# Working Papers



# Firm-level Productivity Spillovers in China's Chemical Industry: A Spatial Hausman-Taylor Approach

Badi H. Baltagi
Peter H. Egger
Michaela Kesina

CESIFO WORKING PAPER NO. 5114
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
DECEMBER 2014



CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5114*

# Firm-level Productivity Spillovers in China's Chemical Industry: A Spatial Hausman-Taylor Approach

## Abstract

This paper assesses the role of intra-sectoral spillovers in total factor productivity across Chinese producers in the chemical industry. We use a rich panel data-set of 12,552 firms observed over the period 2004-2006 and model output by the firm as a function of skilled and unskilled labor, capital, materials, and total factor productivity, which is broadly defined. The latter is a composite of observable factors such as export market participation, foreign as well as public ownership, the extent of accumulated intangible assets, and unobservable total factor productivity. Despite the richness of our data-set, it suffers from the lack of time variation in the number of skilled workers as well as in the variable indicating public ownership. We introduce spatial spillovers in total factor productivity through contextual effects of observable variables as well as spatial dependence of the disturbances. We extend the Hausman and Taylor (1981) estimator to account for spatial correlation in the error term. This approach permits estimating the effect of time-invariant variables which are wiped out by the fixed effects estimator. While the original Hausman and Taylor (1981) estimator assumes homoskedastic error components, we provide spatial variants that allow for both homoskedasticity and heteroskedasticity. Monte Carlo results show, that our estimation procedure performs well in small samples. We find evidence of positive spillovers across chemical manufacturers and a large and significant detrimental effect of public ownership on total factor productivity.

JEL-Code: C230, C310, D240, L650.

Keywords: technology spillovers, spatial econometrics, panel data econometrics, firm-level productivity, Chinese firms.

*Badi H. Baltagi*
*Center for Policy Research*
*Syracuse University*
*USA – Syracuse, NY 13244-1020*
*bbaltagi@maxwell.syr.edu*

*Peter H. Egger**
*ETH Zurich*
*Leonhardstrasse 21*
*Switzerland – 8092 Zurich*
*egger@kof.ethz.ch*

*Michaela Kesina*
*ETH Zurich*
*Leonhardstrasse 21*
*Switzerland – 8092 Zurich*
*kesina@kof.ethz.ch*

*corresponding author
October 2014
The authors gratefully acknowledge numerous helpful comments by the editor (Hashem Pesaran) and two anonymous reviewers. Moreover, they are indebted to the participants at the Annual World Conference of the Spatial Econometrics Association at Toulouse 2011, the European Annual Meeting of the Econometric Society at Oslo 2011, CESifo Venice Summer Institute on China and the Global Economy Post Crisis 2011, the 17th International Conference on Panel Data held at McGill University Montreal 2011, the Singapore Economic Review Conference at Singapore 2011, the Annual Meeting of the European Economic Association at Gothenburg 2013, and the Conference on Cross-sectional Dependence in Panel Data Models at Cambridge 2013.

# 1 Introduction

There is now broad firm-level evidence suggesting that total factor productivity (TFP) is contagious and prone to spillovers which are geographically bounded. Sources of such spillovers are technology gaps between firms, research and development, and access to knowledge through exporting and foreign ownership. Evidence from Europe includes Smarzynska Javorcik (2004) and Lööf (2007), to mention a few. Evidence from the United States includes Keller and Yeaple (2009) and Bloom, Schankerman, and van Reenen (2007), to mention a few. For China, see Hu *et al.* (2005), Chen and Swenson (2006), and Blonigen and Ma (2007), to mention a few examples.

Important transmission channels of TFP spillovers are worker flows, cooperation in research and development across firms, and other (not directly measurable but geographically bounded) forms of dissemination of knowledge and productivity.[1]

The goal of this paper is to assess the presence and relative strength of geographically bound TFP spillovers which originate in observable and unobservable determinants of TFP. The panel data utilized in this study are based on the universe of all firms in China's chemical industry with a turnover exceeding about 700,000 US dollars over the period 2004–2006. The data-set covers 12,552 firms and 37,656 observations. Our focus on the period 2004–2006 is dictated by the quality of data in recent periods and the availability of data on skilled workers. The chemical sector is relatively important in many countries of transition such as China (UNEP Chemicals Branch, 2009; it accounts for about 8 percent of employment of all Chinese manufacturing firms). The average firm in that sector is relatively large (about 257 employees). Although the number of publicly-owned firms account for a little more than 5 percent of all firms in that sector in China over 2004–2006, they employ on average 793 employees as compared to 166 employees for domestically-owned firms; see also Szamosszegi and Kyle (2011).

One main goal of the study is to disentangle the direct role of production factors (skilled and unskilled labor, capital, and material inputs) on the one hand from observable and unobservable determinants of TFP defined in a broad sense. We account for four observable shifters of TFP: export status, foreign ownership, the extent of intangible assets accumulated by a firm, and public ownership status. Export market contact has been found to affect TFP through competitive pressure as well as learning (see Baldwin and Gu, 2006; and Greenaway and Kneller, 2007). Foreign ownership of firms in developing countries and economies in transition offers access to better technologies of foreign subsidiaries and even of other firms in host countries. The beneficiaries include input suppliers (see Smarzynska Javorcik, 2004) and firms which receive worker flows from more productive, foreign-owned units (see Görg and Strobl, 2005). For the foreign subsidiaries themselves, foreign ownership may also lead to specialization on less-advanced production stages. Hence, foreign ownership may impede knowledge transfer, if foreign owners fear the loss of intellectual property through involuntary dissemination of knowledge to other firms (see Puttitanun, 2006), or if foreign entities do not have the absorptive capacity to utilize more advanced technologies available from the parent (see Keller, 2004). The corporate finance literature suggests that the degree of asset (in)tangibility is an important factor in determining access to external credit, opportunities to finance investments with little collateral value such as research and development (R&D), in sustaining growth,

[1] As pointed out by Smarzynska Javorcik (2004) and, in particular, by Bloom, Schankerman, and van Reenen (2007), what is dubbed technology spillovers in empirical work consists of two main components: technology transmission in a narrow sense and interdependence across firms by market structure. Empirical work seldomly disentangles these components and should therefore speak of spillovers in a broad sense.

especially, of young companies (Himmelberg and Petersen, 1994; Hall and Lerner, 2010), and in facilitating knowledge absorption. Finally, public ownership is commonly regarded as a stepping stone to technological advancement due to the lack of competitive pressure (see González-Páramo and Hernández Cos, 2005).

Apart from these shifters, we allow TFP for a particular firm to depend spatially on other firms within a geographical neighborhood. In this paper, we refer to the latter broadly as TFP spillovers. More specifically, we consider immediate effects (also referred to as *local spillovers* or *contextual effects*) of neighboring firms' observable TFP characteristics on a given firm and, at the same time, immediate as well as indirect effects (also referred to as *global spillovers*) from neighboring firms' unobservable TFP characteristics on a given firm. We model the extent of such spillovers to decline with distance. Accordingly, we refer to them as spatial spillovers. This is done by using information on a firm's geographical location as available from the data (by way of a six-digit ZIP code). However, the presence of time-invariant variables – such as skilled labor input and public ownership – and their possible correlation with unobservable firm-specific effects require us to adopt a Hausman and Taylor (1981) estimation approach, hereafter denoted by HT. Unlike the fixed effects estimator, the HT estimator does not wipe out the time-invariant variables. Instead, it uses the between variation of the time-varying exogenous variables to instrument for endogenous time-invariant regressors. Our estimation methodology modifies the HT methodology to allow for spatial correlation in the error term. Moreover, we derive and propose homoskedastic as well as heteroskedasticity-robust variants of the spatial HT estimator (the original, non-spatial HT model assumes homoskedasticity). We provide evidence based on Monte Carlo simulations that both the homoskedastic and the heteroskedastic spatial HT estimators perform well in small samples.

Our empirical findings can be summarized as follows. For China's chemical industry, a high foreign ownership ratio and export-market participation are statistically significant and lead to a positive shift in TFP. Second, public ownership is associated with dramatically lower TFP than private ownership. Clearly, publicly-owned firms are relatively large (in terms of assets and employees), they account for a significant share of the industry's output, but their output is relatively small once controlling for their factor usage. We estimate that the average publicly-owned firm's TFP is about 84 percent lower than the average privately-owned firm's TFP. The data suggest that shocks in TFP are contagious within the industry. Hence, negative TFP shocks do not only affect firms that are hit by those shocks but also their geographic neighbors. Using inverse geographical distance to parameterize the spatial weights matrix across firms, we estimate the spatial dependence parameter at about 0.3 to 0.5. This parameter is statistically significant at 1 percent and is estimated on a spatial weights matrix which is full and of size $12,552 \times 12,552$. One of the merits of the generalized moments approach is that it can cope with spatial problems of this size. We find that restricting the scope of spillovers to firms within a smaller neighborhood – say, within a radius of 60, 100, or 200 miles – does not change the results much.

The remainder of the paper is organized as follows. The next section introduces the econometric model and the spatial Hausman and Taylor (1981) approach (SHT), where we distinguish between homoskedastic and heteroskedastic errors. That section also puts the SHT estimator in context with its spatial random effects (SRE) and spatial fixed effects (SFE) counterparts. Section 3 provides some Monte Carlo experiments on the small sample performance of the estimator. Section 4 summarizes the empirical results for spatial (cum spillover) translog models. The last section concludes with a summary of the key findings.

# 2 Econometric model

In this section, we specify an econometric error components model with spatially autocorrelated disturbances and right-hand side variables that are correlated with the time-invariant firm effects. We allow for two different error term structures. We start with a purely homoskedastic world as in Hausman and Taylor (1981). Since this assumption might be too restrictive in practice, we alternatively allow the idiosyncratic error component to be heteroskedastic. The next subsection describes the econometric model and the notation. Then we outline general assumptions, and afterwards describe the estimation procedures for both scenarios.

## 2.1 Model outline

Consider a large cross section of $N$ spatial units (firms), which are observed repeatedly across a small number of $T$ time periods.[2] We use the subscript $i = 1, ..., N$ to refer to individual units, and $t = 1, ..., T$ to refer to time periods. We specify a Cliff-Ord-type spatial model to describe the interaction between firms at period $t$ as follows:

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\theta} + \mathbf{u}_t = \mathbf{A}_t\boldsymbol{\delta} + \mathbf{u}_t, \quad \mathbf{u}_t = \rho\mathbf{W}\mathbf{u}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\nu}_t \tag{1}$$

where $\mathbf{A}_t = [\mathbf{X}_t, \mathbf{Z}]$, and $\boldsymbol{\delta} = \left[\boldsymbol{\beta}', \boldsymbol{\theta}'\right]'$. Here, $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})'$ is an $N \times 1$ vector of observations on the dependent variable at time $t$, $\mathbf{X}_t$ is an $N \times K$ matrix of time-varying regressors for period $t$, $\mathbf{Z}$ is an $N \times R$ matrix of time-invariant regressors.[3] Some of the regressors in $\mathbf{A}_t$ will be allowed to be correlated with $\boldsymbol{\mu}$. $\mathbf{W}$ is an $N \times N$ observed non-stochastic weights matrix whose properties will be specified below. $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$ is the $N \times 1$ vector of regression disturbances, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ is an $N \times 1$ vector of innovations which consist of a time-invariant error component $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$ and a time-varying idiosyncratic error component $\boldsymbol{\nu}_t = (\nu_{1t}, ..., \nu_{Nt})'$ . The vector $\mathbf{W}\mathbf{u}_t$ represents a spatial lag of $\mathbf{u}_t$. The scalar $\rho$ denotes the spatial autoregressive parameter, while $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are $K \times 1$ and $R \times 1$ vectors of regression parameters.[4]

In matrix form, we sort the data first by time $t$ (the slow running index) and then by firms $i$ (the fast running index) as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\boldsymbol{\iota}_T \otimes \mathbf{Z})\boldsymbol{\theta} + \mathbf{u} = \mathbf{A}\boldsymbol{\delta} + \mathbf{u}, \quad \mathbf{u} = \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \mathbf{Z}_\mu\boldsymbol{\mu} + \boldsymbol{\nu}, \tag{2}$$

where $\boldsymbol{\iota}_T$ denotes a $T \times 1$ vector of ones and $\mathbf{I}_T$ denotes a $T \times T$ identity matrix. $\mathbf{Z}_\mu = \boldsymbol{\iota}_T \otimes \mathbf{I}_N$ is an $NT \times N$ selector matrix of ones and zeroes. For subsequent use, we define the (between) projection matrix $\mathbf{Q}_1 = \overline{\mathbf{J}}_T \otimes \mathbf{I}_N$, where $\overline{\mathbf{J}}_T \equiv T^{-1}\mathbf{J}_T$. The matrix $\mathbf{J}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}'_T$ is a $T \times T$ matrix with unitary elements. The sweeping (within transformation) matrix is given $\mathbf{Q}_0 = \mathbf{I}_{TN} - \mathbf{Q}_1$ (see Baltagi, 2008), where $\mathbf{I}_N$ and $\mathbf{I}_{TN}$ denote identity matrices of dimension $N \times N$ and $TN \times TN$, respectively.

[2]The data available in this study do not permit distinguishing between common factors and spatial correlation in a more narrow sense. The former are in the limelight of a recent literature in theoretical econometrics which focuses on data situations with a relatively large number of time periods (see Pesaran, 2006; Pesaran and Tosetti, 2011; or Pesaran *et al.*, 2013; for a few recent examples of such research).

[3]In terms of the jargon adopted in the introduction, we will consider *local TFP spillovers* (or *contextual effects*) to be part of $\mathbf{A}_t\boldsymbol{\delta}$, while *global TFP spillovers* are reflected by the presence of $\rho\mathbf{W}\mathbf{u}_t$ in (1).

[4]Kapoor *et al.* (2007) use the subscript $N$ to indicate that the elements of all data vectors and matrices as well as of all parameters may depend on the cross section sample size $N$. We skip this subscript to avoid index cluttering.

### 2.2 General assumptions

This subsection contains assumptions that will be maintained throughout the whole paper.

**Assumption 1** *(Assumptions on $\boldsymbol{W}$ and $\rho$)*
*(i) The matrix $\mathbf{W}$ is rowsum normalized. All diagonal elements of $\mathbf{W}$ are zero. (ii) The admissible parameter space for $\rho$ is $\rho \in (-1,1)$. (iii) The matrix $\mathbf{I}_N - \rho\mathbf{W}$ is nonsingular for $\rho \in (-1,1)$. (iv) The row and column sums of $\mathbf{W}$, $(\mathbf{I}_N - \rho\mathbf{W})$ and $(\mathbf{I}_N - \rho\mathbf{W})^{-1}$ are bounded uniformly in absolute value.*

Assumption 1 is a standard normalization of $\mathbf{W}$ which ensures that the shocks in the interdependent system have finite consequences, and it allows the researcher to infer directly from the estimate of $\rho$ whether this is the case.

**Assumption 2** *(Covariates)*
*(i) The covariates $\mathbf{A}$ have full column rank and their elements are bounded uniformly in absolute value. (ii) All columns of $\mathbf{A}$ are assumed to be uncorrelated with $\boldsymbol{\nu}$. (iii) $\lim_{N\to\infty}(NT)^{-1}\mathbf{X}'\mathbf{Q}_0\mathbf{X}$ exists, is finite and nonsingular.*

Assumption 2 summarizes some basic assumptions on all covariates.

**Assumption 3** *(Hausman and Taylor Assumptions)*
*(i) The regressor matrices are decomposed into $\mathbf{X} = [\mathbf{X}_U, \mathbf{X}_C]$ and $\mathbf{Z} = [\mathbf{Z}_U, \mathbf{Z}_C]$, where $\mathbf{X}_f$ is $NT \times K_f$ and $\mathbf{Z}_f$ is $N \times R_f$ for $f = U, C$. In Hausman and Taylor's (1981) notation, $[\mathbf{X}_C, \boldsymbol{\iota}_T \otimes \mathbf{Z}_C]$ are correlated with $\boldsymbol{\mu}$, whereas $\mathbf{H}_I = [\mathbf{X}_U, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U]$ are not. (ii) The Hausman and Taylor (1981) order condition for identification, i.e., $R_C \leq K_U$ holds throughout our analysis.*

Assumption 3 is in line with the decomposition of the covariates in Hausman and Taylor (1981). We have $E(\mathbf{X}'\mathbf{Q}_0\mathbf{u}) = \mathbf{0}$, $E(\mathbf{X}_U'\mathbf{u}) = \mathbf{0}$, $E((\boldsymbol{\iota}_T \otimes \mathbf{Z}_U)'\mathbf{u}) = \mathbf{0}$, but $E((\boldsymbol{\iota}_T \otimes \mathbf{Z}_C)'\mathbf{u}) \neq \mathbf{0}$ and $E(\mathbf{X}'\mathbf{Q}_1\mathbf{u}) \neq \mathbf{0}$ due to $E(\mathbf{X}_C'\mathbf{u}) \neq \mathbf{0}$. Thus the Hausman and Taylor instrument set[5] is given by

$$\mathbf{H}_{HT} = [\mathbf{Q}_0\mathbf{X}, \mathbf{Q}_1\mathbf{X}_U, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U] = [\mathbf{Q}_0\mathbf{X}_C, \mathbf{H}_I] \tag{3}$$

where $\mathbf{H}_I$ is the instrument set used for the initial estimator. The following assumptions hold for the instrument set $\mathbf{H}_{HT}$.

**Assumption 4** *(Instrument set $\mathbf{H}_{HT}$)*
*(i) The instruments are uncorrelated with the error $\boldsymbol{\varepsilon}$. (ii) The matrix $\mathbf{H}_{HT}$ has full column rank. (iii) The elements of $\mathbf{H}_{HT}$ are uniformly bounded in absolute value. (iv) $\lim_{N\to\infty}[(NT)^{-1}\mathbf{H}_I'\mathbf{H}_I]$ exists, is finite and nonsingular. (v) $p\lim_{N\to\infty}[(NT)^{-1}\mathbf{H}_I'\mathbf{Z}]$ exists, is finite and has full column rank.*

Assumption 4 summarizes standard assumptions that need to hold in instrumental variable procedures. Parts (iv) and (v) are needed for the initial Hausman and Taylor estimator described in the next subsection. Furthermore we will make use of the following transformation of (3) $\mathbf{H}^*_{SHT} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})][\mathbf{Q}_0\mathbf{X}, \mathbf{Q}_1\mathbf{X}_U, \mathbf{Z}_U]$, and use the following assumption, where we suppress the subindex $SHT$ for the sake of brevity

[5] Alternative sets of instruments to $\boldsymbol{H}_{HT}$ can be formulated in the spirit of Amemiya and MaCurdy (1986) ($\boldsymbol{H}_{AM}$) and Breusch *et al.* (1989) ($\boldsymbol{H}_{BMS}$): $\boldsymbol{H}^*_{AM} = [\mathbf{Q}_0\mathbf{X}, \mathbf{X}^0_{U1}, ..., \mathbf{X}^0_{UT}, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U]$, $\boldsymbol{H}^*_{SBMS} = [\mathbf{Q}_0\mathbf{X}, \mathbf{Q}_0\mathbf{X}^0_{U1}, ..., \mathbf{Q}_0\mathbf{X}^0_{UT}, \mathbf{Q}_0\mathbf{X}^0_{C1}, ..., \mathbf{Q}_0\mathbf{X}^0_{CT}, \mathbf{Q}_1\mathbf{X}_U, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U]$, where $\mathbf{X}^0_{Ut} = \mathbf{X}_{Ut} \otimes \boldsymbol{\iota}_T$ is a $TN \times K_U$ matrix for all $t = 1, ..., T$.

**Assumption 5** *(Instrument set* $\mathbf{H}^*$*)*
*(i)* $\mathbf{H}^*$ *has full column rank. (ii) The elements of* $\mathbf{H}^*$ *are uniformly bounded in absolute value.*

Note that Assumption 5 can be viewed as a corrollary of Assumptions 1 and 4, and it is mentioned here merely for completeness.

## 2.3 SHT estimation under homoskedasticity

### 2.3.1 Assumptions

We follow the standard assumptions given in Kapoor *et al.* (2007) for the random effects spatial panel model under homoskedasticity.

**Assumption 6** *(Assumptions on the error components)*
*(i) For the idiosnycratic error components we have* $\nu_{it} \sim i.i.d.(0, \sigma_\nu^2)$*, where* $0 < \sigma_\nu^2 < \infty$ *and* $E|v_{it}|^{4+\eta} < \infty$ *for some* $\eta > 0$*. (ii) For the unit-specific error components we have* $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$*, where* $0 < \sigma_\mu^2 < b_\mu < \infty$*, and* $E|\mu_i|^{4+\eta} < \infty$ *for some* $\eta > 0$*. (iii)* $\mu_i$ *and* $\nu_{it}$ *are independent of each other for all* $i$ *and* $t$*.*

Assumption 6 implies the following covariance of $\varepsilon_{it}$ and $\varepsilon_{js}$: $Cov(\varepsilon_{it}\varepsilon_{js}) = \sigma_\mu^2 + \sigma_\nu^2$ for $i = j$ and $t = s$; $Cov(\varepsilon_{it}\varepsilon_{js}) = \sigma_\mu^2$ for $i = j$ and $t \neq s$; and $Cov(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise, see Baltagi (2008).

### 2.3.2 Estimation procedure

**STEP 1 - Estimate $\boldsymbol{\beta}$:** From the assumptions of our model, one can estimate $\boldsymbol{\beta}$ consistently using the fixed effects estimator $\widehat{\boldsymbol{\beta}}_{FE} = (\mathbf{X}'\mathbf{Q}_0\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_0\mathbf{y}$. This provides a consistent estimate of the residuals $\mathbf{d} \equiv (\boldsymbol{\iota}_T \otimes \mathbf{Z})\boldsymbol{\theta} + \mathbf{u}$ given by $\widehat{\mathbf{d}} \equiv \mathbf{y} - \mathbf{X}\widehat{\beta}_{FE}$.

**STEP 2 - Estimate $\boldsymbol{\theta}$:** Now one can retrieve a consistent estimate $\boldsymbol{\theta}$ which was wiped out by the fixed effects estimator. First, we average the residuals in $\widehat{\mathbf{d}}$ over time by computing $\mathbf{Q}_1\widehat{\mathbf{d}}$. Then, we run 2SLS of $\mathbf{Q}_1\widehat{\mathbf{d}}$ on $\mathbf{Z}$ using $\mathbf{H}_I = [\mathbf{X}_U, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U]$ as our instruments. This is the two-step initial Hausman and Taylor estimator and it is given by $\widehat{\boldsymbol{\theta}}_{2SLS} = ((\boldsymbol{\iota}_T \otimes \mathbf{Z})'\mathbf{P}(\boldsymbol{\iota}_T \otimes \mathbf{Z}))^{-1}(\boldsymbol{\iota}_T \otimes \mathbf{Z})'\mathbf{P}\mathbf{Q}_1\widehat{\mathbf{d}}$, where $\mathbf{P} = \mathbf{H}_I(\mathbf{H}_I'\mathbf{H}_I)^{-1}\mathbf{H}_I'$ is the projection matrix on the matrix of instruments $\mathbf{H}_I$. This estimator is consistent (see Hausman and Taylor, 1981), but it is not efficient, since it ignores the error component structure of the data. However the initial estimator yields consistent estimates of $\mathbf{u}$ given by $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{FE} - (\boldsymbol{\iota}_T \otimes \mathbf{Z})\widehat{\boldsymbol{\theta}}_{2SLS}$, which will be used in the subsequent procedure to account for the error component structure of the data in order to improve the efficiency of the estimator.

**STEP 3 - Estimate $\rho$, $\sigma_\nu^2$, and $\sigma_\mu^2$:** Using these $\widehat{\mathbf{u}}$'s we can directly apply the moment conditions in Kapoor *et al.* (2007) to obtain estimates of $\rho$ and the variance components $\sigma_\nu^2$ and $\sigma_\mu^2$.[6] This is true even though the columns in $\mathbf{A}$ are correlated with $\mathbf{u}$ (see Kelejian and Prucha, 2004; Drukker *et al.*, 2013; for a set of assumptions accommodating the

[6]Note that this is different from the pooled OLS residuals used in Kapoor *et al.* (2007). In the Hausman and Taylor (1981) case, the OLS residuals would lead to inconsistent estimates of $\mathbf{u}$ due to the correlation of $[\mathbf{X}_C, \boldsymbol{\iota}_T \otimes \mathbf{Z}_C]$ with $\boldsymbol{\mu}$.

endogenous regressors in $\mathbf{A}$). The six moment conditions are given by

$$E\left[\frac{1}{N(T-1)}\boldsymbol{\varepsilon}'\mathbf{Q}_0\boldsymbol{\varepsilon}\right]=\sigma_\nu^2,\quad E\left[\frac{1}{N(T-1)}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\overline{\boldsymbol{\varepsilon}}\right]=\sigma_\nu^2\frac{1}{N}tr(\mathbf{W}'\mathbf{W}),\quad E\left[\frac{1}{N(T-1)}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\boldsymbol{\varepsilon}\right]=0,$$
$$E\left[\frac{1}{N}\boldsymbol{\varepsilon}'\mathbf{Q}_1\boldsymbol{\varepsilon}\right]=\sigma_1^2,\quad E\left[\frac{1}{N}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_1\overline{\boldsymbol{\varepsilon}}\right]=\sigma_1^2\frac{1}{N}tr(\mathbf{W}'\mathbf{W}),\quad E\left[\frac{1}{N}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_1\boldsymbol{\varepsilon}\right]=0,$$

where $\overline{\boldsymbol{\varepsilon}}\equiv(\mathbf{I}_T\otimes\mathbf{W})\boldsymbol{\varepsilon}$ and $\sigma_1^2=T\sigma_\mu^2+\sigma_\nu^2$. These can be rewritten in terms of $\mathbf{u}$ using the fact that $\boldsymbol{\varepsilon}=(\mathbf{I}_T\otimes[\mathbf{I}_N-\rho\mathbf{W}])\mathbf{u}=\mathbf{u}-\rho\overline{\mathbf{u}}$ whereby $\overline{\mathbf{u}}\equiv(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}$ and $\overline{\boldsymbol{\varepsilon}}\equiv(\mathbf{I}_T\otimes\mathbf{W})(\mathbf{I}_T\otimes[\mathbf{I}_N-\rho\mathbf{W}])\mathbf{u}=\overline{\mathbf{u}}-\rho\overline{\overline{\mathbf{u}}}$ with $\overline{\overline{\mathbf{u}}}\equiv(\mathbf{I}_T\otimes\mathbf{W})\overline{\mathbf{u}}$.

The resulting moment conditions are then stacked and solved as a solution to the system of six equations in three unknowns. More formally, $\boldsymbol{\gamma}-\boldsymbol{\Gamma}\boldsymbol{\alpha}=\mathbf{0}$, where

$$\begin{aligned}
\boldsymbol{\gamma} &= \left(\frac{1}{N(T-1)}\mathbf{u}'\mathbf{Q}_0\mathbf{u},\frac{1}{N(T-1)}\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\mathbf{u}},\frac{1}{N(T-1)}\mathbf{u}'\mathbf{Q}_0\overline{\mathbf{u}},\frac{1}{N}\mathbf{u}'\mathbf{Q}_1\mathbf{u},\frac{1}{N}\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\mathbf{u}},\frac{1}{N}\mathbf{u}'\mathbf{Q}_1\overline{\mathbf{u}}\right)',\\
\boldsymbol{\alpha} &= \left(\rho,\rho^2,\sigma_\nu^2,\sigma_1^2\right)',\\
\boldsymbol{\Gamma} &= \begin{pmatrix}
\frac{2}{N(T-1)}\mathbf{u}'\mathbf{Q}_0\overline{\mathbf{u}} & \frac{-1}{N(T-1)}\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\mathbf{u}} & 1 & 0\\
\frac{2}{N(T-1)}\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}} & \frac{-1}{N(T-1)}\overline{\overline{\mathbf{u}}}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}} & \frac{1}{N}tr\mathbf{W}'\mathbf{W} & 0\\
\frac{1}{N(T-1)}(\mathbf{u}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}}+\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\mathbf{u}}) & \frac{-1}{N(T-1)}\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}} & 0 & 0\\
\frac{2}{N}\mathbf{u}'\mathbf{Q}_1\overline{\mathbf{u}} & \frac{-1}{N}\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\mathbf{u}} & 0 & 1\\
\frac{2}{N}\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}} & \frac{-1}{N}\overline{\overline{\mathbf{u}}}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}} & 0 & \frac{1}{N}tr\mathbf{W}'\mathbf{W}\\
\frac{1}{N}(\mathbf{u}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}}+\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\mathbf{u}}) & \frac{-1}{N}\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}} & 0 & 0
\end{pmatrix}.
\end{aligned}\qquad(4)$$

The matrix $\boldsymbol{\Gamma}$ can be partitioned into two parts, where $\boldsymbol{\Gamma}^0$ corresponds to the first three lines of and $\boldsymbol{\Gamma}^1$ to the last three lines of $\boldsymbol{\Gamma}$. This partition will be used in the subsequent GM procedure. Furthermore, the next assumption is required for consistency of the GM estimator.

**Assumption 7** *The smallest eigenvalues of* $\boldsymbol{\Gamma}^{0\prime}\boldsymbol{\Gamma}^0$ *and* $\boldsymbol{\Gamma}^{1\prime}\boldsymbol{\Gamma}^1$ *are bounded uniformly away from zero i.e.,* $\lambda_{\min}(\boldsymbol{\Gamma}^{i\prime}\boldsymbol{\Gamma}^i)\geq\lambda_*>0$ *for* $i=0,1$ *where* $\lambda_*$ *may depend on* $\rho$, $\sigma_\nu$, *and* $\sigma_1$.

For estimation, $\mathbf{u}$, $\overline{\mathbf{u}}$, and $\overline{\overline{\mathbf{u}}}$ are replaced by their corresponding consistent estimates $\hat{\mathbf{u}}$, $\overline{\hat{\mathbf{u}}}$, and $\overline{\overline{\hat{\mathbf{u}}}}$, yielding $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Gamma}}$. As in Kapoor *et al.* (2007):

$$\hat{\boldsymbol{\alpha}}=(\hat{\rho},\hat{\sigma}_\nu,\hat{\sigma}_1)=\arg\min_{\sigma_\nu^2\in S_\nu,\sigma_1^2\in S_1,\rho\in S_\rho}\left[\left(\hat{\boldsymbol{\gamma}}-\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\alpha}}\right)'\hat{\mathbf{C}}\left(\hat{\boldsymbol{\gamma}}-\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\alpha}}\right)\right],\qquad(5)$$

where $S_\nu$, $S_1$, and $S_\rho$ denote the respective admissible parameter spaces of $\sigma_\nu^2$, $\sigma_1^2$, and $\rho$, and $\hat{\mathbf{C}}$ denotes a suitable estimate of the true weighting matrix of the moment vector, $\mathbf{C}$. In a data-set which is as large as ours and involves a matrix $\mathbf{W}$ of size $12,552\times12,552$, it is advisable to solve the moment conditions in two parts. First, solve the three moment conditions involving $\mathbf{Q_0}$ (using a moments weighting matrix $\mathbf{I}_3$) for $\tilde{\rho}$ and $\tilde{\sigma}_\nu^2$, and subsequently the remaining moment conditions or just one of them for $\tilde{\sigma}_1^2$. These estimates are consistent according to Theorem 1 in Kapoor *et al.* (2007). With these estimates at hand, calculate the weighting matrix matrix $\hat{\boldsymbol{\Upsilon}}$

$$\hat{\boldsymbol{\Upsilon}}=\begin{pmatrix}\frac{1}{T-1}\tilde{\sigma}_\nu^4 & 0\\ 0 & \tilde{\sigma}_1^4\end{pmatrix}\otimes\mathbf{I}_3.\qquad(6)$$

which has bounded elements by Assumption 6, as required. Replacing $\hat{\mathbf{C}}$ with $\hat{\boldsymbol{\Upsilon}}$ and applying nonlinear least squares to (5) using all six moment conditions yields $\hat{\rho}$, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_1^2$, which are consistent according to Theorem 3 in Kapoor *et al.* (2007).

**STEP 4 - The spatial Hausman-Taylor estimator:** Let us denote the variance-covariance matrices of $\mathbf{u}$, $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ and $\boldsymbol{\varepsilon}$, by $\boldsymbol{\Omega}_u$, $\boldsymbol{\Omega}_\mu$, $\boldsymbol{\Omega}_\nu$ and $\boldsymbol{\Omega}_\varepsilon$, respectively. By Assumption 6 we have

$$\begin{aligned} \boldsymbol{\Omega}_\mu &= \sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N), \quad \boldsymbol{\Omega}_\nu = \sigma_\nu^2 \mathbf{I}_{TN}, \quad \boldsymbol{\Omega}_\varepsilon = \sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \sigma_\nu^2 \mathbf{I}_{TN}, \text{ and} \quad (7) \\ \boldsymbol{\Omega}_u &= [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]^{-1}[\sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \sigma_\nu^2 \mathbf{I}_{TN}][\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]^{-1}, \quad (8) \end{aligned}$$

The next transformations aim at filtering out these effects. Premultiplying the model by $[\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]$ leads to spatially Cochrane-Orcutt-transformed variables. These variables are denoted by one star as a superscript, e.g.

$$\mathbf{y}^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{y}, \quad \mathbf{A}^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{A}. \quad (9)$$

Next, we apply the Fuller and Battese transformation which uses the fact that $\boldsymbol{\Omega}_\varepsilon^{-1/2} = \sigma_\nu^{-1}\mathbf{Q}_0 + \sigma_1^{-1}\mathbf{Q}_1$. We denote such variables by two stars as a superscript e.g.

$$\mathbf{A}^{**} = \boldsymbol{\Omega}_\varepsilon^{-1/2}\mathbf{A}^*, \quad \mathbf{y}^{**} = \boldsymbol{\Omega}_\varepsilon^{-1/2}\mathbf{y}^* = \mathbf{A}^{**}\boldsymbol{\delta} + \mathbf{u}^{**}. \quad (10)$$

Since $\boldsymbol{\iota}_T \otimes \mathbf{Z}_C^{**}$ and $\mathbf{X}_C^{**}$ in $\mathbf{A}^{**}$ are still correlated with $\boldsymbol{\mu}$ in $\mathbf{u}^{**}$, we still need an instrumental variable procedure to estimate the double-starred transformed model in (10), where the instruments are transformed in the same way as the other variables. Thus the Hausman and Taylor (1981) transformed set of instruments and its projection matrix are given by:

$$\mathbf{H}_{SHT}^{**} = \boldsymbol{\Omega}_\varepsilon^{-1/2}\mathbf{H}_{SHT}^* = \boldsymbol{\Omega}_\varepsilon^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{H}_{HT}, \quad \mathbf{P}_{SHT}^{**} = \mathbf{H}_{SHT}^{**}(\mathbf{H}_{SHT}^{**\prime}\mathbf{H}_{SHT}^{**})^{-1}\mathbf{H}_{SHT}^{**\prime}. \quad (11)$$

with $\mathbf{H}_{HT}$ defined in (3).

We maintain the following assumptions on the instrument set $\mathbf{H}_{SHT}^*$ to derive the asymptotic properties of our estimator. However we do not index $\mathbf{H}^*$ by $SHT$ to avoid index cluttering.

**Assumption 8** *(Additional assumptions on the instrument set $\mathbf{H}^*$ under homoskedasticity)*
*(i) The matrices $\mathbf{M}^{\mathbf{0}}_{\mathbf{H}^*\mathbf{H}^*} = \lim_{N\to\infty}[(NT)^{-1}\mathbf{H}^{*\prime}\mathbf{Q}_0\mathbf{H}^*]$, $\mathbf{M}^{\mathbf{1}}_{\mathbf{H}^*\mathbf{H}^*} = \lim_{N\to\infty}[(NT)^{-1}\mathbf{H}^{*\prime}\mathbf{Q}_1\mathbf{H}^*]$ exist, are finite and nonsingular. (ii) The matrices $\mathbf{M}^{\mathbf{0}}_{\mathbf{A}^*\mathbf{H}^*} = p\lim_{N\to\infty}[(NT)^{-1}\mathbf{A}^{*\prime}\mathbf{Q}_0\mathbf{H}^*]$, $\mathbf{M}^{\mathbf{1}}_{\mathbf{A}^*\mathbf{H}^*} = p\lim_{N\to\infty}[(NT)^{-1}\mathbf{A}^{*\prime}\mathbf{Q}_1\mathbf{H}^*]$ exist, are finite and have full column rank. (iii) The smallest eigenvalue of $(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*})$ is uniformly bounded away from zero.*

The spatial Hausman-Taylor estimator (SHT) is then defined as a 2SLS estimator of $\mathbf{y}^{**}$ on $\mathbf{A}^{**}$ with the matrix of instruments $\mathbf{H}_{SHT}^{**}$. This estimator removes the error component structure as well as the spatial autocorrelation from the process in (2). The true GLS estimator of $\boldsymbol{\delta}$ is

$$\widehat{\boldsymbol{\delta}}_{SHT} = (\mathbf{A}^{**\prime}\mathbf{P}_{SHT}^{**}\mathbf{A}^{**})^{-1}\mathbf{A}^{**\prime}\mathbf{P}_{SHT}^{**}\mathbf{y}^{**}. \quad (12)$$

The corresponding feasible GLS estimator, which uses the estimators $\hat{\rho}$, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_1^2$ from Step 3 for transforming the model in (10) as well as the instrument set and the projection matrix in (11), is defined as

$$\widehat{\widehat{\boldsymbol{\delta}}}_{SHT} = (\widetilde{\mathbf{A}}^{**\prime}\widetilde{\mathbf{P}}_{SHT}^{**}\widetilde{\mathbf{A}}^{**})^{-1}\widetilde{\mathbf{A}}^{**\prime}\widetilde{\mathbf{P}}_{SHT}^{**}\widetilde{\mathbf{y}}^{**}. \quad (13)$$

Transformations based on the estimated error components are denoted by $\widetilde{\ }$. The distribution of the this estimator is given in the following theorem.

**Theorem 1** *Given that Assumptions 1 – 8 hold. Then*

$$(NT)^{1/2}(\widehat{\widehat{\boldsymbol{\delta}}}_{SHT} - \boldsymbol{\delta}) \xrightarrow{d} N(0, \boldsymbol{\Psi}) \quad as \quad N \to \infty$$

*with*

$$\begin{aligned}
\boldsymbol{\Psi} &= \mathbf{F}(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})\mathbf{F}' \\
\mathbf{F} &= \{(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0'}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1'}_{\mathbf{A}^*\mathbf{H}^*})\}^{-1} \\
&\quad (\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}
\end{aligned}$$

*(ii) And*

$$\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi} \xrightarrow{p} 0 \quad as \quad N \to \infty$$

*with*

$$\begin{aligned}
\widetilde{\boldsymbol{\Psi}} &= \widetilde{\mathbf{F}}((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\varepsilon}}^{-1}\widetilde{\mathbf{H}}^*)\widetilde{\mathbf{F}}' \\
\widetilde{\mathbf{F}} &= \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{A}}^*\}^{-1}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^* \\
&\quad ((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1}
\end{aligned}$$

The first part of the theorem shows that the feasible GLS estimator $\widehat{\widehat{\boldsymbol{\delta}}}_{SHT}$ is consistent with the asymptotic distribution stated in part (i). Part (ii) shows that the variance-covariance matrix can be estimated consistently by $\widetilde{\boldsymbol{\Psi}}$. The proof the theorem will be given in the Appendix.

Our procedure can be contrasted with a spatial fixed effects (SFE) estimator (see Mutl and Pfaffermayr, 2011) defined as

$$\widehat{\boldsymbol{\beta}}_{SFE} = (\mathbf{X}^{*\prime}\mathbf{Q}_0\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Q}_0\mathbf{y}^*. \tag{14}$$

This corresponds to a fixed effects estimator, which is applied to the spatially Cochrane-Orcutt-transformed model. This has an advantage over the standard fixed effects estimator in that it takes into account the spatial correlation. However, it shares the same disadvantage of the standard fixed effects estimator in that it does not deliver an estimate of $\boldsymbol{\theta}$. Such estimates could have been obtained with a spatial random effects estimator (SRE) of the form

$$\widehat{\boldsymbol{\delta}}_{SRE} = (\mathbf{A}^{**\prime}\mathbf{A}^{**})^{-1}\mathbf{A}^{**\prime}\mathbf{y}^{**}, \tag{15}$$

but they would be inconsistent under correlation of some of the covariates with $\boldsymbol{\mu}$. One way of checking whether $\mathbf{A}$ is correlated with $\boldsymbol{\mu}$, is to apply the spatial Hausman test proposed by Mutl and Pfaffermayr (2011),

$$\hat{m}_{SH} = (\widehat{\boldsymbol{\beta}}_{SRE} - \widehat{\boldsymbol{\beta}}_{SFE})'[Var(\widehat{\boldsymbol{\beta}}_{SFE}) - Var(\widehat{\boldsymbol{\beta}}_{SRE})]^-(\widehat{\boldsymbol{\beta}}_{SRE} - \widehat{\boldsymbol{\beta}}_{SFE}), \tag{16}$$

where superscript " $-$ " in (16) refers to the generalized inverse. $\hat{m}_{SH}$ is distributed as $\chi^2(\text{rank}[Var(\widehat{\boldsymbol{\beta}}_{SFE}) - Var(\widehat{\boldsymbol{\beta}}_{SRE})]^-)$ under the null hypothesis of no correlation between $\mathbf{A}$ and $\boldsymbol{\mu}$. If the null is rejected, then $\widehat{\boldsymbol{\delta}}_{SRE}$ is not consistent.
In this case, one can check the choice of $[\mathbf{X}_U, \boldsymbol{\iota}_T \otimes \mathbf{Z}_U]$, i.e., the choice of regressors uncorrelated with $\boldsymbol{\mu}$, by applying the Hausman test based on the contrast between $\widehat{\boldsymbol{\beta}}_{SHT}$ and $\widehat{\boldsymbol{\beta}}_{SFE}$. This is given by

$$\hat{m}_{SHT} = (\widehat{\boldsymbol{\beta}}_{SHT} - \widehat{\boldsymbol{\beta}}_{SFE})'[Var(\widehat{\boldsymbol{\beta}}_{SFE}) - Var(\widehat{\boldsymbol{\beta}}_{SHT})]^-(\widehat{\boldsymbol{\beta}}_{SHT} - \widehat{\boldsymbol{\beta}}_{SFE}) \tag{17}$$

and is distributed as $\chi^2(K_U - R_C)$ where $K_U - R_C$ is the degree of over-identification. Either test is calculated by using the feasible versions of SRE and SFE.

## 2.4 SHT estimation under heteroskedasticity

### 2.4.1 Assumptions

In the following, we allow for heteroskedasticity in the idiosyncratic error component as in Badinger and Egger (2014). We keep Assumptions 1 – 5 but now make a different assumption on the error components.

**Assumption 9** *(Assumptions on the error components):*
*(i) For the idiosnycratic error components we have that $\nu_{it}$ are mutually independently distributed with $E(\nu_{it}) = 0$, $E(\nu_{it}^2) = \sigma_{\nu,it}^2$ where $0 < \sigma_{\nu,it}^2 < \infty$ and $E|v_{it,N}|^{4+\eta} < \infty$ for some $\eta > 0$. Hence, the idiosyncratic disturbances exhibit heteroskedasticity of unknown form. (ii) For the unit-specific error components we have $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$, where $0 < \sigma_\mu^2 < b_\mu < \infty$, and $E|\mu_{i,N}|^{4+\eta} < \infty$ for some $\eta > 0$. (iii) $\mu_i$ and $\nu_{it}$ are independent of each other for all $i$ and $t$.*

Assumption 9 implies the following covariance of $\varepsilon_{it}$ and $\varepsilon_{js}$: $Cov(\varepsilon_{it}\varepsilon_{js}) = \sigma_\mu^2 + \sigma_{\nu,it}^2$ for $i = j$ and $t = s$; $Cov(\varepsilon_{it}\varepsilon_{js}) = \sigma_\mu^2$ for $i = j$ and $t \neq s$; and $Cov(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.

### 2.4.2 Estimation procedure

Step 1 and 2 are the same as above, since the initial Hausman and Taylor estimator is consistent. However, in Step 3 we apply a different GM procedure due to the different error term structure.

**STEP 3 - Estimate $\boldsymbol{\rho}$ and $\boldsymbol{\sigma_\mu^2}$ :** Using the $\widehat{\mathbf{u}}$'s we can apply the moment conditions in Badinger and Egger (2014) to obtain estimates of $\rho$ and $\sigma_\mu^2$. The four moment conditions are given by

$$E\left[\frac{1}{N(T-1)}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\overline{\boldsymbol{\varepsilon}}\right] = \frac{1}{N(T-1)}tr[diag_{n=1}^{NT}E(\nu_n^2)\mathbf{Q}_0(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})], \qquad E\left[\frac{1}{N(T-1)}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\boldsymbol{\varepsilon}\right] = 0$$
$$E\left[\frac{1}{N}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_1\overline{\boldsymbol{\varepsilon}}\right] = \frac{T}{N}\sigma_\mu^2 tr(\mathbf{W}'\mathbf{W}) + \frac{1}{N}tr[diag_{n=1}^{NT}E(\nu_n^2)\mathbf{Q}_1(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})], \quad E\left[\frac{1}{N}\overline{\boldsymbol{\varepsilon}}'\mathbf{Q}_1\boldsymbol{\varepsilon}\right] = 0,$$

where $\overline{\boldsymbol{\varepsilon}} \equiv (\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\varepsilon}$. These can be rewritten in terms of $\mathbf{u}$ using the fact that $\boldsymbol{\varepsilon} = (\mathbf{I}_T \otimes [\mathbf{I}_N - \rho\mathbf{W}])\mathbf{u} = \mathbf{u} - \rho\overline{\mathbf{u}}$ whereby $\overline{\mathbf{u}} \equiv (\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}$ and $\overline{\boldsymbol{\varepsilon}} \equiv (\mathbf{I}_T \otimes \mathbf{W})(\mathbf{I}_T \otimes [\mathbf{I}_N - \rho\mathbf{W}])\mathbf{u} = \overline{\mathbf{u}} - \rho\overline{\overline{\mathbf{u}}}$ with $\overline{\overline{\mathbf{u}}} \equiv (\mathbf{I}_T \otimes \mathbf{W})\overline{\mathbf{u}}$. The resulting moment conditions are then stacked and solved as a solution to the system of four equations in two unknowns. More formally, $\boldsymbol{\gamma} - \boldsymbol{\Gamma}\boldsymbol{\alpha} = \mathbf{0}$, where

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \rho \\ \rho^2 \\ \sigma_\mu^2 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \Gamma_{1,1} & \Gamma_{1,2} & \Gamma_{1,3} \\ \Gamma_{2,1} & \Gamma_{2,2} & 0 \\ \Gamma_{3,1} & \Gamma_{3,2} & \Gamma_{3,3} \\ \Gamma_{4,1} & \Gamma_{4,2} & 0 \end{pmatrix}. \tag{18}$$

with

$$\gamma_1 = \frac{1}{N(T-1)}E\{\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\mathbf{u}} - tr[\mathbf{Q}_0 diag_{n=1}^{NT}(u_n^2)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\}, \qquad \gamma_2 = \frac{1}{N(T-1)}E\{\overline{\mathbf{u}}'\mathbf{Q}_0\mathbf{u}\}$$
$$\gamma_3 = \frac{1}{N}E\{\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\mathbf{u}} - tr[\mathbf{Q}_1 diag_{n=1}^{NT}(u_n^2)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\}, \qquad \gamma_4 = \frac{1}{N}E\{\overline{\mathbf{u}}'\mathbf{Q}_1\mathbf{u}\}$$
$$\Gamma_{1,1} = \frac{2}{N(T-1)}E\{\overline{\mathbf{u}}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}} - tr[\mathbf{Q}_0 diag_{n=1}^{NT}(\bar{u}_n u_n)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\}$$
$$\Gamma_{1,2} = -\frac{1}{N(T-1)}E\{\overline{\overline{\mathbf{u}}}'\mathbf{Q}_0\overline{\overline{\mathbf{u}}} - tr[\mathbf{Q}_0 diag_{n=1}^{NT}(\bar{u}_n^2)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\}, \qquad \Gamma_{1,3} = -\frac{1}{N}tr(\mathbf{W}'\mathbf{W})$$

$$
\begin{aligned}
\Gamma_{2,1} &= \frac{1}{N(T-1)}E\{\overline{\overline{\mathbf{u}}}'\mathbf{Q}_0\ \mathbf{u} + \overline{\mathbf{u}}'\mathbf{Q}_0\overline{\mathbf{u}}\}, \quad \Gamma_{2,2} = -\frac{1}{N(T-1)}E(\overline{\overline{\mathbf{u}}}'\mathbf{Q}_0\overline{\mathbf{u}}), \quad \Gamma_{2,3} = 0 \\
\Gamma_{3,1} &= \frac{2}{N}E\{\overline{\mathbf{u}}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}} - tr[\mathbf{Q}_1 diag_{n=1}^{NT}(\bar{u}_n u_n)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\} \\
\Gamma_{3,2} &= -\frac{1}{N}E\{\overline{\overline{\mathbf{u}}}'\mathbf{Q}_1\overline{\overline{\mathbf{u}}} - tr[\mathbf{Q}_1 diag_{n=1}^{NT}(\bar{u}_n^2)(\mathbf{I}_T \otimes \mathbf{W}'\mathbf{W})]\}, \quad \Gamma_{3,3} = \frac{T-1}{N}tr(\mathbf{W}'\mathbf{W}) \\
\Gamma_{4,1} &= \frac{1}{N}E\{\overline{\overline{\mathbf{u}}}'\mathbf{Q}_1\mathbf{u} + \overline{\mathbf{u}}'\mathbf{Q}_1\overline{\mathbf{u}}\}, \quad \Gamma_{4,2} = -\frac{1}{N}E\{\overline{\overline{\mathbf{u}}}'\mathbf{Q}_1\overline{\mathbf{u}}\}, \quad \Gamma_{4,3} = 0
\end{aligned}
$$

To get consistent estimates on $\rho$ and $\sigma_\mu^2$, the following assumption is required

**Assumption 10** *The smallest eigenvalues of* $\mathbf{\Gamma}'\mathbf{\Gamma}$ *are bounded uniformly away from zero, i.e.,* $\lambda_{\min}(\mathbf{\Gamma}'\mathbf{\Gamma}) \geq \lambda_* > 0$.

For estimation, $\mathbf{u}$, $\overline{\mathbf{u}}$, and $\overline{\overline{\mathbf{u}}}$ are replaced by their corresponding consistent estimates $\hat{\mathbf{u}}$, $\overline{\hat{\mathbf{u}}}$, and $\overline{\overline{\hat{\mathbf{u}}}}$. As in Badinger and Egger (2014):

$$
\hat{\boldsymbol{\alpha}} = (\hat{\rho}, \hat{\sigma}_\mu) = \arg \min_{\sigma_\mu^2 \in S_\mu, \rho \in S_\rho} \left[ \left(\hat{\gamma} - \hat{\mathbf{\Gamma}}\hat{\boldsymbol{\alpha}}\right)' \hat{\boldsymbol{C}} \left(\hat{\gamma} - \hat{\mathbf{\Gamma}}\hat{\boldsymbol{\alpha}}\right) \right], \tag{19}
$$

where $S_\mu$ and $S_\rho$ denote the respective admissible parameter spaces of $\sigma_\mu^2$ and $\rho$, and $\hat{\boldsymbol{C}}$ is a $4 \times 4$ weighting matrix. With less than four time periods, using all moment conditions and an associated weighting to achieve efficiency gains is infeasible with heteroskedasticity as considered here. For this reason, we use only the second moment condition which obtains an estimate of $\rho$ and the third one, which delivers an estimate of $\sigma_\mu^2$. The estimates $\hat{\rho}$ and $\hat{\sigma}_\mu$ are consistent by Theorem 1 in Badinger and Egger (2014).

**STEP 4 - The spatial Hausman-Taylor estimator:** The variance-covariance matrices of $\mathbf{u}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\nu}$ are denoted by $\mathbf{\Omega}_u$, $\mathbf{\Omega}_\varepsilon$, and $\mathbf{\Sigma}$, respectively. By Assumption 9 these are as follows

$$
\begin{aligned}
\mathbf{\Omega}_\varepsilon &= \sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \mathbf{\Sigma}, \quad \mathbf{\Sigma} = E(\boldsymbol{\nu}\boldsymbol{\nu}') = diag_{n=1}^{NT}E(\nu_n^2) = diag_{n=1}^{NT}E(\varepsilon_n^2) - \sigma_\mu^2 \boldsymbol{I}_{TN}, \text{ and} \\
\mathbf{\Omega}_u &= [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]^{-1}[\sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \mathbf{\Sigma}][\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]^{-1}.
\end{aligned}
$$

As before, premultiplying the model by $[\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]$ leads to spatially Cochrane-Orcutt-transformed variables, e.g.

$$
\mathbf{y}^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{y}, \quad \mathbf{A}^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{A} \tag{20}
$$

Notice, that now under the given error term structure no further transformations are possible due to the heteroskedasticity in $\boldsymbol{\nu}$. Thus the estimation will be based on spatially Cochrane-Orcutt transformed variables, where we will calculate a heteroskedasticity-robust variance covariance matrix. Thus we are estimating the following transformed model

$$
\mathbf{y}^* = \mathbf{A}^*\boldsymbol{\delta} + \mathbf{u}^* \tag{21}
$$

Since $\boldsymbol{\iota}_T \otimes \mathbf{Z}_C^*$ and $\mathbf{X}_C^*$ in $\mathbf{A}^*$ are still correlated with $\boldsymbol{\mu}$ in $\mathbf{u}^*$, we need an instrumental variable procedure. The spatially transformed set of instruments and its projection matrix are given by:

$$
\mathbf{H}_{SHT}^* = [\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})]\mathbf{H}_{HT}, \quad \mathbf{P}_{SHT}^* = \mathbf{H}_{SHT}^*(\mathbf{H}_{SHT}^{*\prime}\mathbf{H}_{SHT}^*)^{-1}\mathbf{H}_{SHT}^{*\prime}. \tag{22}
$$

Due to the different error term structure, we maintain the following assumptions to derive the asymptotic properties of the estimator.

**Assumption 11** *(Additional assumptions on the instrument set* $\mathbf{H}^*$ *under heteroskedasticity)*
*(i) The matrix* $\mathbf{M}^J_{\mathbf{H}^*\mathbf{H}^*} = \lim_{N\to\infty}[(NT)^{-1}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \mathbf{I}_N)\mathbf{H}^*]$ *exists, is finite and nonsingular. (ii) The matrix* $\mathbf{M}_{\mathbf{H}^*\mathbf{H}^*} = \lim_{N\to\infty}[(NT)^{-1}\mathbf{H}^{*\prime}\mathbf{H}^*]$ *exists, is finite and nonsingular. (iii) The matrix* $\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*} = p\lim_{N\to\infty}[(NT)^{-1}\mathbf{A}^{*\prime}\mathbf{H}^*]$ *exists, is finite and has full column rank. (iv) The matrix* $\mathbf{M}_{\mathbf{H}^*\boldsymbol{\Sigma}\mathbf{H}^*} = p\lim_{N\to\infty}[\mathbf{H}^{*\prime}\boldsymbol{\Sigma}\mathbf{H}^*]$ *is finite and nonsingular. (v) The smallest eigenvalue of* $(\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*}\mathbf{M}'_{\mathbf{A}^*\mathbf{H}^*})$ *is uniformly bounded away from zero.*

The spatial Hausman-Taylor estimator (SHT) is a 2SLS estimator of $\mathbf{y}^*$ on $\mathbf{A}^*$ with the matrix of instruments $\mathbf{H}^*_{SHT}$. This estimator removes the spatial autocorrelation from the process in (2). The true GLS estimator of $\boldsymbol{\delta}$ is

$$\widehat{\boldsymbol{\delta}}_{SHT} = (\mathbf{A}^{*\prime}\mathbf{P}^*_{SHT}\mathbf{A}^*)^{-1}\mathbf{A}^{*\prime}\mathbf{P}^*_{SHT}\mathbf{y}^*. \tag{23}$$

The corresponding feasible GLS estimator, which uses the estimator $\hat{\rho}$ from Step 3 for transforming the model in (21) as well as the instrument set and the projection matrix in (22), is defined as

$$\widehat{\widehat{\boldsymbol{\delta}}}_{SHT} = (\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{P}}^*_{SHT}\widetilde{\mathbf{A}}^*)^{-1}\widetilde{\mathbf{A}}^{*\prime}\tilde{\mathbf{P}}^*_{SHT}\widetilde{\mathbf{y}}^*. \tag{24}$$

where we again denote transformations based on the estimated error components by $\tilde{}$. The next theorem describes the distribution of the feasible spatial Hausman and Taylor estimator under heteroskedasticity.

**Theorem 2** *Given that Assumptions 1 – 5 and 9 – 11 hold. (i) Then as* $N \to \infty$

$$(NT)^{1/2}(\widehat{\widehat{\boldsymbol{\delta}}}_{SHT} - \boldsymbol{\delta}) \xrightarrow{d} N(0, \boldsymbol{\Psi})$$

*with*

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{D}(\sigma^2_\mu\mathbf{M}^J_{\mathbf{H}^*\mathbf{H}^*} + \mathbf{M}_{\mathbf{H}^*\boldsymbol{\Sigma}\mathbf{H}^*})\mathbf{D}' \\ \mathbf{D} &= (\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*}\mathbf{M}'_{\mathbf{A}^*\mathbf{H}^*})^{-1}\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*}\end{aligned}$$

*(ii) And*

$$\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi} \xrightarrow{p} 0 \quad as \quad N \to \infty$$

*where*

$$\begin{aligned}\widetilde{\boldsymbol{\Psi}} &= \widetilde{\mathbf{D}}(\hat{\sigma}^2_\mu(NT)^{-1}(\widetilde{\mathbf{H}}^{*\prime}(\mathbf{J}_T \otimes \mathbf{I}_N)\widetilde{\mathbf{H}}^*) + (NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Sigma}}^{HR}\widetilde{\mathbf{H}}^*)\widetilde{\mathbf{D}}' \\ \widetilde{\mathbf{D}} &= \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{A}}^*\}^{-1}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1}\end{aligned}$$

The first part of the theorem shows that the feasible GLS estimator $\hat{\hat{\boldsymbol{\delta}}}_{SHT}$ is consistent with the asymptotic distribution stated in part (i). Part (ii) shows that the variance-covariance matrix can be estimated consistently by $\widetilde{\boldsymbol{\Psi}}$. The proof the theorem will be given in the Appendix.

As in the homoskedastic case, one can estimate heteroskedastic versions of the spatial fixed effects and the spatial random effects estimator by using spatially Cochrane transformed variables based on the $\rho$ from the GM procedure under heteroskedasticity and calculating a heteroskedasticity variance-covariance matrix. The spatial fixed effects estimator under heteroskedasticity is retrieved by (14). However, the spatial random effects estimator is now given by

$$\widehat{\boldsymbol{\delta}}_{SRE} = (\mathbf{A}^{*\prime}\mathbf{A}^*)^{-1}\mathbf{A}^{*\prime}\mathbf{y}^*. \tag{25}$$

With these estimates at hand, one can apply Wald test variants of the spatial Hausman test (see Badinger and Egger, 2014) and the spatial Hausman and Taylor test, to check the validity of the assumptions. For the spatial Hausman test the discrepancy vector is given by $\mathbf{q}_{SH} = (\widehat{\boldsymbol{\beta}}'_{SRE}, \widehat{\boldsymbol{\beta}}'_{SFE})'$ and the variance term $\mathbf{V}_{SH}$ is given by

$$\mathbf{V}_{SH} = \begin{pmatrix} Var(\widehat{\boldsymbol{\beta}}_{SRE}) & Cov(\widehat{\boldsymbol{\beta}}_{SRE}\widehat{\boldsymbol{\beta}}_{SFE}) \\ Cov(\widehat{\boldsymbol{\beta}}_{SRE}\widehat{\boldsymbol{\beta}}_{SFE})' & Var(\widehat{\boldsymbol{\beta}}_{SFE}) \end{pmatrix} \tag{26}$$

Furthermore, we define $\mathbf{B} = (\mathbf{I}_K, -\mathbf{I}_K)$, where $K$ denotes the number of time-varying variables. The test statistic of the spatial Hausman test is given by

$$\hat{m}_{SH} = (\mathbf{B}\mathbf{q}_{SH})'(\mathbf{B}\mathbf{V}_{SH}\mathbf{B}')^{-}(\mathbf{B}\mathbf{q}_{SH}), \tag{27}$$

where $\hat{m}_{SH}$ is distributed as $\chi^2(K)$ under the null hypothesis of no correlation between $\mathbf{A}$ and $\boldsymbol{\mu}$. If the null is rejected, then $\widehat{\boldsymbol{\delta}}_{SRE}$ is not consistent and one can check the choice of regressors uncorrelated with $\boldsymbol{\mu}$, by applying the Hausman-Taylor test based on

$$\hat{m}_{SHT} = (\mathbf{B}\mathbf{q}_{SHT})'(\mathbf{B}\mathbf{V}_{SHT}\mathbf{B}')^{-}(\mathbf{B}\mathbf{q}_{SHT}), \tag{28}$$

which is distributed as $\chi^2(K_U - R_C)$ where $K_U - R_C$ is the degree of over-identification and $\mathbf{q}_{SHT} = (\widehat{\boldsymbol{\beta}}'_{SHT}, \widehat{\boldsymbol{\beta}}'_{SFE})'$ and

$$\mathbf{V}_{SHT} = \begin{pmatrix} Var(\widehat{\boldsymbol{\beta}}_{SHT}) & Cov(\widehat{\boldsymbol{\beta}}_{SHT}\widehat{\boldsymbol{\beta}}_{SFE}) \\ Cov(\widehat{\boldsymbol{\beta}}_{SHT}\widehat{\boldsymbol{\beta}}_{SFE})' & Var(\widehat{\boldsymbol{\beta}}_{SFE}) \end{pmatrix}. \tag{29}$$

# 3 Monte Carlo simulations

To illustrate the small sample performance of the estimators and tests presented above, we perform some Monte Carlo experiments.[7] The proposed model is as follows

$$\mathbf{y}_t = \mathbf{X}_{U1,t}\beta_{11} + \mathbf{X}_{U2,t}\beta_{12} + \mathbf{X}_{C,t}\beta_2 + \mathbf{Z}_U\theta_1 + \mathbf{Z}_C\theta_2 + \mathbf{u}_t, \quad \mathbf{u}_t = \rho\mathbf{W}\mathbf{u}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\nu}_t. \tag{30}$$

The covariates are specified as follows: $\mathbf{X}_{U1t} = 0.7\mathbf{X}_{U1,t-1} + \boldsymbol{\vartheta} + \boldsymbol{\zeta}_t$ with the initial value $\mathbf{X}_{U1,1} = \boldsymbol{\zeta}_1/(1-0.7^2)^{1/2} + \boldsymbol{\vartheta}/(1-0.7)$; $\mathbf{X}_{U2t} = 0.7\mathbf{X}_{U2,t-1} + \boldsymbol{\eta} + \boldsymbol{\kappa}_t$ with the initial value $\mathbf{X}_{U2,1} = \boldsymbol{\kappa}_1/(1-0.7^2)^{1/2} + \boldsymbol{\eta}/(1-0.7)$; $\mathbf{X}_{Ct} = 0.7\mathbf{X}_{C,t-1} + \boldsymbol{\mu} + \boldsymbol{\lambda}_t$ with the initial value $\mathbf{X}_{C,1} = \boldsymbol{\lambda}_1/(1-0.7^2)^{1/2} + \boldsymbol{\mu}/(1-0.7)$; $\mathbf{Z}_U = \boldsymbol{\iota}$; $\mathbf{Z}_C = \boldsymbol{\vartheta} + \boldsymbol{\eta} + \boldsymbol{\mu} + \boldsymbol{\xi}$, where all elements of $\boldsymbol{\vartheta}, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\zeta}_t, \boldsymbol{\kappa}_t, \boldsymbol{\lambda}_t$ are $U[-2, 2]$. The true regression coefficients are set to $\beta_{11} = \beta_{12} = \beta_2 = \theta_1 = \theta_2 = 1$. $\mathbf{W}$ is a rowsum normalized weights matrix. It is based on an unnormalized $\mathbf{W}_0$, which has a five-before-five-behind neighborhood structure. We distinguish between two worlds: homoskedastic and heteroskedastic errors . In the homoskedastic world we have $\boldsymbol{\mu}_t \sim N(\mathbf{0}, \mathbf{I}_t)$ and $\boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{I}_t)$, whereby $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, 2\mathbf{I}_t)$. In modeling heteroskedastic innovations we follow Li and Stengos (1994) and let $\boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\omega}_t)$ and $\boldsymbol{\omega}_t = \alpha^2(\boldsymbol{\iota}_t + \ell\boldsymbol{\zeta}_t)^2$. Thus for each value of $\ell$, $\alpha$ is adapted in order to keep the total variance $\sigma^2_{\varepsilon,it} = \sigma^2_\mu + E(\nu^2_{it})$ fixed. We set $\sigma^2_\varepsilon = 2$ and consider a value of $\ell = 4$. Note, that $\ell = 0$ corresponds to homoskedastic idiosyncratic errors.

In our experiments, we consider two sample sizes $N = \{100, 400\}$ each for $T = 3$. We vary the degree of spatial correlation along the true values $\rho_0 = \{0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8\}$. We report bias and RMSE for the estimates of $\rho$, $\beta_{11}$, $\beta_{12}$, $\beta_2$, $\theta_1$, and $\theta_2$. For the

[7] The design is similar to the one in Baltagi *et al.* (2012), who considered the case of homoskedastic disturbances only.

estimates of the regression coefficients we also present the size of the test for $H_0 : \beta_{11} = 1$, $H_0 : \beta_{12} = 1$, $H_0 : \beta_2 = 1$, and $H_0 : \theta_2 = 1$ for a nominal size of 5%. Furthermore, we report the power of the spatial Hausman test (abbreviated as HT in the tables) and the size of the spatial Hausman and Taylor test (abbreviated as HTT in the tables). The results are summarized in Table 1 for homoskedastic time-variant disturbances and in Table 2 for heteroskedastic time-variant disturbances.

– Tables 1–2 about here –

The tables provide the following insights. First, the values of the biases and RMSEs for the parameters are relatively small, even when $N = 100$. Both the biases and the RMSEs tend to be smaller at $N = 400$ as expected. Hence, the multi-step GM estimation procedure works well even in small samples. The same pattern emerges for the test sizes for both parameters, HTT, as well as the power of HT. The power of the HT is remarkable at $N = 400$, and the size of the HTT varies between 4.9 and 6.1 for $N = 400$. These results suggest that the proposed procedure is well suited for an analysis in moderately-sized samples and definitely in samples as large as the one considered in the subsequent application.

# 4 Empirical analysis

## 4.1 Data and descriptive statistics

We use panel data on output, factor usage, ownership characteristics, and export market participation for 12,552 Chinese firms in the chemical industry as compiled by the National Bureau of Statistics of China (NBS) over the period 2004-2006. The data-set covers all Chinese firms with an annual turnover exceeding five million Yuan (about 700,000 US dollars). In particular, we use data on log of sales for firm $i$ at year $t$ ($y_{it}$) as the outcome variable of interest and model it as a function of the following three sets of regressors. First, we employ the following set of variables as *primary production factors* entering the technology specification: log of employment ($l_{it}$); the share of high-skilled workers ($h_i$), which is defined as the fraction of workers with a university or comparable education (it is only available for 2004 and treated as time-invariant as indicated by the single subscript); log of capital used in production ($k_{it}$); and log of material inputs ($m_{it}$). Second, we use a set of *own technology shifters* which only enter as main effects: a binary export status indicator which takes the value one if the firm is an exporter, and zero otherwise ($e_{it}$); foreign ownership which is measured by the share of capital provided by foreign investors ($f_{it}$); an indicator variable for the intensive use of intangible relative to total assets ($i_{it}$); and a binary public ownership status indicator which takes the value one if the firm is publicly-owned (or state-owned) and zero otherwise ($p_i$). The latter is time-invariant for the Chinese chemical industry and the time-period covered by our data. Third, we employ two alternative sets of what we call *spillover technology shifters*. These are spatially weighted covariates which capture contextual effects/spillovers in observed characteristics. We distinguish two sets of contextual effects. Set a consists of $\overline{e}_{it}$, $\overline{f}_{it}$, $\overline{i}_{it}$, and $\overline{h}_{it}$ (i.e., local spillovers from other firms' exporting, being members of a multinational network, or being research intensive, and human capital intensive), which are important channels for spillovers according to the literature. For more flexibility, we use a second Set b, which encompasses Set a plus all other spatially weighted primary production factors of other firms. In general we refer to variables capturing contextual effects by bars.

We are interested in estimating all of the parameters on these variables and, in particular, the one on public ownership, while simultaneously allowing for global spillover effects (i.e., direct plus indirect spillovers) in unspecified (remainder) total factor productivity across firms.

– Table 3 about here –

Table 3 summarizes the descriptive statistics for our data in three panels. The left panel reports the mean and standard deviation for all variables, firms, and years covered. The middle and the right panel reports the mean and standard deviation for state-owned as compared to privately-owned enterprises. Note that 5 percent of the firms in the data are state-owned, and they have on average more employees than privately-owned firms (about 793 versus 166 employees on average). Second, about 24 percent of all firms considered participate in the export market. Also, export market participation is higher for state-owned firms as compared to privately-owned enterprises (29 versus 23 percent). Third, many firms in that sector use intangible assets relatively intensively – which is consistent with the innovative pressure and the relative patent intensity in the manufacturing of chemicals. The relative intangible asset intensity is relatively higher for state-owned units as compared to privately-owned ones (59 compared to 42 percent). Fourth, about 19 percent of all firms are at least partly foreign owned. The ratio of foreign capital to total capital is 0.14 for all firms on average. It is lower for state-owned than for privately-owned ones (0.06 versus 0.14).

This firm level data-set contains information about the postcodes of firms which identify the geographical location of all entities uniquely in terms of the longitude and latitude of a firm's residence. The latter enables calculating great circle distances between all units, using the so-called haversine formula.[8]

– Figures 1a and 1b about here –

We define *counties* as regional aggregates using the first four digits of the postcode. This yields $2,425$ regional aggregates/counties all over China. Figure 1a colors counties according to the *number of firms* in the chemical industry, while Figure 1b colors counties according to *total employment* in that industry. According to Figures 1a and 1b, $1,610$ of the $2,425$ counties in China do not host any chemical producer. Most of all inhabited regions host less than 10 firms. As expected, coastal regions have a higher density of firms. This pattern also holds for the number of employees in the chemical sector. In general, Figure 1b looks quite similar to Figure 1a. However, careful inspection of the data shows that Wujin in the coastal province of Jiangsu is the region with the largest number of chemical producers (175, of which all are privately-owned), while Shijiazhuang in the province of Hebei is the biggest regional employer in the chemical industry (34 firms in total, of which 6 are state-owned).

[8] The haversine formula is particularly suited for calculating great circle distances between two points $i$ and $j$ on the globe, if these two points are very close to each other. Denote the haversine function of an argument $\ell$ by $h(\ell) = 0.5(1 - \cos(\ell))$, and use $\phi_i$, $\phi_j$, and $\Delta\lambda_{ij}$ to refer to the latitude of $i$, the latitude of $j$, and the difference in longitudes between $i$ and $j$ which are all measured in radians. Then, the haversine distance between $i$ and $j$ is defined as $d_{ij} = D \cdot \arcsin(H_{ij}^{1/2})$, where $D$ is the diameter of the globe (e.g., measured in miles) and $H_{ij} = h(\phi_i - \phi_j) + \cos(\phi_i)\cos(\phi_j)h(\Delta\lambda_{ij})$.

## 4.2 Specification

For estimation, we use a flexible translog primary production technology, which nests a more restrictive Cobb Douglas technology. The vectors of inputs $\mathbf{k}$, $\mathbf{l}$, $\boldsymbol{\iota}_T \otimes \mathbf{h}$, and $\mathbf{m}$, represent primary production factors; while the vectors $\mathbf{e}$, $\mathbf{f}$, $\mathbf{i}$, $\boldsymbol{\iota}_T \otimes \mathbf{p}$, and spatially weighted characteristics of neighbors along with the disturbance term $\mathbf{u}$, reflect measurable and unmeasurable aspects of total factor productivity. The right-hand-side regressors for the primal translog production function are as follows:[9]

$$\begin{aligned}[\mathbf{X}, \boldsymbol{\iota}_T \otimes \mathbf{Z}] \quad = \quad & [\mathbf{k}, \mathbf{l}, \boldsymbol{\iota}_T \otimes \mathbf{h}, \mathbf{m}, \mathbf{k}^2, \mathbf{l}^2, \mathbf{m}^2, (\boldsymbol{\iota}_T \otimes \mathbf{h})^2, \mathbf{k} \circ \mathbf{l}, \mathbf{k} \circ \mathbf{m}, \\ & \mathbf{l} \circ \mathbf{m}, (\boldsymbol{\iota}_T \otimes \mathbf{h}) \circ \mathbf{k}, (\boldsymbol{\iota}_T \otimes \mathbf{h}) \circ \mathbf{l}, (\boldsymbol{\iota}_T \otimes \mathbf{h}) \circ \mathbf{m}, \mathbf{e}, \mathbf{f}, \mathbf{i}, \boldsymbol{\iota}_T \otimes \mathbf{p}, \overline{\mathbf{S}}], \end{aligned} \tag{31}$$

where $\circ$ denotes element-wise products.[10] The matrix $\overline{\mathbf{S}}$ contains the spatially lagged covariates from Set a (representing 4 contextual effects) and b (representing 17 contextual effects). With $\mathbf{W} = (w_{ij})$ denoting an $N \times N$ spatial weights matrix, which will be defined below, the $\ell$th element of $\overline{\mathbf{S}}$ is defined as $(\mathbf{I}_T \otimes \mathbf{W})\mathbf{s}_\ell$, where $\mathbf{s}_\ell$ is one vector about which we calculate a contextual counterpart. Hence, $\overline{\mathbf{S}}$ reflects the presence of *contextual effects* (or *local spillovers*).

In general, we do not enforce linear homogeneity of the technology by restricting the parameters on the quadratic and interactive terms to sum up to unity (see Greene, 2008). Linear homogeneity of the production technology is refuted by the data at hand.

In the next subsection, we consider the case where $\mathbf{W}^0$ is full, so that any distance matters and spillovers may occur between all firms in the chemical industry. We consider three alternative cases for the elements of $\mathbf{W}^0$ in a sensitivity analysis.

## 4.3 Estimation results

We report spatial model parameter estimates of the translog production functions in primary form in Tables 4 and 5 for two different specifications of the local spillover variables in $\overline{\mathbf{S}}$. Table 4 is based on the smaller Set a with 4 contextual effects and Table 5 on the larger Set b with 17 contextual effects. We account for heteroskedasticity and spatial correlation in the error term (global unspecified spillovers) by applying the procedure described in Section 2. These tables contain parameters and test statistics for spatial fixed effects (SFE), spatial random effects (SRE) and spatial Hausman and Taylor (SHT) type models. The two tables report parameter estimates and test statistics for specifications assuming that there are global spillovers in $\mathbf{u}$, allowing for $\rho \neq 0$.[11]

– Tables 4 and 5 about here –

[9] The Cobb Douglas counterpart to this model would use $[\mathbf{X}, \boldsymbol{\iota}_T \otimes \mathbf{Z}] = [\mathbf{k}, \mathbf{l}, \boldsymbol{\iota}_T \otimes \mathbf{h}, \mathbf{m}, \mathbf{e}, \mathbf{f}, \mathbf{i}, \boldsymbol{\iota}_T \otimes \mathbf{p}, \overline{\mathbf{S}}]$, which restricts the quadratic and interactive terms to be zero. Since the Cobb Douglas model is generally rejected for the data at hand, we will only report parameter estimates for the more flexible translog model.

[10] Notice that we do not interact vectors $\mathbf{e}, \mathbf{f}, \mathbf{i}, \boldsymbol{\iota}_T \otimes \mathbf{p}$. There is no principal argument against allowing for even more flexibility in interacting these vectors with each other and with the primary production factors. However, $\mathbf{e}, \mathbf{i}, \boldsymbol{\iota}_T \otimes \mathbf{p}$ represent vectors of binary variables so that using an even more flexible form drastically increases the degree of multicollinearity among the regressors. Therefore, we choose the form in (31) (see Burkett and Škegro, 1989, for an earlier example in that vein).

[11] In Tables 4, 6, and 8 the elements of $\mathbf{X_U}$ are $[\mathbf{f}, \overline{\mathbf{f}}]$; the elements of $\mathbf{Z_U}$ are $[\mathbf{h}, \overline{\mathbf{h}}, \mathbf{h} \circ \mathbf{h}]$; the elements of $\mathbf{X_C}$ are $[\mathbf{l}, \mathbf{l} \circ \mathbf{l}, \mathbf{l} \circ \mathbf{k}, \mathbf{l} \circ \mathbf{m}, \mathbf{l} \circ \mathbf{h}, \mathbf{k}, \mathbf{k} \circ \mathbf{k}, \mathbf{k} \circ \mathbf{h}, \mathbf{k} \circ \mathbf{m}, \mathbf{m}, \mathbf{m} \circ \mathbf{m}, \mathbf{m} \circ \mathbf{h}, \mathbf{e}, \overline{\mathbf{e}}, \mathbf{i}, \overline{\mathbf{i}}]$ and the element of $\mathbf{Z_C}$ is $\mathbf{p}$. In Table 5 and 7 the elements of $\mathbf{X_U}$ are $[\mathbf{f}, \overline{\mathbf{f}}]$; the elements of $\mathbf{Z_U}$ are $[\mathbf{h}, \overline{\mathbf{h}}, \mathbf{h} \circ \mathbf{h}, \overline{\mathbf{h} \circ \mathbf{h}}]$; the elements of $\mathbf{X_C}$ are $[\mathbf{l}, \overline{\mathbf{l}}, \mathbf{l} \circ \mathbf{l}, \overline{\mathbf{l} \circ \mathbf{l}}, \mathbf{l} \circ \mathbf{k}, \overline{\mathbf{l} \circ \mathbf{k}}, \mathbf{l} \circ \mathbf{m}, \overline{\mathbf{l} \circ \mathbf{m}}, \mathbf{l} \circ \mathbf{h}, \overline{\mathbf{l} \circ \mathbf{h}}, \mathbf{k}, \overline{\mathbf{k}}, \mathbf{k} \circ \mathbf{k}, \overline{\mathbf{k} \circ \mathbf{k}}, \mathbf{k} \circ \mathbf{h}, \overline{\mathbf{k} \circ \mathbf{h}}, \mathbf{k} \circ \mathbf{m}, \overline{\mathbf{k} \circ \mathbf{m}}, \mathbf{m}, \overline{\mathbf{m}}, \mathbf{m} \circ \mathbf{m}, \overline{\mathbf{m} \circ \mathbf{m}}, \mathbf{m} \circ \mathbf{h}, \overline{\mathbf{m} \circ \mathbf{h}}, \mathbf{e}, \overline{\mathbf{e}}, \mathbf{i}, \overline{\mathbf{i}}]$; and the element of $\mathbf{Z_C}$ is $\mathbf{p}$.

The results in Tables 4 and 5 can be summarized as follows. First, the null hypothesis of homoskedasticity of the remainder (within) error terms is rejected.[12] Accordingly, we only present results of models accounting for heteroskedasticity. Second, the heteroskedasticity-robust Hausman tests reject the absence of correlation of the regressors and the time-invariant error term. Hence, there is evidence of misspecification in the SRE results, and the corresponding parameter estimates are inconsistent and subject to misleading inference. For the SHT instrument set, we find a high joint degree of relevance in the first-stage regression (with triple-digit F-statistics) and the sets of instruments pass the tests in Tables 4 and 5. Third, the estimates of the spatial autocorrelation parameter on $\mathbf{W}\widehat{\mathbf{u}}$, $\widehat{\rho}$, are 0.433 and 0.358 for the SRE estimators and 0.512 and 0.302 for the SHT estimators in Tables 4 and 5, respectively. All estimates of $\rho$ are statistically significant. In general, the inconsistency of the SRE estimator also feeds into estimates of $\rho$. Using a larger set of variables capturing contagious effects in Table 5 relative to Table 4 reduces the spatial autocorrelation of the disturbances, capturing global spillovers associated with unobservable TFP shifters.

Fourth, a comparison of the decomposition of the variance in outcome ($\sigma_y^2$) at the bottom of the Tables 4 and 5 suggests the following. First, as expected, the primary production factors contribute the lion's share ($\sigma_P^2$). Among the other components, the own specified TFP shifters (through $\mathbf{e}$, $\mathbf{f}$, $\mathbf{i}$, $\boldsymbol{\iota}_T \otimes \mathbf{p}$) contribute the biggest share in the SHT models ($\sigma_T^2$). Specified local TFP spillovers from neighbors through effects flowing from their TFP shifters ($\sigma_{\overline{T}}^2$) or their primary production factors ($\sigma_{\overline{P}}^2$) have minor contribution to explaining the overall variance ($\sigma_y^2$). Relative to the just-mentioned specified local spillover terms, global spillovers from time-invariant ($\sigma_\mu^2$) and average heteroskedastic time-variant unspecified components ($\sigma_\nu^2$) are more important.

Fifth, export market participation ($e_{it}$), foreign ownership ($f_{it}$), and state ownership ($p_{it}$) are found to be statistically significant. In fact, export market participation and foreign ownership raise TFP, while state ownership reduces it. The comparison among the SRE and SHT models suggests that state ownership is endogenous and its effect can not be estimated consistently by way of an SRE model.

Sixth, contextual effects associated with $\overline{\mathbf{S}}$ matter jointly. However, their joint contribution to the variance of outcome is much smaller than that of the between component, $\sigma_\mu^2$, or the average within component, $\sigma_\nu^2$, of unspecified TFP, according to the results at the bottom of Tables 4 and 5.

The two SHT models in Tables 4 and 5 suggest that publicly-owned companies' TFP is about $100 \cdot (exp(-1.819) - 1)$ and $100 \cdot (exp(-1.861) - 1)$, or about 84 percent lower than that for privately-owned companies in China's chemical industry. Hence, significant efficiency gains could be had with privatization in that industry. Shocks in TFP display significant spillover effects across firms, and they are geographically bound. Providing a more detailed assessment of the geographical reach of such spillovers is the goal of the next subsection.

## 4.4 Sensitivity checks and quantification of spillovers

First of all, we present results based on even more general models than the ones in Tables 4 and 5, where we permit global spillovers (measured by the estimates of $\rho$) to differ for relationships among privately-owned firms ($\rho_r$), among publicly-owned firms ($\rho_p$), and

[12] For testing against homoskedasticity based on a Breusch-Pagan test, we used the Cochrane-Orcutt-transformed residuals of the spatial FE estimator.

between privately- and publicly-owned firms ($\rho_o$). This process may be viewed as a third-order SHT model. The associated results in Tables 6 and 7 suggest that global spillovers mainly happen among privately-owned firms rather than among other types of firms, irrespective of whether we specify the model otherwise as in Table 4 or 5.

– Tables 6 – 8 about here –

Moreover, we explore the sensitivity of our estimates to alternative weights matrices differing by the specification of the spatial decay function.[13] We report the sensitivity checks in Table 8 for the otherwise same specification as in Table 4, focusing on the SHT estimator. The three columns of Table 8 involve weights matrices that are based on positive cell entries $w_{ij}$ if firms $i$ and $j$ are closer than 60 miles ($\mathbf{W}_{60}$), 100 miles ($\mathbf{W}_{100}$), or 200 miles ($\mathbf{W}_{200}$), respectively. Hence, the results based on $\mathbf{W}_{60}$ correspond to the sparsest weights matrix considered, and in turn, all weights matrices $\mathbf{W}_{60}$, $\mathbf{W}_{100}$, and $\mathbf{W}_{200}$ are sparser than the original $\mathbf{W}$ matrix. With row normalization, it will generally be the case that $\sum_{j=1}^{N} w_{60,ij}^0 \leq \sum_{j=1}^{N} w_{100,ij}^0 \leq \sum_{j=1}^{N} w_{200,ij}^0 \leq \sum_{j=1}^{N} w_{ij}^0$. Therefore, the positive individual cells of the respective matrices have the property $0 < w_{ij} \leq w_{200,ij} \leq w_{100,ij} \leq w_{60,ij} \leq 1$. The associated results suggest that spillovers are stronger among closeby entities. Other than that, our findings from Tables 4 and 5 are qualitatively unaffected by the choice of the weighting scheme considered.

How strong are spillovers in China's chemical industry? There are numerous ways of assessing this question. One of them is to consider the impact of a common shock on total factor productivity of all firms in the sample. Suppose, we considered a shock of one percent in observable and unobservable determinants of TFP. The specification of $\mathbf{W}$ – in particular, its row normalization – suggests that such a uniform shock will trigger total effects inclusive of local and global spillovers together on output across all firms of $\exp\left((\hat{\delta}_e + \hat{\delta}_f + \hat{\delta}_i + \hat{\delta}_p) + (\sum_\ell \hat{\delta}_{\overline{S},\ell}) + \frac{1}{1-\hat{\rho}}\right) - 1$ percent, where we use $[\hat{\delta}_e, \hat{\delta}_f, \hat{\delta}_i, \hat{\delta}_p]$ to denote the parameters on $[\mathbf{e}, \mathbf{f}, \mathbf{i}, \boldsymbol{\iota}_T \otimes \mathbf{p}]$ and $\hat{\delta}_{\overline{S},\ell}$ to denote the parameters on the local spillover terms in $\overline{\mathbf{S}}$. Using the SHT model in Table 4, we would conclude that the response to such spillovers is positive and relatively strong.[14] In the absence of spillovers, TFP would be predicted to decline by about 0.5 percent in response to this shock. The spillover terms together lead to a positive overall shock of about 0.8 percent. Hence, local spillovers from observable TFP shifters and global spillovers from unobservable TFP shifters together mitigate the detrimental effects due to state ownership.

# 5 Conclusions

This paper derives a Hausman and Taylor (1981) type estimator which allows for spatial dependence in the disturbances in the presence of homoskedastic or heteroskedastic disturbances. We discuss the large-sample properties of the estimator and show its small-sample performance by way of Monte Carlo simulations. We then apply the spatial Hausman-Taylor model to the estimation of a primal translog production function for the Chinese chemical industry. The data-set utilized in this study consists of a large cross section of 12,552 firms observed annually over the short period of 2004–2006. The proposed estimator allows some of the regressors to be time-invariant and perhaps correlated with the

[13] However, there is limited scope for sensitivity checks, since estimating spatial models in a data-set as large as ours and with a weighting matrix of size $12,552 \times 12,552$ that is not sparse in the sense of Bickel and Levina (2008) or Fan, Liao, and Mincheva (2013) is very computer intensive.

[14] Notice that the direction of the spillovers is partly driven by the design of the assumed weights matrix.

unobservable firm effects. This estimator has some advantages over the spatial fixed effects or random effects estimators. In fact, the spatial fixed effects estimator does not provide estimates of the time-invariant variables since they are wiped out by the within transformation. Also, the random effects estimates suffer from bias and inconsistency when the regressors are correlated with the firm effects.

The firms in the data are concentrated in the coastal area of China. We estimate a translog production function in primal form and find that the Cobb Douglas production function restrictions are rejected by the data, and so is the assumption of linear homogeneity of the technology. The estimated size of the spatial autocorrelation parameter suggests that there are moderately important *local* spillover effects emerging from observable characteristics and more important *global* spillover effects emerging from unobservable determinants of total factor productivity across firms. The results point to significant detrimental effects of public ownership on total factor productivity. Due to the time-invariant nature of public ownership, this effect could not have been estimated in a spatial fixed effects model.

# References


Amemiya, T, MaCurdy, TE. 1986. Instrumental-variable estimation of an error-components model. *Econometrica* **54**: 869-880. DOI: 10.2307/1912840

Badinger, H, Egger, P. 2014. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroskedastic disturbances. CESifo Working Paper No. 4847. CESifo, Munich.

Baldwin, JR, Gu, W. 2006. Plant turnover and productivity growth in Canadian manufacturing. *Industrial and Corporate Change* **15**: 417-465. DOI: 10.1093/icc/dtj017

Baltagi, BH. 2008. *Econometric Analysis of Panel Data*. Wiley: Chichester.

Baltagi, BH, Egger, PH, Kesina, M. 2012. Small sample properties and pretest estimation of a spatial Hausman-Taylor model. *Advances in Econometrics* **29**: 215-236. DOI: 10.1108/S0731-9053(2012)0000029013

Bickel, PJ, Levina, E. 2008. Covariance regularization by thresholding. *The Annals of Statistics* **36**: 2577-2604. DOI: 10.1214/08-AOS600

Blonigen, B, Ma, A. 2007. Please pass the catch-up: the relative performance of Chinese and foreign firms in Chinese exports. NBER Working paper 13376. DOI: 10.3386/w13376

Bloom, N, Schankerman, M, van Reenen, J. 2007. Identifying technology spillovers and product market rivalry. NBER Working Paper 13060. DOI: 10.3386/w13060

Breusch, TS, Mizon, GE, and Schmidt, P. 1989. Efficient estimation using panel data. *Econometrica* **57**: 695-700. DOI: 10.2307/1911060

Burkett, JP, Škegro, B. 1989. Capitalism, socialism, and productivity: An econometric analysis of CES and translog functions. *European Economic Review* **33**: 1115-1133. DOI: 10.1016/0014-2921(89)90088-3

Chen, H, Swenson, D. 2006. Multinational firms and new Chinese export transactions. Unpublished manuscript. University of California at Davis.

Drukker, D, Egger, PH, Prucha, IR. 2013. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews* **32**: 686-733. DOI:10.1080/07474938.2013.741020

Fan, J, Liao, Y, Mincheva, M. 2013. Large covariance estimation by thresholding principal orthogonal complements. *Journal of Royal Statistical Society, Series B* **75**: 1-44. DOI: 10.1111/rssb.12016

Görg, H, Strobl, E. 2005. Spillovers from foreign firms through worker mobility: an empirical investigation. *Scandinavian Journal of Economics* **107**: 693–709. DOI: 10.1111/j.1467-9442.2005.00427.x

González-Páramo, JM, Hernández Cos, P. 2005. The impact of public ownership and competition on productivity. *Kyklos* **58**: 495-517. DOI: 10.1111/j.0023-5962.2005.00299.x

Greenaway, D, Kneller, R. 2007. Industry differences in the effect of export market entry: learning by exporting? *Review of World Economics (Weltwirtschaftliches Archiv)* **143**: 416-432. DOI: 10.1007/s10290-007-0115-y

Greene, W. 2008. *Econometric Analysis*. Prentice Hall: New Jersey.

Hall, BH, Lerner, J. 2010. The financing of R&D and innovation. In *Handbook of the Economics of Innovation*, Hall, BH, Rosenberg N (eds.). North-Holland: Amsterdam.

Hausman, JA, Taylor, WE. 1981. Panel data and unobservable individual effects. *Econometrica* **49**: 1377-1398. DOI: 10.2307/1911406

Himmelberg, CP, Petersen, BC. 1994. R&D and internal finance: a panel study of small firms in high-tech industries, *Review of Economics and Statistics* **76**: 38-51. DOI: 10.2307/2109824

Hu, A, Jefferson, G, Jinchang, Q. 2005. R&D and technology transfer: firm-level evidence from Chinese industry. *Review of Economics and Statistics* **87**: 780-786. DOI: 10.1162/003465305775098143

Kapoor, M, Kelejian, HH, and Prucha, IR. 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* **140**: 97-130. DOI: 10.1016/j.jeconom.2006.09.004

Kelejian, HH, Prucha, IR. 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* **118**: 27-50. DOI: 10.1016/S0304-4076(03)00133-7

Kelejian, HH, Prucha, IR. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* **157**: 53-67. DOI: 10.1016/j.jeconom.2009.10.025

Keller, W, Yeaple, SR. 2009. Multinational enterprises, international trade and productivity growth: firm level evidence from the United States. *Review of Economics and Statistics* **91**: 821-831. DOI:10.1162/rest.91.4.821

Keller, W. 2004. International technology diffusion. *Journal of Economic Literature* **XLII**: 752-782. DOI: 10.1257/0022051042177685

Li, Q, Stengos, T. 1994. Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. *International Economic Review* **35**: 981–1000. DOI: 10.2307/2527006

Lööf, H. 2007. Technology spillovers and innovation: the importance of domestic and foreign sources. CESIS Electronic Working Paper No. 83, Royal Institute of Technology, Sweden.

Mutl, J, Pfaffermayr, M. 2011. The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* **14**: 48-76. DOI: 10.1111/j.1368-423X.2010.00325.x

Pesaran, MH. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**: 967-1012. DOI: 10.1111/j.1468-0262.2006.00692.x

Pesaran, MH, Tosetti, E. 2011. Large panels with common factors and spatial correlation. *Journal of Econometrics* **161**: 182-202. DOI: 10.1016/j.jeconom.2010.12.003

Pesaran, MH, Smith, LV, Yamagata, T. 2013. Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics* **175**: 94-115. DOI: 10.1016/j.jeconom.2013.02.001

Pötscher, B, Prucha, IR. 2001. Basic elements of asyptotic theory. In: *A Companion to Theoretical Econometrics*, Baltagi, BH (ed.). Blackwell: Oxford.

Puttitanun, T. 2006. Intellectual property rights and multinational firms' modes of entry. *Journal of Intellectual Property Rights* **11**: 269-273.

Smarzynska Javorcik, B. 2004. Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, **94**: 605-627. DOI: 10.1257/aer.102.7.3594

Stock, JH, Watson, MW. 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica* **76**: 155-174. DOI: 10.1111/j.0012-9682.2008.00821.x

Szamosszegi, A, and Kyle, C. 2011. An analysis of state-owned enterprises and state capitalism in China. *U.S.-China Economic and Security Review Commission.*

UNEP Chemicals Branch. 2009. *Global Chemicals Outlook: Rationale and Goals.*

# A. Appendix

## A.1 SHT estimator under homoskedasticity

**Proof of Theorem 1**

The proof utilizes the insights from the proof of Theorem 4 in Kapoor, Kelejian, and Prucha (2007) and the proofs in Badinger and Egger (2014).

**Part (i).**

The feasible spatial Hausman and Taylor estimator under homoskedasticity is given by

$$\widehat{\widehat{\boldsymbol{\delta}}}_{SHT} = (\widetilde{\mathbf{A}}^{**\prime}\widetilde{\mathbf{P}}^{**}_{SHT}\widetilde{\mathbf{A}}^{**})^{-1}\widetilde{\mathbf{A}}^{**\prime}\tilde{\mathbf{P}}^{**}_{SHT}\widetilde{\mathbf{y}}^{**}. \tag{32}$$

To avoid index cluttering, we will suppress the subindex $SHT$. Plugging in (10)-(11) and using the transformations based on the estimators from Step 3 yields

$$\begin{aligned}(NT)^{1/2}(\widehat{\widehat{\boldsymbol{\delta}}} - \boldsymbol{\delta}) \quad = \quad & \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{H}}^{*}((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{H}}^{*})^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{A}}^{*}\}^{-1} \\ & (NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{H}}^{*}((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{H}}^{*})^{-1} \\ & (NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{u}}^{*}. \end{aligned} \tag{33}$$

The terms of the first two lines will be analyzed first. For the first term, we write

$$\begin{aligned}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_{\varepsilon}^{-1}\widetilde{\mathbf{H}}^{*} \quad = \quad & (NT)^{-1}[\mathbf{A}^{*} - (\hat{\rho} - \rho)(\mathbf{I}_T \otimes \mathbf{W})\mathbf{A}]' \\ & \left[(\sigma_{\nu}^{-2}\mathbf{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1) + (\hat{\sigma}_{\nu}^{-2} - \sigma_{\nu}^{-2})\mathbf{Q}_0 + (\hat{\sigma}_1^{-2} - \sigma_1^{-2})\mathbf{Q}_1\right]\end{aligned}$$

$$[\mathbf{H}^* - (\hat\rho-\rho)(\mathbf{I}_T \otimes \mathbf{W})\mathbf{H}]. \tag{34}$$

For $N \to \infty$, using consistency of $\hat\rho$, $\hat\sigma_\nu^{-2}$, and $\hat\sigma_1^{-2}$, and using Assumptions 1, 2, 5, 6, and 8 yields

$$(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^* = \sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*} + o_p(1). \tag{35}$$

Similarly, we get

$$(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{A}}^* = \sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*} + o_p(1) \tag{36}$$

and

$$\begin{aligned}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^* &= \sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*} + o_p(1), &(37)\\ ((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1} &= (\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1} + o_p(1). &(38)\end{aligned}$$

Combining yields

$$(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*}) = O(1) \tag{39}$$

and also

$$\{(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*})\}^{-1} = O(1). \tag{40}$$

Thus we have that

$$\begin{aligned}&\{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{A}}^*\}^{-1}\\ =\ &\{(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*})\}^{-1}\\ +\ &o_p(1) \qquad (41)\end{aligned}$$

Hence, for the first two lines of (33) together we get

$$\begin{aligned}&\{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{A}}^*\}^{-1}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*\\ &((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{H}}^*)^{-1}\\ =\ &\{(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1}(\sigma_\nu^{-2}\mathbf{M}^{0\prime}_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^{1\prime}_{\mathbf{A}^*\mathbf{H}^*})\}^{-1}\\ &(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{A}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{A}^*\mathbf{H}^*})(\sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*})^{-1} + o_p(1) = \mathbf{F} + o_p(1). \qquad (42)\end{aligned}$$

Rewriting the expression in the last line of (33), we obtain

$$\begin{aligned}(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{u}}^* &= (NT)^{-1/2}[\mathbf{H}^* - (\hat\rho-\rho)(\mathbf{I}_T\otimes\mathbf{W})\mathbf{H}]'[(\sigma_\nu^{-2}\mathbf{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)\\ &+ (\hat\sigma_\nu^{-2} - \sigma_\nu^{-2})\mathbf{Q}_0 + (\hat\sigma_1^{-2} - \sigma_1^{-2})\mathbf{Q}_1][\mathbf{I}_{TN} - (\hat\rho-\rho)(\mathbf{I}_T\otimes\mathbf{W})\\ &\quad(\mathbf{I}_T\otimes(\mathbf{I}_N - \rho\mathbf{W}))^{-1}]\boldsymbol{\varepsilon}. \qquad (43)\end{aligned}$$

For $N \to \infty$, using consistency of $\hat\rho$, $\hat\sigma_\nu^{-2}$, and $\hat\sigma_1^{-2}$, and using Assumptions 1, 5, and 6 this expression reduces to

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{u}}^* = (NT)^{-1/2}\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\mathbf{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)\boldsymbol{\varepsilon} + o_p(1). \tag{44}$$

Plugging in $\boldsymbol{\varepsilon} = (\boldsymbol{\iota}_T\otimes\mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\nu}$ and using $\mathbf{Q}_1(\boldsymbol{\iota}_T\otimes\mathbf{I}_N) = (\boldsymbol{\iota}_T\otimes\mathbf{I}_N)$ yields

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{u}}^* = (NT)^{-1/2}\sigma_1^{-2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T\otimes\mathbf{I}_N)\boldsymbol{\mu} + (NT)^{-1/2}\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\mathbf{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)\boldsymbol{\nu} + o_p(1). \tag{45}$$

For the first term of (45), we have that $\sigma_1^{-2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N)$ is uniformly bounded in absolute value by Assumptions 5 and 6, and that for $N \to \infty$

$$(NT)^{-1}(\sigma_1^{-2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N))(\sigma_1^{-2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N))' = \sigma_1^{-4}T\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*} + o_p(1), \quad (46)$$

Now we can apply the theorem in Pötscher and Prucha (2001) and get

$$(NT)^{-1/2}\sigma_1^{-2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} \ \stackrel{d}{\to}\ N(\mathbf{0}, \sigma_\mu^2 T\sigma_1^{-4}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*}). \quad (47)$$

We proceed in the same way for the second term of (45). We have that $(NT)^{-1/2}\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\boldsymbol{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)$ is uniformly bounded in absolute value by Assumptions 5 and 6. Using obvious reformulations, we obtain

$$(NT)^{-1}\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\boldsymbol{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)(\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\boldsymbol{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1))' = \sigma_\nu^{-4}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-4}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*} + o_p(1) \quad (48)$$

and, by the theorem in Pötscher and Prucha (2001), we get

$$(NT)^{-1/2}\mathbf{H}^{*\prime}(\sigma_\nu^{-2}\boldsymbol{Q}_0 + \sigma_1^{-2}\mathbf{Q}_1)\boldsymbol{\nu} \ \stackrel{d}{\to}\ N(\mathbf{0}, \sigma_\nu^2[\sigma_\nu^{-4}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-4}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*}]). \quad (49)$$

Using Assumption 6 and $\sigma_1^2 = \sigma_\nu^2 + T\sigma_\mu^2$, combining (47) and (49) yields

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\boldsymbol{\Omega}}_\varepsilon^{-1}\widetilde{\mathbf{u}}^* \stackrel{d}{\to} N(\mathbf{0}, \sigma_\nu^{-2}\mathbf{M}^0_{\mathbf{H}^*\mathbf{H}^*} + \sigma_1^{-2}\mathbf{M}^1_{\mathbf{H}^*\mathbf{H}^*}). \quad (50)$$

Combining (42) and (50) proves the claim.
**Part (ii).** This follows from (37) and (42) in Part (i).

### A.2 SHT estimator under heteroskedasticity

**Proof of Theorem 2**
The proof utilizes the insights from the proof of Lemma 2 in Badinger and Egger (2014).
**Part (i).**
The feasible SHT estimator under heteroskedasticity is given by

$$\hat{\tilde{\boldsymbol{\delta}}}_{SHT} = (\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{P}}^*_{SHT}\widetilde{\mathbf{A}}^*)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{P}}^*_{SHT}\widetilde{\mathbf{y}}^*. \quad (51)$$

To avoid index cluttering, we will suppress the subindex $SHT$ as above. Plugging in (21) and reformulating yields

$$\begin{aligned}(NT)^{1/2}(\hat{\tilde{\boldsymbol{\delta}}} - \boldsymbol{\delta}) \quad = \quad & \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{A}}^*\}^{-1} \\ & (NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1} \\ & (NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{u}}^*. \quad (52)\end{aligned}$$

When considering obvious modifications, including assumptions, but otherwise analogous arguments as in the case of homoskedasticity, we obtain for the first two lines of (52) that

$$\begin{aligned}& \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{A}}^*\}^{-1}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^*((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^*)^{-1} \\ = \quad & \{(\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*}\mathbf{M}'_{\mathbf{A}^*\mathbf{H}^*}\}^{-1}\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*} + o_p(1) = \mathbf{D} + o_p(1) \quad (53)\end{aligned}$$

where the last line is $O(1)$. For the expression in the last line of (52), upon reformulation, we obtain

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{u}}^* \quad = \quad (NT)^{-1/2}[\mathbf{H}^* - (\hat{\rho} - \rho)(\mathbf{I}_T \otimes \mathbf{W})\mathbf{H}]'[\mathbf{I}_{TN} - (\hat{\rho} - \rho)(\mathbf{I}_T \otimes \mathbf{W})$$

$$(\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W}))^{-1}]\boldsymbol{\varepsilon}. \tag{54}$$

For $N \to \infty$ using consistency of $\hat{\rho}$ and Assumptions 1, 5, and 9 this expression reduces to

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{u}}^{*} \;\; = \;\; (NT)^{-1/2}\mathbf{H}^{*\prime}\boldsymbol{\varepsilon} + o_p(1). \tag{55}$$

Plugging in $\boldsymbol{\varepsilon} = (\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\nu}$ yields

$$(NT)^{-1/2}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{u}}^{*} \;\; = \;\; (NT)^{-1/2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + (NT)^{-1/2}\mathbf{H}^{*\prime}\boldsymbol{\nu} + o_p(1) \tag{56}$$

For the first term of (56) have that $\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N)$ is uniformly bounded in absolute value by Assumption 5 and 9 and that for $N \to \infty$

$$\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \mathbf{I}_N)\mathbf{H}^{*} = \mathbf{M}^{J}_{\mathbf{H}^*\mathbf{H}^*} + o_p(1) \tag{57}$$

using Assumption 11. By the theorem in Pötscher and Prucha (2001) we get

$$(NT)^{-1/2}\mathbf{H}^{*\prime}(\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} \xrightarrow{d} N(\mathbf{0}, \sigma^2_\mu \mathbf{M}^{J}_{\mathbf{H}^*\mathbf{H}^*}). \tag{58}$$

In analyzing the second expression in (56), we apply the CLT for heteroskedastic innovations in Kelejian and Prucha (2010) and get

$$(NT)^{-1/2}\mathbf{H}^{*\prime}\boldsymbol{\nu} \xrightarrow{d} N(\mathbf{0}, \mathbf{M}_{\mathbf{H}^*\boldsymbol{\Sigma}\mathbf{H}^*}) \tag{59}$$

since $\mathbf{H}^{*\prime}$ is uniformly bounded in absolute value by Assumption 5 and for $N \to \infty$ we have $\mathbf{H}^{*\prime}\boldsymbol{\Sigma}\mathbf{H}^{*} = \mathbf{M}_{\mathbf{H}^*\boldsymbol{\Sigma}\mathbf{H}^*} + o_p(1)$ by Assumption 11. Combining (58) and (59) using Assumption 9 yields

$$(NT)^{-1/2}\mathbf{H}^{*\prime}((\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\nu}) \xrightarrow{d} N(\mathbf{0}, \sigma^2_\mu \mathbf{M}^{J}_{\mathbf{H}^*\mathbf{H}^*} + \mathbf{M}_{\mathbf{H}^*\boldsymbol{\Sigma}\mathbf{H}^*}) \tag{60}$$

Combing (53) and (60) proves the claim.

**Part (ii).**

The variance covariance matrix $\boldsymbol{\Psi}$ consists of two parts: one is related to the homoskedastic $\boldsymbol{\mu}$ and the other to the heteroskedastic $\boldsymbol{\nu}$. We know that $\boldsymbol{\mu} \sim N(0, \sigma^2_\mu)$, where $\sigma^2_\mu$ can be estimated consistently by the GM procedure in (19). Regarding the heteroskedastic part, we have that $\boldsymbol{\nu} \sim N(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = diag^{NT}_{n=1}(\nu^2_n)$. However, note that while $\boldsymbol{\Sigma}$ is based on idiosyncratic errors in levels, estimates of idiosyncratic errors can only be retrieved in demeaned form. A similar problem is addressed in Stock and Watson (2008), who estimate a heteroskedasticity-robust variance covariance matrix in a non-spatial fixed effects panel data model by using a bias correction of the variance covariance matrix. Badinger and Egger (2014) adapt the procedure to spatial fixed effects and random effects models, which contain an endogenous spatial lag as well as spatial correlation in the error and heteroskedasticity in the idiosyncratic error component. They suggest using the following heteroskedasticity-robust estimate of $\boldsymbol{\Sigma}$,

$$\widetilde{\boldsymbol{\Sigma}}^{HR} = diag^{NT}_{it=1}[(\underline{\widetilde{\nu}}^{HR}_{it})^2] \quad \text{with} \quad (\underline{\widetilde{\nu}}^{HR}_{it})^2 = \tfrac{T}{T-2}\underline{\widetilde{\nu}}^2_{it} - \tfrac{1}{(T-1)(T-2)}\textstyle\sum^{T}_{r=1}\underline{\widetilde{\nu}}^2_{ir}$$

which are based on within transformed residuals

$$\underline{\widetilde{\boldsymbol{\nu}}} = (\underline{\widetilde{\nu}}_{it}) = \boldsymbol{Q}_0[\mathbf{I}_T \otimes (\mathbf{I}_N - \hat{\rho}\mathbf{W})]\hat{\mathbf{u}} \quad \text{with} \quad \hat{\mathbf{u}} = \mathbf{y} - \mathbf{A}\hat{\tilde{\boldsymbol{\delta}}}_{SHT}. \tag{61}$$

Badinger and Egger (2014) show that

$$\frac{1}{NT}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{\Sigma}}^{HR}\widetilde{\mathbf{H}}^{*} - \frac{1}{NT}\mathbf{H}^{*\prime}\mathbf{\Sigma}\mathbf{H}^{*} = o_p(1) \tag{62}$$

(see their Lemmas C.2 and C.3). We will use this result in our proof.
From (53) we already know that

$$\begin{aligned} & \{(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^{*}((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^{*})^{-1}(NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{A}}^{*}\}^{-1}(NT)^{-1}\widetilde{\mathbf{A}}^{*\prime}\widetilde{\mathbf{H}}^{*}((NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{H}}^{*})^{-1} \\ = \; & \{(\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*}\mathbf{M}'_{\mathbf{A}^*\mathbf{H}^*}\}^{-1}\mathbf{M}_{\mathbf{A}^*\mathbf{H}^*}\mathbf{M}^{-1}_{\mathbf{H}^*\mathbf{H}^*} + o_p(1). \end{aligned} \tag{63}$$

For the term in the middle of $\widetilde{\mathbf{\Psi}}$, we get for $N \to \infty$

$$\{\hat{\sigma}^2_\mu(NT)^{-1}(\widetilde{\mathbf{H}}^{*\prime}(\boldsymbol{\iota}_T\boldsymbol{\iota}'_T \otimes \mathbf{I}_N)\widetilde{\mathbf{H}}^{*}) + (NT)^{-1}\widetilde{\mathbf{H}}^{*\prime}\widetilde{\mathbf{\Sigma}}^{HR}\widetilde{\mathbf{H}}^{*}\} = \sigma^2_\mu\mathbf{M}^J_{\mathbf{H}^*\mathbf{H}^*} + \mathbf{M}_{\mathbf{H}^*\mathbf{\Sigma}\mathbf{H}^*} + o_p(1) \tag{64}$$

using Assumptions 1, 5, 9, and 11 together with consistency of $\hat{\rho}$ and (62). Combining (63) and (64) finally yields the desired result.

Table 1: Homoskedastic world, $N = 100$ and $N = 400$, $T = 3$

| N | $\rho_0$ | $\rho$ | | $\beta_{11}$ | | | $\beta_{12}$ | | | $\beta_2$ | | | $\theta_2$ | | | HT | HTT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias | RMSE | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size | Power | Size |
| 100 | -0.8 | -0.007 | 0.224 | -0.002 | 0.055 | 6.6 | -0.003 | 0.055 | 6 | 0.004 | 0.070 | 6.2 | 0.001 | 0.209 | 6.1 | 73.9 | 4 |
| 100 | -0.6 | -0.013 | 0.218 | 0.001 | 0.055 | 5.7 | 0.002 | 0.056 | 5.9 | -0.001 | 0.074 | 6.9 | -0.007 | 0.212 | 4.9 | 76.2 | 4.8 |
| 100 | -0.4 | -0.019 | 0.205 | 0.001 | 0.055 | 5.8 | 0.002 | 0.056 | 5.5 | -0.001 | 0.074 | 6.8 | -0.007 | 0.214 | 5.1 | 76.6 | 4.6 |
| 100 | -0.2 | -0.023 | 0.190 | 0.001 | 0.055 | 5.3 | 0.002 | 0.056 | 5.6 | -0.001 | 0.075 | 6.9 | -0.008 | 0.215 | 5.2 | 77.2 | 4.6 |
| 100 | 0 | -0.025 | 0.171 | -0.004 | 0.078 | 6 | -0.003 | 0.076 | 5.4 | 0.000 | 0.100 | 5.4 | 0.008 | 0.290 | 4.2 | 50.7 | 3.5 |
| 100 | 0.2 | -0.025 | 0.149 | 0.001 | 0.055 | 5.4 | 0.002 | 0.056 | 5.3 | -0.001 | 0.075 | 7.2 | -0.008 | 0.214 | 5 | 76.2 | 4.7 |
| 100 | 0.4 | -0.024 | 0.124 | 0.001 | 0.055 | 5.1 | 0.002 | 0.056 | 5.3 | 0.000 | 0.074 | 7.3 | -0.008 | 0.213 | 5.3 | 75.3 | 4.5 |
| 100 | 0.6 | -0.020 | 0.092 | -0.005 | 0.077 | 6.3 | -0.003 | 0.079 | 6.2 | 0.007 | 0.104 | 6.3 | 0.009 | 0.301 | 5.7 | 46.9 | 3.6 |
| 100 | 0.8 | -0.019 | 0.060 | -0.001 | 0.076 | 5.6 | 0.000 | 0.077 | 6.4 | 0.004 | 0.097 | 6.7 | -0.010 | 0.291 | 4.2 | 47.1 | 3.2 |
| 400 | -0.8 | -0.005 | 0.109 | 0.000 | 0.026 | 5.2 | 0.000 | 0.026 | 5.1 | -0.001 | 0.035 | 5.9 | -0.002 | 0.102 | 5.3 | 100 | 4.5 |
| 400 | -0.6 | -0.002 | 0.108 | 0.000 | 0.027 | 5.2 | 0.000 | 0.027 | 5.2 | 0.001 | 0.036 | 6.3 | -0.003 | 0.103 | 5.1 | 100 | 4.9 |
| 400 | -0.4 | -0.003 | 0.101 | 0.000 | 0.027 | 5 | 0.000 | 0.027 | 5.2 | 0.001 | 0.037 | 6.2 | -0.003 | 0.104 | 4.8 | 100 | 4.8 |
| 400 | -0.2 | -0.005 | 0.092 | 0.000 | 0.027 | 5.1 | 0.000 | 0.027 | 5.1 | 0.001 | 0.037 | 5.9 | -0.002 | 0.105 | 4.5 | 100 | 5 |
| 400 | 0 | -0.012 | 0.082 | -0.001 | 0.038 | 5.5 | -0.002 | 0.037 | 4.8 | 0.003 | 0.051 | 5.1 | 0.001 | 0.145 | 4.7 | 99.3 | 4.9 |
| 400 | 0.2 | -0.005 | 0.071 | 0.000 | 0.027 | 5.3 | 0.000 | 0.027 | 4.9 | 0.001 | 0.037 | 6 | -0.002 | 0.104 | 5.1 | 100 | 4.6 |
| 400 | 0.4 | -0.005 | 0.058 | 0.000 | 0.027 | 5.3 | 0.000 | 0.027 | 4.9 | 0.001 | 0.037 | 6 | -0.002 | 0.104 | 5.1 | 100 | 4.6 |
| 400 | 0.6 | -0.007 | 0.045 | 0.000 | 0.038 | 5.7 | -0.001 | 0.038 | 4.7 | 0.001 | 0.050 | 5.7 | 0.002 | 0.145 | 5.6 | 98.9 | 5 |
| 400 | 0.8 | -0.005 | 0.027 | 0.001 | 0.037 | 5.6 | 0.001 | 0.036 | 5 | 0.001 | 0.047 | 5.7 | -0.005 | 0.138 | 4.7 | 98.9 | 4.4 |

Table 2: Heteroskedastic world, $N = 100$ and $N = 400$, $T = 3$

| N | $\rho_0$ | $\rho$ | | $\beta_{11}$ | | | $\beta_{12}$ | | | $\beta_2$ | | | $\theta_2$ | | | HT | HTT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias | RMSE | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size | Power | Size |
| 100 | -0.8 | 0.005 | 0.210 | 0.001 | 0.062 | 5.8 | -0.001 | 0.059 | 5.8 | -0.002 | 0.072 | 6.6 | -0.005 | 0.240 | 6.6 | 75.7 | 4.4 |
| 100 | -0.6 | -0.006 | 0.208 | -0.003 | 0.061 | 5.5 | -0.001 | 0.057 | 4.6 | 0.005 | 0.071 | 5.4 | 0.001 | 0.226 | 5.2 | 76.4 | 6.2 |
| 100 | -0.4 | -0.011 | 0.195 | 0.001 | 0.060 | 5.6 | -0.003 | 0.058 | 5.5 | 0.003 | 0.074 | 6.8 | -0.005 | 0.225 | 5 | 78.9 | 5.4 |
| 100 | -0.2 | -0.015 | 0.185 | -0.005 | 0.061 | 4.8 | -0.002 | 0.059 | 5.8 | 0.002 | 0.071 | 5.2 | 0.007 | 0.235 | 6 | 79.1 | 6.4 |
| 100 | 0 | -0.026 | 0.170 | -0.004 | 0.060 | 5.4 | -0.003 | 0.057 | 5 | 0.004 | 0.072 | 6 | 0.008 | 0.222 | 5.3 | 79.6 | 7.1 |
| 100 | 0.2 | -0.026 | 0.151 | -0.005 | 0.061 | 5.9 | -0.004 | 0.057 | 5.3 | 0.004 | 0.072 | 6.2 | 0.010 | 0.226 | 6 | 78.9 | 8.5 |
| 100 | 0.4 | -0.016 | 0.116 | -0.002 | 0.060 | 5.5 | -0.001 | 0.059 | 5.8 | 0.002 | 0.072 | 5.2 | 0.003 | 0.230 | 5.6 | 78.1 | 6.8 |
| 100 | 0.6 | -0.020 | 0.100 | -0.005 | 0.059 | 6.7 | -0.003 | 0.056 | 5.7 | 0.004 | 0.071 | 5.6 | 0.010 | 0.219 | 5.9 | 75.4 | 7.9 |
| 100 | 0.8 | -0.017 | 0.063 | -0.004 | 0.060 | 6.9 | -0.002 | 0.055 | 6 | 0.006 | 0.071 | 5.8 | 0.005 | 0.218 | 6.1 | 70.4 | 7.4 |
| 400 | -0.8 | -0.003 | 0.112 | 0.000 | 0.030 | 5.5 | 0.001 | 0.029 | 6 | 0.001 | 0.035 | 5.7 | -0.004 | 0.108 | 4.9 | 100 | 4.9 |
| 400 | -0.6 | -0.008 | 0.100 | 0.001 | 0.030 | 5 | 0.001 | 0.028 | 4.9 | 0.001 | 0.035 | 5 | -0.005 | 0.109 | 4.8 | 100 | 5.2 |
| 400 | -0.4 | -0.001 | 0.094 | 0.000 | 0.029 | 4.8 | 0.000 | 0.028 | 5.1 | 0.000 | 0.037 | 6.2 | -0.002 | 0.108 | 5.4 | 100 | 6.1 |
| 400 | -0.2 | -0.003 | 0.087 | 0.000 | 0.029 | 4.1 | 0.000 | 0.029 | 4.7 | 0.001 | 0.036 | 4.8 | -0.003 | 0.113 | 5.4 | 100 | 5.4 |
| 400 | 0 | -0.010 | 0.080 | 0.000 | 0.030 | 5 | -0.001 | 0.029 | 5.1 | -0.001 | 0.036 | 5.4 | 0.002 | 0.111 | 5.3 | 100 | 5.3 |
| 400 | 0.2 | -0.009 | 0.070 | 0.000 | 0.030 | 5.1 | 0.000 | 0.029 | 4.9 | -0.001 | 0.036 | 5.4 | 0.001 | 0.110 | 5.5 | 100 | 5.2 |
| 400 | 0.4 | -0.001 | 0.059 | 0.000 | 0.030 | 6.2 | 0.000 | 0.029 | 5.2 | -0.001 | 0.035 | 4.9 | -0.001 | 0.110 | 5.2 | 100 | 5 |
| 400 | 0.6 | -0.007 | 0.043 | 0.000 | 0.029 | 5.5 | 0.001 | 0.028 | 5.2 | 0.001 | 0.036 | 5.5 | -0.003 | 0.109 | 5.8 | 99.9 | 5.5 |
| 400 | 0.8 | -0.005 | 0.028 | 0.000 | 0.029 | 5 | -0.001 | 0.028 | 5.2 | -0.001 | 0.035 | 5.5 | 0.001 | 0.107 | 5.8 | 99.8 | 6.1 |

Table 3: Descriptive statistics

| Variable | All firms Mean | All firms Std.dev. | State-owned firms Mean | State-owned firms Std.dev. | Non state-owned firms Mean | Non state-owned firms Std.dev. |
|---|---|---|---|---|---|---|
| Total sales (in logs) | 10.341 | 1.275 | 10.889 | 1.931 | 10.311 | 1.221 |
| Labor (in logs) | 4.525 | 1.094 | 5.625 | 1.482 | 4.464 | 1.034 |
| Skilled labor ratio (fraction) | 0.175 | 0.188 | 0.226 | 0.184 | 0.172 | 0.187 |
| Capital (in logs) | 8.823 | 1.686 | 10.315 | 2.124 | 8.740 | 1.619 |
| Material (in logs) | 9.828 | 1.349 | 10.203 | 2.089 | 9.807 | 1.293 |
| State-owned (binary indicator) | 0.053 | 0.224 | 1 | 0 | 0 | 0 |
| Foreign-owned-to-total-capital ratio (fraction) | 0.140 | 0.320 | 0.064 | 0.186 | 0.144 | 0.326 |
| Exporter (binary indicator) | 0.236 | 0.425 | 0.292 | 0.455 | 0.233 | 0.423 |
| Intangible asset intensity (binary indicator) | 0.431 | 0.495 | 0.590 | 0.492 | 0.422 | 0.494 |
| Number of firms | 12,552 | | 662 | | 11,890 | |
| Number of observations | 37,656 | | 1,986 | | 35,670 | |

Table 4: Results Set a (4 contextual effects) – heteroskedasticity robust estimators

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| Dependent variable: log sales | | | | |
| *Primary production factors* | | | | |
| Capital | $k_{it}$ | 0.087** | 0.170*** | 0.087** |
| | | (0.032) | (0.022) | (0.032) |
| Labor | $l_{it}$ | 0.632*** | 0.430*** | 0.632*** |
| | | (0.053) | (0.030) | (0.053) |
| Skilled labor ratio | $h_i$ | | 0.766*** | 2.733*** |
| | | | (0.124) | (0.459) |
| Material | $m_{it}$ | −0.164** | 0.182*** | −0.164** |
| | | (0.064) | (0.036) | (0.064) |
| Capital × capital | $k_{it}k_{it}$ | 0.004* | 0.012*** | 0.004* |
| | | (0.002) | (0.002) | (0.002) |
| Capital × skilled labor ratio | $k_{it}h_i$ | −0.025 | 0.061*** | −0.025 |
| | | (0.029) | (0.014) | (0.029) |
| Capital × material | $k_{it}m_{it}$ | −0.022*** | −0.040*** | −0.022*** |
| | | (0.005) | (0.003) | (0.005) |
| Labor × capital | $l_{it}k_{it}$ | 0.021*** | 0.012*** | 0.021*** |
| | | (0.005) | (0.004) | (0.005) |
| Labor × material | $l_{it}m_{it}$ | −0.094*** | −0.074*** | −0.094*** |
| | | (0.007) | (0.005) | (0.007) |
| Labor × labor | $l_{it}l_{it}$ | 0.022*** | 0.032*** | 0.022*** |
| | | (0.005) | (0.004) | (0.005) |
| Labor × skilled labor ratio | $l_{it}h_i$ | 0.173*** | 0.217*** | 0.173*** |
| | | (0.048) | (0.022) | (0.048) |
| Material × material | $m_{it}m_{it}$ | 0.077*** | 0.068*** | 0.077*** |
| | | (0.004) | (0.003) | (0.004) |
| Material × skilled labor ratio | $m_{it}h_i$ | −0.244*** | −0.205*** | −0.244*** |
| | | (0.041) | (0.016) | (0.041) |
| Skilled labor ratio × skilled labor ratio | $h_ih_i$ | | 0.062 | −0.618*** |
| | | | (0.062) | (0.156) |
| *Own specified TFP shifters* | | | | |
| Foreign–owned capital ratio | $f_{it}$ | 0.043* | 0.066*** | 0.044* |
| | | (0.021) | (0.011) | (0.021) |
| Exporter | $e_{it}$ | 0.027*** | 0.016* | 0.027*** |
| | | (0.008) | (0.007) | (0.008) |
| Intangible asset intensity | $i_{it}$ | 0.009 | −0.000 | 0.009 |
| | | (0.006) | (0.006) | (0.006) |
| Publicly owned/state–owned | $p_i$ | | −0.088*** | −1.819*** |
| | | | (0.014) | (0.506) |
| *Specified TFP spillovers from neighbors' TFP shifters* | | | | |
| Neighbors' Foreign–owned capital ratio | $\overline{f}_{it}$ | 0.027 | 0.008 | −0.106 |
| | | (0.175) | (0.046) | (0.082) |
| Neighboring exporters | $\overline{e}_{it}$ | −0.129 | −0.002 | −0.141* |
| | | (0.072) | (0.042) | (0.070) |
| Neighbors' intangible asset intensity | $\overline{i}_{it}$ | −0.227*** | −0.173*** | −0.229*** |
| | | (0.060) | (0.033) | (0.060) |
| *Specified TFP spillovers from neighbors' primary production factors* | | | | |
| Neighbors' skilled labor ratio | $\overline{h}_i$ | | 0.216** | 0.735*** |

*Continued on next page*

Table 4 continued

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| | | | (0.077) | (0.204) |
| *Spatial autocorrelation parameter* | $\rho$ | 0.512*** | 0.433*** | 0.512*** |
| | | (0.017) | (0.016) | (0.017) |
| *Variance components* | | | | |
| Dependent variable | $\sigma_y^2$ | 1.626 | 1.626 | 1.626 |
| Primary production factors | $\sigma_P^2$ | 1.188 | 1.517 | 1.172 |
| Own specified TFP shifters | $\sigma_T^2$ | 0.001 | 0.001 | 0.166 |
| Specified TFP spillovers from neighbors' TFP shifters | $\sigma_{\overline{T}}^2$ | 0.001 | 0.000 | 0.002 |
| Specified TFP spillovers from neighbors' primary production factors | $\sigma_{\overline{P}}^2$ | | 0.000 | 0.002 |
| Between component of unspecified TFP | $\sigma_\mu^2$ | | 0.084 | 0.134 |
| Within component of unspecified TFP (average) | $\sigma_\nu^2$ | 0.034 | 0.036 | 0.034 |
| *Hausman test/Hausman and Taylor test* | | | | |
| Test statistic | | | 312.977 | 0.256 |
| Degrees of freedom | | | 18 | 1 |
| p value | | | 0.000 | 0.613 |
| *First stage* | | | | |
| Wald test statistic | | | | 212.882 |
| Degrees of freedom | | | | 2 |
| p value | | | | 0.000 |

Notes: *, **, and *** refer to significant parameters at 5%, 1% and 0.1%, respectively. SFE, SRE, and SHT refer to spatial fixed effects, spatial random effects, and spatial Hausman and Taylor, respectively. The element sets are: $\mathbf{X}_U = [\mathbf{f}, \bar{\mathbf{f}}]$; $\mathbf{X}_C = [\mathbf{l}, \mathbf{l} \circ \mathbf{l}, \mathbf{l} \circ \mathbf{k}, \mathbf{l} \circ \mathbf{m}, \mathbf{l} \circ \mathbf{h}, \mathbf{k}, \mathbf{k} \circ \mathbf{k}, \mathbf{k} \circ \mathbf{h}, \mathbf{k} \circ \mathbf{m}, \mathbf{m}, \mathbf{m} \circ \mathbf{m}, \mathbf{m} \circ \mathbf{h}, \mathbf{e}, \bar{\mathbf{e}}, \mathbf{i}, \bar{\mathbf{i}}]$; $\mathbf{Z}_U = [\mathbf{h}, \mathbf{h} \circ \mathbf{h}, \overline{\mathbf{h}}]$; and $\mathbf{Z}_C = \mathbf{p}$. There are 12,552 firms and 37,656 observations.

Table 5: Results Set b (15 contextual effects)– heteroskedasticity robust estimators

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| Dependent variable: log sales | | | | |
| *Primary production factors* | | | | |
| Capital | $k_{it}$ | 0.085** | 0.168*** | 0.085** |
| | | (0.032) | (0.023) | (0.032) |
| Labor | $l_{it}$ | 0.617*** | 0.426*** | 0.618*** |
| | | (0.053) | (0.031) | (0.053) |
| Skilled labor ratio | $h_i$ | | 0.748*** | 2.616*** |
| | | | (0.128) | (0.466) |
| Material | $m_{it}$ | −0.151* | 0.183*** | −0.152* |
| | | (0.064) | (0.036) | (0.064) |
| Capital × capital | $k_{it}k_{it}$ | 0.003 | 0.012*** | 0.003 |
| | | (0.002) | (0.002) | (0.002) |
| Capital × skilled labor ratio | $k_{it}h_i$ | −0.019 | 0.060*** | −0.018 |
| | | (0.029) | (0.015) | (0.029) |
| Capital × material | $k_{it}m_{it}$ | −0.020*** | −0.040*** | −0.020*** |
| | | (0.005) | (0.003) | (0.005) |
| Labor × capital | $l_{it}k_{it}$ | 0.022*** | 0.013*** | 0.022*** |
| | | (0.004) | (0.004) | (0.004) |
| Labor × material | $l_{it}m_{it}$ | −0.093*** | −0.074*** | −0.093*** |
| | | (0.007) | (0.005) | (0.007) |
| Labor × labor | $l_{it}l_{it}$ | 0.021*** | 0.032*** | 0.021*** |
| | | (0.005) | (0.004) | (0.005) |
| Labor × skilled labor ratio | $l_{it}h_i$ | 0.165*** | 0.215*** | 0.166*** |
| | | (0.049) | (0.022) | (0.049) |
| Material × material | $m_{it}m_{it}$ | 0.075*** | 0.068*** | 0.075*** |
| | | (0.004) | (0.003) | (0.004) |
| Material × skilled labor ratio | $m_{it}h_i$ | −0.229*** | −0.202*** | −0.229*** |
| | | (0.041) | (0.016) | (0.041) |
| Skilled labor ratio × skilled labor ratio | $h_ih_i$ | | 0.061 | −0.657*** |
| | | | (0.063) | (0.198) |
| *Own specified TFP shifters* | | | | |
| Foreign–owned capital ratio | $f_{it}$ | 0.042* | 0.066*** | 0.048* |
| | | (0.021) | (0.011) | (0.021) |
| Exporter | $e_{it}$ | 0.034*** | 0.016* | 0.034*** |
| | | (0.007) | (0.007) | (0.007) |
| Intangible asset intensity | $i_{it}$ | 0.013* | 0.001 | 0.013* |

*Continued on next page*

Table 5 continued

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| | | (0.006) | (0.006) | (0.006) |
| Publicly owned/state–owned (binary indicator) | $p_i$ | | −0.088*** | −1.861** |
| | | | (0.015) | (0.593) |
| *Specified TFP spillovers from neighbors' TFP shifters* | | | | |
| Neighbors' Foreign–owned capital ratio | $\overline{f}_{it}$ | −0.048 | −0.055 | −0.321** |
| | | (0.151) | (0.049) | (0.107) |
| Neighboring exporters | $\overline{e}_{it}$ | 0.123* | −0.028 | 0.099 |
| | | (0.061) | (0.045) | (0.060) |
| Neighbors' intangible asset intensity | $\overline{i}_{it}$ | −0.132** | −0.163*** | −0.136** |
| | | (0.050) | (0.038) | (0.050) |
| *Specified TFP spillovers from neighbors' primary production factors* | | | | |
| Neighbors' labor | $\overline{l}_{it}$ | −0.890** | 0.028 | −0.934*** |
| | | (0.273) | (0.173) | (0.270) |
| Neighbors' capital | $\overline{k}_{it}$ | 0.643*** | 0.082 | 0.669*** |
| | | (0.172) | (0.139) | (0.172) |
| Neighbors' skilled labor ratio | $\overline{h}_i$ | | −0.877 | 0.751 |
| | | | (0.883) | (1.522) |
| Neighbors' material | $\overline{m}_{it}$ | 0.106 | 0.034 | 0.099 |
| | | (0.179) | (0.157) | (0.179) |
| Neighbors' (capital × capital) | $\overline{k_{it}k_{it}}$ | −0.014 | 0.030** | −0.015 |
| | | (0.011) | (0.010) | (0.011) |
| Neighbors' (capital × skilled labor ratio) | $\overline{k_{it}h_i}$ | −0.409* | 0.091 | −0.415* |
| | | (0.187) | (0.099) | (0.188) |
| Neighbors' (capital × material) | $\overline{k_{it}m_{it}}$ | −0.065*** | −0.057*** | −0.068*** |
| | | (0.017) | (0.016) | (0.017) |
| Neighbors' (labor × capital) | $\overline{l_{it}k_{it}}$ | 0.079* | −0.013 | 0.083* |
| | | (0.033) | (0.022) | (0.033) |
| Neighbors' (labor × material) | $\overline{l_{it}m_{it}}$ | 0.071** | 0.030 | 0.072** |
| | | (0.025) | (0.023) | (0.025) |
| Neighbors' (labor × labor) | $\overline{l_{it}l_{it}}$ | −0.073* | −0.016 | −0.073* |
| | | (0.034) | (0.024) | (0.034) |
| Neighbors' (labor × skilled labor ratio) | $\overline{l_{it}h_i}$ | 0.449 | −0.402** | 0.472 |
| | | (0.242) | (0.154) | (0.242) |
| Neighbors' (material × material) | $\overline{m_{it}m_{it}}$ | 0.019 | 0.018 | 0.020* |
| | | (0.010) | (0.011) | (0.010) |
| Neighbors' (material × skilled labor ratio) | $\overline{m_{it}h_i}$ | 0.010 | 0.194 | 0.001 |
| | | (0.137) | (0.101) | (0.137) |
| Neighbors' (skilled labor ratio × skilled labor ratio) | $\overline{h_ih_i}$ | | 0.145 | 2.436** |
| | | | (0.432) | (0.862) |
| *Spatial autocorrelation parameter* | $\rho$ | 0.302*** | 0.358*** | 0.302*** |
| | | (0.014) | (0.029) | (0.014) |
| *Variance components* | | | | |
| Dependent variable | $\sigma^2_y$ | 1.626 | 1.626 | 1.626 |
| Primary production factors | $\sigma^2_P$ | 1.119 | 1.513 | 1.117 |
| Own specified TFP shifters | $\sigma^2_T$ | 0.001 | 0.001 | 0.174 |
| Specified TFP spillovers from neighbors' TFP shifters | $\sigma^2_{\overline{T}}$ | 0.026 | 0.001 | 0.019 |
| Specified TFP spillovers from neighbors' primary production factors | $\sigma^2_{\overline{P}}$ | 0.000 | 0.001 | 0.002 |
| Between component of unspecified TFP | $\sigma^2_\mu$ | | 0.092 | 0.414 |
| Within component of unspecified TFP (average) | $\sigma^2_\nu$ | 0.033 | 0.036 | 0.033 |
| *Hausman test/Hausman and Taylor test* | | | | |
| Test statistic | | | 517.797 | 0.014 |
| Degrees of freedom | | | 30 | 1 |
| p value | | | 0.000 | 0.905 |
| *First stage* | | | | |
| Wald test statistic | | | | 261.636 |
| Degrees of freedom | | | | 2 |
| p value | | | | 0.000 |

Notes: *, **, and *** refer to significant parameters at 5%, 1% and 0.1%, respectively. SFE, SRE, and SHT refer to spatial fixed effects, spatial random effects, and spatial Hausman and Taylor, respectively. The element sets are: $\mathbf{X}_U = [\mathbf{f}, \overline{\mathbf{f}}]$; $\mathbf{X}_C = [\mathbf{l}, \overline{\mathbf{l}}, \mathbf{l} \circ \mathbf{l}, \overline{\mathbf{l} \circ \mathbf{l}}, \mathbf{l} \circ \mathbf{k}, \overline{\mathbf{l} \circ \mathbf{k}}, \mathbf{l} \circ \mathbf{m}, \overline{\mathbf{l} \circ \mathbf{m}}, \mathbf{l} \circ \mathbf{h}, \overline{\mathbf{l} \circ \mathbf{h}}, \mathbf{k}, \overline{\mathbf{k}}, \mathbf{k} \circ \mathbf{k}, \overline{\mathbf{k} \circ \mathbf{k}}, \mathbf{k} \circ \mathbf{h}, \overline{\mathbf{k} \circ \mathbf{h}}, \mathbf{k} \circ \mathbf{m}, \overline{\mathbf{k} \circ \mathbf{m}}, \mathbf{m}, \overline{\mathbf{m}}, \mathbf{m} \circ \mathbf{m}, \overline{\mathbf{m} \circ \mathbf{m}}, \mathbf{m} \circ \mathbf{h}, \overline{\mathbf{m} \circ \mathbf{h}}, \mathbf{e}, \overline{\mathbf{e}}, \mathbf{i}, \overline{\mathbf{i}}]$; $\mathbf{Z}_U = [\mathbf{h}, \overline{\mathbf{h}}, \mathbf{h} \circ \mathbf{h}, \overline{\mathbf{h} \circ \mathbf{h}}]$; and $\mathbf{Z}_C = \mathbf{p}$. There are 12,552 firms and 37,656 observations.

Table 6: Results Set a (4 contextual effects) with multiple $\rho$ – heteroskedasticity robust estimators

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| Dependent variable: log sales | | | | |
| *Primary production factors* | | | | |
| Capital | $k_{it}$ | 0.083* | 0.170*** | 0.083* |
| | | (0.033) | (0.022) | (0.033) |
| Labor | $l_{it}$ | 0.629*** | 0.430*** | 0.630*** |
| | | (0.053) | (0.030) | (0.053) |
| Skilled labor ratio | $h_i$ | | 0.766*** | 2.657*** |
| | | | (0.124) | (0.455) |
| Material | $m_{it}$ | −0.174** | 0.179*** | −0.175** |
| | | (0.064) | (0.036) | (0.064) |
| Capital × capital | $k_{it}k_{it}$ | 0.004* | 0.013*** | 0.004* |
| | | (0.002) | (0.002) | (0.002) |
| Capital × skilled labor ratio | $k_{it}h_i$ | −0.025 | 0.059*** | −0.023 |
| | | (0.029) | (0.014) | (0.029) |
| Capital × material | $k_{it}m_{it}$ | −0.021*** | −0.041*** | −0.021*** |
| | | (0.005) | (0.003) | (0.005) |
| Labor × capital | $l_{it}k_{it}$ | 0.020*** | 0.012*** | 0.020*** |
| | | (0.004) | (0.004) | (0.005) |
| Labor × material | $l_{it}m_{it}$ | −0.093*** | −0.074*** | −0.093*** |
| | | (0.007) | (0.005) | (0.007) |
| Labor × labor | $l_{it}l_{it}$ | 0.022*** | 0.032*** | 0.022*** |
| | | (0.005) | (0.004) | (0.005) |
| Labor × skilled labor ratio | $l_{it}h_i$ | 0.174*** | 0.219*** | 0.174*** |
| | | (0.048) | (0.022) | (0.048) |
| Material × material | $m_{it}m_{it}$ | 0.077*** | 0.068*** | 0.077*** |
| | | (0.004) | (0.003) | (0.004) |
| Material × skilled labor ratio | $m_{it}h_i$ | −0.243*** | −0.205*** | −0.242*** |
| | | (0.041) | (0.016) | (0.041) |
| Skilled labor ratio × skilled labor ratio | $h_ih_i$ | | 0.056 | −0.557*** |
| | | | (0.061) | (0.145) |
| *Own specified TFP shifters* | | | | |
| Foreign–owned capital ratio | $f_{it}$ | 0.044* | 0.066*** | 0.056** |
| | | (0.021) | (0.011) | (0.020) |
| Exporter | $e_{it}$ | 0.026*** | 0.015* | 0.027*** |
| | | (0.008) | (0.007) | (0.008) |
| Intangible asset intensity | $i_{it}$ | 0.010 | −0.000 | 0.010 |
| | | (0.006) | (0.006) | (0.006) |
| Publicly owned/state–owned | $p_i$ | | −0.089*** | −1.835*** |
| | | | (0.015) | (0.541) |
| *Specified TFP spillovers from neighbors' TFP shifters* | | | | |
| Neighbors' Foreign–owned capital ratio | $\overline{f}_{it}$ | 0.080 | 0.023 | −0.219 |
| | | (0.169) | (0.043) | (0.115) |
| Neighboring exporters | $\overline{e}_{it}$ | −0.152* | −0.001 | −0.178* |
| | | (0.071) | (0.040) | (0.070) |
| Neighbors' intangible asset intensity | $\overline{i}_{it}$ | −0.205*** | −0.173*** | −0.208*** |
| | | (0.059) | (0.031) | (0.059) |
| *Specified TFP spillovers from neighbors' primary production factors* | | | | |
| Spatially lagged skilled labor ratio | $\overline{h}_i$ | | 0.180* | 1.245** |
| | | | (0.071) | (0.380) |
| *Spatial autocorrelation parameters for relationships* | | | | |
| among public/state–owned firms | $\rho_p$ | 0.000 | 0.000 | 0.000 |
| | | (0.102) | (0.096) | (0.102) |
| among private firms | $\rho_r$ | 0.546*** | 0.388*** | 0.546*** |
| | | (0.018) | (0.018) | (0.018) |
| between publicly/state–owned and private firms | $\rho_o$ | 0.000 | 0.000 | 0.000 |
| | | (0.038) | (0.036) | (0.031) |
| *Variance components* | | | | |
| Dependent variable | $\sigma^2_y$ | 1.626 | 1.626 | 1.626 |
| Primary production factors | $\sigma^2_P$ | 1.182 | 1.518 | 1.169 |
| Own specified TFP shifters | $\sigma^2_T$ | 0.001 | 0.001 | 0.169 |
| Specified TFP spillovers from neighbors' TFP shifters | $\sigma^2_{\overline{T}}$ | | 0.000 | 0.007 |
| Specified TFP spillovers from neighbors' primary production factors | $\sigma^2_{\overline{P}}$ | 0.001 | 0.001 | 0.003 |
| Between component of unspecified TFP | $\sigma^2_\mu$ | | 0.085 | 0.134 |

*Continued on next page*

Table 6 continued

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| Within component of unspecified TFP (average) | $\sigma_\nu^2$ | 0.034 | 0.036 | 0.034 |
| *Hausman test/Hausman and Taylor test* | | | | |
| Test statistic | | | 323.269 | 2.679 |
| Degrees of freedom | | | 18 | 1 |
| p value | | | 0 | 0.102 |
| *First stage* | | | | |
| Wald test statistic | | | | 296.093 |
| Degrees of freedom | | | | 2 |
| p value | | | | 0 |

Notes: *, **, and *** refer to significant parameters at 5%, 1% and 0.1%, respectively. SFE, SRE, and SHT refer to spatial fixed effects, spatial random effects, and spatial Hausman and Taylor, respectively. The element sets are: $\mathbf{X}_U = [\mathbf{f}, \bar{\mathbf{f}}]$; $\mathbf{X}_C = [\mathbf{l}, \mathbf{l} \circ \mathbf{l}, \mathbf{l} \circ \mathbf{k}, \mathbf{l} \circ \mathbf{m}, \mathbf{l} \circ \mathbf{h}, \mathbf{k}, \mathbf{k} \circ \mathbf{k}, \mathbf{k} \circ \mathbf{h}, \mathbf{k} \circ \mathbf{m}, \mathbf{m}, \mathbf{m} \circ \mathbf{m}, \mathbf{m} \circ \mathbf{h}, \mathbf{e}, \bar{\mathbf{e}}, \mathbf{i}, \bar{\mathbf{i}}]$; $\mathbf{Z}_U = [\mathbf{h}, \mathbf{h} \circ \mathbf{h}, \bar{\mathbf{h}}]$; and $\mathbf{Z}_C = \mathbf{p}$. There are 12,552 firms and 37,656 observations.

Table 7: Results Set b (15 contextual effects) with multiple $\rho$ – heteroskedasticity robust estimators

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| Dependent variable: log sales | | | | |
| *Primary production factors* | | | | |
| Capital | $k_{it}$ | 0.085** | 0.167*** | 0.085** |
| | | (0.032) | (0.023) | (0.032) |
| Labor | $l_{it}$ | 0.620*** | 0.427*** | 0.621*** |
| | | (0.053) | (0.031) | (0.053) |
| Skilled labor ratio | $h_i$ | | 0.750*** | 2.572*** |
| | | | (0.128) | (0.462) |
| Material | $m_{it}$ | −0.154* | 0.179*** | −0.156* |
| | | (0.064) | (0.036) | (0.064) |
| Capital × capital | $k_{it}k_{it}$ | 0.003 | 0.013*** | 0.003 |
| | | (0.002) | (0.002) | (0.002) |
| Capital × skilled labor ratio | $k_{it}h_i$ | −0.020 | 0.059*** | −0.018 |
| | | (0.029) | (0.015) | (0.029) |
| Capital × material | $k_{it}m_{it}$ | −0.020*** | −0.040*** | −0.020*** |
| | | (0.005) | (0.003) | (0.005) |
| Labor × capital | $l_{it}k_{it}$ | 0.022*** | 0.013*** | 0.022*** |
| | | (0.004) | (0.004) | (0.004) |
| Labor × material | $l_{it}m_{it}$ | −0.093*** | −0.074*** | −0.093*** |
| | | (0.007) | (0.005) | (0.007) |
| Labor × labor | $l_{it}l_{it}$ | 0.021*** | 0.032*** | 0.022*** |
| | | (0.005) | (0.004) | (0.005) |
| Labor × skilled labor ratio | $l_{it}h_i$ | 0.166*** | 0.218*** | 0.167*** |
| | | (0.049) | (0.022) | (0.049) |
| Material × material | $m_{it}m_{it}$ | 0.075*** | 0.068*** | 0.075*** |
| | | (0.004) | (0.003) | (0.004) |
| Material × skilled labor ratio | $m_{it}h_i$ | −0.230*** | −0.202*** | −0.230*** |
| | | (0.041) | (0.016) | (0.041) |
| Skilled labor ratio × skilled labor ratio | $h_ih_i$ | | 0.055 | −0.602** |
| | | | (0.064) | (0.187) |
| *Own specified TFP shifters* | | | | |
| Foreign–owned capital ratio | $f_{it}$ | 0.042* | 0.066*** | 0.058** |
| | | (0.021) | (0.011) | (0.020) |
| Exporter | $e_{it}$ | 0.033*** | 0.016* | 0.034*** |
| | | (0.007) | (0.007) | (0.007) |
| Intangible asset intensity (binary indicator) | $i_{it}$ | 0.013* | 0.001 | 0.013* |
| | | (0.006) | (0.006) | (0.006) |
| Publicly owned/state–owned (binary indicator) | $p_i$ | | −0.089*** | −1.725** |
| | | | (0.015) | (0.577) |
| *Specified TFP spillovers from neighbors' TFP shifters* | | | | |
| Neighbors' Foreign–owned capital ratio | $\bar{f}_{it}$ | −0.026 | −0.047 | −0.386** |
| | | (0.149) | (0.049) | (0.125) |
| Neighboring exporters | $\bar{e}_{it}$ | 0.110 | −0.025 | 0.078 |
| | | (0.061) | (0.044) | (0.061) |
| Neighbors' intangible asset intensity | $\bar{i}_{it}$ | −0.125* | −0.163*** | −0.130** |
| | | (0.050) | (0.037) | (0.050) |

*Continued on next page*

Table 7 continued

| | Acronym | SFE | SRE | SHT |
|---|---|---|---|---|
| *Specified TFP spillovers from neighbors' primary production factors* | | | | |
| Neighbors' capital | $\overline{k}_{it}$ | 0.640*** | 0.047 | 0.674*** |
| | | (0.172) | (0.138) | (0.172) |
| Neighbors' labor | $\overline{l}_{it}$ | −0.843** | 0.077 | −0.902*** |
| | | (0.271) | (0.171) | (0.269) |
| Neighbors' skilled labor ratio | $\overline{h}_{i}$ | | −0.883 | 1.560 |
| | | | (0.861) | (1.544) |
| Neighbors' material | $\overline{m}_{it}$ | 0.085 | 0.090 | 0.078 |
| | | (0.178) | (0.155) | (0.178) |
| Neighbors' (capital × capital) | $\overline{k_{it}k_{it}}$ | −0.013 | 0.035*** | −0.015 |
| | | (0.011) | (0.010) | (0.011) |
| Neighbors' (capital × skilled labor ratio) | $\overline{k_{it}h_{i}}$ | −0.393* | 0.060 | −0.400* |
| | | (0.183) | (0.097) | (0.183) |
| Neighbors' (capital × material) | $\overline{k_{it}m_{it}}$ | −0.063*** | −0.058*** | −0.066*** |
| | | (0.017) | (0.016) | (0.017) |
| Neighbors' (labor × capital) | $\overline{l_{it}k_{it}}$ | 0.072* | −0.022 | 0.077* |
| | | (0.033) | (0.022) | (0.032) |
| Neighbors' (labor × material) | $\overline{l_{it}m_{it}}$ | 0.071** | 0.030 | 0.073** |
| | | (0.025) | (0.023) | (0.025) |
| Neighbors' (labor × labor) | $\overline{l_{it}l_{it}}$ | −0.072* | −0.014 | −0.072* |
| | | (0.033) | (0.024) | (0.033) |
| Neighbors' (labor × skilled labor ratio) | $\overline{l_{it}h_{i}}$ | 0.446 | −0.335* | 0.475* |
| | | (0.236) | (0.150) | (0.236) |
| Neighbors' (material × material) | $\overline{m_{it}m_{it}}$ | 0.019 | 0.015 | 0.020* |
| | | (0.010) | (0.010) | (0.010) |
| Neighbors' (material × skilled labor ratio) | $\overline{m_{it}h_{i}}$ | −0.019 | 0.192 | −0.029 |
| | | (0.134) | (0.099) | (0.134) |
| Neighbors' (skilled labor ratio × skilled labor ratio) | $\overline{h_{i}h_{i}}$ | | 0.111 | 1.788* |
| | | | (0.423) | (0.867) |
| *Spatial autocorrelation parameters for relationships* | | | | |
| among public/state–owned firms | $\rho_p$ | 0.000 | 0.000 | 0.000 |
| | | (0.110) | (0.111) | (0.110) |
| among private firms | $\rho_r$ | 0.323*** | 0.376*** | 0.323*** |
| | | (0.016) | (0.032) | (0.016) |
| between publicly/state–owned and private firms | $\rho_o$ | 0.000 | 0.000 | 0.000 |
| | | (0.053) | (0.049) | (0.052) |
| *Variance components* | | | | |
| Dependent variable | $\sigma_y^2$ | 1.626 | 1.626 | 1.626 |
| Primary production factors | $\sigma_P^2$ | 1.118 | 1.512 | 1.116 |
| Own specified TFP shifters | $\sigma_T^2$ | 0.001 | 0.001 | 0.150 |
| Specified TFP spillovers from neighbors' TFP shifters | $\sigma_{\overline{T}}^2$ | 0.027 | 0.001 | 0.021 |
| Specified TFP spillovers from neighbors' primary production factors | $\sigma_{\overline{P}}^2$ | 0.000 | 0.001 | 0.002 |
| Between component of unspecified TFP | $\sigma_\mu^2$ | | 0.092 | 0.415 |
| Within component of unspecified TFP (average) | $\sigma_\nu^2$ | 0.033 | 0.036 | 0.033 |
| *Hausman test/Hausman and Taylor test* | | | | |
| Test statistic | | | 510.630 | 0.404 |
| Degrees of freedom | | | 30 | 1 |
| p value | | | 0.000 | 0.525 |
| *First stage* | | | | |
| Wald test statistic | | | | 383.230 |
| Degrees of freedom | | | | 2 |
| p value | | | | 0.000 |

Notes: *, **, and *** refer to significant parameters at 5%, 1% and 0.1%, respectively. SFE, SRE, and SHT refer to spatial fixed effects, spatial random effects, and spatial Hausman and Taylor, respectively. The element sets are: $\mathbf{X}_U = [\mathbf{f}, \overline{\mathbf{f}}]$; $\mathbf{X}_C = [\mathbf{l}, \overline{\mathbf{l}}, \mathbf{l} \circ \mathbf{l}, \overline{\mathbf{l} \circ \mathbf{l}}, \mathbf{l} \circ \mathbf{k}, \overline{\mathbf{l} \circ \mathbf{k}}, \mathbf{l} \circ \mathbf{m}, \overline{\mathbf{l} \circ \mathbf{m}}, \mathbf{l} \circ \mathbf{h}, \overline{\mathbf{l} \circ \mathbf{h}}, \mathbf{k}, \overline{\mathbf{k}}, \mathbf{k} \circ \mathbf{k}, \overline{\mathbf{k} \circ \mathbf{k}}, \mathbf{k} \circ \mathbf{h}, \overline{\mathbf{k} \circ \mathbf{h}}, \mathbf{k} \circ \mathbf{m}, \overline{\mathbf{k} \circ \mathbf{m}}, \mathbf{m}, \overline{\mathbf{m}}, \mathbf{m} \circ \mathbf{m}, \overline{\mathbf{m} \circ \mathbf{m}}, \mathbf{m} \circ \mathbf{h}, \overline{\mathbf{m} \circ \mathbf{h}}, \mathbf{e}, \overline{\mathbf{e}}, \mathbf{i}, \overline{\mathbf{i}}]$; $\mathbf{Z}_U = [\mathbf{h}, \overline{\mathbf{h}}, \mathbf{h} \circ \mathbf{h}, \overline{\mathbf{h} \circ \mathbf{h}}]$; and $\mathbf{Z}_C = \mathbf{p}$. There are 12,552 firms and 37,656 observations.

Table 8: Results Robustness Set a (4 contextual effects) – heteroskedasticity robust estimators

| | Acronym | SHT $W_{200}$ | SHT $W_{100}$ | SHT $W_{60}$ |
|---|---|---|---|---|
| Dependent variable: log sales | | | | |
| *Primary production factors* | | | | |
| Capital | $k_{it}$ | 0.098** (0.033) | 0.102** (0.033) | 0.104** (0.033) |
| Labor | $l_{it}$ | 0.638*** (0.053) | 0.637*** (0.053) | 0.636*** (0.052) |
| Skilled labor ratio | $h_i$ | 2.695*** (0.457) | 2.644*** (0.456) | 2.613*** (0.456) |
| Material | $m_{it}$ | −0.176** (0.064) | −0.178** (0.064) | −0.179** (0.064) |
| Capital × capital | $k_{it}k_{it}$ | 0.005** (0.002) | 0.006** (0.002) | 0.006** (0.002) |
| Capital × skilled labor ratio | $k_{it}h_i$ | −0.034 (0.029) | −0.036 (0.030) | −0.038 (0.030) |
| Capital × material | $k_{it}m_{it}$ | −0.023*** (0.005) | −0.024*** (0.005) | −0.024*** (0.005) |
| Labor × capital | $l_{it}k_{it}$ | 0.020*** (0.005) | 0.020*** (0.005) | 0.019*** (0.005) |
| Labor × material | $l_{it}m_{it}$ | −0.093*** (0.007) | −0.093*** (0.007) | −0.093*** (0.007) |
| Labor × labor | $l_{it}l_{it}$ | 0.021*** (0.005) | 0.021*** (0.005) | 0.020*** (0.005) |
| Labor × skilled labor ratio | $l_{it}h_i$ | 0.193*** (0.049) | 0.197*** (0.049) | 0.201*** (0.049) |
| Material × material | $m_{it}m_{it}$ | 0.079*** (0.004) | 0.080*** (0.004) | 0.080*** (0.004) |
| Material × skilled labor ratio | $m_{it}h_i$ | −0.249*** (0.041) | −0.249*** (0.041) | −0.250*** (0.041) |
| Skilled labor ratio × skilled labor ratio | $h_ih_i$ | −0.518*** (0.155) | −0.457** (0.154) | −0.418** (0.153) |
| *Own specified TFP shifters* | | | | |
| Foreign–owned capital ratio | $f_{it}$ | 0.046* (0.020) | 0.048* (0.020) | 0.050* (0.020) |
| Exporter | $e_{it}$ | 0.024** (0.008) | 0.024** (0.008) | 0.023** (0.008) |
| Intangible asset intensity | $i_{it}$ | 0.007 (0.006) | 0.007 (0.006) | 0.006 (0.006) |
| Publicly owned/state–owned | $p_i$ | −1.676*** (0.501) | −1.504** (0.489) | −1.400** (0.480) |
| *Specified TFP spillovers from neighbors' TFP shifters* | | | | |
| Neighbors' Foreign–owned capital ratio | $\bar{f}_{it}$ | −0.095 (0.059) | −0.084 (0.053) | −0.072 (0.048) |
| Neighboring exporters | $\bar{e}_{it}$ | −0.103* (0.045) | −0.076* (0.038) | −0.066 (0.034) |
| Neighbors' intangible asset intensity | $\bar{i}_{it}$ | −0.114** (0.040) | −0.089** (0.032) | −0.087** (0.031) |
| *Specified TFP spillovers from neighbors' primary production factors* | | | | |
| Neighbors' skilled labor ratio | $\bar{h}_i$ | 0.489** (0.152) | 0.406** (0.133) | 0.356** (0.119) |
| *Spatial autocorrelation parameter* | $\rho$ | 0.310*** (0.014) | 0.265*** (0.012) | 0.233*** (0.012) |
| *Variance components* | | | | |
| Dependent variable | $\sigma^2_y$ | 1.626 | 1.626 | 1.626 |
| Primary production factors | $\sigma^2_P$ | 1.224 | 1.237 | 1.245 |
| Own specified TFP shifters | $\sigma^2_T$ | 0.141 | 0.114 | 0.099 |
| Specified TFP spillovers from neighbors' TFP shifters | $\sigma^2_{\overline{T}}$ | 0.002 | 0.001 | 0.001 |
| Specified TFP spillovers from neighbors' primary production factors | $\sigma^2_{\overline{P}}$ | 0.001 | 0.001 | 0.001 |
| Between component of unspecified TFP | $\sigma^2_\mu$ | 0.140 | 0.139 | 0.136 |
| Within component of unspecified TFP (average) | $\sigma^2_\nu$ | 0.034 | 0.035 | 0.035 |
| *Hausman test* | | | | |
| Test statistic | | 267.753 | 257.172 | 251.066 |

*Continued on next page*

Table 8 continued

| | Acronym | SHT $W_{200}$ | SHT $W_{100}$ | SHT $W_{60}$ |
|---|---|---|---|---|
| Degrees of freedom | | 18 | 18 | 18 |
| p value | | 0.000 | 0.000 | 0.000 |
| *Hausman and Taylor test* | | | | |
| Test statistic | | 0.220 | 0.941 | 1.643 |
| Degrees of freedom | | 1 | 1 | 1 |
| p value | | 0.639 | 0.332 | 0.200 |
| *First stage* | | | | |
| Wald test statistic | | 234.779 | 242.544 | 246.990 |
| Degrees of freedom | | 2 | 2 | 2 |
| p value | | 0.000 | 0.000 | 0.000 |

Notes: $^{*}$, $^{**}$, and $^{***}$ refer to significant parameters at 5%, 1% and 0.1%, respectively. SHT refers to spatial Hausman and Taylor. $\mathbf{W}_{200}$, $\mathbf{W}_{100}$, and $\mathbf{W}_{60}$ refer to weights matrices whose elements are zero if distance > 200 miles, distance > 100 miles, and distance > 60 miles, respectively. The element sets are: $\mathbf{X}_U = [\mathbf{f}, \bar{\mathbf{f}}]$; $\mathbf{X}_C = [\mathbf{l}, \mathbf{l} \circ \mathbf{l}, \mathbf{l} \circ \mathbf{k}, \mathbf{l} \circ \mathbf{m}, \mathbf{l} \circ \mathbf{h}, \mathbf{k}, \mathbf{k} \circ \mathbf{k}, \mathbf{k} \circ \mathbf{h}, \mathbf{k} \circ \mathbf{m}, \mathbf{m}, \mathbf{m} \circ \mathbf{m}, \mathbf{m} \circ \mathbf{h}, \mathbf{e}, \bar{\mathbf{e}}, \mathbf{i}, \bar{\mathbf{i}}]$; $\mathbf{Z}_U = [\mathbf{h}, \mathbf{h} \circ \mathbf{h}, \bar{\mathbf{h}}]$; and $\mathbf{Z}_C = \mathbf{p}$. There are 12,552 firms and 37,656 observations.

Figure 1a: Firms per region in China's chemical industry (Map basis: GfK GeoMarketing)

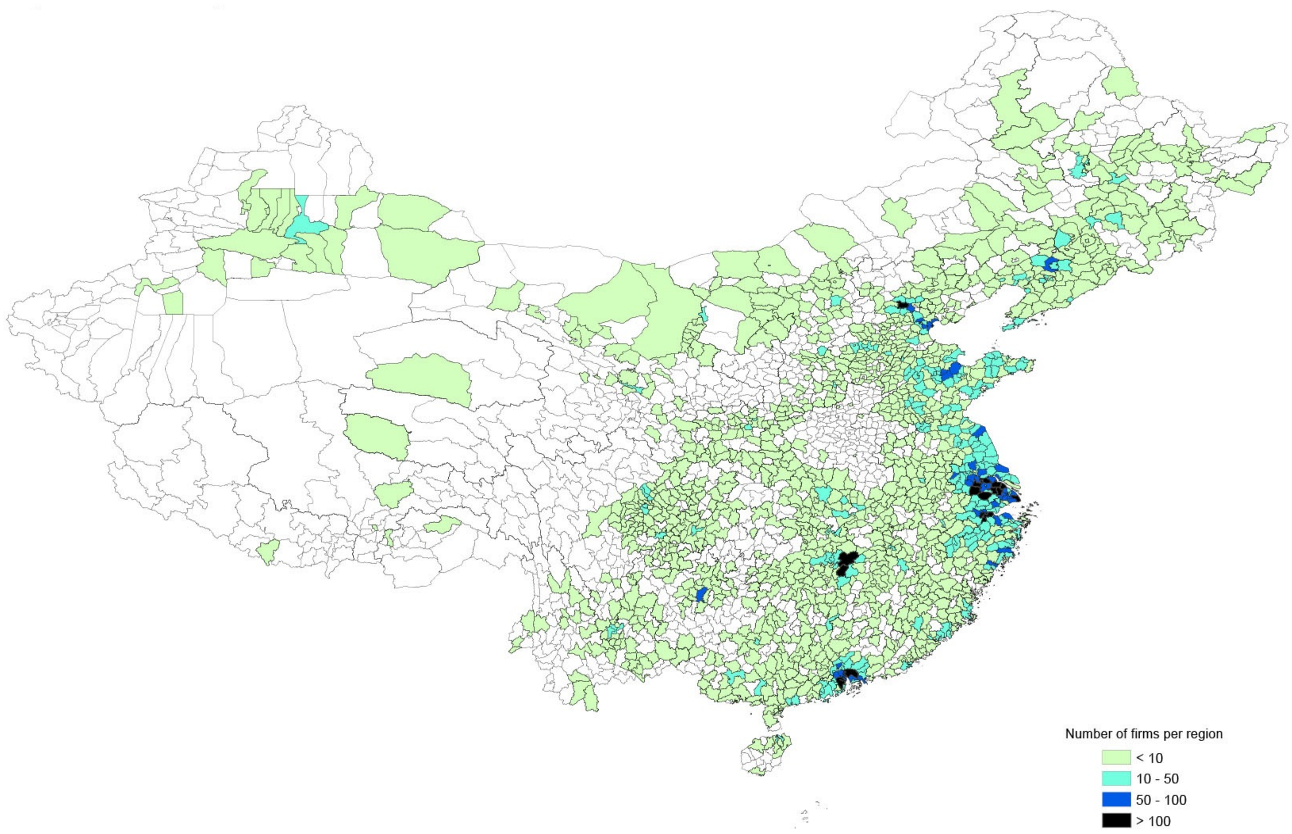


Figure 1b: Total employment per region in China's chemical industry (Map basis: GfK GeoMarketing)