hypothesis that relative performance is influenced or decreases with the level of local dependence, was rejected. Although such tests had been performed during analysis, they are not reported, due to the small number of data repetitions (25) for each model.

The effect of static factor process having a long memory

Models with random and diagonal polynomial matrices $e(z)$ were chosen in order to show the effect of a long-memory static factor process on the estimation results. The specifications can be seen in Section 5.2. The noise-to-signal ratio was again randomly drawn, i.e. $\sigma_i \sim U_{[0.1, 0.9]}$. If the static factor process has a long memory (also a high number of p, the order of $e(z)$ achieves this[39]), then also the observations x_t do have a long memory. This leads to poorer estimates of the covariance matrix $\frac{1}{T} X'X$, which is the basis for the PC-estimator and therefore the initialization step for the TS- and QML-estimator. Very poor estimates of the static factors are the consequence (see tables in Section A). Nevertheless, Figures 12 and 13 show a relative performance gain of the TS- and QML-estimator over the PC-method. The negative effect of noise-misspecification can be clearly seen for small n, it decreases for larger n. The variance of relative performances increases when the process has a long memory, especially for the case $r = q = 1$. The described phenomena are similar for models where $r > q$.

The effect of the noise-to-signal ratio on performances

So far, the noise-to-signal ratio had been uniformly drawn from $[0.1, 0.9]$. The following example shall demonstrate what happens, if the amount of noise is very high compared to the latent signal. For this purpose simulation scenarios with constant $\sigma_i = 0.5, 0.75, 0.9$ were set up. By looking at the absolute performances in Section A, one can observe, that a high noise-to-signal ratio has a negative effect on the absolute performances of all three estimators. This performance loss however decreases with the number of different time-series n.

---

[39]In fact disadvantages of the QML- compared to the PC-estimator could be observed for models A3 if $T = 25$.
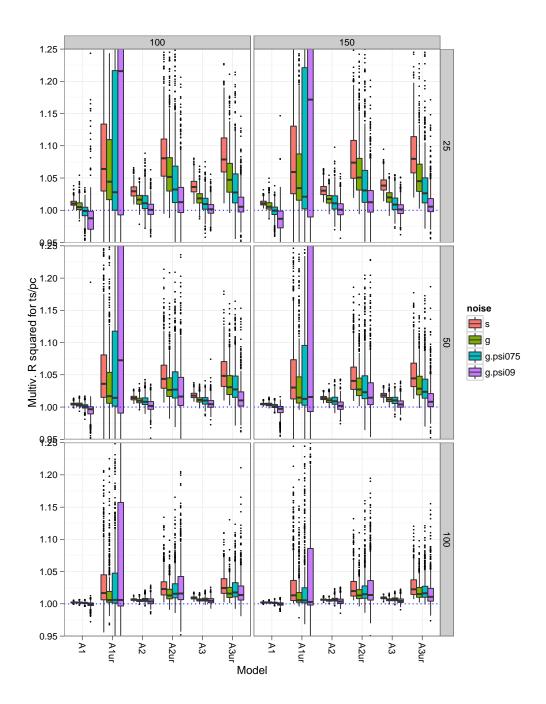
Figure 12: The effect of static factors with a long-memory process on the relative performance of TS compared to PC. Results for the cases where $r = q$ are shown.
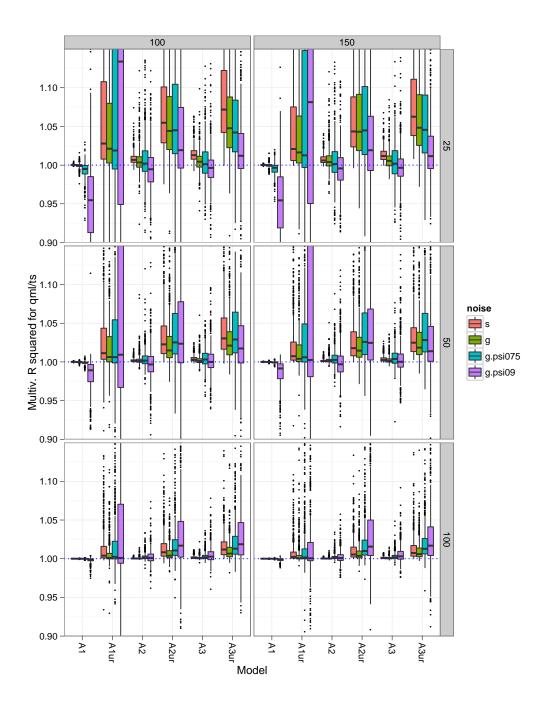
Figure 13: The effect of static factors with a long-memory process on the relative performance of QML compared to TS. Results for the cases where $r = q$ are shown.

Figure 14 shows the effect of very high noise-to-signal ratios on the relative performances of the TS- compared to the PC-estimator. Especially if the latent-signal is very weak, the TS-estimator performs worse than the PC-estimator. The results are even poorer for the QML-estimator compared to the TS-procedure.
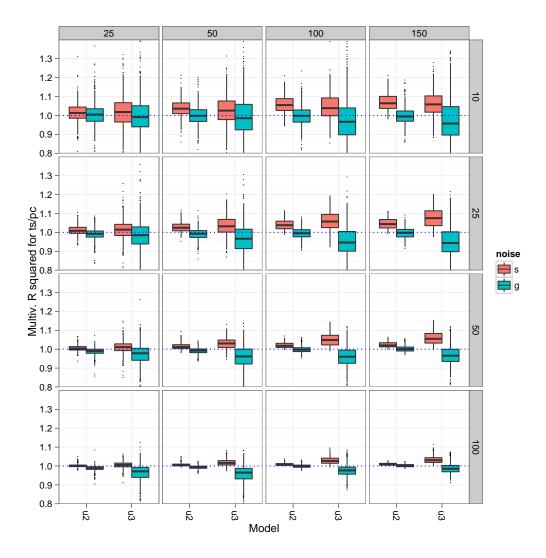
Figure 14: The effect of high noise-to-signal ratios on the relative performance of TS compared to PC. Results are shown for models A2.

The EM-Algorithm

The following paragraph is devoted to the implementation of the EM algorithm. Table 10 shows model specifications where the EM algorithm did not work. The last column of this table shows the relative number of cases[40] where the likelihood did indeed decrease (the likelihood could in fact decrease up to a tolerance level of 0.01 due to numerical considerations). Interestingly also the model whose parameters where determined from real data (see Section 6) for the parameters $r = 4, q = 2$ shows this behavior. This was not the case for the cases where $r = 3, q = 2$ and $r = 3, q = 3$ respectively. Although the models A1 and C1 do have a certain amount of likelihood decreases, they were not excluded from the analysis. Instead those cases of likelihood decreases have been deleted from the analysis.

Model	Noise spec.	q	r	p	psi	LI (%)
A1	general	1	1	2	0.5	75.600
A1	general	1	1	3	0.5	75.170
A1 (near unit root)	general	1	1	3	0.5	92.480
A1	general	1	1	1	0.5	5.410
B1	general	1	3	2	0.5	12.180
B1	general	1	3	2	0.5	12.880
B2	general	2	4	2	0.5	7.840
B2 (Stock and Watson)	general	2	4	1	0.5	53.225
B2 (Stock and Watson)	cross-time	2	4	1	0.5	36.400
B2 (Stock and Watson)	cross	2	4	1	0.5	32.225
B2 (Stock and Watson)	strict	2	4	1	0.5	37.925
C1 (near unit root)	general	2	4	1	0.5	5.325
B2	cross	2	4	2	0.5	14.300
B2	strict	2	4	2	0.5	12.210
B1	cross-time	1	3	2	0.5	18.000
B1	cross	1	3	2	0.5	15.530
B1	strict	1	3	2	0.5	23.670

Table 10: Model specifications (see Section 5.2) where the Likelihood decreased during the EM-Algorithm in more than 5% of the cases.

The number of maximal iterations was set to 25. Although convergence behavior of the EM algorithm is slow, cases where the maximum number of iterations is greater or equal to 25 are considered as cases of either very slow convergence or cases where no convergence behavior could be observed. Although it would be interesting to know, no distinction between those two cases had been made. Their number though is in many cases impressively high. Figure 15 gives examples of convergence behavior for different models. The empirical distribution of the number of observations for two different noise specifications (strict and generalized) can be seen. Note that the graph shows results for simulation settings with different numbers of variables n and observations T, but nevertheless an overall interpretation can be given.

Models A1 and A3 correspond to the models given in [17]. For the A1 model the number of iterations never exceeds 10 (if the case $n = 5$ is excluded). This can also be seen in the first two

---

[40]For each model type the simulation has been repeated 625 times.

histograms in in Figure 15. The number of iterations for model A3 varies in their paper between 4 ($n = T = 100$) and 26 ($n = 10, T = 50$). To compare those results with my results, Figure 15 is not very useful. A more alarming result is the high number of cases (nearly 6000 cases for the general and more than 3000 cases for the strict noise specification) where the number of maximal iterations had been reached.

The number of iterations seems to depend on the noise specification. In general all models, no matter if $r = q$ or $r > q$, show a high amount of cases where the number of maximal iterations had been reached. Figure 15 gives no clear indication that the implementation of the EM Algorithm is problematic for models where $q < r$.
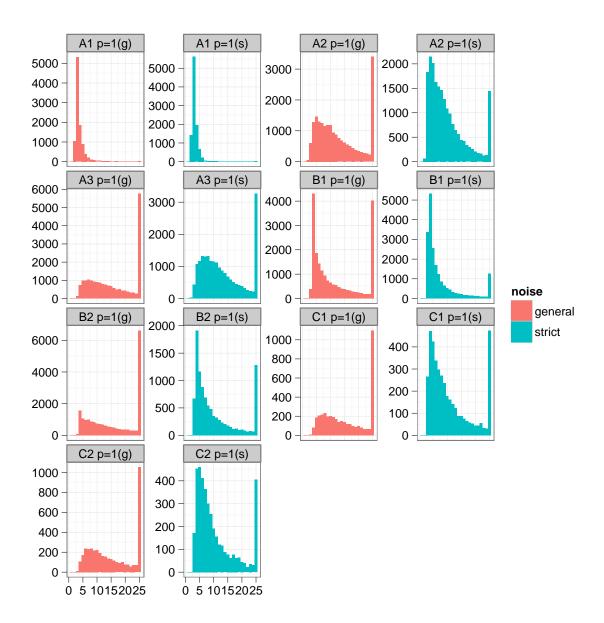
Figure 15: Convergence behavior of EM algorithm for different model configurations. Distinction between strict noise (s) and generalized noise (g) setting.

6   Empirical study

The primary purpose of this empirical analysis has been to find parameter settings for the simulation analysis. In contrast to many other empirical papers [13], the goal is not to do a forecasting analysis for variables like industrial production, GDP growth or the interest rate.

6.1   The data of Stock and Watson

Description of the data.   The data set has been compiled and used by Stock and Watson [50]. The raw data consists of 144 macroeconomic time series observed monthly and quarterly respectively. In total 190 quarters between $1959(III)$ and $2006(IV)$ of data are available. Different types of data transformations are needed in order to make it stationary. To obtain quarterly data from the monthly series, averages have been taken. There are no missing observations.

For factor estimation, Stock and Watson suggest in their paper to use only 110 series. This is done, because some series are related by identities, i.e. an aggregate being the sum of the subaggregates. It is worth mentioning, that for example in [49], the full data set was taken for analysis.

The panel data includes *real variables*, like sectorial industrial production, real investment, real personal consumption expenditures, employment and hours worked, *nominal variables*, like price indices, the oil price, wages, money aggregates, interest rates, *asset prices*, like S&P's stock prices, exchange rates and bond yields, and *survey data*, like data from NAPM[41]. A complete description of the data set can be found in the Data Appendix of [50].

Data transformation.   A table with information about univariate data-transformation can be found in the appendix of [50].

The Great Moderation.   A problem with macroeconomic time-series data is the occurrence of structural breaks which may have an impact on the analysis. Many economists observed, that the structure of the data has changed in the early 1980's [41]. Especially the variance of key macro-economic variables has decreased. Economists call this effect the Great Moderation. Stock and Watson argued in their paper [50] that structural breaks of small magnitude do not have a big effect on static factor estimation, although they did not analyze model selection methods. In the following results are often reported for the whole time frame and for sub-periods related to the period before and after the Great Moderation respectively.

Testing for wide-sense stationarity.   After transforming the data, statistical tests had been conducted in order to see whether the suggested transformations (see [50], Data Appendix) were good enough to obtain data coming from a wide-sense stationary process. Univariate tests like the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and the Elliott, Rothenberg & Stock unit root test had been performed. The testing procedures revealed two more potentially non-stationary time series. They were excluded from subsequent analysis.

Descriptive analysis.   The most important question is whether there is enough collinearity in the data in order to apply dynamic factor models. In order to describe comovements among series, the following relations are helpful

---

[41]National Association of Purchasing Managers.

$$\gamma_x(0) = L\gamma_z(0)L' + \gamma_\xi(0), \quad f_x(\theta) = Lf_z(\theta)L' + f_\xi(\theta), \ \theta \in \Theta$$

One can therefore look at the amount of variance of the panel explained by a certain number of (static and dynamic) factors. The static factors can be estimated by the principal components of $\gamma_x(0)$, whereas dynamic principal components are used as estimates for the dynamic factors (see [26] for more information).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Static	0.22	0.30	0.34	0.38	0.41	0.44	0.47	0.49	0.52	0.54	0.56	0.58	0.60	0.62	0.63
Dynamic	0.22	0.44	0.65	0.72	0.78	0.84	0.87	0.89	0.92	0.93	0.95	0.96	0.97	0.97	0.98

Table 11: Percentage of total variance (of the observations x_{it}) explained by the first q dynamic and r static principal components.

As can be seen in Table 11 quite few principal components explain most of the variance of our 109 variables[42]. Note that the share of variance of same amount of dynamic principal components is higher than the share of the same amount of static principal components. This suggests that the data is not only driven by r static factors, but may be driven by $q < r$ dynamic factors. This panel of macroeconomic data seems to provide the structural features that are necessary for factor analysis. The visual inspection of the cumulative relative variances is however no statistical tool for choosing the unknown model parameters q and r, the numbers of dynamic and static factors respectively.

6.2   Model selection

Number of static factors.   The paper of Bai and Ng [5] describes a procedure (BNg) for consistently estimating the number of static factors r. They use the penalized least squares objective function to develop consistent information criteria

$$IC(r) = lnV(r, \hat{z}^r) + r\, g(n,T), \quad V(r, \hat{z}^r) = \min_{H \in \mathbb{R}^{n \times r}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - H_i^{r\prime} \hat{z}_t^r)^2$$

The estimator $\hat{r} = \arg\min_{0 \le r \le rmax} IC(r)$ is consistent (see [5], Theorem 2), i.e. $\lim_{n,T\to\infty} P[\hat{r} = r] = 1$, if $g(n,T) \to 0$ and $\min\{n,T\}g(n,T) \to \infty$ as $n,T \to \infty$. They propose three criteria

$$IC_1(r) = \ln V(r, \hat{z}^r) + r\left(\frac{n+T}{nT}\right)\ln\left(\frac{nT}{n+T}\right)$$

$$IC_2(r) = \ln V(r, \hat{z}^r) + r\left(\frac{n+T}{nT}\right)\ln\min\{n,T\}$$

$$IC_3(r) = \ln V(r, \hat{z}^r) + r\left(\frac{\ln\min\{n,T\}}{\min\{n,T\}}\right)$$

where $\hat{z}^r$ is the PC-estimator assuming that the dimension of the space spanned by the factors is r. The three information criteria[43] are then defined as $\hat{r}_i = \arg\min_{0 \le r \le rmax} IC_i(r), \ i = 1,2,3$.

---

[42]Note, that the results for the variance explained by dynamic principal components are very sensitive with respect to the estimation of the spectral density of our observations. Therefore these number need to be consumed with care. Nevertheless, the pattern is robust with respect to the spectral density estimation technique.

[43]A deeper analysis that also explains the role of $r_{\max}$ is given in [20] on page 58.

	Full data set	1959–1983	1984–2006
$\hat{r}_1$	4	3	3
$\hat{r}_2$	2	2	1
$\hat{r}_3$	20	20	20

Table 12: Estimates of r, the number of static factors by using the information criteria by Bai and Ng. Maximum number of static factors is set to 20.

Conclusion: No clear conclusion about the number of static factors is possible. For the simulation study $\hat{r} = 3, 4$ had been selected. The former number is the average between $\hat{r}_1$ and $\hat{r}_2$. This selection follows the recommendation of [20], who proposes to choose the average between the first two criteria, if the penalty term for the last one was too small.

Number of dynamic factors.   Different methods of estimating q have been proposed [6], [1], [31], [9]. All three are consistent for both n and T going to infinity, but their finite sample properties depend on the DGP (see for example the results in [1]). I therefore apply all three information criteria to the full and split data. A brief summary shall give a basic feeling of the methods for determining q.

The criterion of Hallin and Liška [31] Their information criterion is based on the general model (17) with one-sided filters. Their method is therefore based on more general assumptions and does for example not need the idea of static factors. The principal idea though is very similar to BNg, where a penalized least squares objective function needs to be minimized. Their information criterion depends on the way the spectral density of the process x_t is estimated and is defined for $0 \leq k \leq q_{\max}$:

$$IC_p(k, \zeta) = \ln \left(\frac{1}{n} \sum_{i=k+1}^{n} \frac{1}{2m_T + 1} \sum_{j=-m_T}^{m_T} \hat{\lambda}_i^p(\theta_j) \right) + \zeta\, k\, p(n, T)$$

$$IC_c(k, \zeta) = \frac{1}{n} \sum_{i=k+1}^{n} \frac{1}{T - 1} \sum_{l=1}^{T-1} \hat{\lambda}_i^c(\omega_l) + \zeta\, k\, p(n, T)$$

where $\theta_j = \pi j/(m_T + 1/2)$, $\omega_l = 2\pi l/T$ and $\zeta \in \mathbb{R}^+$. The subscripts p and c indicate whether the spectral density f_x^n has been estimated by lag-window estimation (correlogram) and periodogram smoothing respectively.

The estimator for q is defined for every $\zeta > 0$ as

$$\hat{q}_{a,\zeta}^{HL} = \arg \min_{0 \leq k \leq q_{\max}} IC_a(k, \zeta), \quad a = p, q$$

Lemma 6.1   Assume that the double-sequence $\boldsymbol{x}$ can be represented as in Defintion 3.4 with one-sided square summable filters $b(z)$, that x_t^n is a linearly regular process, i.e. $x_t^n = \sum_{k=-\infty}^{\infty} c_k \varepsilon_{t-k}$, where ε_t is a full-rank n-dimensional white noise process with finite

fourth-order cumulants, and $\sum_{k \in \mathbb{Z}} |c_{ijk}| |k|^{1/2} < \infty$ for all $1 \leq i, j \leq n$. Further assume weak and strong dependence of the idiosyncratic and the common component respectively. Finally assume that the first q eigenvalues of the latent spectrum diverge linearly to infinity and that $q_{\max}$ is chosen such that the $q_{\max} + 1$-th eigenvalue of the spectral density matrix of x_t^n is bounded away from zero. Then

$$\hat{q}_{a,\zeta}^{HL} \xrightarrow{p} q, \quad n, T \to \infty, \quad \forall \zeta > 0$$

provided that

1. $p(n, T) \to 0$ as $n, T \to \infty$
2. and

   - for $a = c$: $\min(n, m_T^2, m_T^{-1/2} T^{1/2}) \, p(n, T) \to \infty$
   - for $a = p$: $\min(n, B_T^{-2}, B_T^{1/2} T^{1/2}) p(n, T) \to \infty$

The fact that consistency is guaranteed for every $\zeta > 0$ is used to calibrate finite sample behavior. Their practical guide to select q consists of calculating the variance of $\hat{q}_{a,\zeta}^{HL}(n_j, T_j)$, $j = 1, \ldots, J$ over different sample sizes and fixed ζ. As n and T are given $0 < n_1 < n_2 < \ldots < n_J = n$ and $0 < T_1 \leq T_2 \leq \ldots \leq T_j = T$. This sample variance is simply

$$S(\zeta) = \frac{1}{J} \sum_{j=1}^{J} \left(\hat{q}_{a,\zeta}^{HL}(n_j, T_j) - \frac{1}{J} \sum_{j=1}^{J} \hat{q}_{a,\zeta}^{HL}(n_j, T_j) \right)^2$$

A stability region $[\zeta_1, \zeta_2]$ would be an interval, where $S(\zeta) = 0$, $\zeta \in [\zeta_1, \zeta_2]$. In their paper they recommend to choose the ζ_1 of the second stability region. The reason is mainly that the first stability region stands for under-penalization (over-estimation of q) and the higher ζ the more likely over-penalization (under-estimation of q) takes place. By selecting a ζ the estimator for q is determined. The output of this analysis can be seen in Figure 16. I used the smoothed periodogram in order to estimate the spectral density and set $n_j = n - k \cdot j, T_j = T - m \cdot j, \ j = 1, \ldots, 5$, where $k = 1, 5, m = 5, 10$ depended on how large n and T where.

The method of Amengual and Watson [1] Their method is based on the information criteria for the static factors of Bai and Ng. The method assumes that the DGP follows model (24). In order to estimate q, the model is rewritten to

$$\underbrace{x_t + \sum_{j=1}^{p} L e_j z_{t-j}}_{:=y_t} = \underbrace{Lb}_{:=K} u_t + \xi_t, \quad L \in \mathbb{R}^{n \times q}$$

They apply BNg (see above) to the model $y_t = K u_t + \xi_t$. As y_t depends on unknown parameters (L) and data (z_t), the model selection has to be performed on $\hat{y}_t = x_t + \sum_{j=1}^{\hat{p}} \hat{L} \hat{e}_j \hat{z}_{t-j}$. The matrix L and the static factors z_t are estimated by principal components, whereas
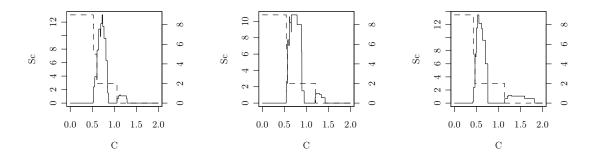
Figure 16: Determining q, the number of dynamic factors.

the coefficients e_j are the result of an OLS regression[44]. The last unknown parameter p is estimated by BIC. The resulting $\hat{q}_1, \hat{q}_2, \hat{q}_3$ will be denoted by $\hat{q}_1^{AW}, \hat{q}_2^{AW}, \hat{q}_3^{AW}$ respectively.

Lemma 6.2 Under suitable assumptions similar to that of [49], [5]

$$\hat{q}_i^{AW} \overset{p}{\to} q, \quad n, T \to \infty, \quad i = 1, 2, 3$$

The criterion of Bai and Ng [6] Their method is based on the same model (24), but they focus on the eigenvalues of the residual covariance matrix of $\nu_t = bu_t$. They seek to determine q by estimating the rank of the variance covariance matrix of the residuals $\hat{\nu}_t$. Their analysis is based on the spectral decomposition of $\Sigma = \mathbb{E}\nu_t\nu_t'$ and their consistent estimators are summarized in Proposition 2. Let $0 \leq c_r \leq \ldots \leq c_2 < c_1$ be the eigenvalues of Σ and define

$$D_{1,k} = \left(\frac{c_{k+1}^2}{\sum_{j=1}^r c_j^2} \right)^{1/2}, \quad D_{2,k} = \left(\frac{\sum_{j=k+1}^r c_j^2}{\sum_{j=1}^r c_j^2} \right)^{1/2}, \quad k = 1, \ldots, r-1$$

The Nullhypothesis $H_0 : \Sigma(k) = \Sigma, \; k = q + 1, \ldots, r$ where $\Sigma(k) = \sum_{j=1}^k c_j v_j v_j'$ and $v_j, \; j = 1, \ldots, r$ are the orthogonal eigenvectors of the spectral decomposition of Σ.

If $\hat{z}_t$ is estimated by principal components and $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^T \hat{\nu}_t\hat{\nu}_t'$, then for $0 < m_i < \infty$ and $0 < \delta < 1/2$

$$\hat{q}_1^{BNg} = \min\{k : \hat{D}_{1,k} < m_1 / \min[n^{1/2-\delta}, T^{1/2-\delta}]\} \tag{63}$$

$$\hat{q}_2^{BNg} = \min\{k : \hat{D}_{2,k} < m_2 / \min[n^{1/2-\delta}, T^{1/2-\delta}]\} \tag{64}$$

---

[44]In their paper they argued that $M_j = He_j$ can also be estimated by regressing the the observations x_t on the static factors z_t.

where $\hat{D}_{i,k}$, $i = 1, 2$ is computed via $\hat{\Sigma}$ and m_i as well as δ are determined in Monte-Carlo simulations, are consistent estimators for q und H_0.

In the simulations of Bai and Ng, $\delta = 0.1$ and if the correlation matrix instead of Σ is used, then $m_1 = 1.25$ and $m_2 = 2.25$. They choose $m_1 = m_2 = 1$ when estimating with Σ.

		Full data set	1959-1984	1984-2006
Hallin and Liška	$\hat{q}^{HL}$	2	2	2
Amengual and Watson	$\hat{q}_1^{AW}$	2	max	max
	$\hat{q}_2^{AW}$	2	max	max
	$\hat{q}_3^{AW}$	3	max	max
Bai and Ng	$\hat{q}_1^{BNg}$	∞	∞	∞
	$\hat{q}_2^{BNg}$	∞	∞	∞

Table 13: Estimates of q, the number of dynamic factors.

Conclusion: The estimators $\hat{q}_1^{BNg}$, $\hat{q}_1^{BNg}$ did not work out, because the parameters m_i were not large enough. This shows how difficult the determination of model parameters in this case is. Also for the criteria by Amengual and Watson the maximum number of factors was reached in few cases. This is indicated by max. I selected $\hat{q} = 2, 3$ for the simulations.

6.3   Parameter estimation

The parameter p had been estimated by BIC, the estimation of the integer parameters r, q, p, the coefficients of the $r \times r$ polynomial matrix $e(z)$ and $b \in \mathbb{R}^{r \times q}$ were estimated by solving Yule-Walker equations (see Section 4.2). The matrix b is equal eigenvectors corresponding to the first q eigenvalues of Σ_ν. Because $q < r$, the minimum norm solution has been chosen in order to estimate $e(z) = (e_1, \ldots, e_p)$.

7   Summary and Conclusions

This thesis analyzed the effect of noise-misspecification for three different estimation methods for the static factors by means of a simulation study. The study was based on the theoretical framework of generalized dynamic factor models, where high-dimensional observations x_t are decomposed into a latent component χ_t and an idiosyncratic component ξ_t. The r-dimensional static factors z_t are linear functions of the latent component, i.e. $\chi_t = L z_t$.

   If one assumes that χ_t has a rational spectrum with constant rank $q \leq r < n$, and that the dimension r does not depend on n, then for an open and dense subspace of the parameter-space that corresponds to the state-space realization of the latent process, the static factors follow a singular vector auto-regressive process. Both, the two-stage (TS) and quasi-maximum likelihood (QML) estimator are based on the assumption that z_t is an autoregressive process. They compute the estimates by means of the Kalman Filter, which computes the least squares estimator of z_{it} as a linear function of all available observations x_{it}. The two methods differ in the way the parameters of the state space model are estimated. The third estimator is based on the spectral decomposition of the observation covariance matrix. It is the famous principal component (PC) estimator.

   In the generalized setting, the process ξ_t is assumed to be weak dependent. The TS and QML estimators are misspecified in the sense that they are based on the assumption that ξ_t is a white-noise process with diagonal covariance matrix. It has been shown, that the three estimators consistently estimate the linear space of the static factors (who are only identified up to a non-singular transformation) with increasing panel size, i.e. $n, T \to \infty$.

   The aim of the simulation study was to assess performance differences for the finite sample case. Performance was measured by the multivariate coefficient of determination. Results show that for the case where ξ_t is indeed white noise with a diagonal covariance matrix, both the TS and the QML estimators outperform the PC estimator in general. This dominance was expected as the simulation models explicitly modeled the autoregressive dynamics of the static factors. The relative performance advantages depend negatively on the degree of local cross-sectional dependence of the idiosyncratic component and the panel size. High local dependence could neutralize performance advantages, and a very high local dependence could even turn the relative advantages into disadvantages. The memory of the static factor process, and therefore the ability to estimate the covariances of the observations well enough, clearly influences the absolute and the relative performance. Whereas the coefficient of determination decreases drastically, relative performance advantages increase if the spectrum of z_t has poles near the unit circle. If the variance of the noise is very high compared to the latent signal, estimation of the static factors becomes more difficult. A very high noise-to-signal ratio even leads to an advantage of the PC over the TS and QML estimator if the process ξ_t is weak dependent.

   An empirical analysis of a US macroeconomic time-series panel of dimension 194×107 for the years 1960–2006 was carried out in order to estimate authentic parameters for the simulation study. The high-dimensional time-series seemed to be driven by a $3 - 4$ dimensional static factor and a $2 - 3$ dimensional dynamic factor process. The estimation of these integer parameters proved to be difficult and inconclusive. Furthermore, for $r = 4$ and $q = 2$ the QML estimation procedure did not work, which could be an indication for structural problems.

   The following remarks maybe a starting point for further analysis. This thesis did not take into account effects of the model selection step, which is a very important point in real data analysis. It did not achieve to uniformly draw from the stability region of the polynomial matrix $e(z)$. Furthermore, no misspecification of the process z_t, e.g. modeling z_t as an ARMA-process,

was studied, and it did not ask what impact performance differences have on forecasting results. Other studies tackled that problem concluding that the way future observations are projected onto this space is more important than the estimation of the linear static factor space (see [13]).

A   Further results

		strict			cross	Noise specification		cross-time			general	
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6116	0.6257	0.627	0.546	0.557	0.5538	0.4971	0.4945	0.482	0.2371	0.2773	0.2982
n=25	0.6629	0.6721	0.6727	0.6714	0.6742	0.6714	0.6374	0.6361	0.6147	0.5102	0.5085	0.4827
n=50	0.6835	0.6862	0.6856	0.7001	0.6976	0.6964	0.6869	0.6873	0.6847	0.6451	0.6325	0.615
n=100	0.6941	0.6957	0.6958	0.7079	0.7073	0.7062	0.7058	0.7016	0.7015	0.705	0.7031	0.6914
T=50												
n=10	0.7805	0.806	0.8168	0.7421	0.7511	0.7461	0.6839	0.6747	0.6605	0.4394	0.4363	0.4226
n=25	0.8387	0.8475	0.8468	0.8247	0.8268	0.8243	0.7957	0.7909	0.7836	0.7346	0.7158	0.6739
n=50	0.8545	0.8583	0.8584	0.8472	0.851	0.8513	0.8341	0.8328	0.8303	0.8225	0.8124	0.7851
n=100	0.8642	0.8669	0.8668	0.8598	0.8614	0.8616	0.8463	0.8451	0.8451	0.8459	0.8429	0.8391
T=100												
n=10	0.843	0.8704	0.8749	0.8224	0.8311	0.8263	0.7876	0.7762	0.7637	0.6194	0.5823	0.5334
n=25	0.8984	0.9083	0.9086	0.8954	0.9018	0.9017	0.8875	0.8852	0.8773	0.8633	0.8397	0.7875
n=50	0.916	0.9204	0.9202	0.9144	0.9182	0.918	0.9092	0.9106	0.9099	0.9147	0.9079	0.8903
n=100	0.9244	0.9264	0.9264	0.9233	0.925	0.9249	0.9194	0.9205	0.9197	0.9326	0.929	0.9263
T=150												
n=10	0.8686	0.9023	0.9061	0.8465	0.8596	0.8586	0.8348	0.8211	0.8075	0.6417	0.6017	0.5586
n=25	0.9236	0.9356	0.9361	0.9134	0.9193	0.9187	0.9149	0.9139	0.9106	0.8834	0.8659	0.8186
n=50	0.9412	0.9458	0.9458	0.9308	0.9359	0.9358	0.9414	0.943	0.9419	0.9307	0.9219	0.9109
n=100	0.9493	0.9512	0.9512	0.9409	0.943	0.943	0.9494	0.9502	0.95	0.948	0.9489	0.9479

Table 14: Simulation results for model A1. Medians for different noise specifications.

		$\psi = 0.5$			$\psi = 0.75$		Cross-section dependence	$\psi = 0.9$			$\psi = 0.95$	
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.4931	0.4982	0.4973	0.3876	0.3934	0.3882	0.2564	0.2858	0.3247	0.1021	0.1252	0.2162
n=25	0.5841	0.5896	0.5829	0.554	0.5517	0.5374	0.4093	0.4153	0.4162	0.1011	0.139	0.2633
n=50	0.6184	0.6182	0.6181	0.5936	0.5924	0.5844	0.5536	0.5457	0.5348	0.1076	0.149	0.3013
n=100	0.6332	0.6315	0.6312	0.6196	0.6198	0.6185	0.6371	0.634	0.6312	0.2217	0.2686	0.3756
T=50												
n=10	0.7141	0.7143	0.7122	0.622	0.6088	0.5956	0.4591	0.4764	0.4873	0.1773	0.2228	0.3288
n=25	0.7742	0.7802	0.7769	0.7627	0.7577	0.7516	0.6822	0.6696	0.6407	0.1923	0.2658	0.3886
n=50	0.7942	0.7943	0.7938	0.792	0.7949	0.7891	0.7801	0.7714	0.7438	0.2404	0.3053	0.4343
n=100	0.8037	0.8032	0.8027	0.8022	0.8033	0.8007	0.8098	0.8071	0.8061	0.5232	0.5178	0.5229
T=100												
n=10	0.8139	0.8153	0.8165	0.7501	0.7499	0.741	0.623	0.6239	0.6041	0.2786	0.3332	0.4167
n=25	0.8699	0.8759	0.8756	0.8661	0.8626	0.8558	0.8218	0.8013	0.7556	0.3484	0.4003	0.4825
n=50	0.8878	0.8918	0.8912	0.8854	0.8856	0.8849	0.8929	0.8822	0.8655	0.5152	0.5188	0.5479
n=100	0.8979	0.8979	0.8976	0.8927	0.8943	0.894	0.9089	0.9079	0.906	0.7962	0.7482	0.6461
T=150												
n=10	0.8496	0.8582	0.8579	0.8023	0.8021	0.7932	0.6547	0.6587	0.641	0.3484	0.3791	0.473
n=25	0.9093	0.9139	0.9138	0.906	0.9094	0.9031	0.8575	0.8419	0.803	0.4402	0.4707	0.5264
n=50	0.9249	0.9285	0.928	0.9252	0.9268	0.9265	0.9193	0.9109	0.8961	0.6434	0.6175	0.5846
n=100	0.9322	0.934	0.9338	0.9308	0.9324	0.9323	0.9341	0.9336	0.9318	0.8667	0.8157	0.7022

Table 15: Simulation results of model A1 (continued). Medians for different cross-sectional dependences.

		general, ntsr=0.75			Noise specification and noise-to-signal ratio general, ntsr=0.9			strict, ntsr=0.75			strict, ntsr=0.9	
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.1936	0.2008	0.2126	0.0604	0.0681	0.0782	0.3248	0.3310	0.3413	0.0800	0.0784	0.0815
n=25	0.3849	0.3830	0.3813	0.1144	0.1246	0.1294	0.4943	0.5002	0.5093	0.1901	0.1979	0.2019
n=50	0.4812	0.4723	0.4663	0.2268	0.2338	0.2359	0.5838	0.5924	0.5921	0.3059	0.3124	0.3271
n=100	0.5574	0.5491	0.5333	0.3284	0.3306	0.3327	0.6409	0.6425	0.6453	0.4310	0.4397	0.4436
T=50												
n=10	0.4040	0.4026	0.4018	0.0580	0.0697	0.0887	0.5304	0.5741	0.6135	0.2261	0.2455	0.2742
n=25	0.6194	0.6093	0.6041	0.1819	0.1955	0.2159	0.6716	0.6969	0.7045	0.4151	0.4455	0.4871
n=50	0.7207	0.7141	0.7121	0.4110	0.4174	0.4145	0.7469	0.7581	0.7612	0.5511	0.5817	0.5992
n=100	0.7669	0.7595	0.7564	0.5652	0.5607	0.5585	0.7795	0.7817	0.7825	0.6588	0.6753	0.6823
T=100												
n=10	0.5475	0.5410	0.5436	0.0779	0.1019	0.1395	0.6635	0.7256	0.7644	0.3975	0.4575	0.5544
n=25	0.7688	0.7623	0.7605	0.3359	0.3641	0.3805	0.7967	0.8234	0.8311	0.6121	0.6765	0.7244
n=50	0.8395	0.8338	0.8296	0.6385	0.6394	0.6359	0.8510	0.8609	0.8623	0.7318	0.7723	0.7880
n=100	0.8733	0.8697	0.8682	0.7573	0.7542	0.7510	0.8789	0.8814	0.8817	0.8086	0.8292	0.8350
T=150												
n=10	0.6233	0.6102	0.6056	0.0781	0.1108	0.1501	0.7094	0.7780	0.8072	0.4464	0.5267	0.6431
n=25	0.8083	0.8075	0.8038	0.4132	0.4331	0.4482	0.8355	0.8633	0.8692	0.6679	0.7436	0.7818
n=50	0.8796	0.8766	0.8756	0.7038	0.7141	0.7176	0.8844	0.8952	0.8961	0.7860	0.8314	0.8447
n=100	0.9080	0.9051	0.9033	0.8226	0.8194	0.8189	0.9112	0.9150	0.9148	0.8596	0.8822	0.8848

Table 16: Simulation results for model A1 (continued). Medians for different noise specifications and noise-to-signal ratios.

| | Noise specification | | | | | | | | | | | |
| | strict | | | cross | | | cross-time | | | general | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.5318	0.5476	0.5577	0.5286	0.5365	0.5296	0.566	0.5681	0.5682	0.6053	0.6029	0.599
n=25	0.6427	0.6534	0.6639	0.6504	0.6548	0.6504	0.6321	0.638	0.6317	0.6255	0.624	0.625
n=50	0.6881	0.6951	0.6984	0.706	0.7097	0.7056	0.6803	0.6848	0.6818	0.6416	0.6425	0.6465
n=100	0.7097	0.7155	0.7169	0.7296	0.7319	0.7282	0.7095	0.7118	0.71	0.6738	0.6788	0.6823
T=50												
n=10	0.6506	0.6865	0.7185	0.5913	0.6018	0.6108	0.6364	0.6414	0.6507	0.6828	0.6843	0.6793
n=25	0.7823	0.8054	0.8152	0.758	0.7671	0.7637	0.7253	0.7342	0.7359	0.7126	0.7145	0.7133
n=50	0.8302	0.8416	0.8452	0.8186	0.8268	0.8275	0.7912	0.8004	0.7986	0.7375	0.7427	0.7458
n=100	0.8547	0.8612	0.8625	0.8429	0.8477	0.848	0.8304	0.8348	0.836	0.7922	0.7959	0.7978
T=100												
n=10	0.7133	0.7575	0.7931	0.6388	0.6554	0.6669	0.6748	0.6851	0.6897	0.742	0.7459	0.7336
n=25	0.8481	0.8733	0.8823	0.8209	0.8331	0.8388	0.7813	0.7894	0.7919	0.7648	0.7676	0.7641
n=50	0.8951	0.9084	0.9102	0.881	0.8895	0.8919	0.8547	0.8631	0.8647	0.8063	0.8102	0.8114
n=100	0.9176	0.9244	0.9252	0.9057	0.9103	0.9103	0.8968	0.9023	0.9058	0.8558	0.8586	0.8594
T=150												
n=10	0.7341	0.786	0.8224	0.6615	0.6744	0.685	0.6901	0.6931	0.7103	0.7656	0.7637	0.7591
n=25	0.8666	0.8959	0.9042	0.8436	0.8594	0.8662	0.7925	0.8051	0.8041	0.7936	0.7894	0.784
n=50	0.9152	0.9285	0.9296	0.9024	0.9132	0.9147	0.8751	0.8862	0.8888	0.8297	0.8325	0.8301
n=100	0.9369	0.9431	0.9431	0.9302	0.9357	0.9362	0.922	0.9295	0.9301	0.8834	0.887	0.8893

Table 17: Simulation results for model A2. Medians for different noise specifications.

| | Cross-section dependence | | | | | | | | | | | |
| | $\psi = 0.5$ | | | $\psi = 0.75$ | | | $\psi = 0.9$ | | | $\psi = 0.95$ | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.7216	0.7354	0.7433	0.7093	0.7189	0.7165	0.7751	0.7725	0.769	0.825	0.8254	0.816
n=25	0.8457	0.8596	0.8689	0.8027	0.8111	0.811	0.8038	0.8092	0.806	0.8551	0.8538	0.8522
n=50	0.901	0.9128	0.9185	0.8648	0.8733	0.8718	0.8388	0.8402	0.844	0.8601	0.8608	0.8582
n=100	0.9318	0.9392	0.9433	0.9108	0.9169	0.9201	0.8923	0.8957	0.8988	0.8622	0.8606	0.8589
T=50												
n=10	0.7318	0.7506	0.7668	0.7025	0.7109	0.7163	0.7798	0.7788	0.7762	0.8357	0.834	0.8272
n=25	0.8695	0.8842	0.8949	0.8061	0.8171	0.8177	0.8062	0.806	0.8057	0.8594	0.8596	0.8601
n=50	0.9275	0.9392	0.9457	0.8871	0.8963	0.9032	0.8478	0.8546	0.852	0.8653	0.8628	0.8593
n=100	0.9564	0.9637	0.9654	0.9412	0.9476	0.9532	0.9113	0.916	0.9203	0.8661	0.8634	0.8582
T=100												
n=10	0.7306	0.7574	0.7806	0.6957	0.7065	0.716	0.7845	0.7839	0.7834	0.8474	0.8424	0.8298
n=25	0.8787	0.9022	0.9149	0.8101	0.8209	0.8277	0.8004	0.8005	0.7991	0.8736	0.8711	0.8673
n=50	0.9389	0.9542	0.9594	0.8988	0.9111	0.9224	0.8496	0.8536	0.8595	0.8791	0.8767	0.8685
n=100	0.9677	0.9747	0.9767	0.9568	0.9651	0.9678	0.9199	0.9265	0.9312	0.8769	0.8719	0.8663
T=150												
n=10	0.7296	0.761	0.7842	0.698	0.7113	0.7185	0.7847	0.7904	0.7863	0.8563	0.8519	0.8406
n=25	0.884	0.9075	0.9194	0.814	0.8262	0.8297	0.807	0.811	0.8076	0.8809	0.8777	0.8734
n=50	0.9445	0.9579	0.9635	0.9059	0.9197	0.9295	0.8542	0.8578	0.86	0.8839	0.883	0.8778
n=100	0.971	0.9782	0.9799	0.9614	0.9687	0.9713	0.9243	0.9306	0.9358	0.8858	0.8806	0.8733

Table 18: Simulation results of model A2 (continued). Medians for different cross-sectional dependences.

	strict			general, $\psi = 0.5$			general, $\psi = 0.75$			general, $\psi = 0.9$		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6569	0.6695	0.6342	0.6699	0.6792	0.6627	0.6652	0.6651	0.6577	0.7145	0.7122	0.7015
n=25	0.8017	0.816	0.7972	0.7954	0.8054	0.7841	0.7577	0.7743	0.7574	0.7275	0.7355	0.7338
n=50	0.8553	0.8651	0.8557	0.8457	0.8537	0.8428	0.8271	0.8354	0.8252	0.7719	0.7802	0.7773
n=100	0.8873	0.8921	0.8871	0.881	0.8847	0.8781	0.8753	0.8825	0.8795	0.8271	0.8311	0.8355
T=50												
n=10	0.6959	0.7285	0.7317	0.6646	0.677	0.6844	0.6634	0.674	0.6708	0.7022	0.7037	0.6965
n=25	0.854	0.8787	0.8779	0.8304	0.8499	0.8433	0.7755	0.784	0.7805	0.7354	0.7389	0.736
n=50	0.9046	0.9196	0.9186	0.8982	0.9121	0.9086	0.867	0.8762	0.8781	0.78	0.7891	0.7978
n=100	0.9312	0.9384	0.9394	0.9312	0.9374	0.9376	0.9212	0.9299	0.9329	0.8686	0.8738	0.8766
T=100												
n=10	0.7279	0.7773	0.8094	0.668	0.6872	0.7066	0.6459	0.6593	0.6744	0.7042	0.7035	0.7046
n=25	0.8799	0.9098	0.9136	0.8523	0.8723	0.8763	0.7813	0.7943	0.8019	0.718	0.7266	0.7341
n=50	0.9306	0.9452	0.9456	0.9232	0.9352	0.935	0.9005	0.9133	0.9155	0.7865	0.7922	0.7987
n=100	0.9564	0.9635	0.9636	0.9543	0.9605	0.96	0.9511	0.9579	0.9584	0.8962	0.905	0.9144
T=150												
n=10	0.7406	0.7951	0.8344	0.6671	0.6853	0.7176	0.6474	0.6607	0.676	0.7061	0.7082	0.7107
n=25	0.891	0.9212	0.9244	0.8647	0.8936	0.898	0.7843	0.7992	0.8098	0.7261	0.7259	0.7291
n=50	0.9406	0.9544	0.9549	0.9354	0.9479	0.948	0.9117	0.9244	0.9275	0.7807	0.7888	0.801
n=100	0.9646	0.9715	0.9717	0.9637	0.9699	0.9698	0.9587	0.965	0.9649	0.9162	0.9258	0.9312

Table 19: Simulation results of model A2 (continued) with estimated parameters from the Stock and Watson data-set. Medians for different cross-sectional dependences.

	Noise specification and noise-to-signal ratio											
	general, ntsr=0.75			general, ntsr=0.9			strict, ntsr=0.75			strict, ntsr=0.9		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.4970	0.5014	0.4811	0.2550	0.2560	0.2466	0.5245	0.5291	0.5239	0.2787	0.2798	0.2725
n=25	0.6504	0.6402	0.6205	0.2926	0.2860	0.2705	0.6886	0.6922	0.6980	0.3961	0.3958	0.3975
n=50	0.7412	0.7311	0.7101	0.3505	0.3411	0.3144	0.7980	0.7986	0.8012	0.5124	0.5175	0.5255
n=100	0.8032	0.7954	0.7814	0.3984	0.3892	0.3543	0.8733	0.8726	0.8726	0.6431	0.6495	0.6498
T=50												
n=10	0.4620	0.4591	0.4517	0.2073	0.1961	0.1819	0.5331	0.5549	0.5869	0.2652	0.2715	0.2759
n=25	0.6348	0.6266	0.6168	0.2556	0.2427	0.2144	0.7261	0.7448	0.7622	0.4135	0.4262	0.4481
n=50	0.7641	0.7560	0.7477	0.3655	0.3481	0.3207	0.8374	0.8464	0.8531	0.5603	0.5727	0.5925
n=100	0.8509	0.8454	0.8405	0.4805	0.4617	0.4308	0.9052	0.9098	0.9123	0.7058	0.7181	0.7271
T=100												
n=10	0.4534	0.4536	0.4418	0.1624	0.1560	0.1478	0.5434	0.5819	0.6314	0.2626	0.2748	0.2914
n=25	0.6507	0.6430	0.6393	0.2385	0.2248	0.2007	0.7437	0.7714	0.7928	0.4467	0.4701	0.5000
n=50	0.7932	0.7891	0.7851	0.4037	0.3852	0.3506	0.8497	0.8665	0.8746	0.6082	0.6369	0.6602
n=100	0.8869	0.8846	0.8841	0.5838	0.5717	0.5543	0.9181	0.9250	0.9284	0.7555	0.7758	0.7865
T=150												
n=10	0.4511	0.4478	0.4389	0.1480	0.1432	0.1302	0.5468	0.5899	0.6415	0.2727	0.2859	0.3144
n=25	0.6548	0.6508	0.6460	0.2271	0.2134	0.1972	0.7512	0.7850	0.8051	0.4708	0.5000	0.5343
n=50	0.8049	0.8045	0.8043	0.4202	0.4059	0.3883	0.8556	0.8746	0.8823	0.6310	0.6658	0.6926
n=100	0.8985	0.8999	0.9007	0.6346	0.6311	0.6238	0.9215	0.9294	0.9319	0.7740	0.7978	0.8102

Table 20: Simulation results for model A2 (continued). Medians for different noise specifications and constant noise-to-signal ratios.

	strict			general			general, $\psi = 0.75$			general, $\psi = 0.9$		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.0325	0.0338	0.038	0.0422	0.0439	0.0499	0.0467	0.0468	0.0517	0.0608	0.0604	0.0617
n=25	0.0443	0.046	0.0506	0.0519	0.0522	0.0579	0.0494	0.0512	0.0544	0.058	0.0589	0.0583
n=50	0.0523	0.0536	0.0561	0.0628	0.0648	0.0659	0.0597	0.0612	0.063	0.0559	0.0571	0.0592
n=100	0.0588	0.0601	0.0604	0.068	0.0685	0.0683	0.0672	0.0684	0.0701	0.0622	0.0626	0.0635
T=50												
n=10	0.0658	0.0714	0.0837	0.0723	0.0788	0.0882	0.0774	0.0792	0.0847	0.0964	0.0967	0.1022
n=25	0.0926	0.0985	0.1056	0.092	0.0953	0.1014	0.0846	0.0863	0.0917	0.0872	0.0886	0.094
n=50	0.1086	0.1136	0.1175	0.1119	0.1157	0.1182	0.0948	0.0972	0.1033	0.0822	0.0859	0.0935
n=100	0.1199	0.1214	0.1227	0.1261	0.1265	0.1272	0.1108	0.1127	0.1167	0.1014	0.1032	0.1047
T=100												
n=10	0.1145	0.1278	0.1567	0.1105	0.1195	0.1326	0.1289	0.1349	0.145	0.1528	0.157	0.1585
n=25	0.157	0.1709	0.1812	0.1405	0.1517	0.1628	0.1383	0.1423	0.1522	0.1522	0.1542	0.1588
n=50	0.1747	0.1839	0.191	0.1727	0.1793	0.1814	0.167	0.174	0.1782	0.1523	0.1548	0.1622
n=100	0.1909	0.1967	0.1981	0.1914	0.1945	0.1964	0.1908	0.1939	0.1979	0.1728	0.1758	0.1787
T=150												
n=10	0.1747	0.1959	0.2258	0.1478	0.1575	0.1723	0.169	0.1734	0.1837	0.1995	0.1997	0.2069
n=25	0.217	0.2288	0.2478	0.2015	0.2113	0.2193	0.1877	0.1936	0.2032	0.1987	0.1977	0.2074
n=50	0.2365	0.2481	0.2523	0.2311	0.2383	0.2425	0.2136	0.2192	0.2275	0.2125	0.2135	0.2209
n=100	0.2478	0.2541	0.2563	0.2468	0.2503	0.2518	0.2403	0.2427	0.2456	0.233	0.2345	0.2385

Table 21: Simulation results for model A2 (continued). Medians for processes with a long memory.

						Noise specification						
	strict			cross			cross-time			general		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.5324	0.5385	0.5422	0.5845	0.5789	0.5774	0.5694	0.5689	0.5657	0.6259	0.6199	0.6122
n=25	0.6317	0.6438	0.6459	0.6519	0.6578	0.6518	0.6322	0.6353	0.631	0.6406	0.644	0.6347
n=50	0.6841	0.6947	0.6979	0.6881	0.6909	0.6892	0.6726	0.6767	0.6733	0.6593	0.6612	0.6618
n=100	0.7132	0.7206	0.7223	0.7177	0.7189	0.7189	0.7091	0.7124	0.7104	0.6814	0.6854	0.6858
T=50												
n=10	0.5975	0.6167	0.6508	0.6514	0.6614	0.6671	0.6374	0.6411	0.6351	0.7176	0.7124	0.7093
n=25	0.7367	0.7572	0.7686	0.7481	0.7584	0.7562	0.718	0.7274	0.7248	0.7457	0.7459	0.7432
n=50	0.7981	0.8113	0.8135	0.8038	0.813	0.8136	0.7746	0.7802	0.7859	0.7624	0.767	0.7706
n=100	0.8311	0.8411	0.842	0.8374	0.8421	0.841	0.8236	0.828	0.8326	0.7937	0.796	0.8001
T=100												
n=10	0.6557	0.6864	0.7332	0.6791	0.6893	0.7051	0.6695	0.6744	0.6749	0.7501	0.7471	0.7381
n=25	0.8107	0.8403	0.8527	0.7944	0.8112	0.8171	0.768	0.7753	0.7774	0.78	0.7852	0.7815
n=50	0.8728	0.8882	0.8936	0.8664	0.8761	0.88	0.8316	0.841	0.8456	0.8145	0.8177	0.819
n=100	0.9039	0.9122	0.9136	0.903	0.9088	0.9094	0.8848	0.8913	0.8932	0.8545	0.8576	0.8634
T=150												
n=10	0.6731	0.7063	0.7575	0.7022	0.7174	0.7255	0.6776	0.6817	0.6871	0.7708	0.767	0.7599
n=25	0.8356	0.8686	0.88	0.8162	0.8322	0.8399	0.7806	0.7916	0.7922	0.7996	0.7985	0.7972
n=50	0.897	0.9134	0.9166	0.8927	0.9035	0.9056	0.8509	0.8604	0.8641	0.8275	0.8308	0.8363
n=100	0.9276	0.9353	0.9368	0.9269	0.9329	0.9336	0.9072	0.9143	0.9181	0.8772	0.8818	0.8893

Table 22: Simulation results for model A3. Medians for different noise specifications.

| | | Cross-section dependence | | | | | | | | | | | |
| | $\psi = 0.5$ | | | $\psi = 0.75$ | | | $\psi = 0.9$ | | | $\psi = 0.95$ | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.7094	0.7191	0.7255	0.7369	0.7394	0.7348	0.814	0.8154	0.8004	0.8845	0.8803	0.8706
n=25	0.8257	0.8376	0.8441	0.818	0.825	0.8189	0.8345	0.8394	0.8325	0.9089	0.907	0.9038
n=50	0.8751	0.8844	0.891	0.8623	0.8675	0.8705	0.8515	0.8574	0.856	0.9112	0.9091	0.9036
n=100	0.9136	0.9211	0.9259	0.8974	0.9036	0.9091	0.8815	0.8877	0.8915	0.9004	0.9007	0.8974
T=50												
n=10	0.6927	0.7049	0.7281	0.7234	0.7265	0.7272	0.807	0.8101	0.7964	0.8977	0.892	0.877
n=25	0.8335	0.8554	0.8679	0.8064	0.8133	0.8138	0.8178	0.8202	0.8179	0.9241	0.921	0.9142
n=50	0.8991	0.9121	0.921	0.8702	0.878	0.8862	0.8435	0.8491	0.8527	0.9234	0.9207	0.9139
n=100	0.9399	0.9491	0.9554	0.9209	0.9295	0.9374	0.8891	0.8964	0.9005	0.9122	0.9109	0.9067
T=100												
n=10	0.6806	0.7004	0.7225	0.7171	0.7194	0.7212	0.8043	0.8106	0.7908	0.902	0.8982	0.8836
n=25	0.8384	0.8617	0.8823	0.8083	0.8132	0.8216	0.8098	0.8131	0.807	0.9326	0.9303	0.9247
n=50	0.9122	0.9281	0.939	0.8794	0.8918	0.9017	0.8319	0.8395	0.8464	0.9317	0.9291	0.9208
n=100	0.9564	0.9656	0.9695	0.939	0.949	0.9569	0.892	0.8989	0.9076	0.9197	0.9182	0.9149
T=150												
n=10	0.676	0.6966	0.7317	0.7133	0.7207	0.7257	0.8063	0.8104	0.7967	0.9073	0.9038	0.888
n=25	0.8443	0.8692	0.89	0.8019	0.8076	0.818	0.8044	0.8077	0.8065	0.9347	0.934	0.9275
n=50	0.9171	0.9344	0.9446	0.8825	0.8939	0.907	0.8286	0.8378	0.8506	0.9376	0.9349	0.9265
n=100	0.9608	0.9705	0.9737	0.9451	0.9559	0.9636	0.8967	0.9039	0.9091	0.9221	0.9201	0.9146

Table 23: Simulation results of model A3 (continued). Medians for different cross-sectional dependences.

	strict			cross			cross-time			general		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6647	0.672	0.6809	0.7009	0.7052	0.6986	0.7119	0.5886	0.632	0.7787	0.5181	0.5999
n=25	0.7287	0.7511	0.7567	0.7413	0.7903	0.7757	0.7305	0.7187	0.7257	0.7907	0.6553	0.676
n=50	0.7563	0.7776	0.7748	0.7611	0.8331	0.8055	0.7463	0.7799	0.765	0.7973	0.7473	0.7428
n=100	0.7689	0.7937	0.7874	0.774	0.8445	0.8224	0.7654	0.8212	0.806	0.809	0.8203	0.8044
T=50												
n=10	0.7607	0.7767	0.804	0.7719	0.7698	0.784	0.791	0.621	0.6786	0.8546	0.4768	0.6065
n=25	0.8367	0.8586	0.8632	0.8322	0.8527	0.8548	0.8271	0.8093	0.8114	0.8661	0.6547	0.74
n=50	0.8678	0.879	0.8801	0.8573	0.8763	0.8713	0.8468	0.8613	0.8605	0.8718	0.8271	0.8361
n=100	0.877	0.8862	0.8865	0.8699	0.8884	0.8841	0.8645	0.8847	0.8819	0.8839	0.89	0.8821
T=100												
n=10	0.8173	0.8333	0.8697	0.825	0.832	0.8431	0.8457	0.6724	0.7559	0.9032	0.4522	0.6643
n=25	0.8932	0.9105	0.9156	0.8955	0.9127	0.9157	0.8851	0.8761	0.8802	0.9077	0.692	0.8031
n=50	0.919	0.9283	0.9295	0.9181	0.9317	0.9315	0.914	0.9244	0.925	0.9117	0.8905	0.8954
n=100	0.9294	0.9357	0.9363	0.9294	0.9376	0.9379	0.9315	0.9373	0.9382	0.9292	0.9375	0.9396
T=150												
n=10	0.8409	0.843	0.8993	0.8452	0.8489	0.8686	0.8694	0.7031	0.785	0.9158	0.4615	0.6845
n=25	0.9164	0.9351	0.9393	0.9054	0.9241	0.9272	0.9023	0.8993	0.903	0.92	0.7146	0.8261
n=50	0.9415	0.9522	0.9535	0.9314	0.9405	0.9403	0.9321	0.9433	0.9471	0.9268	0.9107	0.9171
n=100	0.9521	0.9576	0.9574	0.9405	0.9474	0.9475	0.9524	0.9575	0.9582	0.9432	0.9525	0.9578

Table 24: Simulation results for model B1. Medians for different noise specifications.

	Cross-section dependence											
	$\psi = 0.5$			$\psi = 0.75$			$\psi = 0.9$			$\psi = 0.95$		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.8283	0.8485	0.8678	0.8716	0.8466	0.8475	0.8593	0.6864	0.7359	0.9412	0.567	0.616
n=25	0.8961	0.9213	0.9289	0.8955	0.9032	0.9092	0.8954	0.8417	0.8543	0.9501	0.6008	0.6895
n=50	0.934	0.9481	0.9526	0.9309	0.9368	0.9452	0.9126	0.9009	0.9089	0.9535	0.7087	0.7788
n=100	0.9507	0.9596	0.9623	0.9473	0.9547	0.9589	0.936	0.9344	0.939	0.9499	0.7991	0.8304
T=50												
n=10	0.8356	0.8608	0.8931	0.8798	0.8615	0.8524	0.8752	0.6923	0.7369	0.9507	0.4503	0.5359
n=25	0.9064	0.9391	0.9502	0.9017	0.9153	0.9273	0.9054	0.8638	0.8614	0.9647	0.4622	0.6152
n=50	0.9474	0.9661	0.9705	0.9451	0.956	0.965	0.9252	0.9179	0.9281	0.9669	0.6071	0.7408
n=100	0.9681	0.9764	0.979	0.9636	0.9729	0.9775	0.9491	0.9509	0.9566	0.964	0.7514	0.8061
T=100												
n=10	0.8478	0.88	0.9094	0.8828	0.8772	0.8598	0.8984	0.6979	0.7467	0.9586	0.3757	0.5075
n=25	0.9122	0.9499	0.9606	0.9099	0.9255	0.941	0.9235	0.8874	0.879	0.9713	0.3569	0.5866
n=50	0.9558	0.9751	0.9794	0.9495	0.9664	0.9745	0.9426	0.9358	0.9474	0.9746	0.5358	0.7382
n=100	0.9746	0.9852	0.9871	0.9702	0.9819	0.9853	0.9617	0.9648	0.9752	0.9741	0.7296	0.8116
T=150												
n=10	0.8516	0.8816	0.9137	0.8865	0.8794	0.8673	0.898	0.7008	0.7414	0.9637	0.3581	0.4919
n=25	0.9144	0.9534	0.9631	0.9121	0.9301	0.9449	0.9275	0.8963	0.8806	0.9735	0.3153	0.5801
n=50	0.9579	0.9782	0.9823	0.9533	0.9703	0.9784	0.9454	0.9424	0.9514	0.9771	0.4855	0.7297
n=100	0.977	0.9876	0.99	0.9732	0.985	0.988	0.9637	0.9689	0.9789	0.9765	0.7091	0.8022

Table 25: Simulation results of model B1 (continued). Medians for different cross-sectional dependences.

	Noise specification and noise-to-signal ratio											
	general, ntsr=0.75			general, ntsr=0.9			strict, ntsr=0.75			strict, ntsr=0.9		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.5689	0.5443	0.5490	0.3285	0.2274	0.2346	0.6038	0.6272	0.6209	0.3521	0.2846	0.2945
n=25	0.7207	0.7238	0.7220	0.3977	0.3412	0.3218	0.7626	0.8031	0.8049	0.4923	0.4597	0.4803
n=50	0.8146	0.8215	0.8089	0.4804	0.4476	0.4126	0.8503	0.8797	0.8819	0.6407	0.6279	0.6458
n=100	0.8649	0.8668	0.8585	0.5658	0.5543	0.5099	0.9020	0.9234	0.9254	0.7548	0.7482	0.7737
T=50												
n=10	0.5605	0.5521	0.5258	0.2715	0.1581	0.1611	0.6171	0.6699	0.6852	0.3549	0.3132	0.3318
n=25	0.7397	0.7525	0.7632	0.3585	0.3196	0.2902	0.7841	0.8437	0.8466	0.5412	0.5592	0.5762
n=50	0.8400	0.8600	0.8588	0.5057	0.5164	0.5001	0.8689	0.9091	0.9089	0.6834	0.7242	0.7420
n=100	0.9025	0.9130	0.9131	0.6601	0.6738	0.6717	0.9212	0.9456	0.9474	0.8006	0.8377	0.8461
T=100												
n=10	0.5561	0.5592	0.5499	0.2460	0.1091	0.1298	0.6321	0.7071	0.7310	0.3634	0.3908	0.4190
n=25	0.7456	0.7729	0.7969	0.3566	0.3453	0.3623	0.8025	0.8649	0.8723	0.5625	0.6322	0.6579
n=50	0.8583	0.8884	0.8954	0.5582	0.5930	0.6217	0.8843	0.9267	0.9287	0.7136	0.7819	0.7989
n=100	0.9180	0.9397	0.9428	0.7230	0.7656	0.7770	0.9327	0.9595	0.9599	0.8235	0.8768	0.8840
T=150												
n=10	0.5517	0.5542	0.5542	0.2247	0.1037	0.1203	0.6417	0.7182	0.7429	0.3678	0.4147	0.4538
n=25	0.7507	0.7819	0.8168	0.3449	0.3566	0.3957	0.8099	0.8740	0.8792	0.5798	0.6577	0.6839
n=50	0.8635	0.8974	0.9073	0.5731	0.6164	0.6581	0.8918	0.9321	0.9343	0.7254	0.8015	0.8171
n=100	0.9238	0.9470	0.9510	0.7547	0.7996	0.8095	0.9379	0.9633	0.9637	0.8349	0.8911	0.8982

Table 26: Simulation results for model B1 (continued). Medians for different noise specifications and noise-to-signal ratios.

	strict			general, $\psi = 0.5$			general, $\psi = 0.75$			general, $\psi = 0.9$		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6746	0.6964	0.6771	0.7066	0.7132	0.6995	0.7198	0.7134	0.7132	0.7618	0.7428	0.7358
n=25	0.8116	0.8359	0.821	0.8139	0.8214	0.8186	0.8091	0.7994	0.7955	0.7835	0.7652	0.764
n=50	0.8747	0.8866	0.8798	0.8704	0.8767	0.8692	0.864	0.8691	0.8656	0.8153	0.802	0.8031
n=100	0.9019	0.9063	0.9066	0.9011	0.9018	0.9035	0.9012	0.9032	0.8997	0.8548	0.8438	0.8403
T=50												
n=10	0.7222	0.7571	0.7604	0.6967	0.7101	0.7209	0.701	0.6766	0.6887	0.7536	0.6971	0.6989
n=25	0.8614	0.8983	0.8973	0.8403	0.8615	0.8678	0.802	0.7892	0.8006	0.7974	0.735	0.7619
n=50	0.9141	0.9318	0.9328	0.9116	0.9279	0.9291	0.8843	0.8941	0.8998	0.829	0.7942	0.8075
n=100	0.9402	0.9486	0.9489	0.9409	0.9486	0.9505	0.9288	0.9425	0.9455	0.8822	0.8724	0.8894
T=100												
n=10	0.7415	0.802	0.8305	0.6953	0.6992	0.7336	0.6954	0.6454	0.6798	0.7634	0.6752	0.6902
n=25	0.8899	0.929	0.9317	0.8634	0.8981	0.9068	0.8083	0.7943	0.812	0.8031	0.7146	0.7321
n=50	0.9378	0.9598	0.9603	0.9292	0.9505	0.9526	0.8979	0.9206	0.9318	0.8416	0.7803	0.7961
n=100	0.9641	0.974	0.9741	0.9578	0.9669	0.9679	0.9513	0.9636	0.9664	0.8989	0.9087	0.9329
T=150												
n=10	0.7458	0.8199	0.8457	0.694	0.7037	0.7334	0.6884	0.6465	0.6722	0.7698	0.6747	0.6858
n=25	0.8958	0.9364	0.9388	0.8694	0.9099	0.9197	0.813	0.7819	0.816	0.8036	0.709	0.7294
n=50	0.9443	0.9651	0.9654	0.937	0.9576	0.9595	0.9057	0.9329	0.9445	0.8466	0.7704	0.7904
n=100	0.9683	0.9783	0.9785	0.966	0.9759	0.9765	0.9596	0.9724	0.9739	0.903	0.9214	0.9506

Table 27: Simulation results of model B12 ($r = 3, q = 2$) with estimated parameters from the Stock and Watson data-set. Medians for different cross-sectional dependences.

	Noise specification											
	strict			cross			cross-time			general		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6019	0.5874	0.6024	0.6278	0.5968	0.6159	0.6511	0.5848	0.6021	0.6988	0.6056	0.6254
n=25	0.6817	0.6807	0.6861	0.6909	0.6981	0.6874	0.695	0.6637	0.6623	0.7155	0.6381	0.6517
n=50	0.7192	0.7273	0.7267	0.7128	0.7378	0.7192	0.7203	0.719	0.7204	0.717	0.6682	0.6818
n=100	0.7361	0.746	0.7446	0.7248	0.775	0.7311	0.7335	0.765	0.745	0.7304	0.7272	0.7308
T=50												
n=10	0.7008	0.6728	0.7162	0.7214	0.6419	0.6901	0.7325	0.601	0.6465	0.7804	0.6184	0.6456
n=25	0.7861	0.7995	0.8035	0.8005	0.8005	0.8032	0.7903	0.7117	0.7394	0.8095	0.6526	0.6922
n=50	0.8256	0.8366	0.84	0.8273	0.8429	0.8326	0.8262	0.8072	0.8139	0.8194	0.7013	0.7381
n=100	0.8439	0.8544	0.8572	0.8449	0.866	0.8511	0.8437	0.8555	0.8544	0.8314	0.7908	0.8027
T=100												
n=10	0.76	0.7319	0.807	0.7716	0.6806	0.7368	0.7878	0.6242	0.6607	0.8597	0.6222	0.6679
n=25	0.8506	0.8714	0.8773	0.8591	0.8626	0.8692	0.8423	0.7416	0.7832	0.8822	0.6648	0.7308
n=50	0.8913	0.902	0.9043	0.8904	0.9035	0.9038	0.8795	0.8537	0.8716	0.8973	0.73	0.8064
n=100	0.9113	0.9174	0.9186	0.9109	0.922	0.9204	0.9001	0.9057	0.9046	0.9103	0.8649	0.8912
T=150												
n=10	0.7809	0.7417	0.835	0.7924	0.6982	0.7675	0.8133	0.6346	0.677	0.8715	0.6237	0.6647
n=25	0.8721	0.8932	0.9009	0.8789	0.8844	0.892	0.8669	0.7613	0.8129	0.896	0.6609	0.7419
n=50	0.9117	0.9238	0.9262	0.9139	0.924	0.9262	0.9053	0.8861	0.9085	0.9138	0.7291	0.8191
n=100	0.9328	0.9379	0.9394	0.9346	0.9437	0.9431	0.931	0.9371	0.941	0.9234	0.8795	0.9131

Table 28: Simulation results for model B2. Medians for different noise specifications.

	Cross-section dependence											
	$\psi = 0.5$			$\psi = 0.75$			$\psi = 0.9$			$\psi = 0.95$		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.7606	0.7569	0.7749	0.7701	0.7276	0.7407	0.8442	0.7668	0.7801	0.9104	0.795	0.8041
n=25	0.8549	0.8689	0.8751	0.8248	0.8093	0.8229	0.8646	0.8133	0.8329	0.9248	0.8262	0.8303
n=50	0.8996	0.9093	0.9145	0.8792	0.8816	0.8896	0.8817	0.8565	0.8621	0.9223	0.8351	0.8465
n=100	0.9279	0.934	0.9369	0.9123	0.9165	0.9209	0.9042	0.8965	0.9046	0.9135	0.8315	0.8464
T=50												
n=10	0.7594	0.7569	0.8	0.7513	0.7059	0.7193	0.8415	0.7504	0.7575	0.9232	0.7758	0.7907
n=25	0.8589	0.8821	0.8969	0.8253	0.8096	0.8341	0.8636	0.7955	0.8155	0.9392	0.8059	0.8219
n=50	0.9143	0.9311	0.9387	0.8911	0.8998	0.9099	0.8833	0.8518	0.8683	0.9364	0.817	0.8337
n=100	0.9475	0.9571	0.9607	0.9324	0.9397	0.9461	0.9097	0.9062	0.9227	0.9195	0.8177	0.8372
T=100												
n=10	0.7507	0.7603	0.8089	0.7493	0.7011	0.7248	0.8373	0.7461	0.7576	0.9289	0.7638	0.7868
n=25	0.8626	0.8951	0.9169	0.8198	0.8135	0.8385	0.864	0.7943	0.8189	0.9492	0.7974	0.8098
n=50	0.9253	0.9484	0.9553	0.8992	0.9124	0.9276	0.882	0.8463	0.8709	0.9455	0.8036	0.8208
n=100	0.9594	0.9714	0.9743	0.9438	0.957	0.9655	0.9118	0.9111	0.9314	0.9305	0.8108	0.8288
T=150												
n=10	0.7497	0.7621	0.8131	0.7478	0.7028	0.7173	0.8434	0.7494	0.7591	0.9304	0.7645	0.7815
n=25	0.8639	0.902	0.9228	0.8231	0.8117	0.8402	0.8612	0.7905	0.8168	0.9516	0.7946	0.8075
n=50	0.9308	0.9536	0.9611	0.9013	0.9185	0.9368	0.88	0.8475	0.8706	0.9491	0.7988	0.8282
n=100	0.964	0.9773	0.9798	0.9484	0.9632	0.9724	0.9114	0.9118	0.9358	0.9325	0.8082	0.8298

Table 29: Simulation results of model B2 (continued). Medians for different cross-sectional dependences.

	Noise specification											
	strict			cross			cross-time			general		
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.5837	0.6245	0.6614	0.677	0.6909	0.7379	0.6164	0.6363	0.6719	0.6885	0.716	0.7458
n=25	0.6924	0.7266	0.7483	0.7677	0.7885	0.8203	0.6911	0.7117	0.7324	0.7534	0.7793	0.7924
n=50	0.7513	0.7721	0.7858	0.8276	0.8466	0.8642	0.7375	0.7541	0.7693	0.796	0.8093	0.8172
n=100	0.7847	0.7966	0.8036	0.8697	0.8798	0.8932	0.7723	0.7849	0.794	0.8238	0.8315	0.8348
T=50												
n=10	0.6188	0.6802	0.7361	0.6833	0.7026	0.776	0.6334	0.6536	0.7137	0.6922	0.7324	0.7783
n=25	0.7453	0.7884	0.8089	0.7914	0.8188	0.8582	0.7279	0.7506	0.7903	0.7783	0.8107	0.8355
n=50	0.8026	0.8304	0.8376	0.8567	0.8777	0.8987	0.7821	0.8029	0.8221	0.8313	0.8501	0.8628
n=100	0.8314	0.8457	0.8489	0.8982	0.913	0.9212	0.8197	0.8335	0.841	0.8639	0.876	0.881
T=100												
n=10	0.6596	0.7268	0.798	0.7315	0.7662	0.8406	0.6914	0.7106	0.7829	0.7172	0.756	0.8135
n=25	0.7914	0.8406	0.8586	0.8408	0.8719	0.9099	0.7873	0.8108	0.8491	0.8146	0.8512	0.8771
n=50	0.8467	0.8766	0.881	0.9045	0.9268	0.9448	0.842	0.8653	0.878	0.8686	0.8904	0.9008
n=100	0.8732	0.8886	0.8898	0.9406	0.9553	0.9599	0.8745	0.8889	0.8933	0.9	0.914	0.9184
T=150												
n=10	0.6889	0.7604	0.8314	0.7427	0.7806	0.854	0.7206	0.7417	0.811	0.7365	0.7788	0.8326
n=25	0.821	0.8707	0.8867	0.8507	0.8856	0.9205	0.8131	0.8348	0.8751	0.8341	0.8688	0.8973
n=50	0.8739	0.9041	0.9071	0.9125	0.9361	0.9521	0.8669	0.8897	0.9013	0.8877	0.9083	0.9209
n=100	0.8994	0.9144	0.9153	0.9479	0.963	0.9671	0.8962	0.9106	0.9138	0.9186	0.9316	0.9361

Table 30: Simulation results for model C1. Medians for different noise specifications.

| | | Cross-section dependence | | | | | | | | | |
| | $\psi = 0.5$ | | | $\psi = 0.75$ | | | $\psi = 0.9$ | | | $\psi = 0.95$ | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6142	0.6401	0.6685	0.6911	0.7003	0.738	0.7681	0.75	0.777	0.7842	0.7522	0.7633
n=25	0.6778	0.7082	0.7137	0.7382	0.7554	0.7786	0.7884	0.7641	0.792	0.7932	0.7621	0.774
n=50	0.72	0.7418	0.7451	0.7866	0.8014	0.8127	0.7956	0.7791	0.806	0.7906	0.7662	0.7776
n=100	0.749	0.7592	0.76	0.8183	0.8283	0.8326	0.8223	0.818	0.8352	0.7816	0.7517	0.7721
T=50												
n=10	0.6242	0.6522	0.7022	0.7039	0.7215	0.7783	0.7731	0.7491	0.7899	0.8299	0.7922	0.8095
n=25	0.7081	0.7348	0.7593	0.751	0.7681	0.8065	0.7899	0.7413	0.7966	0.8384	0.7997	0.8148
n=50	0.7617	0.7817	0.7937	0.8141	0.8262	0.8464	0.8004	0.7677	0.8102	0.836	0.8027	0.8184
n=100	0.7943	0.8068	0.8107	0.8569	0.8673	0.8802	0.8312	0.8187	0.8506	0.8197	0.7779	0.8105
T=100												
n=10	0.6674	0.6961	0.7568	0.7177	0.74	0.7948	0.7945	0.7605	0.8109	0.8645	0.8289	0.8431
n=25	0.7588	0.7869	0.8159	0.7674	0.7864	0.8293	0.8168	0.7498	0.8258	0.8729	0.8304	0.8491
n=50	0.8152	0.8365	0.8481	0.8353	0.8408	0.8736	0.8277	0.7787	0.8424	0.8717	0.8292	0.852
n=100	0.846	0.8582	0.8639	0.8848	0.8918	0.9106	0.8611	0.8384	0.8859	0.8583	0.7998	0.8475
T=150												
n=10	0.68	0.7083	0.7708	0.7311	0.7558	0.8168	0.8073	0.7789	0.8302	0.8809	0.8434	0.8574
n=25	0.7738	0.8052	0.8311	0.7815	0.8009	0.8496	0.8312	0.7695	0.8431	0.8893	0.8431	0.8641
n=50	0.8295	0.8505	0.8616	0.8525	0.8553	0.8908	0.8408	0.7928	0.8545	0.8884	0.8435	0.8702
n=100	0.8586	0.8711	0.8748	0.9009	0.9092	0.9269	0.8743	0.8464	0.8974	0.876	0.8077	0.8658

Table 31: Model C1 (continued). Medians for near unit root setting and different cross-sectional dependences.

| | | Noise specification and noise-to-signal ratio | | | | | | | | | |
| | general, ntsr=0.75 | | | general, ntsr=0.9 | | | strict, ntsr=0.75 | | | strict, ntsr=0.9 | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.4361	0.4463	0.4839	0.2770	0.2736	0.2815	0.4487	0.4945	0.5386	0.2561	0.2625	0.2737
n=25	0.5167	0.5315	0.5749	0.3109	0.3430	0.3385	0.6017	0.6585	0.6961	0.3118	0.3346	0.3668
n=50	0.6088	0.6258	0.6253	0.3478	0.3635	0.3416	0.7117	0.7632	0.7807	0.3950	0.4335	0.4802
n=100	0.6776	0.6842	0.6775	0.3799	0.3935	0.3685	0.8079	0.8325	0.8473	0.5309	0.5502	0.6137
T=50												
n=10	0.4289	0.4401	0.5168	0.2234	0.2032	0.2172	0.4537	0.5396	0.6552	0.2301	0.2389	0.2970
n=25	0.5434	0.5832	0.6491	0.2552	0.2463	0.2646	0.6440	0.7216	0.7758	0.3312	0.3606	0.4415
n=50	0.6645	0.6894	0.7194	0.3139	0.3276	0.3454	0.7698	0.8326	0.8533	0.4594	0.5009	0.5864
n=100	0.7694	0.7745	0.7848	0.4212	0.4148	0.4265	0.8553	0.8885	0.9052	0.5988	0.6307	0.7083
T=100												
n=10	0.4491	0.4575	0.5959	0.2088	0.1613	0.1993	0.4876	0.6034	0.7418	0.2357	0.2751	0.3695
n=25	0.6122	0.6401	0.7249	0.2552	0.2328	0.3132	0.6904	0.7889	0.8412	0.3811	0.4689	0.5842
n=50	0.7393	0.7681	0.8041	0.3807	0.3867	0.4890	0.8085	0.8843	0.9024	0.5383	0.6220	0.7161
n=100	0.8334	0.8544	0.8675	0.5514	0.5592	0.6255	0.8880	0.9342	0.9410	0.6926	0.7504	0.8058
T=150												
n=10	0.4615	0.4850	0.6206	0.2159	0.1492	0.2109	0.5071	0.6354	0.7745	0.2487	0.3081	0.4351
n=25	0.6281	0.6733	0.7513	0.2625	0.2410	0.3762	0.7083	0.8167	0.8628	0.4183	0.5286	0.6611
n=50	0.7584	0.7974	0.8308	0.4322	0.4371	0.6150	0.8287	0.9029	0.9178	0.5836	0.6881	0.7612
n=100	0.8621	0.8853	0.8983	0.6060	0.6347	0.7029	0.9029	0.9483	0.9524	0.7269	0.8093	0.8473

Table 32: Simulation results for model C1 (continued). Medians for different noise specifications and noise-to-signal ratios.

| | Noise specification | | | | | | | | | | | |
| | strict | | | cross | | | cross-time | | | general | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6039	0.6345	0.6626	0.6333	0.6357	0.6705	0.5475	0.551	0.5768	0.6302	0.6598	0.6797
n=25	0.7076	0.7312	0.7464	0.7168	0.7371	0.7575	0.6186	0.6328	0.6544	0.7127	0.736	0.7418
n=50	0.7559	0.7742	0.7855	0.771	0.7871	0.8023	0.6712	0.6845	0.7014	0.7476	0.765	0.7659
n=100	0.7864	0.7984	0.8041	0.804	0.8153	0.8238	0.7031	0.7164	0.7235	0.7693	0.7818	0.782
T=50												
n=10	0.6383	0.6811	0.7347	0.6795	0.687	0.7453	0.5977	0.5997	0.6484	0.6729	0.7007	0.7441
n=25	0.7552	0.7938	0.8117	0.7765	0.8006	0.8294	0.6808	0.7002	0.729	0.7698	0.7939	0.8137
n=50	0.8031	0.8275	0.8358	0.8328	0.8544	0.8687	0.7367	0.7556	0.7727	0.8139	0.8344	0.8417
n=100	0.8313	0.8465	0.849	0.8647	0.8788	0.8847	0.7673	0.7844	0.7896	0.8405	0.8536	0.8569
T=100												
n=10	0.687	0.739	0.7991	0.7189	0.7228	0.7883	0.6579	0.6558	0.7238	0.7066	0.7271	0.7867
n=25	0.8071	0.8502	0.8639	0.8193	0.8476	0.8812	0.7473	0.7644	0.7989	0.8095	0.8317	0.8564
n=50	0.852	0.8777	0.8822	0.8779	0.9018	0.9143	0.8039	0.8225	0.837	0.8539	0.8707	0.8821
n=100	0.8765	0.8903	0.8917	0.9081	0.9221	0.9262	0.8323	0.8453	0.8498	0.8802	0.8916	0.8956
T=150												
n=10	0.7117	0.7613	0.8281	0.7357	0.7408	0.8127	0.6922	0.6917	0.7617	0.7246	0.7509	0.8083
n=25	0.8333	0.8791	0.8904	0.8407	0.8717	0.9072	0.78	0.799	0.8354	0.8261	0.8503	0.8748
n=50	0.8773	0.9035	0.9071	0.9011	0.9259	0.9373	0.8348	0.8558	0.8677	0.8706	0.888	0.8988
n=100	0.9007	0.9145	0.9154	0.9314	0.9457	0.9492	0.8608	0.8737	0.8775	0.8958	0.9075	0.911

Table 33: Simulation results for model C2. Medians for different noise specifications.

| | Cross-section dependence | | | | | | | | | | | |
| | $\psi = 0.5$ | | | $\psi = 0.75$ | | | $\psi = 0.9$ | | | $\psi = 0.95$ | | |
	PC	TS	QML	PC	TS	QML	PC	TS	QML	PC	TS	QML
T=25												
n=10	0.6991	0.7162	0.7436	0.6667	0.6599	0.6926	0.7374	0.7149	0.7385	0.8215	0.785	0.8023
n=25	0.7755	0.7898	0.8033	0.7118	0.7235	0.7408	0.753	0.7349	0.7588	0.8316	0.8052	0.8115
n=50	0.8172	0.8319	0.8362	0.7486	0.7596	0.7676	0.7663	0.7578	0.7784	0.8288	0.7958	0.8125
n=100	0.8417	0.8476	0.8533	0.7764	0.7841	0.7888	0.7869	0.7884	0.8016	0.8232	0.7964	0.8087
T=50												
n=10	0.7169	0.7339	0.772	0.7107	0.697	0.7453	0.7694	0.7346	0.7773	0.8432	0.8017	0.8167
n=25	0.8072	0.8231	0.8476	0.7598	0.7534	0.7916	0.7831	0.7433	0.785	0.8552	0.8119	0.8315
n=50	0.858	0.8764	0.8885	0.801	0.801	0.8252	0.7994	0.7673	0.8107	0.8531	0.8062	0.831
n=100	0.8884	0.8999	0.9047	0.8343	0.8428	0.8539	0.8228	0.8097	0.8411	0.8452	0.7935	0.8233
T=100												
n=10	0.736	0.7547	0.8042	0.7556	0.7282	0.7916	0.8035	0.7605	0.8077	0.8696	0.833	0.8441
n=25	0.8384	0.8586	0.8882	0.8065	0.7766	0.8386	0.8151	0.7562	0.8176	0.8823	0.8365	0.8569
n=50	0.8924	0.9131	0.9263	0.8453	0.838	0.8726	0.8354	0.7785	0.8425	0.8803	0.8323	0.8565
n=100	0.9239	0.9361	0.9408	0.8812	0.8833	0.8995	0.8608	0.8355	0.8772	0.8721	0.8096	0.8512
T=150												
n=10	0.7459	0.764	0.8172	0.7825	0.7456	0.8229	0.8243	0.7735	0.8272	0.8784	0.8419	0.8544
n=25	0.8528	0.8749	0.9052	0.8349	0.7955	0.8662	0.8354	0.7721	0.8365	0.8915	0.8436	0.866
n=50	0.9086	0.9294	0.9436	0.8713	0.8614	0.8989	0.8563	0.793	0.8637	0.8898	0.8296	0.8633
n=100	0.939	0.9521	0.9557	0.906	0.9076	0.9243	0.8821	0.8531	0.8953	0.8803	0.8066	0.8588

Table 34: Simulation results of model C2 (continued). Medians for near unit root setting and different cross-sectional dependences.

B Software

For empirical- and simulation analysis the statistical software $\mathbf{R}^{45}$ was used. An overview of the functionality of this statistical software in the field of multivariate time series analysis is given by the Time Series Task View, the Econometrics Task View and the Finance Task View. The "Views" can be found on the `R`-homepage.

The author wrote thousands lines of computational and user-interface code. For principal component analysis the function `prcomp`, for computing the Kalman smoother, the package `KFAS` and for the EM-algorithm, modified code from Shumway and Stoffer was used. The estimation of the AR-coefficients $(e(z), b)$ of the static factors, was either done by means of the function `ar` with the 'yw' option, or by code of the author in order to calculate the minimum norm solution of the singular Yule-Walker equations. A brief summary of different `R`-packages that implemented the Kalman Filter and -smoother is given by [51]. Shumway and Stoffer provide `R`-code on the homepage of their book [48]. The time-series data was generated with the help of the `dse` package. This package also supported the author in generating random non-diagonal stable polynomial matrices $e(z)$. Realizations from multivariate normal distributions have been generated with the help of the package `mvtnorm`. In order to compare results and debug the code respectively, the package `R.matlab` helped a lot. Tables and figures were produced with the packages `xtable` and `reshape2` as well as `ggplot2` respectively.

The simulations would not have been possible without the kind help of my friend Philip Dobner, who supported me with computational power.

C Theoretical backups

Remark C.1 (From rational spectral density to Wold representation). Let (x_t) be a wide sense stationary n-dimensional stochastic process defined on $\mathbb{Z}$ such that his spectral density $f(\theta)$ exists and $\mathrm{rk} f = q$ a.e. on Θ. Then according to Theorem 2.3

$$f(\theta) = \frac{1}{2\pi}\, w(e^{-i\theta})\, w^*(e^{-i\theta})$$

with $w(z)$ being rational and having neither zeros nor poles on $\{z \in \mathbb{C} : |z| \le 1\}$. Therefore $w(z)$ can be expanded for $|z| < r$, $r > 1$:

$$w(z) = \sum_{j=0}^{\infty} k_j\, z^j, \quad k_j \in \mathbb{R}^{n \times q}.$$

Every $n \times q$ rational matrix of rank q can be written as $w(z) = u(z)d(z)v(z)$, where u and v are unimodular and $d(z)$ has a diagonal form, with rational functions on the diagonal (see [32], Lemma 2.4.3). Define

$$w^- = v^{-1}(d'd)^{-1}d'u^{-1}$$

as a left-inverse of w. Then w^- is also causal as it has no poles and zeros in and on the complex unit circle. Define

---

$$\varepsilon_t := w^- x_t,$$

then ε_t is white noise with covariance matrix $\frac{1}{2\pi} I_q$. This can be shown easily with the transformation theorem for spectral densities. As w^- is causal $\mathcal{H}_x = \mathcal{H}_\varepsilon$ and because $f(\cdot)$ is integrable, $\sum_{j=0}^\infty ||k_j||^2 < \infty$.

All this leads to the conclusion, that $x_t = w(z)\varepsilon_t$ corresponds to a Wold decomposition (see Theorem 2.2). $\hfill \square$

Linear Algebra

Theorem C.2 (Spectral theorem for Hermitian matrices, [35]): Let $A \in \mathbb{C}^{n \times n}$. Then A is Hermitian if and only if there is a unitary matrix $U \in \mathbb{C}^{n \times n}$ and a real diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that $A = U\Lambda U^*$. Moreover, A is real symmetric if and only if there is an orthogonal matrix $O \in \mathbb{R}^{n \times n}$ and a real diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that $A = O\Lambda O'$.

Theorem C.3 (Rayleigh-Ritz, [35]): Let $A \in \mathbb{C}^{n \times n}$ be Hermitian, i.e. $A = A^*$, and let the eigenvalues of A be arranged in decreasing order, i.e. $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$. Then for all $x \in \mathbb{C}^n$

$$\lambda_n x^* x \leq x^* A x \leq \lambda_1 x^* x$$

$$\lambda_{\max} = \lambda_1 = \max_{x \neq 0} \frac{x^* A x}{x^* x} = \max_{x^* x = 1} x^* A x$$

$$\lambda_{\min} = \lambda_n = \min_{x \neq 0} \frac{x^* A x}{x^* x} = \min_{x^* x = 1} x^* A x$$

Theorem C.4 (Weyl, [35]): Let $A, B \in \mathbb{C}^{n \times n}$ and let the eigenvalues $\lambda_i(A), \lambda_i(B)$, and $\lambda_i(A + B)$ be arranged in decreasing order. For each $k = 1, 2, \ldots, n$ we have

$$\lambda_k(A) + \lambda_n(B) \leq \lambda_k(A + B) \leq \lambda_k(A) + \lambda_1(B)$$

Lemma C.5 Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric matrices. Then the convergence $||A - B|| \to 0$ implies the convergence of the eigenvalues of A to the eigenvalues of B.

Proof. As all norms are equivalent in euclidean spaces, we can choose $||A - B|| = \lambda_{\max}(A - B)$. It follows from Theorem C.4 that $\lambda_{\max}(A - B) \leq \lambda_k(A) - \lambda_k(B)$. Theorem C.3 reveals the relation $\lambda_i(-A) = -\lambda_{n-i+1}(A)$ and by using the left inequality of Weyl's Theorem, $\lambda_n(B - A) = -\lambda_1(A - B) \leq \lambda_k(B) - \lambda_k(A)$. We get for $k = 1, \ldots, n$ that $|\lambda_k(B) - \lambda_k(A)| \leq ||A - B|| \to 0$. $\hfill \square$

Lemma C.6 (see [18], Lemma 1) Let $\boldsymbol{x} = (x_t^n)_{n \in \mathbb{N}}$ be a double sequence of wide sense stationary processes x_t^n. For a given $n \in \mathbb{N}$, denote Σ_x^n as the covariance-matrix and $\Sigma_x^n(\theta)$, $\theta \in \Theta$ as the

spectral density of x_t^n. If Assumption (gdfm b) holds, then any eigenvalue of Σ_x^n belongs to $[2\pi m, 2\pi M]$, with

$$m = \inf\{\lambda_{\min}(\Sigma_x^n(\theta)), \ \theta \in \Theta \setminus N\}, \ M = \sup\{\lambda_{\max}(\Sigma_x^n(\theta)), \ \theta \in \Theta \setminus N\},$$

where N is a Lebesgue-nullset. The existence of the nullset is motivated in Remark 3.3.

Proof. See [18], page 197. $\hfill\square$

Lemma C.7 For matrices A, U, C, V of correct dimension the following holds:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{65}$$

This identity is also often called the *Matrix Inversion Lemma (MIL)*.

Definition C.8 (Kronecker Product):   Let $A = (a_{ij})$ and B be $(m \times n)$ and $(p \times k)$ matrices. Then

$$A \otimes B := \begin{pmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{pmatrix}$$

is a $(mp \times nk)$ matrix and called the *Kronecker product* or *direct product* of A and B.

Definition C.9 (Vec operator):   Define $\mathrm{vec}(A)$ of a matrix $A = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^{m \times n}$ as the stacked columns of matrix A:

$$\mathrm{vec}(A) := (a_1', a_2', \ldots, a_n')'$$

Lemma C.10 Let A, B, C, D be matrices of suitable dimensions. Then for the Kronecker product and the vec operator the following rules hold.

1. $(A \otimes B)' = (A' \otimes B')$

2. $A \otimes (B + C) = A \otimes B + A \otimes C$

3. $(A \otimes B)(C \otimes D) = AC \otimes BD$

4. If A and B are invertable, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

5. $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$

6. $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B) = (I \otimes AB)\mathrm{vec}(C) = (C'B' \otimes I)\mathrm{vec}(A)$

7. $\mathrm{vec}(AB) = (I \otimes A)\mathrm{vec}(B) = (B' \otimes I)\mathrm{vec}(A)$

8. If A and B are square matrices with eigenvalues λ_A, λ_B, respectively, and associated eigenvectors v_A, v_B, then $\lambda_A \lambda_B$ is an eigenvalue of $A \otimes B$ with eigenvector $v_A \otimes v_B$

9. If A and B are $m \times m$ and $n \times n$ square matrices, respectively, then $\det(A \otimes B) = \det(A)^n \det(B)^m$

Proof. Simple calculations using the definitions or see [36], Chapter 4 (eventually Notes and Further Readings, p. 287). $\qquad\square$

Lemma C.11 Consider the partitioned matrix $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where A, B, C, D are matrices of conformable dimensions. If M and A and D are non-singular, then with $S := (D - CA^{-1}B)$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

Probablity theory

Lemma C.12 Let $Z \sim \mathcal{N}(\mu, \Sigma)$ be an m dimensional random vector and $W = AZ + a$, where both $A \in \mathbb{R}^{m \times k}$ and $a \in \mathbb{R}^m$ are non-stochastic. If W has full row rank, then $W \sim \mathcal{N}(A\mu + a, A\Sigma A')$.

Definition C.13 (Multivariate Normal distribution): An n dimensional random vector x with mean μ and covariance matrix Σ is said to be normally distributed or distributed according to $\mathcal{N}(\mu, \Sigma)$, if there is an affine transformation $x = a + Ay$, $a \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times m}$, where y is an $m \leq n$ dimensional random vector, the number of rows of A is n, the number of columns is equal to the rank of Σ, and where y has a density

$$(2\pi)^{-m/2} \det(Q) \, e^{-\frac{1}{2}(y-\nu)'Q^{-1}(y-\nu)}$$

Theorem C.14. Let x be a n-dimensional random vector with mean μ and covariance matrix Σ which is partitioned according to $x = (x_1, x_2)$ where x_1 is n_1-dimensional, x_2 is n_2-dimensional and $n = n_1 + n_2$. Then $\mu = (\mu_1', \mu_2')'$ and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

If the distribution of x is normal, then the conditional distribution of x_1 given $x_2 = y$ is normal with mean $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2)$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof. See [3], Theorem 2.5.1, p. 35. $\qquad\square$

Definition C.15 (Order in probability of a sequence of random variables, [43]): Let Z_n be a sequence of random variables, and let c_n be a sequence of positive real numbers. We then say

Z_n is at most of order c_n in probability, and write $Z_n = O_p(c_n)$, if for every $\varepsilon > 0$ there exists a constant $M_\varepsilon < \infty$ such that $P(c_n^{-1}|Z_n| \geq M_\varepsilon) \leq \varepsilon$. We say that Z_n is of smaller order in probability than c_n, and write $Z_n = o_p(c_n)$, if $c_n^{-1}|Z_n| \xrightarrow{p} 0$ as $n \to \infty$. (The definition extends to vectors and matrices by applying the definition to each element or, equivalently, to the norm.)

References

[1] D. Amengual and M. W. Watson. Consistent Estimation of the Number of Dynamic Factors in a Large N And T Panel. *Journal of Business and Economic Statistics*, 25(1), 2007.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., New York, 3rd edition, 2003.

[4] J. Bai. Inferential Theory for Factor Models of large dimensions. *Econometrica*, 71:135–171, 2003.

[5] J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

[6] J. Bai and S. Ng. Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25:52–60, 2007.

[7] E. R. Beadle and P. M. Djurić. Uniform Random Parameter Generation of Stable Minimum-Phase Real ARMA (p,q) Processes. *IEEE Signal Processing Letters*, 4(9), 1997.

[8] B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics*, 120:387–422, 2005.

[9] J. Breitung and U. Pigorsch. A canonical correlation approach for selecting the number of dynamic factors. November 2010.

[10] D. R. Brillinger. *Time Series: Data Analysis and Theory*. Holden-Day, 1981.

[11] G. Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):1305–1323, September 1983.

[12] G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

[13] A. D'Agostino and D. Giannone. Comparing Alternative Predictors Based on Large-Panel Factor Models. *Oxford Bulletin of Economics and Statistics*, 74:306–326, 2012.

[14] M. Deistler, B. Anderson, A. Filler, C. Zinner, and W. Chen. Generalized linear dynamic factor models - an approach via singular autoregressions. *European Journal of Control*, 3:1–15, 2010.

[15] M. Deistler and C. Zinner. Modelling high-dimensional time series by Generalized Linear Dynamic Factor Models: An introductory survey. *Communications in Information and Systems*, 7(2):153–166, 2007.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[17] C. Doz, D. Giannone, and L. Reichlin. A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. ECARES working paper 2008_034, September 2006.

[18] C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164:188–205, 2011.

[19] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press Inc., New York, 2001.

[20] A. Filler. *Verallgemeinerte dynamische Faktormodelle, Darstellung und Vergleich zweier Schätzmethoden*. PhD thesis, Vienna University of Technology, 2007.

[21] A. Filler. *Generalized Dynamic Factor Models, Structure Theory and Estimation for Single Frequency and Mixed Frequency Data*. PhD thesis, Vienna University of Technology, 2010.

[22] P. Filzmoser and K. Hron. Correlation analysis for compositional data. *Mathematical Geosciences*, 41:905–919, 2009.

[23] M. Forni and L. Gambetti. The dynamic effects of monetary policy: A structural factor model approach. November 2008.

[24] M. Forni and L. Gambetti. Macroeconomic Shocks and the Business Cycle: Evidence from a Structural Factor Model. February 2010.

[25] M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the Black Box: Structural Factor Models with large cross sections. *Econometric Theory*, 25:1319–1347, 2009.

[26] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, 2000.

[27] M. Forni and M. Lippi. The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17(6):1113–1141, 2001.

[28] B. Funovits. Analyticity properties of the decomposition of the spectral density matrix in the context of dynamic PCA. Master's thesis, Vienna University of Technology, 2009.

[29] A. Graef, C. Flamm, M. Hartmann, and M. Deistler. Spectral Estimation in Time-Discrete Digital Signal Processing. Vienna University of Technology, October 2011.

[30] W. H. Greene. *Econometric Analysis*. Prentice-Hall, 2003.

[31] M. Hallin and R. Liška. Determining the Number of Factors in the General Dynamic Factor Model. *Journal of the American Statistical Association*, 102:603–617, 2007.

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

[33] H. Havlicek. *Lineare Algebra für Technische Mathematiker*. Heldermann Verlag, 2006.

[34] M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods*. John Wiley & Sons, 1973.

[35] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

[36] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1999.

[37] T. Hothorn, K. Hornik, M. A. van de Wiel, and A. Zeileis. Implementing a Class of Permutation Tests: The coin Package. *Journal of Statistical Software*, 28(8):1–23, 11 2008.

[38] I. T. Jolliffe. *Principal Component Analysis*. Spinger, 1986.

[39] D. N. Lawley and A. E. Maxwell. *Factor analysis as a statistical method*. Butterworths, 1971.

[40] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, 2007.

[41] M. M. McConnell and G. Perez-Quiros. Output Fluctuations in the United States: What Has Changed Since the Early 1980's? *The American Economic Review*, 90(5):1464–1476, 2000.

[42] R. J. Meinhold and N. D. Singpurwalla. Understanding the kalman filter. *The American Statistician*, 37:123–127, 1983.

[43] B. M. Pötscher and I. R. Prucha. Basic Elements of Asymptotic Theory. Electronic Working Papers 99-001, University of Maryland, Department of Economics, Jan. 1999.

[44] J. A. Rozanov. *Stationary random processes*. Holden-Day, 1967.

[45] G. Rünstler, K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze. Short-term forecasting of gdp using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting*, 28(7):595–611, 2009.

[46] W. Scherrer and M. Deistler. A structure theory for linear dynamic errors-in-variables models. *SIAM J. Control Optim.*, 36(6):2148–2175, 1998.

[47] P. H. Schönemann. On the Formal Differentiation of Traces and Determinants. *Multivariate Behavioral Research*, 20:113–139, 1985.

[48] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications, With R Examples*. Springer, 3rd edition, 2011.

[49] J. H. Stock and M. W. Watson. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, Dec. 2002.

[50] J. H. Stock and M. W. Watson. Forecasting in Dynamic Factor Models Subject to Structural Instability. In J. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press, 2008.

[51] F. Tusell. Kalman Filtering in R. *Journal of Statistical Software*, 39, 2011.

[52] M. W. Watson and R. F. Engle. Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics*, 23:385–400, 1983.

D   Deutschsprachige Zusammenfassung

Verallgemeinerte lineare dynamische Faktormodelle (GDFM's) dienen der Analyse hoch-dimensionaler Zeitreihen, wo die Anzahl der einzelnen Zeitreihen relativ groß im Verhältnis zur Stichprobengröße ist. Sie wurden in Forni et al. (2000), Forni und Lippi (2001) sowie Stock und Watson (2002a) eingeführt, und verallgemeinern lineare dynamische Faktormodelle mit strikt idiosynkratischen Fehlern (Geweke (1977)) bzw. kombinieren diese mit verallgemeinerten statischen Faktormodellen (Chamberlain (1983), Chamberlain und Rothschild (1983)). Erfolgreiche Anwendungsgebiete dieser Modelle sind die Prognose sowie die Strukturanalyse von länderübergreifenden disaggregierten Finanz- und makroökonomischen Daten.

GDFM's erlauben zwar eine schwache Abhängigkeit in den Fehlern, sind dadurch jedoch nur noch asymptotisch identifizierbar. Diese Diplomarbeit betrachtet drei Schätzmethoden für den nicht beobachtbaren statischen Faktorprozess, die auf vereinfachten identifizierbaren Modellen beruhen. Das Ziel dieser Arbeit ist es in einer Simulationsstudie die Rolle der Missspezifikation bei Anwendung der Schätzer im verallgemeinerten Kontext herauszuarbeiten, sowie herauszufinden, unter welchen Umständen eine Schätzmethode die andere dominiert. Der Vergleich im Rahmen einer Simulation ist durch fehlende analytische Aussagen gerechtfertigt. Zwar wurde in Stock und Watson (2002a) und Doz, Giannone und Reichlin (2008, 2011) gezeigt, dass alle drei Schätzer den Raum der statischen Faktoren konsistent schätzen, doch gibt es meines Wissens nur ein Paper von Bai (2003), welches die asymptotische Verteilung eines der drei Schätzer betrachtet.

Die Arbeit ist in vier Teile gegliedert. Im ersten werden GDFM's definiert und mit Hilfe von Ergebnissen von Deistler et al. (2010) das Basismodell der Simulation hergeleitet. Der zweite Teil definiert die drei Schätzer und betrachtet deren Konsistent und Berechnung.

Der principal component estimator (PC-Schätzer) entsteht aus der Spektralzerlegung der Kovarianzmatrix der Beobachtungen. Er schätzt die Faktoren durch die Berechnung der ersten Hauptkomponenten des Beobachtungsprozesses. Der two-stage estimator (TS-Schätzer) berechnet jene Linearkombination aller Beobachtungen, die den quadratischen Abstand zu den statischen Faktoren komponentenweise minimieren. Er entspricht also der komponentenweise orthogonalen Projekten der statischen Faktoren auf den linearen Raum, der von den Komponenten aller Beobachtungen aufgespannt wird. Berechnet wird dieser durch die Aufstellung eines Zustandsraum-Modells und die Anwendung der Kalman Filter Rekursionen. Die Parameter werden durch die PC-Schätzer sowie die Lösung der Yule-Walker Gleichungen bestimmt. Der quasi-maximum likelihood estimator (QML-Schätzer) entspricht ebenfalls der orthogonalen Projektion der statischen Faktoren auf den Raum aller Beobachtungen, jedoch unter anderen geschätzten Parametern für das Zustandsraum-Modell.

Die Beschreibung der Simulationsstudie, sowie deren Resultate, bilden den dritten Teil meiner Arbeit. Es wird gezeigt, dass im Falle von white-noise Fehlern mit diagonaler Kovarianzmatrix, der zu erwartende relative Vorteil der TS- und QML-Schätzer gegenüber dem PC-Schätzer eintritt. Kommen die Daten aus einem verallgemeinerten dynamischen Faktormodell, welches durch Missspezifikation des Fehlerspektrums gekennzeichnet ist, so geht dieser Vorteil zurück bzw. verloren, und bei hoher lokaler Abhängigkeit zwischen den idiosynkratischen Komponenten kann er sich sogar in einen Nachteil verwandeln. Ein langes Gedächtnis der statischen Faktoren hat ebenso Einfluss auf die absolute wie die relative Schätzgenauigkeit, wie das Ausmaß des Rausch-Anteils am latenten Signal.

Der vierte und letzte Teil betrachtet 107 makroökonomische Zeitreihen der US-Wirtschaft für den Zeitraum von 1960 bis 2006. Es werden Methoden zur Schätzung der Dimension der dy-

namischen und statischen Faktoren angewendet, sowie die Parameter des Faktormodells aus den Daten geschätzt. Diese dienen als Grundlage für weitere Simulationsmodelle. In meiner Arbeit stellte sich heraus, dass durch eine falsche Wahl der Dimensionen (der statischen Faktoren) der EM-Algorithmus zur Berechnung der QML-Methode nicht mehr funktionierte.

Die Simulationsstudie ist ausbaufähig. Auf die Analyse des Einflusses der Schätzung der Dimension der dynamischen und statischen Faktoren auf die Schätzgenauigkeit (Einfluss von Modellselektion) wurde ebenso verzichtet, wie auf eine Missspezifikation der Dynamik der statischen Faktoren. Der Mehrwert dieser Arbeit liegt vor allem in dem Versuch die Schätzmethoden für einen Großteil der Parameter des Modells zu vergleichen. Weitere Aufgaben für die Zukunft wären zum Beispiel die Erstellung eines Algorithmus, um gleichverteilt aus dem Stabilitäts-bereich eines multivariaten Lag-Polynoms des AR-Prozesses der statischen Faktoren ziehen zu können.

E Curriculum Vitae

<div align="right">

Alexander Braumann
Curriculum Vitae

</div>

<u>Home Address</u>	Rudolfsplatz 2/32 A-1010 Vienna Austria		Date of birth	Feb 9, 1984
			Place of birth	Vienna, Austria
Telephone	+43 699 195 295 59		Citizenship	Austrian
Email	alex@nagual.at		Sex	Male

<u>Languages</u>	German (Native)	English (Fluent)	Spanish (Advanced)

EDUCATION

Mag.soc.oec. der Volkswirtschaftslehre, *University of Vienna*, 10/2003–09/2012
  Equivalent to: M.Sc. in Economics
  Specialisation: Econometrics

B.Sc./M.Sc. for Statistics and Mathematics for Econ., *Vienna University of Technology*, since 10/2006
  Thesis (B.Sc.): On the construction of Brownian Motion
  Advisor: Wolfgang Wertz
  (B.Sc. Expected Winter 2012, M.Sc. Expected Summer 2013)

Erasmus-Student at the *University of Salamanca*, Department of Economics, 10/2005 – 07/2006

HTL Donaustadt, Department of Informatics (EDVO), 1998–2003

Gymnasium Billrothstrasse 26-30, Vienna, 1995–1998

THESIS

Comparing three estimation methods in the context of generalized dynamic factor models; Supervisor: Em.O.Univ.Prof. Dipl-Ing. Dr. techn. Manfred Deistler

---

FIELDS OF INTEREST

Primary Field	**Econometrics** – Generalized Dynamic Factor Models, Model Selection
Secondary Field	International Macroeconomics, Stochastics

RELEVANT EMPLOYMENT

2007–2012 Volunteer at the Austrian Service Abroad – Österreichischer Auslandsdienst, Department of Development Cooperation (www.auslandsdienst.at).

2006–2012 Technical and Consulting Assistant, Brainbows Informationsmanagement Gmbh (www.brainbows.com). Assisting in consulting projects in the field of renewable energy and energy policies.

2004–2005 Non-executive director of the NGO Austrian Service Abroad – Österreichischer Auslandsdienst.
Volunteer since 2003.

1999–2003 Internships at Siemens, Msedv and Brainbows Informationsmanagement Gmbh where I participated in programming projects and gained experience in project development.

SPECIAL KNOWLEDGE AND PERSONAL INTERESTS

Computer	MAC/PC/Unix, Microsoft Office, LaTeX, HTML, Programming (C, C++, PHP), Database (MySQL, Oracle), Matlab, Maple
Statistics	R (www.r-project.org), eViews
Sports	Tennis, Rowing, Paragliding, Badminton, Swimming, Running, Cross-Country Skiing

Vienna, September 12, 2012