



University of
Salford
MANCHESTER

Semiparametric smooth coefficient estimation of a production system

Kumbhakar, S, Sun, K and Zhang, R

<http://dx.doi.org/10.1111/1468-0106.12193>

Title	Semiparametric smooth coefficient estimation of a production system
Authors	Kumbhakar, S, Sun, K and Zhang, R
Publication title	Pacific Economic Review
Publisher	Wiley
Type	Article
USIR URL	This version is available at: http://usir.salford.ac.uk/id/eprint/39040/
Published Date	2016

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: library-research@salford.ac.uk.

Semiparametric smooth coefficient estimation of a production system*

Subal C. Kumbhakar¹, Kai Sun^{†2}, and Rui Zhang³

¹Department of Economics, State University of New York at Binghamton, NY 13902, USA

¹University of Stavanger Business School, Stavanger, Norway

²Salford Business School, University of Salford, Greater Manchester, M5 4WT, UK

³Department of Financial Engineering, Sichuan University, 610064, P.R.China

Abstract

This paper addresses endogeneity of inputs in estimating a semiparametric smooth coefficient production function using a system approach. The system consists of a translog production function and the first-order conditions (FOC's) of profit maximization. Each coefficient of the production function is an unknown function of some exogenous environmental variables. This makes the production function observation-specific so long as the environmental variables are observation-specific. The estimation of the system involves applying the functional coefficient instrumental variable method (Cai, Das, Xiong and Wu 2006) for the endogeneity of inputs in the first step, and the semiparametric smooth coefficient seemingly unrelated regression method (Henderson, Kumbhakar, Li and Parmeter 2015) in the second step. Using a Chinese food industry data set, we show that the semiparametric system approach gives most economically meaningful input elasticity estimates, compared with alternative models. We also calculate the returns to scale along with the technical and allocative inefficiency estimates.

Keywords: Semiparametric smooth coefficient model; Technical efficiency; Seemingly unrelated regression; Endogeneity

JEL Classification Number: C14, C36, C44, D21, D22, D24, Q18

*This project is funded by the National Natural Science Foundation of China (Grant ID number: 71171137).

[†]Corresponding author. Address: Salford Business School, University of Salford, Greater Manchester, M5 4WT, UK. Email: k.sun@salford.ac.uk. Tel: +44 161 295 3663.

1 Introduction

The semiparametric smooth coefficient (SPSC) model was proposed by Hastie and Tibshirani (1993) and Chen and Tsay (1993). It was further studied/extended by Fan and Zhang (1999), Cai, Fan and Li (2000), Li, Huang, Li and Fu (2002), and Li and Racine (2010). Because of its flexibility and interpretability of regression coefficients, the SPSC model has been widely applied in the literature, e.g., Zhang, Sun, Delgado and Kumbhakar (2012), Heshmati, Kumbhakar and Sun (2014), Bhaumik, Kumbhakar and Sun (2015), among others.

Recent development of the SPSC model includes accommodating the endogeneity of the right-hand-side variables. Cai et al. (2006) proposed an instrumental variable (IV) method to handle the endogeneity of regressors via a two-stage procedure, where the fitted values of the endogenous variables are obtained in the first stage and are then used in the second stage as IV's. Cai and Li (2008) further suggested a one-step estimator via a nonparametric generalized method of moments (GMM) as an alternative. These methods of controlling for endogeneity have also been applied in the literature, e.g., Delgado, McCloud and Kumbhakar (2014), Liu (2014), and Polemis and Stengos (2015), among others.

In this paper we apply the SPSC model in a production system where we explicitly consider the endogeneity of inputs. More specifically, we look at the endogeneity problem from economic point of view in the sense that the quantities of variable inputs used by a firm are chosen by the firm based on some economic behavior such as profit/revenue maximization or cost minimization. Here we consider profit maximization as the criterion for choosing optimal quantities of variable inputs. Thus, instead of using the production function alone (which is the standard practice), we also use the first-order conditions (FOC's) of profit maximization to estimate the production technology. This gives us a system of equations consisting of the production function and the FOC's of profit maximization (Kumbhakar 1987). The system automatically takes into account endogeneity of input and output variables. Since we use a translog production function, the FOC's are expressed in terms of input cost share equations. Each coefficient of the production function and the share equations is a nonparametric function of some exogenous environmental variables.

We estimate this system (which we call primal as opposed to the dual cost and profit systems) in two steps.¹ The first step is to run a nonparametric regression of each input on a number of instruments, including environmental variables as well as input prices. This step follows Cai et al.'s (2006) IV method for treating the endogeneity of regressors in a smooth coefficient framework. The second step is to replace the input quantities with their fitted values obtained in the first step, and then apply Henderson et al.'s (2015) SPSC seemingly unrelated regression (SUR) estimation method to estimate all the smooth coefficients. Finally, technical inefficiencies are calculated using the stochastic frontier analysis (SFA) and allocative inefficiencies are defined in terms of the error terms in the share equations. In other words, allocative inefficiencies are interpreted as deviations of the actual input shares from the optimal shares arising from non-fulfillment of the FOC's of profit maximization (Kumbhakar 1987, Kumbhakar and Lovell 2000).

As an empirical example, we use a firm-level Chinese food industry panel data set that covers 3,922 firms over 9 years. To account for firm heterogeneity, we use the time trend as well as some firm characteristics as exogenous environmental variables. For comparison purposes, in addition to the semiparametric system approach, we also estimate a single equation SPSC model and a traditional parametric SFA model without endogeneity correction. The results show that the semiparametric system approach gives most economically meaningful results in terms of estimated input elasticities, while the other two models generate many negative elasticity estimates that are not economically meaningful in the sense that negative input elasticities mean negative marginal products. We then calculate the technical and allocative inefficiencies, as well as the technical efficiency change (TEC) and technical change (TC), and compare the temporal behaviors of the input elasticities, returns to scale (RTS), and inefficiencies based on our system approach.

The rest of this paper is organized as follows. Section 2 motivates a primal system via a profit maximization problem before proceeding to details about how to estimate a SPSC SUR model with endogenous regressors. Section 3 describes the data set used as an empirical example. Section 4 presents and interprets the estimation results, and section 5 concludes.

¹Such a procedure has been used in the past, e.g., Kumbhakar (1987), but not for the SPSC model. To the best of our knowledge, the SPSC model in a simultaneous equation setup has not been used before.

2 Semiparametric Primal System Approach

We write the production technology as

$$y = f(x; \beta(z)), \quad (1)$$

where y is the output, x is a vector of K traditional inputs such as capital and labor, and z is a vector of S policy variables or external factors (often referred to as facilitating inputs). In specifying the production function in (1) we separate the traditional inputs (x) from the environmental (policy) variables (z). The z variables affect output via the technology parameters β . It is a very flexible specification and allows technology to vary with the z variables even when $f(\cdot)$ is Cobb-Douglas.

We assume that the firm chooses to maximize its profit to decide on the input levels as well as output. Thus a firm's optimization problem can be stated as:

$$\begin{aligned} \max_{y,x} \pi &= py - w'x \\ \text{st. } y &= f(x; \beta(z)), \end{aligned} \quad (2)$$

where π denotes profit, p is the output price, and w is a vector of K input prices. The solution to the above optimization problem is:

$$w_j = pf_j(x; \beta(z)), \quad \forall j = 1, \dots, K, \quad (3)$$

where w_j is the j -th input price, $f_j(x; \beta(z))$ is the marginal product of the j -th input, viz., $f_j(x; \beta(z)) = \partial f(x; \beta(z)) / \partial x_j$. If the above FOC's are satisfied exactly, then there is no optimization error or allocative inefficiency. That is, the presence of allocative inefficiency would mean that the FOC's are not satisfied exactly.

If we multiply both sides of the above FOC's in (3) by $x_j/(py)$, we get:

$$\frac{w_j x_j}{py} = \frac{f_j(x; \beta(z)) x_j}{y}. \quad (4)$$

The right-hand-side of (4) is equivalent to $\partial \ln f(x; \beta(z)) / \partial \ln x_j$, and therefore we can say that $(w_j x_j) / (py)$ (cost share of input x_j in the total revenue) is the same as the j -th input elasticity, $\forall j = 1, \dots, K$.

To empirically estimate the production technology, we take the natural logarithm of (1) and get $\ln y = \ln f(x; \beta(z))$, and then specify $\ln f(x; \beta(z))$ as:

$$\ln f(x, z) = \beta_0(z) + \sum_j \beta_j(z) \ln x_j + \frac{1}{2} \sum_j \sum_k \beta_{jk}(z) \ln x_j \ln x_k. \quad (5)$$

That is, the production function is specified as a translog function of traditional inputs, x , but the associated coefficients are unknown functions of external factors, z . Given this specification, we can find the cost share of the j -th input as $(w_j x_j) / (py) = \partial \ln f(x; \beta(z)) / \partial \ln x_j$, viz.,

$$q_j = \beta_j(z) + \sum_k \beta_{jk}(z) \ln x_k, \quad \forall j = 1, \dots, K, \quad (6)$$

where $q_j = (w_j x_j) / (py)$ is the cost share of the j -th input in the total revenue. If we add ε_j to (6) and let $q_j^* = \beta_j(z) + \sum_k \beta_{jk}(z) \ln x_k$, $\forall j$, be the optimal shares, then

$$q_j = q_j^* + \varepsilon_j, \quad \forall j = 1, \dots, K, \quad (7)$$

where ε_j is interpreted as an input misallocation (allocative inefficiency), i.e., deviation of the actual cost share from optimal cost share for the j -th input. Note that the allocative inefficiency comes from non-fulfillment of the FOC's in (4). We follow Schmidt and Lovell (1979, 1980) and Kumbhakar (1987) for this interpretation.² If $\varepsilon_j > 0$ (< 0), then the actual share is higher (lower) than optimal share by a proportion of ε_j .

To estimate the smooth coefficient functions in (5), we consider the following $K + 1$ system of

²In defining the production function in (1), we have not introduced either noise or inefficiency. If noise is added as in (8a) or both noise and inefficiency is added as in (10), the FOC's based on expected (or median) profit maximization conditional on inefficiency, when expressed as input cost shares, will depend on the noise term but not on inefficiency (follows from the derivation in Kumbhakar (1987)). In such a case ε_j will be affected by noise, i.e., ε_j cannot be interpreted as pure input misallocation.

equations. Adding subscripts i and t to all variables, the system becomes:

$$\ln y_{it} = \beta_0(z_{it}) + \sum_j \beta_j(z_{it}) \ln x_{jit} + \frac{1}{2} \sum_j \sum_k \beta_{jk}(z_{it}) \ln x_{jit} \ln x_{kit} + v_{it} \quad (8a)$$

$$\begin{aligned} q_{1it} &= \beta_1(z_{it}) + \sum_k \beta_{1k}(z_{it}) \ln x_{kit} + \varepsilon_{1it}, \\ &\vdots \\ q_{Kit} &= \beta_K(z_{it}) + \sum_k \beta_{Kk}(z_{it}) \ln x_{kit} + \varepsilon_{Kit}, \end{aligned} \quad (8b)$$

where (8a) is the first equation (i.e., production function) of the system,³ and (8b) consists of the rest of the K share equations of the system. v_{it} is the noise term that reflect the impact of uncontrollable/unforeseen factors (such as weather, machine break down, etc.) on output. We also interpret the error terms $\varepsilon_j, \forall j = 1, \dots, K$, in the share equations as allocative inefficiency for each input.

Recall that in the profit maximization problem, the firm chooses the output and inputs simultaneously. This means both output and inputs are endogenous. This is captured in the system of $(K + 1)$ equations in (8), which consists of (8a) and (8b). However, since x_{jit} and y_{it} are hidden in q_{jit} , the endogenous variables appear everywhere in (8) – not just on the right hand side of (8). Thus, although we have a system of $(K + 1)$ equations, we cannot solve them analytically for the $(K + 1)$ endogenous variables (which will give us the unconditional input demand and output supply equations as in the Cobb-Douglas case of Kumbhakar (1987)). We take a different approach to get around the problem.

Although no analytical solution is possible for the unconditional input demand functions, we can write them implicitly as:

$$\ln x_{jit} = m_j(z_{it}, w_{it}) + \eta_{jit}, \quad \forall j = 1, \dots, K, \quad (9)$$

where η_{jit} is a function of v_{it} and $\varepsilon_{1it}, \dots, \varepsilon_{Kit}$. In the absence of these errors, one gets the textbook

³We replace (8a) with (11) in estimation for the purpose of estimating the technical inefficiency.

type deterministic solutions of input demand functions. Our goal is to use these input demand functions to construct two-stage least squares (2SLS) type instruments for $\ln x_{jit}$. For this, we first run a nonparametric regression for each j using (9), where $E(\eta_{jit}|z_{it}, w_{it}) = 0$, and obtain $\widehat{\ln x_{jit}}$. Following the spirit of the 2SLS procedure for smooth coefficient models (Cai et al. 2006), we replace $\ln x_{jit}$ in the system with $\widehat{\ln x_{jit}} = m_j(\widehat{z_{it}}, w_{it})$, $\forall j = 1, \dots, K$ to avoid the endogeneity problem associated with $\ln x_{jit}$ being correlated with ε_{jit} . The smooth coefficients can then be estimated using Henderson et al.'s (2015) SPSC SUR estimator. In addition to taking into account the cross-equation correlations in the error terms in the system, we also need to impose the cross-equation restrictions on the smooth coefficients, such that $\beta_j(z_{it})$, $\forall j = 1, \dots, K$, in the first equation of the system, are equal to the corresponding intercepts in the share equations; and similarly, $\beta_{jk}(z_{it})$, $\forall j = 1, \dots, K$, in the first equation of the system, are equal to the corresponding slopes in the share equations. Appendix B gives a brief introduction of the SPSC SUR estimator with our empirical example, as well as its first order gradient estimator for estimating marginal effects.

Our analysis so far does not include technical inefficiency. It is not obvious how technical inefficiency should be modeled in the SPSC models. Note that the technology is observation-specific when the environmental (z) variables are observation-specific. This raises the issue of what the reference technology is. Conditional on the inputs, the technology varies with the environmental variables, and one cannot find a single reference technology that is invariant to the z variables. One solution is to think of measuring inefficiency for each firm relative to its own frontier. That is, a firm at a point in time might not be able to attain its potential output given by its own frontier. Under this scenario, one can simply add a one-sided inefficiency term which may or may not be dependent on the z variables. If we do this and the production function in (8a) can be written as:

$$\ln y_{it} = \beta_0(z_{it}) + \sum_j \beta_j(z_{it}) \ln x_{jit} + \frac{1}{2} \sum_j \sum_k \beta_{jk}(z_{it}) \ln x_{jit} \ln x_{kit} + v_{it} - u_{it} \quad (10)$$

where $u_{it} \geq 0$ is the technical inefficiency. To allow the possibility that the environmental variables can affect the inefficiency, we assume that $u_{it} = g(z_{it}) \times u^*$ where $g(\cdot)$ is the scaling function, $u^* \sim iidN^+(0, \sigma_u^2)$ is the basic function and $v_{it} \sim iidN(0, \sigma_v^2)$. Using these specifications, we can

rewrite the above equation as:

$$\ln y_{it} = \theta_0(z_{it}) + \sum_j \beta_j(z_{it}) \ln x_{jit} + \frac{1}{2} \sum_j \sum_k \beta_{jk}(z_{it}) \ln x_{jit} \ln x_{kit} + [v_{it} - (u_{it} - E(u_{it}))], \quad (11)$$

where $\theta_0(z_{it}) = \beta_0(z_{it}) - E(u_{it}) = \beta_0(z_{it}) - g(z_{it})\sqrt{(2/\pi)}\sigma_u$. Thus the steps we discussed earlier still work⁴ because the new error term $[v_{it} - (u_{it} - E(u_{it}))]$ has a zero mean.⁵ The only difference is that the smooth coefficient for the so-called intercept is now a different function, $\theta_0(\cdot)$, instead of $\beta_0(\cdot)$.⁶ The residuals are computed from $\ln y_{it} - \hat{\theta}_0(z_{it}) - \sum_j \hat{\beta}_j(z_{it}) \ln x_{jit} - \frac{1}{2} \sum_j \sum_k \hat{\beta}_{jk}(z_{it}) \ln x_{jit} \ln x_{kit}$.

To estimate inefficiency, we consider the residuals $e_{it} = v_{it} - u_{it} + E(u_{it}) = v_{it} - u_{it} + g(z_{it})\sqrt{(2/\pi)}\sigma_u$ which can be viewed as a stochastic frontier except that the frontier part of it, $g(z_{it})\sigma_u\sqrt{(2/\pi)}$, is nonlinear and it has no additional parameters. That is, the parameters in the frontier function are exactly the same as that appear in $E(u_{it})$. With these changes in place one can estimate the function $e_{it} = v_{it} - u_{it} + g(z_{it})\sigma_u\sqrt{(2/\pi)}$ using the maximum likelihood (ML) method which will give estimates of the parameters in the $g(\cdot)$ function, σ_u^2 and σ_v^2 . These estimated values can then be used to obtain estimates of u_{it} using the Jondrow, Lovell, Materov and Schmidt (1982) formula. Note that it is not necessary to assume that u_{it} depends on all the z_{it} variables. In fact, the z_{it} variables in u_{it} can be completely different from those in the smooth coefficients.

3 Data

The data set is based on the Annual Surveys of Industrial Production (ASIP) conducted by the National Bureau of Statistics (NBS) of China. It covers all the state-owned enterprises (SOE's) and the non-SOE's with annual sales exceeding 5 million renminbi (RMB) in the agricultural product and food industries. This is an unbalanced panel data set covering 3,922 firms over 9 years. Output quantity is measured as the monetary value of output (in thousands of RMB) divided by regional Producer Price Index (PPI) of food. The price of capital (w_1) is calculated as the sum of the loan

⁴Since the mean inefficiency is a function of the z variables, the input demand functions based on expected profit maximization will depend on input prices and the z variables. Thus, there is no need to change anything in the input demand functions in (9) for the construction of IV's.

⁵For this, we do not need to have a scaling function representation of inefficiency.

⁶In estimation we use (11) together with the share equations in (8b), replacing inputs by their predicted values.

interest of capital and depreciation rate of fixed assets. The capital inventory (x_1) is measured as the net value of fixed assets in thousands of RMB. Labor quantity (x_2) is measured as the number of employees, and wage (w_2) is the annual income (in thousands of RMB), including wages and benefits, per employee. Based on the regional input-output table of China in 2007, intermediate inputs are divided into three categories: agricultural products, food products, and other materials (e.g, packaging materials, simple gadgets, etc). Because of competitive markets of the intermediate inputs, we take the regional PPI of each intermediate input as its input price. The quantities of these intermediate inputs are calculated as the monetary value (in thousands of RMB) of each intermediate input divided by its regional PPI.

We use four z variables in the model. z_1 is a marketization index. It is a general index published by the National Economic Research Institute (NERI) of China that measures the degrees of markets' competitiveness of provinces in China. China is still switching from a centrally planned economy to a market economy, and the extent of market freedom varies greatly across regions. z_2 is the time trend that goes from 1 (year 2000) to 9 (year 2008). z_3 is a firm ownership variable that discretizes types of ownership and z_4 is a binary regional variable for the eastern and western regions of China. Table 1 contains summary statistics of all the variables used.

4 Estimation results

For comparison purposes, we estimate the following three models:

Model 1: the SPSC SUR model consisting of (11) and (8b),⁷

Model 2: the single equation SPSC model, that is (8a), without taking into account the endogeneity of inputs, and

Model 3: a standard translog parametric model (Kumbhakar and Sun 2013):

$$\ln y_{it} = \beta_0 + \sum_j \beta_j \ln x_{jit} + \frac{1}{2} \sum_j \sum_k \beta_{jk} \ln x_{jit} \ln x_{kit} + v_{it} - u_{it} \quad (12)$$

⁷We set z_1 (marketization index) as a continuous variable, z_2 (time trend) as an ordered categorical variable, and z_3 (ownership type) and z_4 (region) as unordered categorical variables.

where $v_{it} \sim iidN(0, \sigma_v^2)$ and $u_{it} \sim N^+(\mu(z_{it}), \sigma_u^2(z_{it}))$.⁸ No endogeneity (of inputs) correction is made in Models 2 and 3.

Table 2 reports the mean and quartile values (Q1-Q3) of the input elasticity as well as RTS estimates for all the three models. We calculate RTS as the sum of input elasticities, i.e., $RTS = \sum_j \partial \ln y / \partial \ln x_j = \sum_j \hat{q}_j^*$. Note that RTS is observation-specific. It can be seen that the SPSC SUR model has only positive input elasticity estimates for these values, suggesting positive marginal products, while the SPSC single equation and parametric model has negative input elasticity estimates for some inputs.⁹ This might be because of the fact that for the SPSC SUR model we take into account the endogeneity of inputs, while inputs are assumed to be exogenous in the other two models. Meanwhile, the SPSC SUR model generates more precise estimates: it gives the smallest standard errors for the median (Q2) estimates for all input elasticities. This may be because by estimating a system of equations, we take into account more information via the cost share equations in the system approach compared to the other two models which does not take into account information on cost shares and input prices.

Figure 1 illustrates Table 2 via kernel density plots. We can see that the distributions of the input elasticity estimates from the SPSC SUR model lie on the right-hand-side of the zero vertical line, and the dispersions of the elasticity estimates of labor, agriculture products, food products, and other materials are much wider for the SPSC single equation and parametric models than those for the SPSC SUR counterpart. Finally, the RTS estimates from the SPSC SUR model, inclining towards decreasing returns to scale, are stochastically dominated by those from the other two models which show stronger evidence for constant returns to scale. However, these estimates are not appropriate because they include negative elasticities.¹⁰

Since the SPSC SUR model gives more reliable and precise estimates, Figure 2 further plots the mean input elasticity estimates over time based on it, and compare the trends of the estimates between the eastern and western regions of China. The results of the rest of this paper will also

⁸We set z_1 as a continuous variable, and z_2 - z_4 as dummy variables in this model.

⁹Thus RTS estimates from Models 2 and 3 are not meaningful because these estimates include negative input elasticities also. There are too many violations for the input elasticity of x_3 in Model 2, and x_1 in Model 3. Dropping them would lead to dropping most of the observations.

¹⁰See footnote 9.

be based on this model. It can be seen that the mean input elasticity estimates of capital, labor, and food products, are generally higher in the eastern region, while the elasticities of agricultural products and other materials are generally higher in the western region. This is because firms in the eastern region tend to replace agricultural products with food products and/or use advanced machineries instead of simpler equipments. These firms tend to have a higher level of labor specialization than those in the western region. The RTS is higher in the eastern region because most input elasticities are higher in the eastern region. With an increasing amount of accumulated capital during the years of rapid development of China, the marginal productivity of capital diminishes over time. The labor productivity increases after 2004 because of increases in labor efficiency in state-owned firms as well as effects of one-child policy and education.

Figure 3 plots the mean marginal effects of marketization index on the output ($\ln y$) over time for each region. A higher marketization index indicates a higher level of market freedom. These estimates are calculated using the first order gradient estimator of the SPSC SUR estimator presented in Appendix B. It can be seen that there is a decreasing trend of the effects over time. This is probably because of government control or intervention that can not be easily quantified. It is obvious that the western region benefits more from a market economy than the eastern region. This is probably because the western region currently is at a lower level of market freedom, and enjoys a higher level of marginal benefit of an increase in the level of market freedom.

Table 3 reports the summary statistics of the technical (u) and allocative ($\varepsilon_j, \forall j$) inefficiency estimates. The mean technical inefficiency is 0.0586. This means that an average firm could have gained 5.86% more output if there had been no inefficiency. The median estimates of the allocative inefficiencies¹¹ for the first three inputs (i.e., ε_1 - ε_3) are all less than zero. This means that the actual shares of these inputs are lower than their respective optimal shares. For example, the actual share of capital is 2.64% lower than the optimal share of capital. Therefore, given the distributions of the estimated allocative inefficiencies, it seems that most firms in the food industry in China are

¹¹Since the presence of unobserved noise term v in the production function is likely to affect the cost shares, the error terms ε_j will also be affected by v under expected or median profit maximization (following the derivation in Kumbhakar (1987) for the Cobb-Douglas model). Because of this, one may be careful in interpreting the estimated residuals as pure allocative errors, although the noise term might be washed away in the median values. We owe this point to an anonymous referee.

still not doing their best in terms of optimizing input use. To maximize profits, an average firm should increase using those inputs that are currently under-used, for example, capital, labor, and agricultural products, and decrease the over-used inputs such as food products and other materials.

Figure 4 plots the mean technical (u) and allocative ($\varepsilon_j, \forall j$) inefficiencies over time for each region. It can be seen that the technical inefficiency (u) generally follows a decreasing trend. This is probably because of the restructuring of the state-owned enterprises in the western region, as well as of the benefit from the reform and opening-up policy in the eastern region. The western region gradually catches up with the eastern region in terms of the technical inefficiency. This is probably because a large number of the firms in the food industry have been migrating from the eastern to the western region. The technical inefficiency gradually increases in 2004 possibly because of the increases in the labor cost as well as an influx of low-quality food. The allocative inefficiencies of capital and labor show that the two inputs are under-used (i.e., their actual shares are less than their respective optimal shares) in China spanning most years of this study. However, the allocative inefficiencies of almost all inputs seem to be improving over time, approaching the horizontal lines at zero. The western region uses more capital and agricultural products than the eastern region. The eastern region uses more labor, food products, and other materials than the western region. This is probably because most firms in the eastern region are family-owned business or township enterprises.

Figure 5 plots the mean marginal effects of marketization index on technical and allocative inefficiencies over time for each region. These estimates are calculated by running a separate nonparametric regression of each estimated inefficiencies on the marketization index.¹² The median marginal effect on technical inefficiency is -0.0359 (i.e., 3.59% decrease for a one unit increase in the marketization index), and is statistically significant at the 1% level. This indicates that a higher level of market freedom may reduce technical inefficiency in most cases. It is obvious that the marginal effects on the allocative inefficiencies of all inputs except for labor are mostly negative. This means that a higher level of market freedom decreases these allocative inefficiencies as well. For the case of labor, sometimes the marginal effects are positive probably because of corruption

¹²Kumbhakar and Sun (2013) gives an alternative method of obtaining the marginal effects on technical inefficiency based on the Jondrow et al. (1982) estimator.

during the labor hiring processes.

Finally, Figures 6 and 7 plots the mean technical efficiency change ($TEC = -\partial\hat{u}/\partial t$) and technical change ($TC = \partial\widehat{\ln y}/\partial t$). They are obtained by running a nonparametric regression of the estimated technical inefficiency and output, respectively, on a continuous time trend and the binary region variable.¹³ It can be seen that using the entire sample, the technical efficiency decreases (i.e., technical inefficiency increases) in 2000 and 2004, and it increases (i.e., technical inefficiency decreases) in all other years. This is consistent with the first plot of Figure 4. Furthermore, the western region enjoys a more positive TEC during most years than eastern region. This is also consistent with the first plot of Figure 4 in that the technical inefficiency decreases faster in the western region than eastern region. The TC 's are positive for most years of the study. The overall mean TC is 0.0353 (i.e., 3.53% per year) with a standard error of 0.0006. The western region has a much higher level of TC during the earlier years of the study and then gradually follows the trajectory of the eastern region. The western region was under-developed before this century, and the Chinese government launched the Western Development Program in 2000. One of the main aims of this program is to balance the development between the eastern and western regions of China. The TC becomes negative in 2008 probably because of the effect of the financial crisis.

5 Conclusion

This paper addresses the endogeneity of inputs in estimating a semiparametric smooth coefficient (SPSC) production function using a system approach. The system consists of a translog production frontier and the first-order conditions (FOC's) of profit maximization. Each coefficient of the production function is an unknown function of some exogenous environmental variables. The estimation of the system involves applying the functional coefficient instrumental variable (IV) method (Cai et al. 2006) for the endogeneity of inputs in the first step, and the semiparametric smooth coefficient seemingly unrelated regression method (Henderson et al. 2015) in the second step. Technical inefficiencies are calculated using the stochastic frontier analysis (SFA), and allocative

¹³Time trend is used as one of the categorical z variables in (11) and (8b), but a continuous variable when calculating the TEC and TC for the partial derivatives to be well-defined.

inefficiencies are defined as the error terms of the share equations, which can be interpreted as deviations of the actual input shares from optimal shares. We then apply a firm-level Chinese food industry panel data set as an empirical example, and find that the system approach gives more sensible input elasticity estimates than single equation SPSC or traditional SFA models. We highlight the differences in the input elasticity, RTS, and inefficiency estimates between the eastern and western regions of China. Technical efficiency change (TEC) and technical change (TC) are also computed.

Appendix A

We use the Lagrangian multiplier method to solve for the profit maximization problem in (2). The Lagrangian function is:

$$L = py - w'x + \lambda[f(x; \beta(z)) - y]. \quad (13)$$

We maximize this function with respect to y and x_j , $\forall j = 1, \dots, K$. The first order conditions are:

$$\frac{\partial L}{\partial y} = p - \lambda = 0 \quad (14)$$

and

$$\frac{\partial L}{\partial x_j} = -w_j + \lambda f_j(x; \beta(z)) = 0, \quad \forall j = 1, \dots, K, \quad (15)$$

where $f_j(x; \beta(z)) = \partial f(x; \beta(z)) / \partial x_j$. Replace λ in (15) with p from (14), and we would have:

$$w_j = pf_j(x; \beta(z)), \quad \forall j = 1, \dots, K. \quad (16)$$

Appendix B

Following Henderson et al. (2015), the general form of the SPSC SUR model is:

$$Y = X\beta(Z) + e, \quad (17)$$

where Y is a $G \sum_{i=1}^N T_i \times 1$ stack vector, G is the number of equations, N is the number of cross-sectional units, and T_i is the maximum number of years for the i -th unit. X is a $G \sum_{i=1}^N T_i \times \tilde{K}$ data matrix, and Z is a $\sum_{i=1}^N T_i \times S$ data matrix. Note that all coefficients in all equations have the same Z matrix.

Given our dataset and model, we define the following data matrices for the SPSC SUR estimator:

$Y' = [\ln y' \ q'_1 \ \cdots \ q'_K]$, where $\ln y$ and $q_j, \forall j = 1, \dots, K$, are $\sum_{i=1}^N T_i \times 1$ data vectors;

$$X = \begin{bmatrix} X_0 \\ \mathbf{0} \ \mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_1} \ \widehat{\ln x_2} \ \widehat{\ln x_3} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \\ \mathbf{0} \ \mathbf{0} \ \mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_2} \ \widehat{\ln x_3} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \\ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_2} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_3} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \\ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_2} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_3} \ \mathbf{0} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \ \mathbf{0} \\ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_1} \ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_2} \ \mathbf{0} \ \mathbf{0} \ \widehat{\ln x_3} \ \mathbf{0} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \end{bmatrix},$$

where $X_0 = [1 \ \widehat{\ln x_1} \ \widehat{\ln x_2} \ \widehat{\ln x_3} \ \widehat{\ln x_4} \ \widehat{\ln x_5} \ 0.5\widehat{\ln x_1}^2 \ \widehat{\ln x_1}\widehat{\ln x_2} \ \widehat{\ln x_1}\widehat{\ln x_3} \ \widehat{\ln x_1}\widehat{\ln x_4} \ \widehat{\ln x_1}\widehat{\ln x_5} \ 0.5\widehat{\ln x_2}^2 \ \widehat{\ln x_2}\widehat{\ln x_3} \ \widehat{\ln x_2}\widehat{\ln x_4} \ \widehat{\ln x_2}\widehat{\ln x_5} \ 0.5\widehat{\ln x_3}^2 \ \widehat{\ln x_3}\widehat{\ln x_4} \ \widehat{\ln x_3}\widehat{\ln x_5} \ 0.5\widehat{\ln x_4}^2 \ \widehat{\ln x_4}\widehat{\ln x_5} \ 0.5\widehat{\ln x_5}^2]$; $\widehat{\ln x_j}, \forall j = 1, \dots, K$, are vectors of fitted values of $\ln x_j$; $\mathbf{0}$ and $\mathbf{1}$ are $\sum_{i=1}^N T_i \times 1$ zero and one vectors, respectively;¹⁴ and $Z = [z_1, z_2, z_3, z_4]$. While we set z_1 (marketization index) as a continuous variable, z_2 (time trend) is an ordered categorical variable, and z_3 (ownership type) and z_4 (region) are unordered categorical variables. Finally, $e' = [v' \ \varepsilon'_1 \ \cdots \ \varepsilon'_K]$.

The SPSC SUR estimator for $\beta(Z)$ is:

$$\hat{\beta}(Z) = [X' \mathbf{K}(Z)^{1/2} \Omega^{-1} \mathbf{K}(Z)^{1/2} X]^{-1} X' \mathbf{K}(Z)^{1/2} \Omega^{-1} \mathbf{K}(Z)^{1/2} Y, \quad (18)$$

where $\mathbf{K}(Z)$ is a $G \sum_{i=1}^N T_i \times G \sum_{i=1}^N T_i$ block diagonal matrix. Each diagonal element of $\mathbf{K}(Z)$ is a $\sum_{i=1}^N T_i \times \sum_{i=1}^N T_i$ kernel diagonal matrix, and the it -th element of the kernel diagonal matrix is a product kernel function (Li and Racine 2007),

$$k_{it}(Z) = \prod_{l=1}^L k_{it}^c(Z_l^c) \prod_{m=1}^M k_{it}^o(Z_m^o) \prod_{r=1}^R k_{it}^u(Z_r^u), \quad (19)$$

¹⁴Note that the X matrix is designed to impose the cross-equation restrictions on the coefficients.

where

$$k_{it}^c(Z_l^c) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{Z_{lit}^c - Z_l^c}{h_l^c}\right)^2\right) \quad (20)$$

is the kernel function for the l -th continuous Z variable (i.e., $Z_l^c = z_1$) and h_l^c is the bandwidth for this variable. Following Aitchison and Aitken (1976), and letting \tilde{c} denote the number of categories the discrete variable can take,

$$k_{it}^o(Z_m^o) = \binom{\tilde{c}}{j} h_m^{oj} (1 - h_m^o)^{\tilde{c}-j}, \text{ when } |Z_{mit}^o - Z_m^o| = j \quad (21)$$

is the kernel function for the m -th ordered categorical Z variable (i.e., $Z_m^o = z_2$) and h_m^o is the bandwidth for this variable.

$$k_{it}^u(Z_r^u) = \begin{cases} 1 - h_r^u, & \text{if } Z_{rit}^u = Z_r^u \\ h_r^u / (\tilde{c} - 1), & \text{otherwise} \end{cases} \quad (22)$$

is the kernel function for the r -th unordered categorical Z variable (i.e., $Z_r^u = z_3$ or z_4) and h_r^u is the bandwidth for this variable.

Let $h' = [h^c \ h^{o'} \ h^{u'}]$ be the bandwidth vector for all types of Z variables, and we estimate h using the least-squares cross-validation method. The cross-validation criterion function is given by:

$$CV(h) = \min_h \sum_{i=1}^N \sum_{t=1}^{T_i} [Y_{it} - X'_{it} \hat{\beta}(Z_{it})_{-it}]^2, \quad (23)$$

in which $X'_{it} \hat{\beta}(Z_{it})_{-it}$ is the leave-one-out estimator of the conditional mean, and $\hat{\beta}(Z_{it})_{-it}$ is the leave-one-out SPSC SUR estimator where the it -th observation is removed from the g -th equation, ignoring any cross-equation correlation.

Finally, $\Omega = \omega \otimes I_{\sum_{i=1}^N T_i}$, where ω is the $(K + 1) \times (K + 1)$ variance-covariance matrix of the error terms in (8). In practice, ω is replaced by $\hat{\omega}$ containing estimates of the variances and covariances, calculated as means of the cross products of the residuals, $\hat{v}, \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_K$.

In addition to presenting the SPSC SUR estimator, we also derive its first order gradient

estimator, which facilitates the estimation of marginal effects of any continuous Z variable. Let $S(Z) = \mathbf{K}(Z)^{1/2}\Omega^{-1}\mathbf{K}(Z)^{1/2}$, then the SPSC SUR estimator is written as:

$$\hat{\beta}(Z) = [X'S(Z)X]^{-1}X'S(Z)Y. \quad (24)$$

Let $A(Z) = X'S(Z)X$ and $B(Z) = X'S(Z)Y$, then the derivative with respect to the l -th continuous Z , Z_l^c , is:

$$\frac{\partial \hat{\beta}(Z)}{\partial Z_l^c} = \frac{\partial A(Z)^{-1}}{\partial Z_l^c} B(Z) + A(Z)^{-1} \frac{\partial B(Z)}{\partial Z_l^c}, \quad (25)$$

where $\frac{\partial A(Z)^{-1}}{\partial Z_l^c} = -A(Z)^{-1} \frac{\partial A(Z)}{\partial Z_l^c} A(Z)^{-1}$, $\frac{\partial A(Z)}{\partial Z_l^c} = X' \frac{\partial S(Z)}{\partial Z_l^c} X$, $\frac{\partial B(Z)}{\partial Z_l^c} = X' \frac{\partial S(Z)}{\partial Z_l^c} Y$,

$$\frac{\partial S(Z)}{\partial Z_l^c} = \frac{\partial \mathbf{K}(Z)^{1/2}}{\partial Z_l^c} \Omega^{-1} \mathbf{K}(Z)^{1/2} + \mathbf{K}(Z)^{1/2} \Omega^{-1} \frac{\partial \mathbf{K}(Z)^{1/2}}{\partial Z_l^c}, \quad (26)$$

and

$$\frac{\partial \mathbf{K}(Z)^{1/2}}{\partial Z_l^c} = \text{diag} \left[\frac{\partial \mathbf{k}(Z)^{1/2}}{\partial Z_l^c}, \dots, \frac{\partial \mathbf{k}(Z)^{1/2}}{\partial Z_l^c} \right] \quad (27)$$

is a block diagonal matrix, where each element

$$\frac{\partial \mathbf{k}(Z)^{1/2}}{\partial Z_l^c} = \text{diag} \left[\frac{\partial k_1(Z)^{1/2}}{\partial Z_l^c}, \dots, \frac{\partial k_{\sum_{i=1}^N T_i}(Z)^{1/2}}{\partial Z_l^c} \right], \quad (28)$$

where a typical element of this diagonal matrix is a function of $k_{it}(Z)$. Finally,

$$\frac{\partial k_{it}(Z)^{1/2}}{\partial Z_l^c} = \frac{1}{2} \frac{Z_{lit}^c - Z_l^c}{h_l^c} k_{it}(Z)^{1/2}. \quad (29)$$

References

- Aitchison, J. and Aitken, C. (1976), ‘Multivariate binary discrimination by kernel method’, *Biometrika* **63**, 413–420.
- Bhaumik, S. K., Kumbhakar, S. C. and Sun, K. (2015), ‘A note on a semiparametric approach to estimating financing constraints in firms’, *European Journal of Finance* **21**, 992–1004.
- Cai, Z., Das, M., Xiong, H. and Wu, X. (2006), ‘Functional coefficient instrumental variables models’, *Journal of Econometrics* **133**(1), 207–241.

- Cai, Z., Fan, J. and Li, R. (2000), 'Efficient estimation and inferences for varying-coefficient models', *Journal of the American Statistical Association* **95**(451), 888–902.
- Cai, Z. and Li, Q. (2008), 'Nonparametric estimation of varying coefficient dynamic panel data models', *Econometric Theory* **24**, 1321–1342.
- Chen, R. and Tsay, R. (1993), 'Functional-coefficient autoregressive models', *Journal of the American Statistical Association* **88**, 298–308.
- Delgado, M. S., McCloud, N. and Kumbhakar, S. C. (2014), 'A generalized empirical model of corruption, foreign direct investment, and growth', *Journal of Macroeconomics* **42**, 298–316.
- Fan, J. and Zhang, W. (1999), 'Statistical estimation in varying-coefficient models', *The Annals of Statistics* **27**, 1491–1518.
- Hastie, T. and Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4), 757–796.
- Henderson, D. J., Kumbhakar, S. C., Li, Q. and Parmeter, C. F. (2015), 'Smooth coefficient estimation of a seemingly unrelated regression', *Journal of Econometrics* **189**(1), 148–162.
- Heshmati, A., Kumbhakar, S. C. and Sun, K. (2014), 'Estimation of productivity in Korean electric power plants: A semiparametric smooth coefficient model', *Energy Economics* **45**, 491–500.
- Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**, 233–238.
- Kumbhakar, S. C. (1987), 'The specification of technical and allocative inefficiency in stochastic production and profit frontiers', *Journal of Econometrics* **34**(3), 335–348.
- Kumbhakar, S. C. and Lovell, C. A. K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press.
- Kumbhakar, S. C. and Sun, K. (2013), 'Derivation of marginal effects of determinants of technical inefficiency', *Economics Letters* **120**(2), 249–253.
- Li, Q., Huang, C., Li, D. and Fu, T. (2002), 'Semiparametric smooth coefficient models', *Journal of Business and Economic Statistics* **20**(3), 412–422.
- Li, Q. and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Li, Q. and Racine, J. S. (2010), 'Smooth varying-coefficient estimation and inference for qualitative and quantitative data', *Econometric Theory* **26**, 1607–1637.
- Liu, W. (2014), 'Modeling gasoline demand in the United States: A flexible semiparametric approach', *Energy Economics* **45**, 244–253.
- Polemis, M. L. and Stengos, T. (2015), 'Does market structure affect labour productivity and wages? Evidence from a smooth coefficient semiparametric panel model', *Economics Letters* **137**, 182–186.

Schmidt, P. and Lovell, C. A. K. (1979), ‘Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers’, *Journal of Econometrics* **9**(3), 343–366.

Schmidt, P. and Lovell, C. A. K. (1980), ‘Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated’, *Journal of Econometrics* **13**(1), 83–100.

Zhang, R., Sun, K., Delgado, M. S. and Kumbhakar, S. C. (2012), ‘Productivity in China’s high technology industry: Regional heterogeneity and R&D’, *Technological Forecasting and Social Change* **79**, 127–141.

Table 1: **Summary Statistics of the Variables**

Symbol	Variable Name	Mean	Sd.	Min.	Max.
y	Output quantity	82403.21	229317.38	2.06	5323079.00
x_1	Capital stock	15239.90	42113.53	1.00	1085352.00
x_2	Number of employee	165.65	286.12	2.00	6023.00
x_3	Agricultural products	25748.34	68948.43	0.44	1796366.00
x_4	Food products	13333.13	41604.01	0.18	1312121.00
x_5	Other materials	20819.63	57604.14	0.33	1707589.00
w_1	Capital price	0.14	0.21	0.00	4.97
w_2	Labor price	21.53	24.80	0.01	446.70
w_3	Agricultural products price	1.11	0.23	0.85	1.63
w_4	Food price	1.02	0.08	0.90	1.18
w_5	Other materials price	1.01	0.07	0.92	1.17
z_1	Marketization index	8.03	1.79	4.41	11.04
z_2	Time trend	5.68	2.53	1.00	9.00
z_3	Ownership type
z_4	Region

1. Total number of observations = 13361.
2. t is defined as year - 1999, where year varies from 2000 to 2008.
3. Ownership type= 1 if stated-owned firm, = 2 if collectively-owned firm, = 3 if legal person-owned firm, = 4 if private firm, = 5 if foreign firm, = 6 if firm of Hongkong, Taiwan and Macau, = 7 if others.
4. Region= 1 if eastern region, = 2 if western region. The eastern region is represented by the Pearl River Delta, and the western region is represented by Sichuan province.

Table 2: Summary Statistics of the Input Elasticity Estimates and RTS

	$\frac{\partial \ln y}{\partial \ln x_1}$	$\frac{\partial \ln y}{\partial \ln x_2}$	$\frac{\partial \ln y}{\partial \ln x_3}$	$\frac{\partial \ln y}{\partial \ln x_4}$	$\frac{\partial \ln y}{\partial \ln x_5}$	RTS
Model 1: SPSC SUR						
Mean	0.0471 (0.0005)	0.0849 (0.0009)	0.3483 (0.0031)	0.1568 (0.0014)	0.2587 (0.0022)	0.8991 (0.0082)
Q1	0.0312 (0.0001)	0.0464 (0.0003)	0.3341 (0.0001)	0.1276 (0.0002)	0.2378 (0.0002)	0.8391 (0.0006)
Q2	0.0464 (0.0002)	0.0774 (0.0003)	0.3486 (0.0002)	0.1561 (0.0004)	0.2573 (0.0002)	0.8879 (0.0005)
Q3	0.0613 (0.0028)	0.1172 (0.0038)	0.3626 (0.0034)	0.1846 (0.0036)	0.2789 (0.0030)	0.9513 (0.0070)
Model 2: SPSC single equation						
Mean	0.0226 (0.0004)	0.0561 (0.0011)	-0.2074 (0.0084)	0.7565 (0.0103)	0.3581 (0.0104)	0.9859 (0.0087)
Q1	0.0024 (0.0002)	-0.0108 (0.0005)	-0.7936 (0.0089)	0.0314 (0.0025)	-0.3949 (0.0075)	0.9594 (0.0002)
Q2	0.0205 (0.0002)	0.0513 (0.0007)	-0.3112 (0.0054)	0.9615 (0.0110)	0.1492 (0.0050)	0.9880 (0.0004)
Q3	0.0405 (0.0005)	0.1107 (0.0012)	0.3545 (0.0049)	1.5293 (0.0082)	0.7266 (0.0066)	1.0127 (0.0083)
Model 3: Parametric						
Mean	-0.1064 (0.0010)	0.0776 (0.0011)	-0.0220 (0.0082)	0.2080 (0.0032)	0.8544 (0.0115)	1.0117 (0.0090)
Q1	-0.1315 (0.0048)	0.0059 (0.0005)	-0.7867 (0.0060)	-0.0193 (0.0015)	0.0810 (0.0220)	0.9901 (0.0002)
Q2	-0.1061 (0.0003)	0.0757 (0.0007)	0.1155 (0.0080)	0.2326 (0.0032)	0.6466 (0.0047)	1.0126 (0.0003)
Q3	-0.0814 (0.0002)	0.1449 (0.0014)	0.6953 (0.0074)	0.4182 (0.0047)	1.6349 (0.0079)	1.0340 (0.0045)

Standard errors are in the parentheses.

Figure 1: Kernel Density Plots for Input Elasticity Estimates

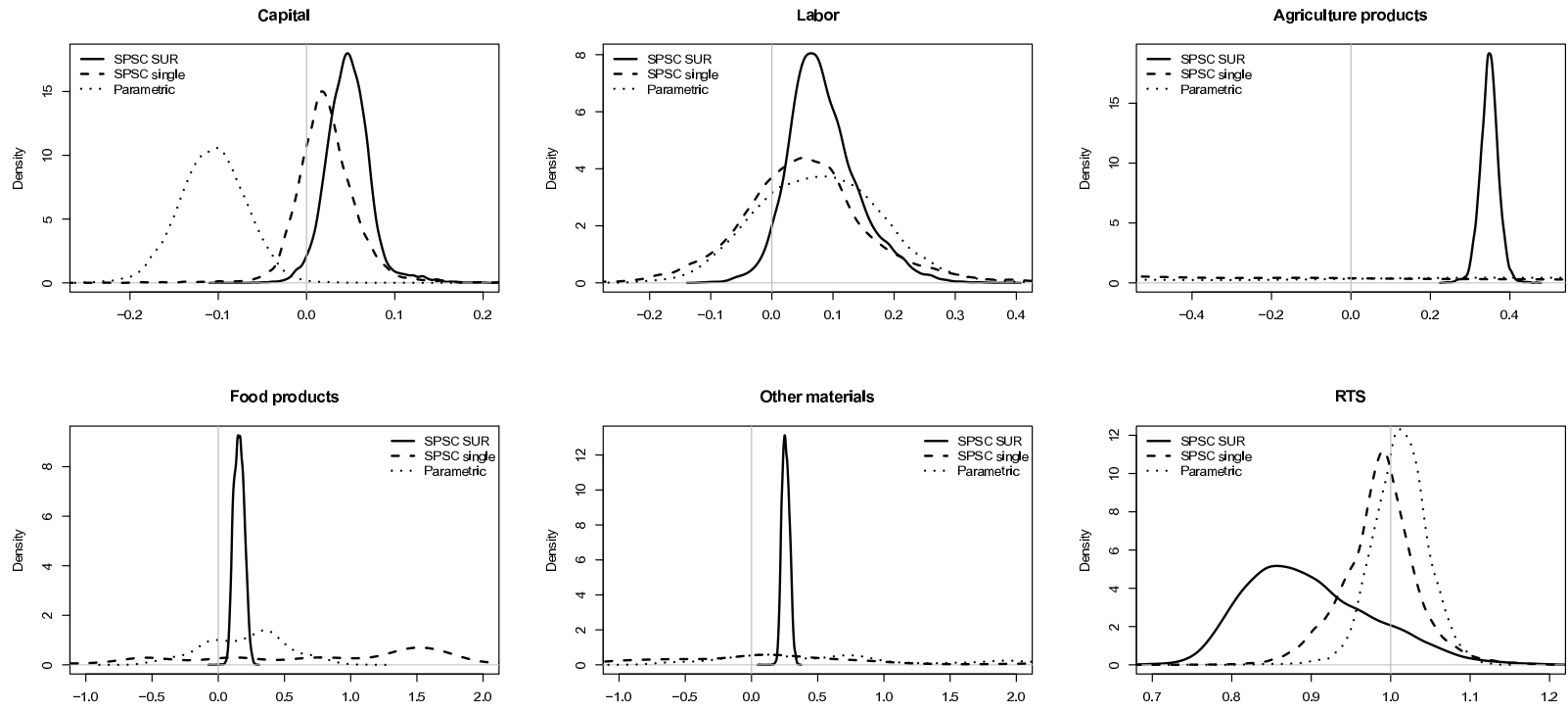


Figure 2: Input Elasticity Estimates by Region

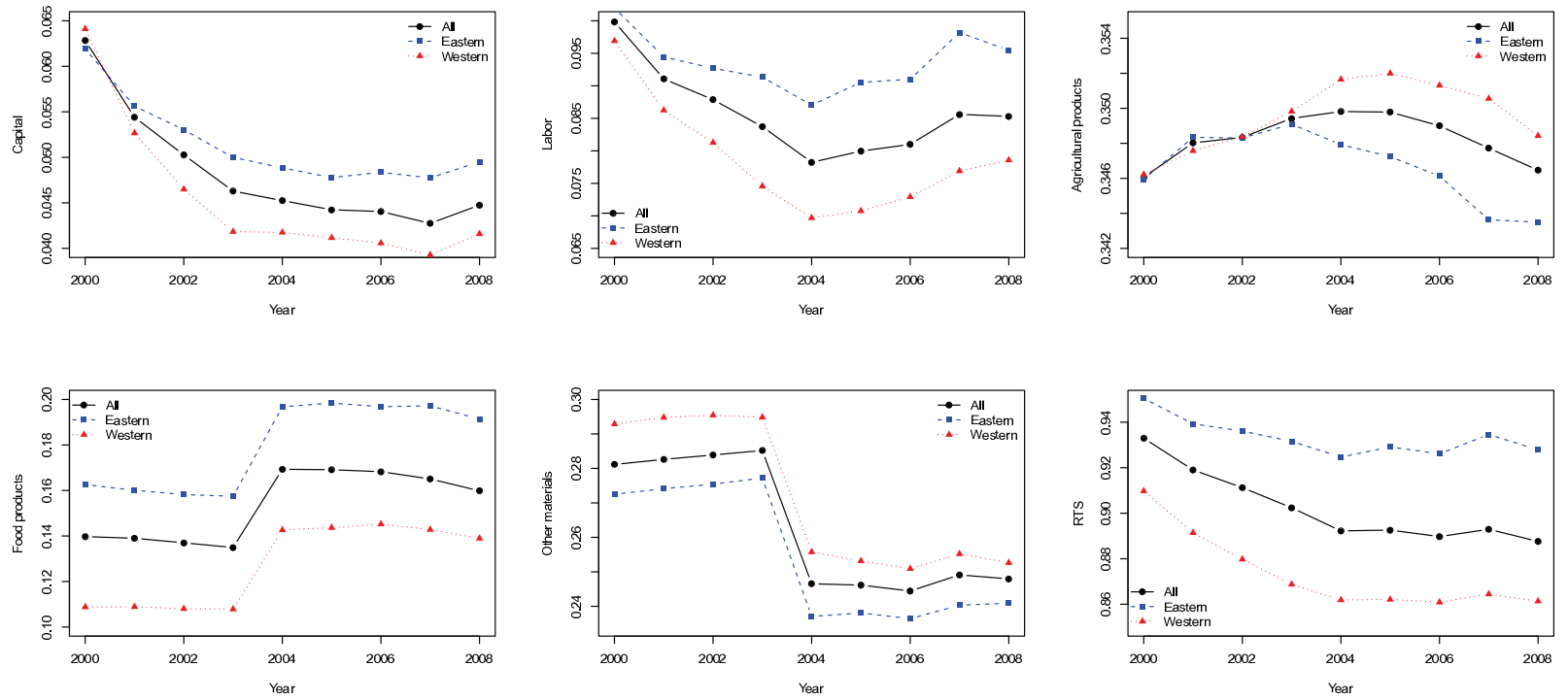


Figure 3: Marginal Effects of Marketization Index on Output

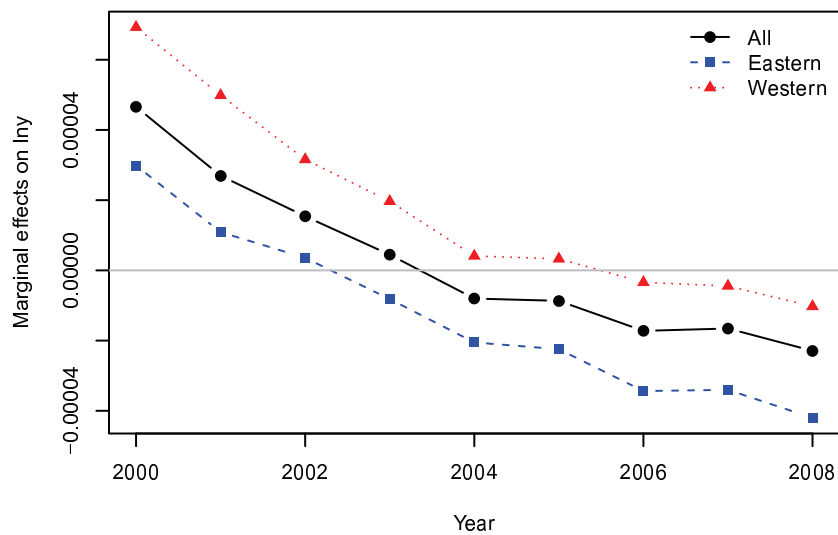


Table 3: Summary Statistics of Inefficiencies

	u	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5
	Output	Capital	Labor	Agricultural products	Food products	Other materials
Mean	0.0586 (0.0007)	-0.0091 (0.0033)	-0.0045 (0.0015)	-0.000003 (0.0008)	0.000038 (0.0004)	0.0008 (0.0006)
Q1	0.0208 (0.0003)	-0.0424 (0.0010)	-0.0629 (0.0007)	-0.0314 (0.0004)	-0.0239 (0.0003)	-0.0285 (0.0003)
Q2	0.0414 (0.0001)	-0.0264 (0.0002)	-0.0209 (0.0005)	-0.0019 (0.0003)	0.0004 (0.0002)	0.0010 (0.0003)
Q3	0.0844 (0.0005)	-0.0063 (0.0001)	0.0273 (0.0005)	0.0300 (0.0004)	0.0226 (0.0003)	0.0293 (0.0004)

Standard errors are in the parentheses.

Figure 4: Inefficiencies by Region

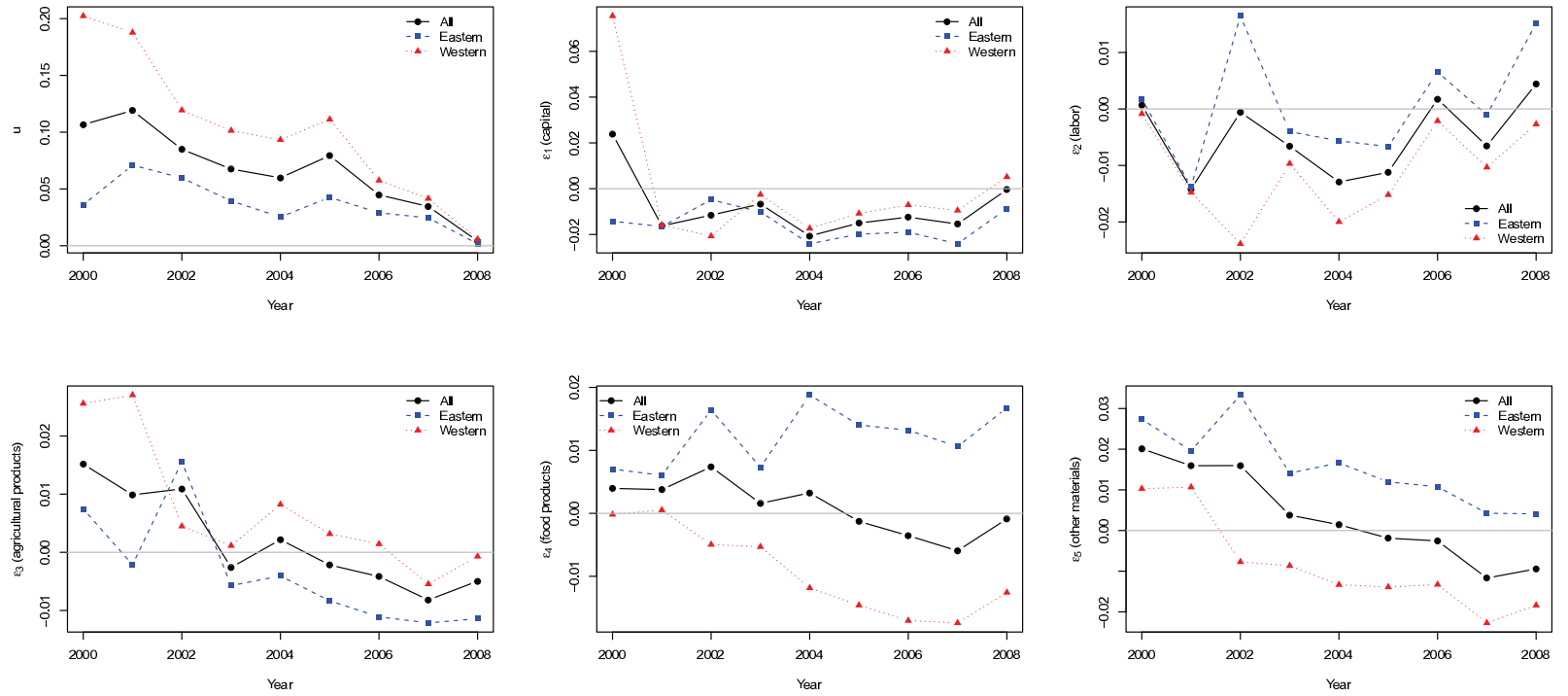


Figure 5: Marginal Effects of Marketization Index on Inefficiencies

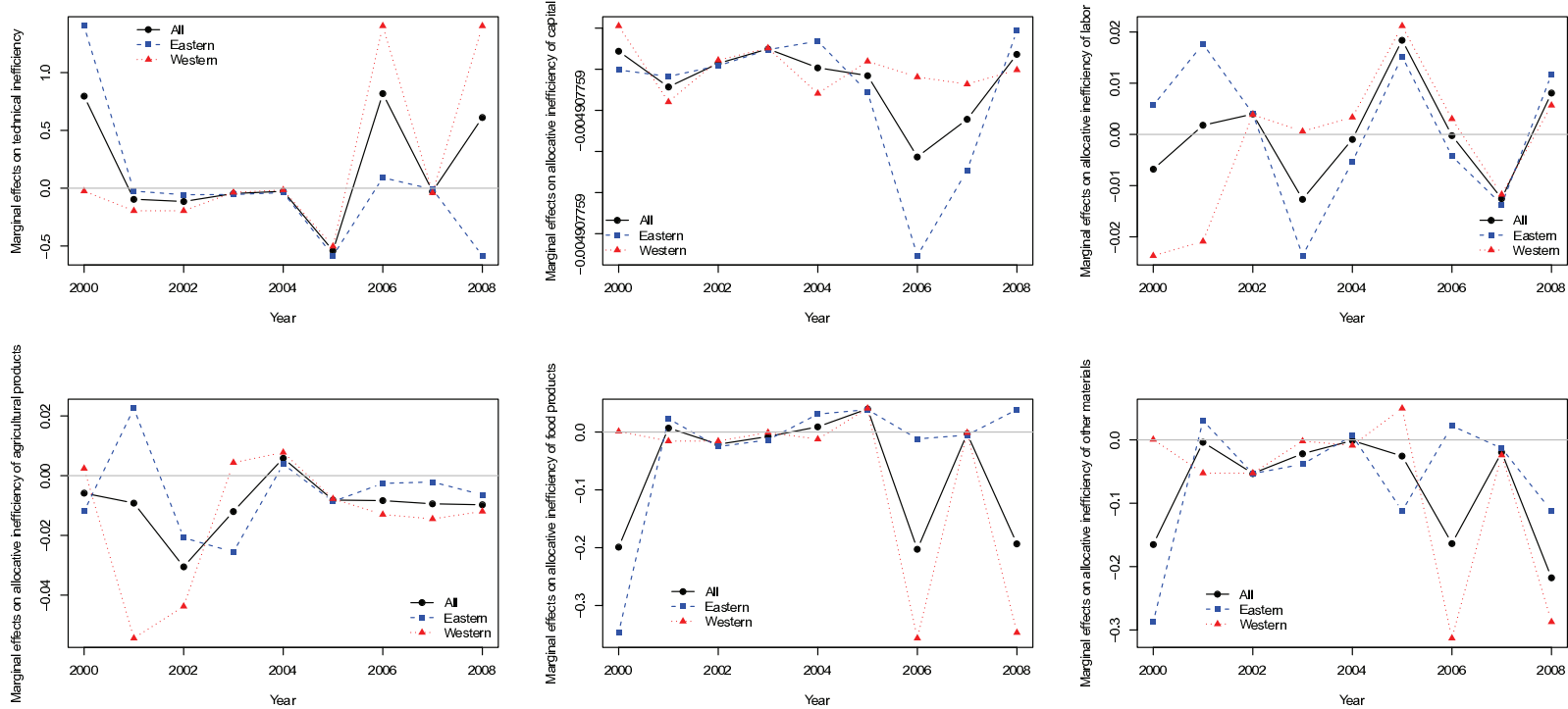


Figure 6: Technical efficiency change

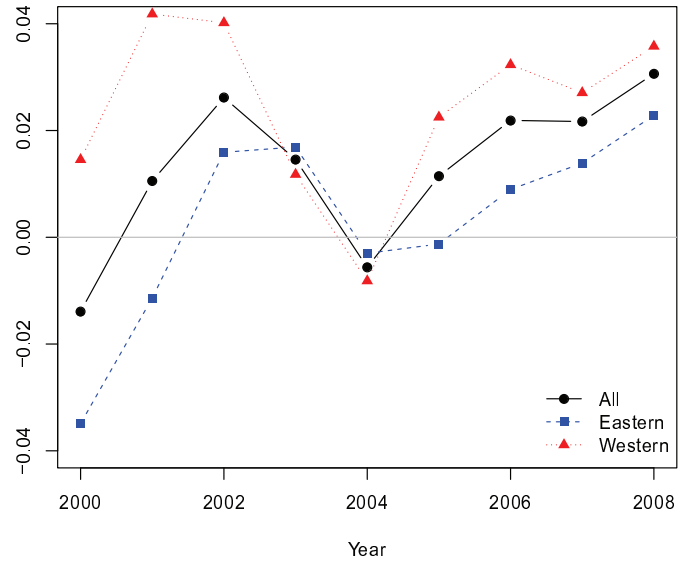


Figure 7: Technical change

