Common factors in a panel with two cross-sectional dimensions

Håvard Hungnes
Statistics Norway
POB 8131 Dep.
NO-0033 Oslo

Email: hhu@ssb.no

Tel.: +47 21 09 48 18

June 20, 2017

Abstract

The paper suggest how to extend the common correlations effect (CCE) introduced by Pesaran (2015, Econometrica) to a panel with two cross-sectional dimensions. It also shows how to use the CCE technique when there is cross-sectional dependence in only one of the cross-sectional dimensions. A Monte Carlo simulation indicates that the CCE-estimator also has good properties when the variables follow non-stationary processes. The approach is illustrated by estimating the elasticity of substitution in a panel with input factors along one dimension and industries along the other dimension.

JEL codes: C33, E23.

Keywords: common correlated effects, cross-sectional averages

Acknowledgments: Thanks to Neil Ericsson, Bruce Hansen and Terje Skjerpen for valuable comments.

1 Introduction

An increasing number of analyses take into account cross-sectional dependence in analyzing panels. Pesaran (2015, p. 750) states that the presence of some cross-sectional correlation of errors in panel data applications in econometrics is likely to be the rule rather than the exception. Ignoring cross-sectional dependence can lead to both misleading inference and inconsistent estimators, see Pesaran (2015, p. 750).

To account for interdependence between observational units common factors are often used. Two approaches are applied in the presence of unobserved common factors: the principal component (PC) approach proposed by Coakley et al. (2005) and refined by Bai (2009); and the common correlated effects (CCE) approach presented in Pesaran (2006) and shown to also apply to non-stationary variables in Kapetanios et al. (2011). The principal component approach in Coakley et al. (2005) assumes that there is no correlation between the common factors and the other regressors. Bai (2009) suggests an extension with an iterative method and shows that the corresponding estimator is consistent even if the common factors are correlated with the regressors. The CCE approach in Pesaran (2006) implies a consistent estimator in the presence of a correlation between the common factors and the regressors, without applying an iterative method. Furthermore, Urbain and Westerlund (2011) show that the cross-sectional averages approach generally yields more precise estimates of the effect of the exogenous variables than the principal component approach (see also Pesaran, 2015, p. 711).

Both the PC and the CCE approach is based one cross-sectional dimension. However, some data sets might have two cross-sectional dimensions. If one of the cross-sectional dimensions is small, the framework in Pesaran et al. (2004) — also denoted the GVAR (global model vector-autoregressive) model — can be applied. The interdependence in the small cross-sectional dimension can then be taken into account directly by analysing this cross-sectional dimension as a VAR model. The cross-sectional dependence in the other dimension can be approximated by using cross-sectional averages across this dimension. However, since this approach involves estimating the full covariance structure of the smallest dimension, it entails estimating many parameters if both cross-sectional dimensions are large. Chudik and Pesaran (2016) and Pesaran (2015) indicate that, when applying the Global VAR, the smaller of the two cross-sectional dimensions is typically in the range of four to six variables.

This paper presents a generalization of the CCE approach to two cross-sectional dimensions. One example of a data set with two cross-sectional dimensions is a data set of factor demand, where different input factors are measured in one dimension and different industries in the other dimension. In addition, the data set has a time dimension. Here, one can have some common factors that are industry-specific and others that are input-type-specific. The present paper illustrates the use of common factors with two cross-sectional dimensions by considering the demand of different input factors over all industries in Norway.

Bilateral trade is another example of a data set with two cross-sectional dimensions. Here, the first cross-sectional dimension corresponds to the exporting country and the second cross-sectional dimension is the importing country.

The rest of the paper is organised as follows. In Section 2, I present a common factor model with two cross-sectional dimensions and demonstrate that they can be approximated by cross-sectional averages in both of these dimensions. In Section 3 a Monte Carlo simulation is applied and it is found that the suggested CCE approach also works well for non-stationary time series. In Section 4 the approach is illustrated by considering the elasticity of substitution between different input factors in all non-government industries in mainland Norway (i.e., excluding Oil and gas extraction, and Ocean transport). Section 5 concludes.

2 Common factors

In this section I present a heterogeneous model with two cross-sectional dimensions and with common factors.

$$Y_{ijt} = \alpha'_{ij} D_t + \beta'_{ij} X_{ijt} + E^f_{ijt}, \quad \begin{array}{c} i = 1, \dots, N^A, \\ j = 1, \dots, N^B, \\ t = 1, \dots, T. \end{array} \tag{1}$$

Here, Y_{ijt} is the observation of the endogenous variable for unit i in the first cross-sectional dimension and unit j in the second cross-sectional dimension at time t. For example, the first cross-sectional dimension can be country and the second cross-sectional dimension can be industry. Here, however, I will refer to the first cross-sectional dimension as industry and the second cross-sectional dimension as input factors. Hence, for each time period t, we have observations of the endogenous variable for different input factors in different industries.

The vector D_t contains n deterministic variables such as an intercept and a trend. In addition, it can contain macro variables that are equal across both cross-sectional dimensions. The oil price could be an example of such a variable.

The vector X_{ijt} contains k variables that we assume differ in at least one of the cross-sectional dimensions. In most of the presentation, I will assume that all observations in X_{ijt} are unique in both cross-sectional dimensions, since this will simplify the presentation. However, it will be convenient to partition this vector as $X'_{ijt} = \left(x^{A\prime}_{it}, x^{B\prime}_{jt}, x'_{ijt}\right)$, where x^A_{it} is a vector of the k_1 variables that only differs in the first dimension, x^B_{jt} is a vector of k_2 variables that only differs in the second dimension, and x_{ijt} is a vector of the k_3 variables that differs in both dimensions; $k = k_1 + k_2 + k_3$. The coefficient vector β_{ij} is partitioned similarly; $\beta'_{ij} = \left(\beta^{A\prime}_{ij}, \beta^{B\prime}_{ij}, \beta^{C\prime}_{ij}\right)$. When including both industry-specific and input factor-specific common factors, β^A_{ij} and β^B_{ij} are not identifiable because the effect from the exogenous variables x^A_{it} and x^B_{jt} cannot be distinguished from the common factors. Hence, only β^C_{ij} can be identified.

The errors in (1) includes the common factors (hence, the top script "f"). The most general formulations of these common factors is

$$E^f_{ijt} = \gamma'_{ij} f_t + \gamma^{A\prime}_j f^A_{it} + \gamma^{B\prime}_i f^B_{jt} + E_{ijt},$$

where γ_{ij}, γ^A_j and γ^B_i — which are vectors of dimension m_0, m_1 and m_2, respectively — are the coefficient vectors for how the common factors affect the endogenous variable.

The exogenous variables follow the process

$$X_{ijt} = A'_{ij} D_t + V^f_{ijt}, \tag{2}$$

with the most general formulation of the error structure given by

$$V^*_{ijt} = \Gamma'_{ij} f_t + \Gamma^{A\prime}_j f^A_{it} + \Gamma^{B\prime}_i f^B_{jt} + V_{ijt}$$

Here, Γ_{ij}, Γ^A_j and Γ^B_i — which are matrices of dimension $m_0 \times k$, $m_1 \times k$ and $m_2 \times k$, respectively

— are the coefficient matrices for how the common factors affect the exogenous variables.

Combining equations (1) and (2) yields

$$Z_{ijt} = \begin{pmatrix} Y_{ijt} \\ X_{ijt} \end{pmatrix} = B'_{ij} D_t + U^f_{ijt} \tag{3}$$

where

$$U^f_{ijt} = \begin{pmatrix} E^f_{ijt} + \beta'_{ij} V^f_{ijt} \\ V^f_{ijt} \end{pmatrix} \text{ and } B_{ij} = \begin{pmatrix} \alpha_{ij} & A_{ij} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_{ij} & I_k \end{pmatrix}.$$

Hence, the combined multi-factor structure in its most general form is given by

$$U^f_{ijt} = C'_{ij} f_t + C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + U_{ijt}, \tag{4}$$

where

$$\begin{aligned} C_{ij} &= \begin{pmatrix} \gamma_{ij} & \Gamma_{ij} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_{ij} & I_k \end{pmatrix} \\ C^A_{ij} &= \begin{pmatrix} \gamma^A_j & \Gamma^A_j \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_{ij} & I_k \end{pmatrix}, \text{ and} \\ C^B_{ij} &= \begin{pmatrix} \gamma^B_i & \Gamma^B_i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_{ij} & I_k \end{pmatrix}. \end{aligned}$$

The system in (3) and (4) implies that the exogenous variables are allowed to be correlated with the common factors. Now we consider 5 different formulations of the multi-factor structure:

- **Alternative 0:** $U^f_{ijt} = U_{ijt}$, i.e. $C_{ij} = C^A_{ij} = C^B_{ij} = 0$. In this alternative, no common factors are included.

- **Alternative I:** $U^f_{ijt} = C'_{ij} f_t + U_{ijt}$, i.e. $C^A_{ij} = C^B_{ij} = 0$. This multi-factor structure is similar to the one considered in Pesaran (2006). This formulation of multi-factor structure implies that we do not consider the two cross-sectional dimensions explicitly. Hence, we could have stacked the two cross-sectional dimensions into one cross-sectional dimension.

- **Alternative II:** $U^f_{ijt} = C^{A\prime}_{ij} f^A_{it} + U_{ijt}$, i.e. $C_{ij} = C^B_{ij} = 0$. This multi-factor structure is also similar to the one considered in Pesaran (2006) when each of the N^A cross-sectional data sets is considered separately.

- **Alternative III:** $U^f_{ijt} = C'_{ij} f_t + C^{A\prime}_{ij} f^A_{it} + U_{ijt}$, i.e. $C^B_{ij} = 0$. This multi-factor structure implies that we combine overall common factors (f_t) with common factors that differ across the first cross-sectional dimension.[1] This multi-factor structure is a combination of I and II.

- **Alternative IV:** $U^f_{ijt} = C'_{ij} f_t + C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + U_{ijt}$, i.e. no additional restrictions. This multi-factor structure implies that common factors that are specific to both of the two cross-sectional dimensions are included. These are included in addition to some overall common factors.

Note that the multi-factor structure $U^*_{ijt} = C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + U_{ijt}$ (i.e., the multi-factor structure IV with $C'_{ij} = 0$) is not included above. It turns out that proxies for the common factors should be the same as in the case of multi-structure IV (although it can be simplified if both of the cross-sectional dimensions are large). This multi-factor structure is therefore not considered separately.

Under some assumptions, which are set out below, observable proxies can be derived for the common factors. These proxies are constructed as weighted averages of the observable variables. Let w^A_j define the weights in the first cross-sectional dimension (here, of input factors within an industry) with $\sum_j w^A_j = 1$; and let w^B_i define the weights in the second cross-sectional dimension (here, of an input factor across industries) with $\sum_i w^B_i = 1$. Additional conditions for these weights are given in Assumption 2.5.

Summary 2.1 *Observable proxies for the common factors with the various multi-factor structures are*

[1] For example, these could be country-specific factors. This choice of multi-factor structure could be appropriate if the first cross-time dimension is countries and the second is individuals or firms. Then we would not expect there to be a particular common factor between individual (or firm) j in countries 1 and 2.

given as:

- *With multi-factor structure I — i.e.,* $U^f_{ijt} = C'_{ij} f_t + U_{ijt}$ *— the vector of observable variables* $\left(D'_t, \overline{Z}'_t\right)'$ *with* $\overline{Z}_t = \sum_{j=1}^{N^B} w_j^A \sum_{i=1}^{N^A} w_i^B Z_{ijt}$ *can be used as a proxy for the common factors. This implies that* $k+1$ *additional regressors are included in the regression to approximate for the common factors for each cross-combination of the two cross-sectional dimensions, and — therefore —* $N^A N^B(k+1)$ *additional regressors in total.*

- *With multi-factor structure II —* $U^f_{ijt} = C^{A\prime}_{ij} f^A_{it} + U_{ijt}$ *— the vector of observable variables* $\left(D'_t, \overline{Z}'_{i.t}\right)'$ *with* $\overline{Z}_{i.t} = \sum_{j=1}^{N^B} w_j^A Z_{ijt}$ *can be used as a proxy for the common factors. This result follows from Pesaran (2006). Note, however, that* $\sum_{j=1}^{N^B} w_j^A x^A_{it} = x^A_{it}$*, so these* k_1 *cross-sectional means are already included in the regressions. Hence, this implies that we are only including* $N^A N^B(k_2+k_3+1)$ *additional variables in the regressions to proxy for the common factors.*

- *With multi-factor structure III —* $U^f_{ijt} = C'_{ij} f_t + C^{A\prime}_{ij} f^A_{it} + U_{ijt}$ *— the vector of observable variables* $\left(D'_t, \overline{Z}'_t, \overline{Z}_{i.t}\right)'$*, with* $\overline{Z}_t$ *and* $\overline{Z}_{i.t}$ *defined above, can be used as a proxy for the common factors. This result is shown below. Note that* $\sum_{j=1}^{N^B} w_j^A \sum_{i=1}^{N^A} w_i^B x^B_{jt} = \sum_{j=1}^{N^B} w_j^A x^B_{jt}$*, where the* k_2 *averages on the left-hand side are included in* $\overline{Z}_t$ *and the* k_2 *averages on the right-hand side are included in* $\overline{Z}_{i.t}$*. In addition to the fact that* $\sum_{j=1}^{N^B} w_j^A x^A_{it} = x^A_{it}$ *(see the bullet point above), this implies that* $N^A N^B(k+k_3+1)$ *additional averages are included here to serve as proxies for the common factors.*

- *With multi-factor structure IV —* $U^f_{ijt} = C'_{ij} f_t + C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + U_{ijt}$ *— the vector of observable variables* $\left(D'_t, \overline{Z}'_t, \overline{Z}_{i.t}, \overline{Z}_{.jt}\right)'$*, with* $\overline{Z}_{.jt} = \sum_{i=1}^{N^A} w^B Z_{ijt}$ *and with* $\overline{Z}_t$ *and* $\overline{Z}_{i.t}$ *defined above, can be used as a proxy for the common factors. This result is shown below. This implies that* $N^A N^B(k+2k_3+1)$ *additional averages are included to serve as proxies for the common factors. The same proxies for the common factors can be used with the multi-factor structure* $U^f_{ijt} = C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + U_{ijt}$ *(i.e., when* $\gamma_{ij} = 0$ *and* $\Gamma_{ij} = 0$*).*

Remark 2.0.1 *Note that* D_t *is a part of the proxies for the common factor. This implies that, when including the proxies for the common factors, we cannot distinguish between the direct effect of the variables in* D_t *and the effect through the proxies, see Pesaran (2006). A similar argument implies that we cannot identify the direct effect of* x^A_{it} *(when* $\overline{Z}_{i.t}$ *is used as part of the proxies) and* x^B_{it} *(when* $\overline{Z}_{.jt}$ *is used as part of the proxies).*

2.1 Deriving the proxies for the common factors

In this section, I consider the most general formulation of the multi-factor structure and derive the proxies from this formulation. Based on the expressions for the proxies, we can see how they change when one considers simpler forms of the multi-factor error structure.

Combining equation (3) with multi-factor error structure IV in (4) yields

$$Z_{ijt} = \begin{pmatrix} Y_{ijt} \\ X_{ijt} \end{pmatrix} = B'_{ij} D_t + C^{A\prime}_{ij} f^A_{it} + C^{B\prime}_{ij} f^B_{jt} + C'_{ij} f_t + U_{ijt}. \tag{5}$$

Pesaran (2006) presents five assumptions for his formulation of the heterogeneous panel with multi-factor error structure. These assumptions are summarised below and extended in the present model by two cross-section dimensions. In this section $\|A\| = \left(tr(AA')\right)^{1/2}$ denotes the Euclidean norm of the matrix A; A^- denotes a generalized inverse of A; and $\xrightarrow{p}$ denotes convergence in probability.

Assumption 2.1 ***Common effects:*** *$\left(D'_t, f'_t, f^{A\prime}_{it}, f^{B\prime}_{jt}\right)'$ is covariance stationary with absolute summable auto-covariances and distributed independently of the errors $E_{ijt'}$ and $V_{ijt'}$ for all i, j, t and t'.*

Assumption 2.2 ***Errors:*** *The errors E_{ijt} and $V_{ijt'}$ are distributed independently for all i, j, t and t'. For each i and j, E_{ijt} and $V_{ijt'}$ follows linear stationary processes with absolute summable autocovariances, $E_{ijt} = \sum_{\ell=0}^{\infty} a_{ij\ell}\zeta_{ij,t-\ell}$ and $V_{ijt} = \sum_{\ell=0}^{\infty} S_{ij\ell}\nu_{ij,t-\ell}$, where $\left(\zeta'_{ijt}, \nu'_{ijt}\right)'$ are $(k+1) \times 1$ vectors of identically, independently distributed random variables with zero mean, covariance matrix, I_{k+1}, and finite fourth order cumulations. In particular, $Var(E_{ijt}) = \sum_{\ell=0}^{\infty} a^2_{ij\ell} = \sigma^2_{ij} \leq \overline{\sigma}^2 < \infty$, and $Var(V_{ijt}) = \sum_{\ell=0}^{\infty} S_{ij\ell}S'_{ij\ell} = \Sigma^2_{ij} \leq \overline{\Sigma}^2 < \infty$ for all i and j and some constants $\overline{\sigma}^2$ and $\overline{\Sigma}$, where Σ_{ij} is a positive definite matrix.*

Assumption 2.3 ***Factor-loadings:*** *The unobserved factor loadings are independently and identically distributed as*

$$\begin{aligned}
\gamma_{ij} &= \gamma + \eta^0_{ij}, & \eta^0_{ij} &\sim IID\left(0, \Omega_{\eta^0}\right) & &\text{for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B, \\
\gamma^A_j &= \gamma^A + \eta^A_j, & \eta^A_j &\sim IID\left(0, \Omega_{\eta^A}\right) & &\text{for } j = 1, \ldots, N^B, \\
\gamma^B_i &= \gamma^B + \eta^B_i, & \eta^B_i &\sim IID\left(0, \Omega_{\eta^B}\right) & &\text{for } i = 1, \ldots, N^A,
\end{aligned}$$

where Ω_η is an $m_0 \times m_0$ symmetric non-negative definite matrix; Ω_{η^A} is an $m_1 \times m_1$ symmetric non-negative definite matrix; and Ω_{η^B} is an $m_2 \times m_2$ symmetric non-negative definite matrix. The vectors η_{ij}^0, η_j^A, η_i^B are distributed independently of each other and independently of the errors E_{ijt} and V_{ijt} and the common factors $\left(D_t', f_t', f_{it}^{A\prime}, f_{jt}^{B\prime}\right)'$ for all i, j, t. Furthermore, $\|\gamma\| < K$, $\|\gamma^A\| < K$, $\|\gamma^B\| < K$, $\|\Omega_{\eta^0}\| < K$, $\|\Omega_{\eta^A}\| < K$, and $\|\Omega_{\eta^B}\| < K$ for some positive constant $K < \infty$. Similarly, vec (Γ_{ij}), vec $\left(\Gamma_j^A\right)$ and vec (Γ_i^B) (with dimension km_0, km_1 and km_2, respectively) are also independently and identically distributed with the same properties as γ_{ij}, γ_j^A and γ_i^B.

Assumption 2.4 ***Random slope coefficients:*** *The slope coefficients β_{ij} follow the random coefficient model*

$$\beta_{ij} = \beta + v_{ij}^0, v_{ij}^0 \sim IID\left(0, \Omega_{v^0}\right) \text{ for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B,$$

where Ω_{v^0} is a $k \times k$ symmetric non-negative definite matrix, and the random deviations v_{ij}^0 are distributed independently of $\gamma_{ij}, \gamma_j^A, \gamma_i^B, \Gamma_{ij}, \Gamma_j^A, \Gamma_i^B, E_{ijt}, V_{ijt}$, and $\left(D_t', f_t', f_{it}^{A\prime}, f_{jt}^{B\prime}\right)'$ for all i, j, t. Finally, $\|\beta\| < K$, $\|\Omega_{v^0}\| < K$, $\|\Omega_{v^1}\| < K$, and $\|\Omega_{v^2}\| < K$ for some positive constant $K < \infty$.

Remark 2.1.1 *The assumptions for the distribution of γ_{ij} and β_{ij} above imply that $\frac{1}{N^B}\sum_j \beta_{ij} - \frac{1}{N^B}\sum_j \beta_{i'j} \xrightarrow{p} 0$ for $i' \neq i$, i.e., the mean of the β-vector will converge to the same vector for all industries i. These assumptions could be refined such that*

$$\gamma_{ij} = \gamma + \eta_{ij}^0 + \eta_i^1 + \eta_j^2, \quad \begin{cases} \eta_{ij}^0 \sim IID\left(0, \Omega_{\eta^0}\right), \\ \eta_i^1 \sim IID\left(0, \Omega_{\eta^1}\right), \\ \eta_j^2 \sim IID\left(0, \Omega_{\eta^2}\right), \end{cases} \quad \text{for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B,$$

and

$$\beta_{ij} = \beta + v_{ij}^0 + v_i^1 + v_j^2, \quad \begin{cases} v_{ij}^0 \sim IID\left(0, \Omega_{v^0}\right), \\ v_i^1 \sim IID\left(0, \Omega_{v^1}\right), \\ v_j^2 \sim IID\left(0, \Omega_{v^2}\right), \end{cases} \text{for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B.$$

The derived proxies for the common factors will be the same with these more general assumptions, as is shown in Hungnes (2016, Appendix C).

Assumption 2.5 ***Identification of*** β_{ij} ***and*** β***:*** *The weights used to generate cross-sectional averages in the two cross-sectional dimensions satisfy the conditions*

$$w_j^A = O\left(\frac{1}{N^B}\right), \quad \sum_{j=1}^{N^B} w_j^A = 1, \quad \sum_{j=1}^{N^B} |w_j^A| < K,$$
$$w_i^B = O\left(\frac{1}{N^A}\right), \quad \sum_{i=1}^{N^A} w_i^B = 1, \quad \sum_{i=1}^{N^A} |w_i^B| < K,$$

Let

$$M_{wij} = I_T - H_{wij}\left(H'_{wij}H_{wij}\right)^{-} H'_{wij}, \text{ and} \tag{6}$$
$$M_{gij} = I_T - G_{ij}\left(G'_{ij}G_{ij}\right)^{-} G'_{ij}, \tag{7}$$

where $H_{wij} = \begin{pmatrix} D & \overline{Z} & \overline{Z}_{i.} & \overline{Z}_{.j} \end{pmatrix}$, $G_{ij} = \begin{pmatrix} D & F & F_{.i} & F_{i.} \end{pmatrix}$,

$$D = \begin{pmatrix} D_1 & D_2 & \dots & D_T \end{pmatrix}' \tag{8}$$

$$\overline{Z} = \begin{pmatrix} \overline{Z}_0 & \overline{Z}_1 & \dots & \overline{Z}_T \end{pmatrix}', \overline{Z}_{i.} = \begin{pmatrix} \overline{Z}_{i.0} & \overline{Z}_{i.1} & \dots & \overline{Z}_{i.T} \end{pmatrix}', \overline{Z}_{.j} = \begin{pmatrix} \overline{Z}_{.j0} & \overline{Z}_{.j1} & \dots & \overline{Z}_{.jT} \end{pmatrix}', \tag{9}$$

$$F = \begin{pmatrix} f_1 & f_2 & \dots & f_T \end{pmatrix}', F_{i.} = \begin{pmatrix} f_{i1}^A & f_{i2}^A & \dots & f_{iT}^A \end{pmatrix}', F_{.j} = \begin{pmatrix} f_{j1}^B & f_{j2}^B & \dots & f_{jT}^B \end{pmatrix}', \tag{10}$$

with D *being a* $T \times n$ *matrix of observed common factors;* F*,* $F_{i.}$ *and* $F_{.j}$ *all being* $T \times (m_0 + m_1 + m_2)$ *matrices of unobservable common factors; and* $\overline{Z}$*,* $\overline{Z}_{i.}$ *and* $\overline{Z}_{.j}$ *all being* $T \times 3(k+1)$ *matrices of cross-sectional averages. Finally, let* $X_{ij} = \left(X_{ij1}, X_{ij2}, \dots, X_{ijT}\right)'$ *denote the* $T \times k$ *matrix of individual-specific regressors.*

(a) Identification of β_{ij}***:*** *The* $k \times k$ *matrices* $\hat{\Psi}_{ijT} = T^{-1}\left(X'_{ij}M_{wij}X_{ij}\right)$ *and* $\hat{\Psi}_{ijg} = T^{-1}\left(X'_{ij}M_{gij}X_{ij}\right)$ *are non-singular, and* $\hat{\Psi}_{ijT}^{-1}$ *and* $\hat{\Psi}_{ijg}^{-1}$ *have finite second-order moments for all* i, j*.*

(b) Identification of β***:*** *The* $k \times k$ *pooled observation matrix* $\hat{\Psi}_{N^A,N^B,T}$ *defined by*

$$\hat{\Psi}_{N^A,N^B,T} = \sum_{j=1}^{N^B} \theta_j^A \sum_{i=1}^{N^A} \theta_i^B \hat{\Psi}_{ijT} \tag{11}$$

is non-singular for the scalar weights θ_j^A and θ_i^B that satisfy the conditions

$$\theta_j^A = O\left(\frac{1}{N^B}\right), \quad \sum_{j=1}^{N^B} \theta_j^A = 1, \quad \sum_{j=1}^{N^B} |\theta_j^A| < K,$$
$$\theta_i^B = O\left(\frac{1}{N^A}\right), \quad \sum_{i=1}^{N^A} \theta_i^B = 1, \quad \sum_{i=1}^{N^A} |\theta_i^B| < K.$$

Remark 2.1.2 *The assumptions for the factor-loading parameter (Assumption 2.3) and the random slope coefficients (Assumption 2.4) imply*

$$\overline{C}_w \equiv \sum_i w_i^B \sum_j w_j^A C_{ij} \xrightarrow{p} C, \quad \overline{C}_{iw} \equiv \sum_j w_j^A C_{ij} \xrightarrow{p} C, \quad \overline{C}_{wj} \equiv \sum_i w_i^B C_{ij} \xrightarrow{p} C,$$
$$\overline{C}_w^A \equiv \sum_i w_i^B \sum_j w_j^A C_{ij}^A \xrightarrow{p} C^A, \quad \overline{C}_{iw}^A \equiv \sum_j w_j^A C_{ij}^A \xrightarrow{p} C^A, \quad \overline{C}_{wj}^A \equiv \sum_i w_i^B C_{ij}^A \xrightarrow{p} C_j^A,$$
$$\overline{C}_w^B \equiv \sum_i w_i^B \sum_j w_j^A C_{ij}^B \xrightarrow{p} C^B, \quad \overline{C}_{iw}^B \equiv \sum_j w_j^A C_{ij}^B \xrightarrow{p} C_i^B, \quad \overline{C}_{wj}^B \equiv \sum_i w_i^B C_{ij}^B \xrightarrow{p} C^B,$$

where

$$C = \begin{pmatrix} \gamma & \Gamma \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix},$$
$$C^A = \begin{pmatrix} \gamma^A & \Gamma^A \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix}, \quad C_j^A = \begin{pmatrix} \gamma_j^A & \Gamma_j^A \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix},$$
$$C^B = \begin{pmatrix} \gamma^B & \Gamma^B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix}, \quad C_i^B = \begin{pmatrix} \gamma_i^B & \Gamma_i^B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta_i & I_k \end{pmatrix}.$$

Then

$$\sum_i w_i^B \sum_j w_j^A \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}^{\prime} f_t\right] - \left[C^{A\prime} f_t^A + C^{B\prime} f_t^B + C^{\prime} f_t\right] \xrightarrow{p} 0, \tag{12}$$
$$\sum_j w_j^A \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}^{\prime} f_t\right] - \left[C^{A\prime} f_{it}^A + C_i^{B\prime} f_t^B + C^{\prime} f_t\right] \xrightarrow{p} 0, \tag{13}$$
$$\sum_i w_i^B \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}^{\prime} f_t\right] - \left[C_j^{A\prime} f_t^A + C^{B\prime} f_{jt}^B + C^{\prime} f_t\right] \xrightarrow{p} 0, \tag{14}$$

where $f_t^A = \sum_i w_i^B f_{it}^A$ and $f_t^B = \sum_j w_j^A f_{jt}^B$.

Pesaran (2006) derives the cross-sectional weighted averages of a system that is similar to

(5). Here, since we have two cross-sectional dimensions, $N^A + N^B + 1$ averages of the vector Z_{ijt} are generated: N^A averages for each different unit in the first cross-sectional dimension; another N^B averages for each unit in the second cross-sectional dimension; and, finally, one overall weighted average. Applying these averages to (5) yields

$$\begin{aligned}
\overline{Z_t} &= \overline{B}'_w D_t + \sum_i \sum_j w_j^A w_i^B \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}' f_t \right] + \overline{U}_t, \\
\overline{Z}_{i.t} &= \overline{B}'_{iw} D_t + \sum_j w_j^A \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}' f_t \right] + \overline{U}_{it}^A, \\
\overline{Z}_{.jt} &= \overline{B}'_{wj} D_t + \sum_i w_i^B \left[C_{ij}^{A\prime} f_{it}^A + C_{ij}^{B\prime} f_{jt}^B + C_{ij}' f_t \right] + \overline{U}_{jt}^B,
\end{aligned}$$

where

$$\begin{aligned}
&\overline{B}'_w = \textstyle\sum_i \sum_j w_j^A w_i^B B'_{ij}, \quad && \overline{U}_t = \textstyle\sum_i \sum_j w_j^A w_i^B U_{ijt}, \\
&\overline{B}'_{iw} = \textstyle\sum_j w_j^A B'_{ij}, && \overline{U}_{it}^A = \textstyle\sum_j w_j^A U_{ijt}, \\
&\overline{B}'_{wj} = \textstyle\sum_i w_i^B B'_{ij}, && \overline{U}_{jt}^B = \textstyle\sum_i w_i^B U_{ijt}.
\end{aligned}$$

By applying the convergence in probability properties derived in Remark 2.1.2, we have

$$\overline{Z_t} - \overline{B}'_w D_t - \left[C^{A\prime} f_t^A + C^{B\prime} f_t^B + C' f_t \right] \xrightarrow{p} 0, \tag{15}$$

$$\overline{Z}_{i.t} - \overline{B}'_{iw} D_t - \left[C^{A\prime} f_{it}^A + C_i^{B\prime} f_t^B + C' f_t \right] \xrightarrow{p} 0, \tag{16}$$

$$\overline{Z}_{.jt} - \overline{B}'_{wj} D_t - \left[C_j^{A\prime} f_t^A + C^{B\prime} f_{jt}^B + C' f_t \right] \xrightarrow{p} 0. \tag{17}$$

Now, based on the proxy derived in Pesaran (2006), the following "solutions" are conjectured

$$f_t - A \left[\overline{Z}_t - \overline{B}'_w D_t \right] \xrightarrow{p} 0, \tag{18}$$

$$f_{it}^A - A_{i0}^A \left[\overline{Z}_{i.t} - \overline{B}'_{iw} D_t \right] - A_{i1}^A \left[\overline{Z}_t - \overline{B}'_w D_t \right] \xrightarrow{p} 0, \tag{19}$$

$$f_{jt}^B - A_{j0}^B \left[\overline{Z}_{.jt} - \overline{B}'_{wj} D_t \right] - A_{j1}^B \left[\overline{Z}_t - \overline{B}'_w D_t \right] \xrightarrow{p} 0, \tag{20}$$

for some matrices $A, A_{i0}^A, A_{i1}^A, A_{j0}^B, A_{j1}^B$, where A is of dimension $m_0 \times (k+1)$; A_{i0}^A and A_{i1}^A are both of dimension $m_1 \times (k+1)$; and A_{j0}^B and A_{j1}^B are both of dimension $m_2 \times (k+1)$. By taking the average of (19) over i and the average of (20) over j, it follows that

$$f_t^A - \left(A_0^A + A_1^A\right)\left[\overline{Z}_t - \overline{B}_w' D_t\right] \xrightarrow{p} 0, \tag{21}$$

$$f_t^B - \left(A_0^B + A_1^B\right)\left[\overline{Z}_t - \overline{B}_w' D_t\right] \xrightarrow{p} 0, \tag{22}$$

where $A_0^A = \sum_i w_i^B A_{i0}^A$, $A_1^A = \sum_i w_i^B A_{i1}^A$, $A_0^B = \sum_j w_j^A A_{j0}^B$, and $A_1^B = \sum_j w_j^A A_{j1}^B$.

From (18) it follows that m_0 linear combinations of the averages in Z_t (adjusted for the deterministic variables in D_t) express the common factors. Hence, it is only necessary to investigate the expression in (15) given by the space spanned by the m_0 row vectors in A. To find the expression for A we pre-multiply (15) by the unknown A, apply (18), (21) and (22), and set this equal to zero. This yields the identity

$$A - AC^{A\prime}\left(A_0^A + A_1^A\right) - AC^{B\prime}\left(A_0^B + A_1^B\right) - AC'A = 0,$$

which implies

$$A = \left(CC'\right)^{-} C\left[I - C^{A\prime}\left(A_0^A + A_1^A\right) - C^{B\prime}\left(A_0^B + A_1^B\right)\right]. \tag{23}$$

Similarly, pre-multiplying (16) with A_{i0}^A and applying (18)–(22) yields the following indirect solutions for A_{i0}^A and A_{i1}^A:

$$\begin{aligned} A_{i0}^A - A_{i0}^A C^{A\prime} A_{i0}^A &= 0, \\ A_{i0}^A C^{A\prime} A_{i1}^A + A_{i0}^A C_i^{B\prime}\left(A_0^B + A_1^B\right) + A_{i0}^A C' A &= 0, \end{aligned}$$

which gives

$$A_{i0}^A = \left(C^A C^{A\prime}\right)^{-} C^A, \tag{24}$$

$$A_{i1}^A = -\left(C^A C^{A\prime}\right)^{-} C^A\left[C_i^{B\prime}\left(A_0^B + A_1^B\right) + C_i' A\right]. \tag{25}$$

Finally, applying the same procedure to (17) yields

$$A_{j0}^{B} = \left(C^{B}C^{B\prime}\right)^{-} C^{B}, \tag{26}$$

$$A_{j1}^{B} = -\left(C^{B}C^{B\prime}\right)^{-} C^{B}\left[C_{j}^{A\prime}\left(A_{0}^{A}+A_{1}^{A}\right)+C^{\prime}A\right]. \tag{27}$$

Hence, this gives explicit solutions for A_{i0}^{A} and A_{j0}^{B} and, hence, also for A_{0}^{A} and A_{0}^{B}:

$$A_{0}^{A} = \left(C^{A}C^{A\prime}\right)^{-} C^{A}$$

$$A_{0}^{B} = \left(C^{B}C^{B\prime}\right)^{-} C^{B}$$

These expressions are now applied to derive the proxies for the different multi-factor error structures:

- With the multi-factor structure in alternative I — $U_{ijt}^{*} = C_{ij}^{\prime}f_{t} + U_{ijt}$, which implies that $C_{ij}^{A} = C_{ij}^{B} = 0$ — it follows that $A = (CC^{\prime})^{-}C$ and $A_{i0}^{A} = A_{i1}^{A} = A_{j0}^{B} = A_{j1}^{B} = 0$. Therefore, $\left(D_{t}^{\prime}, \overline{Z}_{t}^{\prime}\right)^{\prime}$ can be used as an observable proxy. This expression corresponds to the one found in Pesaran (2006).

- With the multi-factor structure in alternative II, it follows that $C = 0$ and $C^{B} = 0$. Applying this in (23)–(27) yields $A_{i0}^{A} = \left(C^{A}C^{A\prime}\right)^{-}C^{A}$ and $A = A_{i1}^{A} = A_{j0}^{B} = A_{j1}^{B} = 0$, which implies using $\left(D_{t}^{\prime}, \overline{Z}_{i.t}^{\prime}\right)^{\prime}$ as an observable proxy for the common factors.

- With the multi-factor structure in alternative III, it follows that $C^{B} = 0$. Applying this in (23)–(27) yields $A_{j0}^{B} = A_{j1}^{B} = 0$, $A_{i0}^{A} = \left(C^{A}C^{A\prime}\right)^{-}C^{A}$, and indirect solutions for A and A_{i0} given as

$$A = (CC^{\prime})^{-} C\left[I - C^{A\prime}\left(A_{0}^{A}+A_{1}^{A}\right)\right], \tag{28}$$

$$A_{i1}^{A} = -\left(C^{A}C^{A\prime}\right)^{-} C^{A}C^{\prime}A. \tag{29}$$

  Hence, $\left(D_{t}^{\prime}, \overline{Z}_{t}^{\prime}, \overline{Z}_{i.t}\right)^{\prime}$ can be used as an observable proxy for the common factors.

- With the multi-factor structure in alternative IV; the coefficient matrices for A, A_{i0}^{A}, A_{i1}^{A},A_{j0}^{B} and A_{j1}^{B} are indirectly given in (23)–(27). With none of the coefficient matrices equal to zero, all of the proxies suggested in (18)–(20) must be included. Hence, $\left(D_t', \overline{Z}_t', \overline{Z}_{i.t}, \overline{Z}_{.jt}\right)'$ can be used as an observable proxy for the common factors.

- With the multi-factor structure in alternative IV with the restriction $C = 0$, it follows that $A = 0$. However, since A_{i0}^A, A_{i1}^A,A_{j0}^B and A_{j1}^B generally will be non-zero, all the proxies in (19) and (20) must be included. Hence, the vector of proxies is the same as for the multi-factor structure IV without a zero-restriction on C imposed, see above.

The results in this section are derived as though all variables in X_{ijt} vary in both cross-sectional dimensions. If they do, all parameters in the vector β_{ij} can be estimated. However, if they do not, some proxies for the common factors can be identical to an exogenous variable (i.e. a variable in X_{ijt}). Then we cannot distinguish between the direct effect of X_{ijt} and the effect from the common factor. This is similar to the problem of interpreting the coefficients for the deterministic variables in D, as also noted by Pesaran (2006). For example, with multi-factor structure in alternative II — i.e. $U_{ijt}^f = C_{ij}^{A\prime} f_{it}^A + U_{ijt}$ — then β_{ij}^A cannot be estimated, because we cannot distinguish between the exogenous variables x_{it}^A and their cross-sectional averages $\sum_{j=1}^{N^B} w_j^A x_{it}^A = x_{it}^A$ used to compose the common factors. However, in this example, both β_{ij}^B and β_{ij}^C can be estimated and interpreted as the direct effect of the exogenous variable on the endogenous variable. In the empirical section, the elasticity of substitution is the only parameter that can be interpreted as a direct effect of the exogenous variables on the endogenous variable, as the corresponding variable in X_{ijt} (the relative factor price) is the only variable in X_{ijt} that varies in both cross-sectional dimensions.

2.2 The number of common factors

The common factors are here approximated by different cross-sectional averages. That implies that there is a limit to how many common factors can be approximated. In our multi-factor structure I, it follows from Pesaran (2006) that, to be able to approximate the common factors with this approach, the following assumption must be fulfilled: $rank(\overline{C}_w) = m_0$ (where m_0 is the number of common factors). One implication of this assumption is that $m_0 \leq k+1$, as $\overline{C}_w$ is a $(k+1) \times m_0$ matrix. Hence, the number of common factors cannot exceed the number of cross-sectional averages that we add to the regression to proxy the common factors. Another implication of the assumption that $\overline{C}_w$ must have full rank is that there must be enough linearly independent variation in the cross-sectional averages to proxy the common factors.

For the multi-factor structure in alternative II, a sufficient assumption is $rank(\overline{C}_{iw}^A) = m_1 \leq k_2 + k_3 + 1$.

For the multi-factor structure in alternative III, sufficient rank restrictions are

$$rank \begin{pmatrix} \overline{C}_w \\ \overline{C}_w^A \end{pmatrix} = m_0 + m_1 \leq k + 1$$

and $rank(\overline{C}_{iw}^A) \leq k_2 + k_3 + 1$.

For the multi-factor structure in alternative IV, the following assumption must hold to make it possible to proxy the common factors:

Assumption 2.6 *We assume*

(a) $rank \begin{pmatrix} \overline{C}_w \\ \overline{C}_w^A \\ \overline{C}_w^B \end{pmatrix} = m_0 + m_1 + m_2 \leq k + 1$

(b) $rank(\overline{C}_{iw}^A) \leq k_2 + k_3 + 1$, *and*

(c) $rank(\overline{C}_{wj}^B) \leq k_1 + k_3 + 1$.

Assumption 2.6 can be used for all the alternative multi-factor structures. For example, with the multi-factor structure in alternative III, we have $\overline{C}_w^B = 0$ and $m_2 = 0$, which implies that conditions (a) and (b) in 2.6 are identical to this multi-factor structure presented above, whereas condition (c) in 2.6 is obviously fulfilled.

2.3 Asymptotic property of the estimator

Here, I give a limiting result for the estimator of β_{ij} under the most general multi-factor structure (i.e. alternative IV) when the appropriate proxies for the common factors are included in the regression. In matrix notation, the model with the proxies for the common factors with multi-factor structure IV can be written as

$$Y_{ij} = X_{ij}\beta_{ij} + H_{wij}\Theta_{ij}^* + \varepsilon_{ij}^*, \tag{30}$$

where

$$Y_{ij} = \left(Y_{ij1}, Y_{ij2}, \ldots, Y_{ijT}\right)' \text{ with dimension } T \times 1, \tag{31}$$

$$X_{ij} = \left(X_{ij1}, X_{ij2}, \ldots, X_{ijT}\right)' \text{ with dimension } T \times k, \tag{32}$$

$$E^*_{ij} = \left(E^*_{ij1}, E^*_{ij2}, \ldots, E^*_{ijT}\right)' \text{ with dimension } T \times 1 \tag{33}$$

are white noise and H_{wij} (and G_{ij} used below) defined in Assumption 2.5. The estimator of β_{ij} is

$$\widehat{b}_{ij} = \left(X'_{ij} M_w X_{ij}\right)^{-1} \left(X'_{ij} M_w Y_{ij}\right), \tag{34}$$

where $M_w = I_T - H_w \left(H'_w H_w\right)^{-1} H'_w$.

Proposition 2.1 *Under the multi-factor structure in alternative IV and Assumptions 2.1–2.6, the following limiting result for the estimator in (34) is given by:*

$$\begin{aligned} \widehat{b}_{ij} - \beta_{ij} &= \left(\frac{X'_{ij} M_{gij} X_{ij}}{T}\right)^{-1} \left(\frac{X'_{ij} M_{gij} E_{ij}}{T}\right) \\ &\quad + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right), \end{aligned} \tag{35}$$

where $M_{gij} = I_T - G_{ij}\left(G'_{ij} G_{ij}\right)^{-} G'_{ij}$. Since $E_{ij} = \left(E_{ij1}, E_{ij2}, \ldots, E_{ijT}\right)'$ is distributed independently of X_{ij} and G_{ij}, then, for a fixed T and $N^A \to \infty$ and $N^A \to \infty$, we have $E\left(\widehat{b}_{ij} - \beta_{ij}\right) = 0$.

The proof is given in Appendix A.

The proposition shows that the proposed estimator is asymptotically unbiased.

3 Monte Carlo

In a Monte Carlo experiment, I consider the following data-generating process with only one exogenous variable:

$$y_{ijt} = \alpha + \beta_{ij} x_{ijt} + \gamma^A_{yij} f^A_{it} + \gamma^B_{yij} f^B_{jt} + \varepsilon_{ijt}, \tag{36}$$

$$x_{ijt} = \gamma^A_{xij} f^A_{it} + \gamma^B_{xij} f^B_{jt} + v_{ijt}, \tag{37}$$

for $i = 1, \ldots, N^A$, $j = 1, \ldots, N^B$ and $t = 1, \ldots, T$ and where the common factors (f_{it}^A and f_{jt}^B) and the error terms (v_{ijt} and ε_{ijt}) all follow AR(1) processes given by

$$\begin{aligned}
f_{it}^A &= \rho_{fi}^A f_{it-1}^A + v_{fit}^A, \text{ with } v_{fit}^A \sim NIID\left(0,1\right) \text{ for } t = 1, \ldots, T \\
&\quad \text{with } f_{i0}^A \sim NIID\left(0, \frac{1}{\left(1-\rho_{i0}^A\right)^2}\right) \text{ if } |\rho_{fi}^A| < 1 \text{ for } i = 1, \ldots, N^A \\
f_{jt}^B &= \rho_{fj}^B f_{jt-1}^B + v_{fjt}^B, \text{ with } v_{fjt}^B \sim NIID\left(0,1\right) \text{ for } t = 1, \ldots, T \\
&\quad \text{with } f_{j0}^B \sim NIID\left(0, \frac{1}{\left(1-\rho_{j0}^B\right)^2}\right) \text{ if } |\rho_{fj}^B| < 1 \text{ for } j = 1, \ldots, N^A \\
v_{ijt} &= \rho_{ij} v_{ijt-1} + v_{ijt}, \text{ with } v_{ijt} \sim NIID\left(0,1\right) \text{ for } t = 1, \ldots, T \\
&\quad \text{with } v_{ij0} \sim NIID(0, \frac{1}{\left(1-\rho_{ij}\right)}) \text{ if } |\rho_{ij}| < 1 \text{ for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B
\end{aligned}$$

and

$$\begin{aligned}
\varepsilon_{ijt} &= \rho_{\varepsilon i} \varepsilon_{ijt-1} + v_{\varepsilon ijt}, \text{ with } v_{\varepsilon jt} \sim NIID\left(0,1\right) \text{ for } t = 1, \ldots, T \\
&\quad \text{with } \varepsilon_{ij0} \sim NIID(0, \frac{1}{\left(1-\rho_{\varepsilon ij}\right)^2}) \text{ if } |\rho_{\varepsilon ij}| < 1 \text{ for } i = 1, \ldots, N^A \text{ and } j = 1, \ldots, N^B.
\end{aligned}$$

This data-generating process implies that there is one common factor for each i ($i = 1, \ldots, N^A$) and one common factor for each j ($j = 1, \ldots, N^B$). Hence, in the full system $N^A + N^B$ independent common factors are included.

The set-up of the data-generating process above allows for both stationary and non-stationary processes. If all the ρ-variables have absolute value less than unity (i.e. ...), all processes are stationary. However, if some ρ-parameters are equal to unity, some of the variables follows non-stationary processes. Section 2 assumed that all variables are stationary. With the set-up of the Monte Carlo above, also the case for non-stationary variables can be investigated. Hence, both the stationary and non-stationary case will be considered. However, the results are almost identical between the stationary and non-stationary case. Therefore, only the non-stationary case is presented here, whereas the result from the stationary case is reported in Table 4 in Appendix C.

In the non-stationary case with variables integrated of order $I(1)$ we have: $\rho_{fi}^A = \rho_{fj}^B =$

Table 1: Results from simulations: $\hat{\beta}_{MG}$ — non-stationary case

	infeasible	naïve	overall c.f.	c.f. in the 1. dim.	both	all	special
Proxies	$\begin{pmatrix} f_{it}^A \\ f_{jt}^B \end{pmatrix}$	0	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \end{pmatrix}$	$\begin{pmatrix} \bar{y}_{it} \\ \bar{x}_{it} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \\ \bar{y}_{it} \\ \bar{x}_{it} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \\ \bar{y}_{it} \\ \bar{x}_{it} \\ \bar{y}_{jt} \\ \bar{x}_{jt} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_{it} \\ \bar{x}_{it} \\ \bar{y}_{jt} \\ \bar{x}_{jt} \end{pmatrix}$
Asympt. bias	0	2/3	2/3	1/2	1/2	0	0

In the table, "c.f." is used for "common factors" and '1. dim" is short for "first cross-sectional dimension", "both" is the joint of "overall c.f." and "c.f. in 1. dim.'

	infeasible	naïve	overall c.f.	c.f. in the 1. dim.	both	all	special
mean	1.0000	1.6671	1.6648	1.5002	1.5002	1.0000	1.0195
min	0.9976	1.6161	1.6321	1.4546	1.4614	0.9970	1.0045
max	1.0024	1.6994	1.6945	1.5490	1.5402	1.0030	1.0595
st.d.*100	0.0760	1.1934	0.9910	1.4416	1.2710	0.0962	0.7768

In this simulation: 1000 replications, $N^A = N^B = 100$, $T = 100$, $\alpha = 0$, $\rho_{fi}^A = \rho_{fj}^B = \rho_{ij} = \rho_{\varepsilon ij} = 1$, and $\rho_{\varepsilon ij} = 0.5, \forall i, j$, $\beta_{ij} = 1, \forall i, j$, $\gamma_{yij}^A = \gamma_{yij}^B = \gamma_{xij}^A = \gamma_{xij}^B = 1, \forall i, j$.

	infeasible	naïve	overall c.f.	c.f. in the 1. dim.	both	all	special
mean	1.0019	1.6666	1.6445	1.4991	1.4971	1.0001	1.1482
min	0.8278	1.4313	1.4619	1.2347	1.2036	0.5601	0.8721
max	1.2190	1.8680	1.8667	1.7943	1.8376	1.5572	1.4509
st.d.*100	5.6764	5.6071	5.6310	7.3794	8.3871	12.165	8.4186

In this simulation: As above, except $N^A = N^B = 10$, $T = 10$.

$\rho_{ij} = 1$ for $i = 1, \ldots, N^A$ and $j = 1, \ldots, N^B$ and $\rho_{\varepsilon ij} = 0.5$. Furthermore, in the non-stationary case the initial values, f_{i0}^A, f_{j0}^B, v_{ij0}, can be set equal to zero (as the results will be independent of the starting values in this particular set-up of the Monte Carlo process).

In these simulations, the parameters are fixed across replications. For each replication, the "Common Correlated Effects Mean Group Estimator" (CCEMG) $\hat{\beta}_{MG} = \left(N^A N^B\right)^{-1} \sum_{i=1}^{N^A} \sum_{j=1}^{N^B} \hat{\beta}_{ij}$ is estimated from (36) without the common factors but with the proxies for the common factor included as extra regressors. In Table 1, the mean over all replications is reported for the stationary case. In addition, the smallest estimate (min) and the highest estimate (max) of all replications are reported. Finally, the standard deviation of the distribution of β_{MG} over the replications is reported. These numbers are reported for different specifications of the proxies used for the common factors.

In Table 1, I consider seven different estimators of β. They all differ with respect to which variables are included as proxies for the common factors. The first estimator — denoted the 'infeasible' estimator — is the estimator where the common factors themselves are included in the regression. It is denoted 'infeasible' as we assume that these common factors are unobservable, and — hence — cannot be included directly in the regression. These results are included to provide a benchmark for how well we could estimate β if all information were available. The asymptotic bias for this estimator is zero, as reported in the first column in the upper part of Table 1.

In the middle part of Table 1, simulation results with a relatively large data set are reported. In the lower part of the table, simulation results with the same parameters in the data-generating process are used with a relatively small data set. Simulations with other parameter values are also conducted, but the results from them show a similar picture as the one reported in the table.[2]

The second estimator I consider is denoted 'naïve' in the table. These are the results from the regression where no proxies for the common factors are included in the regression. Hence, the endogenous variable is here only regressed on the exogenous variable in addition to an intercept. In Appendix B, it is shown that the asymptotic bias with this estimator is 2/3 with the data-generating process considered here. The simulation results confirm that the estimates are very biased and close to the asymptotic bias for both sample sizes, showing that neglecting the common factors can lead to large estimation biases.

The results for the third estimator I consider are reported in the column 'overall c.f.' in the table. Here the overall common factors are included, i.e. $\overline{y}_t = \left(N^A N^B\right)^{-1} \sum_i \sum_j y_{ijt}$ and $\overline{x}_t = \left(N^A N^B\right)^{-1} \sum_i \sum_j x_{ijt}$ are included in addition to x_{ijt} and an intercept. With a relatively large data set, see the upper part of the table, the bias is approximately as for the 'naïve' estimator. With a smaller data set, see the lower part of the table, the bias is somewhat smaller. In a large data set — i.e. when N^A and N^B are large — the variation in $\overline{y}_t$ and $\overline{x}_t$ is negligible compared to the variation in x_{ijt}. Hence, the correlation between these two proxies and x_{ijt} is small and including them does not matter much for the estimated coefficient of x_{ijt}. However, in a small data set $\overline{y}_t$ is correlated with $\overline{y_i}_t$ and $\overline{y_j}_t$ (and similarly for $\overline{x}_t$), which reduces the bias somewhat.[3]

[2]The Monte Carlo simulation is programmed in Ox Professional, see Doornik (2013), and the code is available from www.hungnes.net.

[3]For example, if, say, $N^A = 1$, then $\overline{x}_t = \overline{x}_{it}$ and $\overline{y}_t = \overline{y}_{it}$ for $i = 1$. Hence, the proxies for the overall common

The forth estimator is considered in the column 'industry-sp. c.f.' where $(\overline{y_{i}}_t, \overline{x_{i}}_t)'$ is used as a proxy for f_i^A. It can be shown that the bias will converge to 1/2 if the proxy converges to f_i^A. This is indeed the case, both in the large and in the small data set, indicating that the proxy is very good for f_i^A. However, the omission of a proxy for f_j^B still leads to a large bias.

The fifth estimator is considered in the column 'both', where both $(\overline{y_{i}}_t, \overline{x_{i}}_t)'$ and $(\overline{y}_t, \overline{x}_t)'$ are used as proxies. The bias is almost identical as in the case when only $(\overline{y_{i}}_t, \overline{x_{i}}_t)'$ is used as a proxy.

The sixth estimator is considered in the column 'all', where $(\overline{y_{i}}_t, \overline{x_{i}}_t, \overline{y_{j}}_t, \overline{x_{j}}_t, \overline{y}_t, \overline{x}_t)'$ is used as a proxy for f_{it}^A and f_{jt}^B. Here the estimator is almost as good as the infeasible estimator in the relatively large data set. Also in the small sample, the estimator has a small bias, measured as the deviation of the mean of the estimates over the replications and the true value. However, the standard deviation is twice that for the infeasible estimator, and, for one of the replications, the estimate is as low as 0.239. The reason for the relatively high uncertainty in the estimates is that very many parameters need to be estimated compared to the number of observations. For each replication and combination of industry and input factors, we have 10+1 observations (including the initial observation) to estimate the parameter for x_{ijt} and the intercept and six variables used to proxy the common factors.

Finally, in the column 'special' the results for the seventh estimator are reported. Here $(\overline{y_{i}}_t, \overline{x_{i}}_t, \overline{y_{j}}_t, \overline{x_{j}}_t)'$ is used as a proxy for f_{it}^A and f_{jt}^B, i.e., $\overline{y}_t$ and $\overline{x}_t$ are not used in the set of variables proxying the common factors. Excluding the overall averages leads to a bias, but the bias decreases with the size of the data set. The reason is that, in our data-generating process, both $\overline{y}_t$ and $\overline{x}_t$ become closer and closer to a constant as the data set increases, meaning that the variation in $(\overline{y}_t, \overline{x}_t)'$ is negligible in relation to the remaining proxies. Therefore, they become less and less important to include as proxies. Hence, this indicates that, if $N^A \cdot N^B$ is large, $(\overline{y}_t, \overline{x}_t)'$ can be excluded from the proxy for the common factors when no overall common factors are included. On the other hand, if we are only interested in the mean-group estimator β_{MG}, the gain of doing so is negligible since the degree of freedom is large. However, if we are interested in the estimates of individual β_{ij}'s, excluding $\overline{y}_t$ and $\overline{x}_t$ from the proxy may yield

factors are equal to the proxies for the industry-specific common factors. Therefore, the bias would be the same as the one reported in the column 'industry-sp. c.f.', which is equal to 1/2.

more precise estimates.

4 Empirical illustration

Specification of the production function is important when estimating the elasticity of substitution between input factors. Berndt (1976), applying a constant elasticity of scale (CES) function with labour and capital as input factors, finds an elasticity close to unity for US aggregate production when only allowing for Hicks-neutral technological changes. Antràs (2004) shows that the estimated elasticity of substitution decreases when allowing for biased technological changes. However, Antràs (2004) only considers production functions where the technological changes follow deterministic processes.

Diamond et al. (1978) showed that joint identification of the elasticity of substitution and factor-biased technological changes can be infeasible, also known as the impossibility theorem. One approach to circumventing this problem is to assume a certain functional form for the growth rates of efficiency levels for the input factors, see, e.g., Klump et al. (2012). Typically, these efficiency levels are assumed to follow deterministic trends, see, e.g., the overview in Leon-Ledesma et al. (2010). However, a steady trend might not reflect technological changes in a good manner, since technological changes can follow a process with large and unpredictable shifts.

Another approach to tackling the impossibility theorem of Diamond et al. (1978) is to consider a system with more than two input factors where the growth rates of the efficiency levels are restricted to follow a reduced number of stochastic trends, as in Hungnes (2011). However, Hungnes (2011) assumes that relative factor prices are given outside the model, so that they are weakly exogenous. This implies that shifts in the use of input factors due to technological changes will not lead to changes in the relative input prices. This is a necessary assumption in Hungnes (2011) in order to obtain unbiased estimates of the elasticity of substitution: if this assumption does not hold, the estimates may be downward biased. To understand this, consider a technological shift that increases the productivity of one input factor. More demand for this input factor may lead to a higher price for the input factor. Hence, we will observe increased use of an input factor with increased price.

Here, I do not assume that the relative input prices are independent of the technological changes. This is achieved by including common factors, which are allowed to be correlated with other variables in the system.

Within an industry, the common factors can capture processes such as factor neutral technological progress. Without the common factors, the technological progress will usually only be explained by a deterministic trend. When common factors are included, they can pick up both stationary and non-stationary processes, depending on the order of integration of the observable variables included in the analysis. Combined with a deterministic trend, these common factors can express the process of the technological progress better than the deterministic trend alone. Similarly, common factors within an industry can capture factor-biased technological changes that are only present in that industry.

Technological changes can also change the optimal composition of factor use in more than one industry. For example, a technological change can lead to more use of some input factors in many (or all) industries and reduced use of other input factors in the same industries. To capture such technological changes variables that are composed as averages over industries are used as proxies for common factors. Controlling for technological changes both within and between industries, we can obtain unbiased estimates of the substitution elasticity in each industry.

4.1 Theory

The demand function is based on cost minimising given a constant elasticity of substitution (CES) production function. In industry i ($i = 1, \ldots, N^A$) the log of the demand for input factor j ($j = 1, \ldots, N^B$) at time t ($t = 1, \ldots, T$), v_{ijt}, is a linear function of the log of the relative factor price ($p_{ijt} - p_{iAt}$), log of production in the industry (x_{it}) and some time-varying parameters (δ_{ijt} and θ_{it}) explained below:[4]

$$v_{ijt} = \sigma_i \ln \delta_{ijt} - \frac{1}{\kappa_i}\theta_{it} - \sigma_i \left(p_{ijt} - p_{iAt}\right) + \frac{1}{\kappa_i} x_{it}. \tag{38}$$

[4]See Appendix A in Hungnes (2011) for how the factor demand function is derived.

The formulation in (38) implies the same elasticity of substitution between all input factors within each industry, denoted σ_i.

In (38) $\delta_{i1t}, \ldots, \delta_{iN^B t}$ are time-varying distribution parameters for industry i, where $\delta_{ijt} \geq 0$ $(\forall i, j, t)$ and $\sum_{k=1}^{N^B} \delta_{ikt} = 1$ $(\forall i, t)$. With a Cobb-Douglas technology, i.e. when $\sigma_i = 1$, these time-varying distribution parameters express the optimal cost shares for the input factors. The time-dependence of the δ's is interpreted as capturing factor-biased (or factor-augmenting) technological changes.[5] The latent stochastic variable θ_{it} represents the factor-neutral technology level. The parameter κ_i denotes the elasticity of scale in industry i.

In general, the expression of the weighted factor price, p_{iAt}, is rather complicated. However, if $\sigma = 1$ (i.e. with a Cobb-Douglas production function), it is simply the weighted average of the different input factors, where the weight is equal to the optimal cost share. In order to calculate the weighted factor prices p_{iAt}, I use

$$p_{iAt} = \sum_{j=1}^{N^B} \zeta_{ik} p_{ikt}, \tag{39}$$

where ζ_{ij} is the weight of input factor j in industry i, where $\zeta_{ij} \geq 0$ $(\forall i, j)$ and $\sum_{k=1}^{N^B} \zeta_{ik} = 1$ $(\forall i, t)$. These weights are reported in Table 2 and are also used to construct the the averages used to derive the proxies for the common factors, see also Hungnes (2016).

The joint process of the factor-neutral technological level and the distribution parameters follows a deterministic trend and some common factors:

$$\sigma_i \ln \delta_{ijt} - \frac{1}{\kappa_i} \theta_{it} = \mu_{ij} + \gamma_{ij} t + \lambda_{ij}' f_{ijt}^*, \tag{40}$$

where f_{ijt}^* is a vector of common factors and λ_{ij} is the corresponding vector of parameters. The vector f_{ijt}^* includes subscripts for both the industry i and the input factor j as the vector can include both industry and input factor-specific common factors.

Table 2: Cost shares ($\zeta_{i|j}$) = weights ($w_{j|i}$), in per cent

Industry	L	E	F	FT	M	K10	K30	K40	K50	K60
01: Agriculture etc.	35.8	1.5	0.5	1.8	27.4	21.6	—	0.7	10.7	0.0
02: Fishing and hunting	49.4	—	0.2	7.3	25.6	—	16.0	—	1.4	0.1
03: Aquaculture	10.3	—	—	0.4	82.8	2.0	1.8	0.7	1.2	0.9
04: Consumer goods	15.5	0.9	0.2	0.3	75.6	2.9	—	0.3	3.9	0.5
05: Intermediate goods etc.	24.1	1.3	0.6	0.7	61.7	3.9	—	0.4	5.7	1.5
06: Energy-intensive goods	11.2	7.6	0.8	0.1	64.7	5.5	—	0.1	9.0	1.0
07: Petroleum products	1.8	0.3	3.8	—	88.2	3.4	—	0.0	2.4	0.1
08: Engineering products	26.3	0.5	0.1	0.1	66.6	1.6	—	0.1	2.6	2.1
09: Construction	28.6	0.2	0.2	9.9	64.9	2.9	—	0.8	1.3	0.1
10: Banking and insurance	35.3	0.4	—	0.0	50.8	10.0	—	1.2	0.4	1.7
11: Electricity	13.0	4.9	—	1.0	16.9	31.1	—	0.3	32.1	0.7
12: R & D	42.2	—	—	—	28.1	4.0	—	—	1.6	24.1
13: Domestic transport	31.0	0.4	1.0	6.4	48.6	4.2	3.2	3.7	1.4	0.2
14: Merchandising	42.9	1.1	0.4	0.7	48.4	3.6	—	0.7	1.8	0.4
15: Information services	33.4	0.3	0.1	0.3	54.4	2.4	—	0.2	5.9	2.9
16: Other private services	43.3	0.8	0.2	0.4	46.9	4.3	—	0.8	2.6	0.6
17: Leasing com. buildings	13.7	2.2	0.5	0.3	43.1	39.1	—	—	1.0	0.1

L: man-hours (sum of employed and employees); E: intermediate consumption of electricity; F: intermediate consumption of heating oil; FT: intermediate consumption of transport oil; M: other intermediate consumption; K10: real capital, buildings and constructions; K30: real capital, ships and fishing boats; K40: real capital, cars; K50; real capital, machinery and equipment; K60: real capital, R&D and other intangible assets. The symbol '—' indicates that the input factor is not used in the industry (in at least one time period), according to the national accounts.

4.2 Estimation results

In this section, I apply the estimators considered in Section 2 on quarterly data from the Norwegian national accounts to estimate the elasticity of substitution in (38) for 17 industries in Norway. The estimation period is 1980q1 – 2013q4 and is conducted using PcGive, see Doornik and Hendry (2013). The estimated elasticity of substitution in the separate industries under different specifications of the common factors are reported in Table 3.

The first column shows the results when no proxies for common factors are included. In the estimation, I have imposed the additional restriction that the elasticity of substitution is equal across input factors within each industry.[6] Hence, for each industry, the following regression

[5]However, as also pointed out in Hungnes (2011), the parameter instability may also be due to other reasons, such as aggregation (over firms) effects.

[6]However, I have not imposed the restriction that the coefficient for the elasticity of scale is the same across input factors within an industry. The reason for the latter is that this restriction is impossible to impose when proxies for the common factors are included (as the cross-sectional average of production in an industry is equal to the production, and — hence — makes it impossible to distinguish between the direct effect of production, which is given by the inverse of the elasticity of scale, and the effect of production as a proxy for common factors).

is estimated:

$$v_{ijt} = \alpha_{0ij} + \alpha_{1ij}t + \alpha_{2ij}x_{it} - \sigma_i \left(p_{ijt} - p_{iAt}\right) + \varepsilon_{ijt}, \tag{41}$$

for $i = 1, \ldots, N^A$, $j = 1, \ldots, N^B$ and $t = 1, \ldots, T$.

In the table, the estimate for the elasticity of substitution, σ_i, is reported for each industry. As can be seen from the table, the estimate of the elasticity of substitution is negative in three industries: these are the Fishing and hunting industry (02); the Petroleum products industry (07); and the R & D industry (12). A negative elasticity of substitution implies that the industry will use relatively more of an input factor that increases in price, contrary to economic theory.

In the second column in Table 3, the estimate of the elasticity of substitution is reported when proxies for overall common factors are included. Here, the mean of production over industries and the mean of the input factor use over all combinations of industries and input factor types are used as proxies. The results are similar to the case without proxies for common factors; in the same three industries, the estimated elasticity of substitution is negative.

The third column in Table 3 shows the estimation results for the case where it is assumed that the common factors are industry-specific. In this case, the mean of the input factors within each industry is used as a proxy for the common factor. With this formulation of the common factor, there are still sign problems with the elasticity of substitution in two of the industries (02, Fishing and hunting; and 07, Petroleum product). In the R & D industry (12), the estimated elasticity of substitution is positive, though very close to zero (and not significantly different from zero) — implying almost no substitution possibilities.

The fourth column in Table 3 shows the results when common factors are assumed to be either overall or industry-specific (but no common factors that are input-specific). These estimated elasticities of substitution are quite similar to the estimates when only industry-specific common factors are considered (cf. the results in the third column).

In the final column of Table 3, all types of common factors are considered. That means that in this column input-specific common factors are also allowed. Now, five variables are used as proxies for the common factors, see the top right corner of Table 3. The reported estimates of the elasticity of substitution are all positive, ranging from 0.2 (in the Fishing and

Table 3: Estimated elasticity of substitution

	no	overall	industry-specific	both	all types
Proxies	0	$\begin{pmatrix} x_{.t} \\ v_{..t} \end{pmatrix}$	$\begin{pmatrix} v_{i.t} \end{pmatrix}$	$\begin{pmatrix} x_{.t} \\ v_{..t} \\ v_{i.t} \end{pmatrix}$	$\begin{pmatrix} x_{.t} \\ p_{.jt} - p_{.At} \\ v_{i.t} \\ v_{..t} \\ v_{.jt} \end{pmatrix}$
Industry	Est. (std. err.)	Est. (std. err.)	Est. (std. err.)	Est. (std. err.)	Est. (std. err.)
01	0.522 (0.052)	0.478 (0.052)	0.451 (0.048)	0.399 (0.046)	0.216 (0.081)
02	-0.382 (0.047)	-0.277 (0.045)	-0.338 (0.049)	-0.121 (0.048)	0.195 (0.061)
03	0.241 (0.052)	0.227 (0.050)	0.232 (0.047)	0.218 (0.045)	0.952 (0.071)
04	0.566 (0.046)	0.531 (0.046)	0.553 (0.043)	0.484 (0.043)	0.458 (0.096)
05	0.328 (0.044)	0.292 (0.044)	0.340 (0.041)	0.284 (0.041)	0.271 (0.082)
06	0.662 (0.044)	0.663 (0.044)	0.689 (0.041)	0.613 (0.042)	1.054 (0.059)
07	-0.141 (0.024)	-0.099 (0.023)	-0.104 (0.023)	-0.053 (0.022)	0.211 (0.023)
08	0.823 (0.039)	0.691 (0.039)	0.813 (0.038)	0.626 (0.038)	0.939 (0.064)
09	0.356 (0.053)	0.336 (0.052)	0.332 (0.052)	0.264 (0.051)	0.521 (0.122)
10	0.439 (0.066)	0.473 (0.065)	0.676 (0.064)	0.444 (0.062)	0.836 (0.095)
11	0.597 (0.048)	0.551 (0.048)	0.271 (0.057)	0.236 (0.056)	0.402 (0.064)
12	-0.134 (0.103)	-0.206 (0.103)	0.011 (0.097)	0.028 (0.097)	0.552 (0.147)
13	0.137 (0.044)	0.119 (0.043)	0.129 (0.043)	0.120 (0.039)	0.403 (0.106)
14	0.554 (0.052)	0.540 (0.050)	0.641 (0.049)	0.524 (0.045)	0.357 (0.091)
15	0.306 (0.053)	0.390 (0.055)	0.383 (0.052)	0.438 (0.050)	0.850 (0.089)
16	0.790 (0.046)	0.851 (0.047)	0.809 (0.044)	0.860 (0.042)	0.428 (0.092)
17	0.678 (0.073)	0.615 (0.070)	0.558 (0.074)	0.631 (0.069)	0.837 (0.092)
log.lik.	-5525.6	-3487.21	-3596.77	-1082.02	4478.2
no. of par.	446	732	589	875	1161
LR vs 'all types'	20008[0.000]**	15930[0.000]**	16149[0.000]**	11120[0.000]**	

For name of industries, see Table 2.

hunting industry, 02) to 1.05 (in the Energy-intensive goods industry, 06). In all industries, the estimated elasticity of substitution is significantly different from zero.

5 Conclusions

In this paper, I have presented a procedure for estimating a system with two cross-sectional dimensions and interdependence in both of these dimensions. The procedure is an extension of the one in Pesaran (2006), where only one cross-sectional dimension is considered. The procedure is applied to estimate the elasticity of substitution between input factors where the two cross-sectional dimensions are industry and input factor. Hence, the approach allows some type of interdependence between input factors within one industry, but also interdependence between the same input factors in different industries. These types of interdependencies can be due to technological shifts. Such technological shifts can lead to changes in the the relative use of input factors in just one industry or be common across multiple industries. If the relative prices are correlated with the process for the technological change, we get biased estimates of the elasticity of substitution when not controlling for the technological changes.

The approach allows for a factor-neutral technological process. This process can differ between industries. A factor-neutral technological process will have the same effect on the optimal use of all input factors within each industry.

When estimating the substitution elasticity between up to 10 input factors within 17 industries, I find negative estimates in three industries when not including common factors to account for technological changes. However, when controlling for all types of technological changes, i.e. by including common factors both within and between industries, we get positive estimates of the substitution elasticities in all industries. Hence, these results illustrate the importance of taking into account technological changes that can also work across industries.

References

Abadir, K. M. and J. R. Magnus (2005), *Matrix Algebra*, Cambridge University Press, New York.

Antràs, P. (2004), 'Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution', *Contributions in Macroeconomics* **4**(1), ISSN (Online) 1534-6005, DOI: 10.2202/1534-6005.1161, April 2004

Bai, J. (2009), 'Panel Data Models With Interactive Fixed Effects', *Econometrica* **77**(4), 1229–1279. DOI: 10.3982/ECTA6135

Berndt, E. R . (1976), 'Reconciling Alternative Estimates of the Elasticity of Substitution', *The Review of Economics and Statistics* **58**(1), 59–68. DOI: 10.2307/1936009

Chudik, A. and M. H. Pesaran (2016), 'Theory and Practice of GVAR Modelling', *Journal of Economic Surveys* **30**(1), 165–197. DOI: 10.1111/joes.12095

Coakley, J., N. Kellard and S. Snaith (2005), 'The PPP Debate: Price Matters!', *Economics Letters* **88**(2), 209–213. DOI: 10.1016/j.econlet.2005.01.020

Diamond, P., D. McFadden, and M. Rodriguez (1978), 'Measurement of the Elasticity of Factor Substitution and the Bias of Technical Change'. In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, Amsterdam: North-Holland Publishing Company

Doornik, J. A. (2013), *Object-Oriented Matrix Programming using Ox*, Timberlake Consultants Press, London.

Doornik, J. A. and D. F. Hendry (2013), *Empirical Econometric Modelling Using PcGive 14: Volume I*, Timberlake Consultants Press, London.

Hungnes, H. (2011), 'A Demand System for Input Factors when there are Technological Changes in Production', *Empirical Economics* **40**(3), 581–600. DOI: 10.1007/s00181-010-0346-y

Hungnes, H. (2016), 'Using common factors to identify substitution possibilities in a factor demand system with technological changes', Discussion Papers 849, Statistics Norway http://www.ssb.no/en/forskning/discussion-papers/_attachment/284509

Kapetanios, G., M. H. Pesaran and T. Yamagata (2011), 'Panels with Non-Stationary Multifactor Error Structures', *Journal of Econometrics* **160**(2), 326–348. DOI: 10.1016/j.jeconom.2010.10.001

Klump, R., P. Mcadam and A. Willman (2012), 'The Normalized CES Production Function: Theory and Empirics', *Journal of Economic Surveys* **26**(5), 769–799. DOI: 10.1111/j.1467-6419.2012.00730.x

Leon-Ledesma, M. A., P. McAdam and A. Willman (2010), 'Identifying the Elasticity of Substitution with Biased Technical Change', *American Economic Review* **100**(4), 1330–1357. DOI: 10.1257/aer.100.4.1330

Pesaran, M. H. (2006), 'Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure', *Econometrica* **74**(4), 967–1012. DOI: 10.1111/j.1468-0262.2006.00692.x

Pesaran, M. H. (2015), *Time Series and Panel Data Econometrics*, Oxford University Press.

Pesaran, M. H., T. Schuermann and S. M. Weiner (2004), 'Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model', *Journal of Business & Economic Statistics* **22**(2), 129–162. DOI: 10.1198/073500104000000019

Urbain, J.-P. and J. Westerlund (2011), 'Least Squares Asymptotics in Spurious and Cointegrated Panel Regressions with Common and Idiosyncratic Stochastic Trends', *Oxford Bulletin of Economics and Statistics* **73**(1), 119–139. DOI: 10.1111/j.1468-0084.2010.00605.x

A Asymptotic properties

A.1 Asymptotic result for β_{ij}

Below A^{-} denotes a generalized inverse of A.

Lemma A.1 *Let Q, G and P be matrixes such that $Q = GP$ and where P has full column rank. Then $Q\left(Q'Q\right)^{-}Q' = G\left(G'G\right)^{-}G'$.*

Proof. First, $Q\left(Q'Q\right)^{-}Q' = Q\left(Q'Q\right)^{+}Q'$, where Q^{+} is the Moore-Penrose inverse of Q; see e.g. Abadir and Magnus (2005, Exercise 10.54). Then $Q\left(Q'Q\right)^{+}Q' = QQ^{+}$ follows from Abadir and Magnus (2005, Exercise 10.29). Inserting for Q we have $QQ^{+} = GP\left(GP\right)^{+}$, and applying that P has full row rank, it follows from Abadir and Magnus (2005, Exercise 10.40) that $GP\left(GP\right)^{+} = GG^{+}$. Applying all these results, we have the result in the lemma. ■

In matrix notation the model with the proxies for the common factors with multi-factor structure IV can be written as

$$Y_{ij} = X_{ij}\beta_{ij} + H_{wij}\Theta^{*}_{ij} + E^{*}_{ij}, \tag{42}$$

where Y_{ij}, X_{ij} and E^{*}_{ij} are defined in (31)–(33); and $H_{wij} = \left(D, \overline{Z}, \overline{Z}_{i.}, \overline{Z}_{.j}\right)$, where D, $\overline{Z}$, $\overline{Z}_{i.}$ and $\overline{Z}_{.j}$ are defined in (8)–(9). The estimator of β_{ij} is

$$\widehat{b}_{ij} = \left(X'_{ij}M_{wij}X_{ij}\right)^{-1}\left(X'_{ij}M_{wij}Y_{ij}\right), \tag{43}$$

where $M_{wij} = I_T - H_{wij}\left(H'_{wij}H_{wij}\right)^{-1}H'_{wij}$

The model with the common factors in matrix notation is given by

$$Y_{ij} = X_{ij}\beta_{ij} + G_{ij}\Theta_{ij} + E_{ij} \tag{44}$$

where $E_{ij} = \left(E_{ij1}, E_{ij2}, \ldots, E_{ijT}\right)'$ with dimension $T \times 1$ and $G_{ij} = \left(D, F, F_{i.}, F_{.j}\right)$ where F, $F_{i.}$ and $F_{.j}$ are defined in (10) and $\Theta_{ij} = \left(\alpha'_{ij}, \gamma'_{ij}, \gamma^{A\prime}_{ij}, \gamma^{B\prime}_{ij}\right)'$.

Inserting this model into the estimator yields a similar expression as in Pesaran (2006, eq.

29);

$$\begin{aligned}
\widehat{b}_{ij} - \beta_{ij} &= \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} F}{T}\right) + \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} F_{i.}}{T}\right) + \\
& \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} F_{.j}}{T}\right) + \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} E_{ij}}{T}\right) \\
&= \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} F_{ij}^*}{T}\right) + \left(\frac{X_{ij}' M_{wij} X_{ij}}{T}\right)^{-1} \left(\frac{X_{ij}' M_{wij} E_{ij}}{T}\right), \qquad (45)
\end{aligned}$$

where $F_{ij}^* = \left(F, F_{i.}, F_{.j}\right)$.

To evaluate this expression we need to apply the process for the exogenous variables

$$X_{ij} = G_{ij}^* \Pi_{ij} + V_{ij} \qquad (46)$$

and the proxies for the unobserved common factors

$$H_{wij} = G_{ij}^* P_{wij} + U_{wij}^* \qquad (47)$$

where $G_{ij}^* = \left(D, F, F^A, F^B, F_{i.}, F_{.j}\right)$, $\Pi_{ij} = \left(A_{ij}', 0_{k\times m^1}, 0_{k\times m^2}, \Gamma_{ij}', \Gamma_{ij}^{A\prime}, \Gamma_{ij}^{B\prime}\right)'$, $V_{ij} = \left(v_{ij1}, v_{ij2}, \ldots, v_{ijT}\right)'$,

$$P_{wij} = \begin{pmatrix} I_n & \overline{B}_w & \overline{B}_{iw} & \overline{B}_{wj} \\ 0 & \overline{C}_w & \overline{C}_{iw} & \overline{C}_{wj} \\ 0 & \overline{C}_w^A & \overline{C}_{iw}^A & \overline{C}_{wj}^A \\ 0 & \overline{C}_w^B & \overline{C}_{iw}^B & \overline{C}_{wj}^B \\ 0 & 0 & \overline{C}_{iw}^A & 0 \\ 0 & 0 & 0 & \overline{C}_{wj}^B \end{pmatrix}$$

and $U_{wij}^* = \left(0, U_w, U_{iw}, U_{jw}\right)$ with

$$\begin{aligned}
U_w &= \left(\overline{U}_1, \overline{U}_2, \ldots, \overline{U}_T\right)' \qquad (48) \\
U_{iw} &= \left(\overline{U}_{i1}^A, \overline{U}_{i2}^A, \ldots, \overline{U}_{iT}^A\right)' \qquad (49) \\
U_{jw} &= \left(\overline{U}_{j1}^B, \overline{U}_{j2}^B, \ldots, \overline{U}_{jT}^B\right)' \qquad (50)
\end{aligned}$$

Note that under Assumption 2.6 and assuming that both $\overline{C}^A_{iw}$ and $\overline{C}^B_{wj}$ have full rank, P_{wij} has full row rank, i.e., $rank(P_{wij}) = n + m_0 + 2m_1 + 2m_2$, which follows from repeatedly using that

$$\mathbf{Z} = rank \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \geq rank(\mathbf{A}) + rank(\mathbf{B}),$$

see, e.g., Abadir and Magnus (2005, Exercise 5.4), which holds with equality if both $\mathbf{A}$ and $\mathbf{D}$ have full row rank since the rank of $\mathbf{Z}$ cannot exceed its number of rows. If either $\overline{C}^A_{iw}$ or $\overline{C}^B_{wj}$ does not have full row rank, we can modify P_{wij} to a matrix with $n + m_0 + m_1 + m_2 + rank(\overline{C}^A_{iw}) + rank(\overline{C}^B_{wj})$ rows and full row rank, and adjust G^*_{ij} accordingly.[7]

Using the results from the Lemma 2 and 3 in Pesaran (2006), but adjusted to the formulation applied here, we have:

$$\frac{X'_{ij}H_{wij}}{T} = \frac{X'_{ij}G^*_{ij}}{T}P_{wij} + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \quad (51)$$

$$\frac{H'_{wij}H_{wij}}{T} = P'_{wij}\frac{G^{*\prime}_{ij}G^*_{ij}}{T}P_{wij} + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \quad (52)$$

$$\frac{H'_{wij}F^*_{ij}}{T} = P'_{wij}\frac{G^{*\prime}_{ij}F^*_{ij}}{T} + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \quad (53)$$

Proof of Proposition 2.1. Applying the results in (51)–(53) gives

$$\frac{X_{ij}M_{wij}F^*_{ij}}{T} = \frac{X_{ij}M_{qij}F^*_{ij}}{T} + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \quad (54)$$

[7] These adjusted matrices are

$$\tilde{P}_{wij} = \begin{pmatrix} I_n & \overline{B}_w & \overline{B}_{iw} & \overline{B}_{wj} \\ 0 & \overline{C}_w & \overline{C}_{iw} & \overline{C}_{wj} \\ 0 & \overline{C}^A_w & \overline{C}^A_{iw} & \overline{C}^A_{wj} \\ 0 & \overline{C}^B_w & \overline{C}^B_{iw} & \overline{C}^B_{wj} \\ 0 & 0 & \tilde{\overline{C}}^A_{iw} & 0 \\ 0 & 0 & 0 & \tilde{\overline{C}}^B_{wj} \end{pmatrix} \text{ and } \tilde{G}^*_{ij} = \begin{pmatrix} D & F & F^A & F^B & \tilde{F}_{i.} & \tilde{F}_{.j} \end{pmatrix},$$

with $\tilde{F}_{i.} = F_{i.}M_{i.}\left(M'_{i.}M_{i.}\right)^{-1}$, $\tilde{F}_{.j} = F_{.j}M_{.j}\left(M'_{.j}M_{.j}\right)^{-1}$, $\tilde{\overline{C}}^A_{iw} = M'_{i.}\overline{C}^A_{iw}$, and $\tilde{\overline{C}}^B_{wj} = M'_{.j}\overline{C}^B_{wj}$, with $M_{i.}$ and $M_{.j}$ being matrices of dimension $m_1 \times rank(\overline{C}^A_{iw})$ and $m_2 \times rank(\overline{C}^B_{wj})$, respectively, such that both $\tilde{\overline{C}}^A_{iw}$, and $\tilde{\overline{C}}^B_{wj}$ have full row rank. This implies that $M_{i.\perp}\overline{C}^A_{iw} = 0$ and $M_{.j\perp}\overline{C}^B_{wj} = 0$ (where $M_\perp$ denotes the orthogonal complement to M, see, e.g., Abadir and Magnus (2005, p. 46), such that we have $\tilde{G}^*_{ij}\tilde{P}_{wij} = G^*_{ij}P_{wij}$, where $\tilde{P}_{wij}$ has full rank.

where $M_{qij} = I_T - Q_{wij}\left(Q'_{wij}Q_{wij}\right)^{-} Q'_{wij}$ with $Q_{wij} = G^*_{ij}P_{wij}$. When the rank conditions in Assumption 2.6 hold, we have $M_{qij} = M_{gij} = I_T - G^*_{ij}\left(G^{*\prime}_{ij}G^*_{ij}\right)^{-} G^{*\prime}_{ij}$, see Lemma A.1. In addition, since $F^*_{ij} \subset G^*_{ij}$, then $M_{qij}F^*_{ij} = M_{gij}F^*_{ij} = 0$, and

$$\frac{X_{ij}M_{wij}F^*_{ij}}{T} = O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \tag{55}$$

Similarly, we have

$$\frac{X_{ij}M_{wij}X^*_{ij}}{T} = \frac{X_{ij}M_{qij}X^*_{ij}}{T} + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right) + O_p\left(\frac{1}{\sqrt{N^A T}}\right) + O_p\left(\frac{1}{\sqrt{N^B T}}\right) \tag{56}$$

and

$$\frac{X_{ij}M_{wij}\varepsilon_{ij}}{T} = \frac{X_{ij}M_{qij}\varepsilon_{ij}}{T} + O_p\left(\frac{1}{N^A}\right) + O_p\left(\frac{1}{N^B}\right), \tag{57}$$

where we again can replace M_{qij} with M_{gij} when Assumption 2.6 holds. Finally, $T^{-1}X'_{ij}M_{qij}X_{ij} = O_p(1)$, and we have the results in the proposition. ■

B Asymptotic biases in the Monte Carlo experiment

B.1 The naïve estimator

Here the estimator is given by

$$\widehat{\beta}_{ij}^{naive} = \left(x'_{ij}M_1x_{ij}\right)^{-1}\left(x'_{ij}M_1y_{ij}\right) \text{ with } M_1 = I_T - 1_{T\times 1}\left(1'_{T\times 1}1_{T\times 1}\right)^{-1}1'_{1\times 1} = I_T - T^{-1}1_{T\times T}$$

and the data generating process is given by

$$y_{ij} = \alpha_{ij} + x_{ij}\beta_{ij} + f^A_i\gamma^A_{yij} + f^B_j\gamma^B_{yij} + \varepsilon_{ij} \text{ with } \alpha_{ij} = 0.$$

Inserting the latter into the former yields an expression for the estimation bias as

$$\begin{aligned}&\widehat{\beta}_{ij}^{naive} - \beta_{ij} \\ = \;& \left(T^{-1}x_{ij}'M_1x_{ij}\right)^{-1}\left[\left(T^{-1}x_{ij}'M_1f_i^A\right)\gamma_{yij}^A + \left(T^{-1}x_{ij}'M_1f_j^B\right)\gamma_{yij}^B + \left(T^{-1}x_{ij}'M_1\varepsilon_{ij}\right)\right]. \quad (58)\end{aligned}$$

Inserting from the DGP for unobservable common factors the nominator of the bias in (58) becomes

$$\begin{aligned}&\left(T^{-1}x_{ij}'M_1f_i^A\right)\gamma_{yij}^A + \left(T^{-1}x_{ij}'M_1f_j^B\right)\gamma_{yij}^B + \left(T^{-1}x_{ij}'M_1\varepsilon_{ij}\right) \\ = \;& \gamma_{xij}^A\left(T^{-1}f_i^{A\prime}M_1f_i^A\right)\gamma_{yij}^A + \gamma_{xij}^B\left(T^{-1}f_j^{B\prime}M_1f_i^A\right)\gamma_{yij}^A + \left(T^{-1}v_{ij}'M_1f_i^A\right)\gamma_{yij}^A \\ & + \gamma_{xij}^A\left(T^{-1}f_i^{A\prime}M_1f_j^B\right)\gamma_{yij}^B + \gamma_{xij}^B\left(T^{-1}f_j^{B\prime}M_1f_j^B\right)\gamma_{yij}^B + \left(T^{-1}v_{ij}'M_1f_j^B\right)\gamma_{yij}^B \\ & + \left(T^{-1}x_{ij}'M_1\varepsilon_{ij}\right),\end{aligned}$$

where the terms $\gamma_{xij}^A\left(T^{-1}f_i^{A\prime}M_1f_i^A\right)\gamma_{yij}^A$ and $\gamma_{xij}^B\left(T^{-1}f_j^{B\prime}M_1f_j^B\right)\gamma_{yij}^B$ converges to unity by construction of the data generating process and remaining terms converges to zero due to Assumption 2.2.

The denominator of the bias in (58) becomes (by applying (37))

$$\begin{aligned}\left(T^{-1}x_{ij}'M_1x_{ij}\right) = \;& \left(T^{-1}f_i^{A\prime}M_1f_i^A\right) + \left(T^{-1}f_i^{A\prime}M_1f_j^B\right) + \left(T^{-1}f_i^{A\prime}M_1v_{ij}\right) \\ & + \left(T^{-1}f_j^{B\prime}M_1f_i^A\right) + \left(T^{-1}f_j^{B\prime}M_1f_j^B\right) + \left(T^{-1}f_j^{B\prime}M_1v_{ij}\right) \\ & + \left(T^{-1}v_{ij}'M_1f_i^A\right) + \left(T^{-1}v_{ij}'M_1f_j^B\right) + \left(T^{-1}v_{ij}'M_1v_{ij}\right),\end{aligned}$$

and by construction of the data generating process the terms $\left(T^{-1}f_i^{A\prime}M_1f_i^A\right)$, $\left(T^{-1}f_j^{B\prime}M_1f_j^B\right)$, and $\left(T^{-1}v_{ij}'M_1v_{ij}\right)$ converges to unity; and $\left(T^{-1}f_i^{A\prime}M_1f_j^B\right)$ and $\left(T^{-1}f_j^{B\prime}M_1f_i^A\right)$ converges to zero. By Assumption 2.2 the terms $\left(T^{-1}f_i^{A\prime}M_1v_{ij}\right)$, $\left(T^{-1}f_j^{B\prime}M_1v_{ij}\right)$, $\left(T^{-1}v_{ij}'M_1f_i^A\right)$, and $\left(T^{-1}v_{ij}'M_1f_j^B\right)$ converges to zero. Applying these in the bias yields

$$\widehat{\beta}_{ij}^{naive} - \beta_{ij} \xrightarrow{p} \frac{2}{3}.$$

The mean group estimator will have the same bias.

Table 4: Results from simulations: $\hat{\beta}_{MG}$ — stationary case

	infeasible	naïve	overall c.f.	c.f. in the 1. dim.	both	all	special
Proxies	$\begin{pmatrix} f_{it}^A \\ f_{jt}^B \end{pmatrix}$	0	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \end{pmatrix}$	$\begin{pmatrix} \bar{y}_{it} \\ \bar{x}_{it} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \\ \bar{y}_{it} \\ \bar{x}_{it} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_t \\ \bar{x}_t \\ \bar{y}_{it} \\ \bar{x}_{it} \\ \bar{y}_{jt} \\ \bar{x}_{jt} \end{pmatrix}$	$\begin{pmatrix} \bar{y}_{it} \\ \bar{x}_{it} \\ \bar{y}_{jt} \\ \bar{x}_{jt} \end{pmatrix}$
Asympt. bias	0	2/3	2/3	1/2	1/2	0	0

In the table, "c.f." is used for "common factors" and '1. dim" is short for "first cross-sectional dimension", "both" is the joint of "overall c.f." and "c.f. in 1. dim.'

	infeasible	naïve	overall c.f.	c.f. in the 1.dim.	both	all	special
mean	1.0001	1.6665	1.6643	1.4999	1.4999	1.0001	1.0191
min	0.9952	1.6567	1.6539	1.4826	1.4830	0.9950	1.0100
max	1.0036	1.6754	1.6729	1.5151	1.5144	1.0038	1.0315
st.d.*100	0.1301	0.2933	0.3004	0.4689	0.4723	0.1314	0.3575

In this simulation: 1000 replications, $N^A = N^B = 100$, $T = 100$, $\alpha = 0$, $\rho_{fi}^A = \rho_{fj}^B = \rho_{ij} = \rho_{\varepsilon ij} = 0.5, \forall i, j$, $\beta_{ij} = 1, \forall i, j$, $\gamma_{yij}^A = \gamma_{yij}^B = \gamma_{xij}^A = \gamma_{xij}^B = 1, \forall i, j$.

	infeasible	naïve	overall c.f.	c.f. in the 1. dim.	both	all	special
mean	1.0004	1.6672	1.6444	1.5022	1.4994	0.9965	1.1460
min	0.8304	1.5095	1.4789	1.2824	1.2152	0.4344	0.8711
max	1.1901	1.8213	1.8380	1.6901	1.7274	1.3104	1.3502
st.d.*100	5.5142	4.5134	4.9570	6.4732	7.4627	10.169	7.8590

In this simulation: As above, except $N^A = N^B = 10$, $T = 10$.

C Monte Carol result with stationary variables

Table 4 reports the results for the stationary Monte Carlo simulation.