# Estimation of a partially linear seemingly unrelated regressions model: Application to a translog cost system

Xin Geng[a] and Kai Sun[*b]

[a]School of Finance, Nankai University, Tianjin 300350, China
[b]School of Economics, Shanghai University, Shanghai 200444, China

**Abstract**

This paper studies a partially linear seemingly unrelated regressions (SUR) model to estimate a translog cost system that consists of a partially linear translog cost function and input share equations. The parametric component is estimated via a simple two-step feasible SUR estimation procedure. We show that the resulting estimator achieves root-$n$ convergence and is asymptotically normal. The nonparametric component is estimated with a nonparametric SUR estimator based on the Cholesky decomposition. We show that this estimator is consistent, asymptotically normal, and more efficient relative to the ones that ignore cross-equation correlation. A model specification test for parametric functional form is proposed. An Italian banking data set is used to estimate the translog cost system. Results show that marginal effects of risks on cost of production are heterogeneous, but increase with risk levels.

*Keywords:* translog cost system; partially linear model; seemingly unrelated regressions

*JEL codes:* C14; D24

[*]Corresponding author.
*E-mail addresses:* x.geng@outlook.com (X. Geng), ksun1@shu.edu.cn (K. Sun).

# 1 Introduction

Robinson (1988) proposed a partially linear single-equation model

$$y_i = \theta(z_i) + x_i'\beta + u_i,$$

where the subscript $i = 1, \ldots, n$ denotes observation, $y_i$ is a response, $\theta(\cdot)$ is a nonparametric function of a vector of covariates $z_i$, $x_i$ is a vector of regressors with a parameter vector $\beta$, and $u_i$ is a scalar error term with $\mathbb{E}(u_i \mid z_i, x_i) = 0$. This model is more flexible than one with $\theta(\cdot)$ being replaced by a parametric function of $z_i$, and thus has been used to estimate many economic relationships, such as hedonic price equation (Bontemps et al. 2006), firm production technology (Bhaumik et al. 2015), environmental Kuznets curve (Millimet et al. 2003), and Engel curve (Blundell et al. 1998).

However, investigating economic relationships often involves estimating multiple equations. For example, in production theory, a cost function is usually estimated along with conditional factor demand functions (or factor share equations) derived from Shephard's Lemma; a profit function can be estimated along with either unconditional factor demand or output supply functions derived from Hotelling's Lemma. The factor demand functions are correlated because of input substitutability, and the output supply functions are correlated because a firm must make production decisions based on the production possibilities frontier. In consumer theory, the expenditure share equations can be estimated in a multiple-equation system for the analysis of Engle curves for different consumption goods, which might also be correlated when a consumer makes choice decisions.

Recently, Henderson et al. (2015) proposed a semiparametric smooth coefficient (SPSC) seemingly unrelated regressions (SUR) model, in which all the regression coefficients in the system are nonparametric functions of $z_i$. While this specification is more flexible than the parsimonious specification of our partially linear SUR model, all the smooth-varying coefficients, including intercept and slopes, of the SPSC SUR model are subject to the "curse of dimensionality". If information is given *a priori* that the coefficients in their translog system are constants, it would be inefficient to estimate their translog system with varying coefficients and the coefficient estimator cannot achieve the faster convergence rate of

$\sqrt{n}$. Subject to similar issues, Xu et al. (2008) and Su et al. (2013) considered the estimation of fully nonparametric SUR without any parametric component in their regression functions.

In this paper, we propose a straightforward extension of Robinson (1988), such that estimating a system of partially linear equations is allowed—we call it partially linear SUR. The parametric component is estimated via a simple two-step feasible SUR estimation procedure. In the first step, cross-equation correlation is ignored and the partially linear models are estimated equation by equation as in Robinson (1988); the second step employs the generalized least squares (GLS) estimator for the linear part, analogous to Zellner (1962), and uses the residuals from the first step to capture contemporaneous correlations among the equations. We show that the resulting SUR estimator for the linear part achieves root-$n$ convergence and is asymptotically normal. The nonparametric component can be consistently estimated using either Robinson (1988) single-equation estimator or one that is based on the above SUR estimator for the linear part. While these two nonparametric estimators are asymptotically equivalent, we propose a more efficient nonparametric SUR estimator based on the Cholesky decomposition. Estimation of both the parametric and nonparametric components can be easily implemented. To the best of our knowledge, this more efficient nonparametric SUR estimation has not been implemented in any of the partially linear SUR setting in the literature.

When a fully parametric SUR is correctly specified, it would yield more efficient estimates than its partially linear counterpart. For this reason, we propose a model specification test for the null hypothesis of a parametric functional form, following Cai et al. (2000) goodness-of-fit test procedure, by comparing the residual sums of squares (RSS) from the restricted (i.e., parametric) and unrestricted (i.e., partially linear) fittings. A novel cluster bootstrap procedure is proposed to obtain the empirical distribution of the test statistic and the $p$-value of the test. You & Zhou (2010, 2014) proposed alternative versions of the semiparametric SUR model, and employed a combination of profile least squares and local polynomial estimation techniques to estimate the parametric and nonparametric components, respectively. However, they did not recognize the potential advantage of a parametric SUR over its semiparametric counterpart, and did not provide a test procedure that helps researchers determine which specification is preferred in practice. Furthermore, no real application was given in You & Zhou (2014). By contrast, the estimation procedure in this paper is simpler and easier to implement in practice.

The SUR estimator for the linear part and nonparametric SUR estimator are then applied seamlessly to the estimation of a translog cost system. The cost system consists of a partially linear translog cost function and a set of share equations. The nonparametric component of the cost function is interpreted as a productivity parameter, which is an unknown nonparametric function of $z_i$—a vector of non-traditional inputs or environmental factors, e.g., firm age, size, risk exposure, policy variables, that describe the environment in which production takes place. Recently, attention has been paid to more flexible modeling of these environmental factors in estimating technology (Bhaumik et al. 2015, Baležentis et al. 2020, among others). The parametric component of the cost function is translog in input prices, e.g., labor and capital prices, as well as in outputs—it includes all the higher order and interaction terms of these variables.[1] The share equations are derived from the cost function itself by Shephard's Lemma, and therefore share the same parameters as the cost function. This indicates that estimating the cost system is more efficient than estimating a single cost function only, because with the help of the share equations, more data are used in the SUR without an increase in the number of parameters (Kumbhakar 1991, Kumbhakar & Tsionas 2005).

As an empirical example, an Italian banking data set is used to demonstrate the methodology. The three banking risks (i.e., credit risk, solvency risk, and liquidity risk), along with a time trend, are modeled as the environmental variables, which affect total cost in flexible manners. It is important to include risk-taking behavior of banks when investigating their performances (Hughes & Mester 1998). On one hand, risks could increase costs for some banks, because of additional non-interest expenses in administering the loan portfolios and managing financial capital and liquidity assets, but on the other hand, risks could also decrease costs for the other banks, if managers of these banks skimp on the resources contributed to risk management (Berger & DeYoung 1997). Therefore, it would be desirable to

[1] The translog parameters have few economic meanings because they are not elasticities on their own. Therefore, it would be costly to make these parameters unknown functions of $z_i$ as in Henderson et al. (2015). The input price and output elasticities derived from the translog parameters are more economically meaningful, and are observation-specific so long as the cost function is translog in input prices and outputs, even if the slope translog coefficients are constants rather than nonparametric functions.

examine the marginal impact of risks on bank performance, especially when a cost function is employed to represent a bank's technology. Briefly, we find that 1) the partially linear SUR model produces more accurate parameter estimates than the partially linear single-equation and linear SUR counterparts, is more successful in predicting the input shares than the partially linear single-equation model, and yields more reasonable RTS estimates than the linear SUR model; 2) the marginal effect estimates obtained from the nonparametric SUR estimation have much smaller variations than those from the single-equation estimation; and 3) the heterogeneous marginal effects of banking risks and time on cost of production obtained from the nonparametric SUR estimator are useful in making regulations aiming to reduce the probability of bank failure without profoundly harming banks' profitability.

As a further topic, in light of potential heterogeneity in firm production decisions or degrees of input substitutability, how to account for the resultant heteroscedasticity in the context of the partially linear production system is discussed. From the econometric perspective, between-equation heteroscedasticity exists when the contemporaneous cross-equation correlation is observation-specific, and within-equation heteroscedasticity exists when the variance of the error term of each equation of the system is a function of inputs ($x_i$) and environmental factors ($z_i$). We describe the steps of estimating the cross- and within-equation scedastic functions nonparametrically, and also of detecting these two types of heteroscedasticity.

The rest of this paper is organized as follows. Section 2 motivates our partially linear SUR model using a translog cost/profit system that is popular in production theory. Section 3 describes the model and its estimation in detail, establishes consistency and asymptotic normality of the parametric and nonparametric estimators, suggests a more efficient nonparametric SUR estimator via the Cholesky decomposition, and provides a testing procedure for the parametric functional form. Section 4 provides some simulation results that highlight the relative efficiency gains of both our SUR estimator for the linear part and nonparametric SUR estimator over their single-equation counterparts. Section 5 illustrates the methodology with an Italian banking data set. Section 6 discusses the heteroscedasticity issue in estimating a production system. Section 7 concludes.

## 2 Motivational examples

The proposed SUR estimator for the linear part and nonparametric SUR estimator can be used to estimate many economic models. This section provides two motivational examples in production theory. If a firm's objective is to minimize cost, a partially linear cost function can be estimated, where the total cost is a nonparametric function of the $z$ variables. The cost function can be a translog function of input prices and outputs. Applying Shephard's Lemma would give us the cost share equations that are naturally linked to the cost function itself and share the same translog parameters as the cost function. Similarly, for a profit-maximizing firm, a partially linear profit function can be estimated, where profit is a nonparametric function of the $z$ variables, and the profit function can be a translog function of input and output prices. Applying Hotelling's Lemma would give us the cost share equations in profit. These share equations are, again, linked to the profit function itself and share the same translog parameters as the profit function. It would be desirable to estimate the cost or profit function with its share equations as a cost or profit system, respectively. In what follows we describe a translog cost and profit system as examples of our partially linear SUR model.

### 2.1 A translog cost system

Consider a cost function

$$\mathcal{C} = \Theta(z) \cdot c(W, O), \tag{2.1}$$

where $\mathcal{C}$ denotes total cost and $\Theta(\cdot)$ is a productivity parameter that depends on $z$—a set of environmental factors that affect the total cost in fully flexible manners. $W$ is a $J$-vector of input prices and $O$ is a $P$-vector of outputs. Imposing the restriction of homogeneous of degree one in input prices on the cost function in (2.1) and using the first input price, $W_1$, as the numeraire, we would get

$$\widetilde{\mathcal{C}} = \Theta(z) \cdot c(\widetilde{W}, O), \tag{2.2}$$

where $\widetilde{\mathcal{C}} = \mathcal{C}/W_1$ and $\widetilde{W}$ is a $(J-1)$-vector of input price ratios with $\widetilde{W}_j = W_j/W_1$ for $j = 2, \ldots, J$. Shephard's Lemma implies that

$$X_j = \partial \mathcal{C}/\partial W_j = \partial \widetilde{\mathcal{C}}/\partial \widetilde{W}_j, \tag{2.3}$$

where $X_j$ is the $j^{\text{th}}$ conditional input demand. Multiplying both sides of (2.3) by $W_j/\mathcal{C}$, or equivalently, by $\widetilde{W}_j/\widetilde{\mathcal{C}}$, gives

$$S_j = \partial \ln \mathcal{C}/\partial \ln W_j = \partial \ln \widetilde{\mathcal{C}}/\partial \ln \widetilde{W}_j, \quad \text{for } j = 2, \cdots, J, \tag{2.4}$$

where $S_j = W_j X_j/\mathcal{C}$ is the cost share of the $j^{\text{th}}$ input, and $\sum_{j=1}^{J} S_j = 1$, i.e., the sum of the cost shares equals unity. The cost system consists of the $J$ equations in (2.2) and (2.4). In fact, the share equations are derived from the cost function and provide additional information in estimating the cost function parameters. To facilitate the estimation of the system, we take the natural log of both sides of (2.2),

$$\ln \widetilde{\mathcal{C}} = \ln \Theta(z) + \ln c(\widetilde{W}, O). \tag{2.5}$$

Following Kumbhakar (1991) and Kumbhakar & Tsionas (2005), $\ln c(\widetilde{W}, O)$ can be specified by a translog function, and thus with subscript $i$, (2.5) is rewritten as

$$\begin{aligned} \ln \widetilde{\mathcal{C}}_i = {} & \theta(z_i) + \sum_{j=2}^{J} \beta_j \ln \widetilde{W}_{ji} + \sum_{p=1}^{P} \alpha_p \ln O_{pi} + \frac{1}{2} \sum_{j=2}^{J} \sum_{k=2}^{J} \beta_{jk} \ln \widetilde{W}_{ji} \ln \widetilde{W}_{ki} \\ & + \sum_{j=2}^{J} \sum_{p=1}^{P} \rho_{jp} \ln \widetilde{W}_{ji} \ln O_{pi} + \frac{1}{2} \sum_{p=1}^{P} \sum_{q=1}^{P} \alpha_{pq} \ln O_{pi} \ln O_{qi} + u_{1i}, \end{aligned} \tag{2.6}$$

where $\theta(z) \equiv \ln \Theta(z)$ is a nonparametric function of $z$, $u_1$ is a noise term, and symmetry restrictions are imposed such that $\beta_{jk} = \beta_{kj}$ for any $j, k = 2, \cdots, J$, and $\alpha_{pq} = \alpha_{qp}$ for any $p, q = 1, \cdots, P$. Rewriting (2.4) for the remaining $J-1$ equations in a similar manner, we have

$$S_{ji} = \beta_j + \sum_{k=2}^{J} \beta_{jk} \ln \widetilde{W}_{ki} + \sum_{p=1}^{P} \rho_{jp} \ln O_{pi} + u_{ji}, \quad \text{for } j = 2, \cdots, J, \tag{2.7}$$

where $u_j$ are the noise terms of the share equations. We call the $J$ equations in (2.6) and (2.7) a translog cost system.[2] This system can be estimated using the estimation procedure described in Section 3. It is worth noting that the translog cost function (2.6) and its $J-1$ share equations (2.7) have the same slope parameters—$\beta_j$, $\beta_{jk}$, and $\rho_{jp}$ for $j, k = 2, \ldots, J$ and $p = 1, \ldots, P$. This indicates that cross-equation restrictions must be imposed on the translog cost system. The parameter estimates based on this system of equations should be more accurate than those from a single partially linear translog cost function such as (2.6) because more information (i.e., cost share data) is used while no additional parameters are estimated. The cross-equation restrictions can be easily imposed by adding variables taking zero values into (2.7) such that each equation in (2.7) has the same number and the same order of right-hand-side terms as (2.6).[3]

## 2.2 A translog profit system

Consider a profit function

$$\Pi = \Theta^{\pi}(z) \cdot \pi(\mathcal{P}, W), \tag{2.8}$$

where $\Pi$ denotes profit and $\Theta^{\pi}(\cdot)$ is the profit productivity parameter that depends on the $z$ variables in fully flexible manners. $W$ is again a $J$-vector of input prices, and $\mathcal{P}$ is a scalar of output price (Kumbhakar 2001). Imposing the restriction of homogeneous of degree one in $(\mathcal{P}, W)$ on the profit function in (2.8) and using the output price as the numeraire would give us

$$\widetilde{\Pi} = \Theta^{\pi}(z) \cdot \pi(\widetilde{\mathcal{W}}), \tag{2.9}$$

[2]The first share equation, $S_1$, is dropped because the sum of the cost shares equals unity. If the $z_i$ in (2.6) is viewed as fixed, then the partially linear SUR is equivalent to a parametric SUR model. The equation-by-equation OLS estimator that we use in the first step is invariant to the choice of the numeraire, and also to the equation dropped (Chavas & Segerson 1987). Therefore, the residuals from the first-step estimation, and also the estimated variance-covariance matrix, would be invariant to the choice of the numeraire. The case of $z_i$ being viewed as stochastic and its impact on the choice of numeraire is saved for future research.

[3]For example, if we have three inputs and three outputs, we can write the parameter vector of the translog cost function, (2.6), as $(\beta_2, \beta_3, \alpha_1, \alpha_2, \alpha_3, \beta_{22}, \beta_{23}, \rho_{21}, \rho_{22}, \rho_{23}, \beta_{33}, \rho_{31}, \rho_{32}, \rho_{33}, \alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{22}, \alpha_{23}, \alpha_{33})'$, then the right-hand-side variable vector of the share equations, (2.7), can be written as: $(1, 0, 0, 0, 0, \ln \widetilde{W}_{2i}, \ln \widetilde{W}_{3i}, \ln O_{1i}, \ln O_{2i}, \ln O_{3i}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$ for $j = 2$, and $(0, 1, 0, 0, 0, 0, \ln \widetilde{W}_{2i}, 0, 0, 0, \ln \widetilde{W}_{3i}, \ln O_{1i}, \ln O_{2i}, \ln O_{3i}, 0, 0, 0, 0, 0, 0)'$ for $j = 3$. The $R$ codes of estimating the partially linear SUR model are available from the authors upon request.

where $\widetilde{\Pi} = \Pi/\mathcal{P}$ and $\widetilde{\mathcal{W}}$ is a $J$-vector of price ratios with $\widetilde{\mathcal{W}}_j = W_j/\mathcal{P}$ for $j = 1, \ldots, J$. Hotelling's Lemma implies that

$$X_j = -\partial \Pi/\partial W_j = -\partial \widetilde{\Pi}/\partial \widetilde{\mathcal{W}}_j, \tag{2.10}$$

where $X_j$ is the $j^{\text{th}}$ unconditional input demand. Multiplying both sides of (2.10) by $W_j/\Pi$, or equivalently, by $\widetilde{\mathcal{W}}_j/\widetilde{\Pi}$, gives

$$S_j^\pi = -\partial \ln \Pi/\partial \ln W_j = -\partial \ln \widetilde{\Pi}/\partial \ln \widetilde{\mathcal{W}}_j, \quad \text{for } j = 1, \cdots, J, \tag{2.11}$$

where $S_j^\pi = W_j X_j/\Pi$ is the cost share of the $j^{\text{th}}$ input in profit, and $\sum_{j=1}^J S_j^\pi = \mathcal{C}/\Pi$—the sum of the cost shares in profit equals total cost per unit of profit. The profit system consists of the $J+1$ equations in (2.9) and (2.11). Taking the natural log of both sides of (2.9) gives

$$\ln \widetilde{\Pi} = \ln \Theta^\pi(z) + \ln \pi(\widetilde{\mathcal{W}}). \tag{2.12}$$

Using the translog specification for $\ln \pi(\widetilde{\mathcal{W}})$ and adding the subscript $i$, we rewrite (2.12) as

$$\ln \widetilde{\Pi}_i = \theta^\pi(z_i) + \sum_{j=1}^J \beta_j^\pi \ln \widetilde{\mathcal{W}}_{ji} + \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^J \beta_{jk}^\pi \ln \widetilde{\mathcal{W}}_{ji} \ln \widetilde{\mathcal{W}}_{ki} + u_{0i}^\pi, \tag{2.13}$$

where $\theta^\pi(z) \equiv \ln \Theta^\pi(z)$ is a nonparametric function of $z$, $u_0^\pi$ is a noise term, and symmetry restrictions are imposed such that $\beta_{jk}^\pi = \beta_{kj}^\pi$ for any $j, k = 1, \cdots, J$. Rewriting (2.11) for all the $J$ equations, we have

$$-S_{ji}^\pi = \beta_j^\pi + \sum_{k=1}^J \beta_{jk}^\pi \ln \widetilde{\mathcal{W}}_{ki} + u_{ji}^\pi, \quad \text{for } j = 1, \cdots, J, \tag{2.14}$$

where $u_j^\pi$ are the noise terms of the share equations. Unlike the case of the translog cost system described in the previous section where the sum of the cost shares equals unity—therefore one of the share equations is dropped, the sum of the cost shares in profit does not equal unity in the profit system, and therefore there would be no need to drop any of the share equations of the profit system (Kumbhakar 2001). Finally, we call the $J+1$ equations in (2.13) and (2.14) a translog profit system, which can again be estimated using the same procedure detailed in the following section. The translog profit function (2.13) and its $J$ share equations (2.14) have the same slope parameters of $\beta_j^\pi$ and $\beta_{jk}^\pi$, for $j, k = 1, \ldots, J$. These cross-equation restrictions can be imposed in a manner that is similar to the case of the translog cost system.[4]

# 3 A partially linear SUR model

Motivated by the estimation of the translog cost/profit system, we consider in this section estimation of the following system of equations

$$y_{si} = \theta_s(z_{si}) + x_{si}'\beta_s + u_{si}, \tag{3.1}$$

where the subscript $i = 1, \ldots, n$ denotes observation and $s = 1, \ldots, m$ indexes equation. $y_{si}$ is a scalar response, $\theta_s(\cdot)$ is an unknown function of a $p_s$-vector of covariates $z_{si}$, $x_{si}$ is a $q_s$-vector of regressors in the linear parametric part with a parameter vector $\beta_s$. $u_{si}$ is a scalar error with $\mathbb{E}\left(u_{si} \mid z_{si}\right) = 0$, and with contemporaneous correlation, i.e., $\mathbb{E}\left(u_{si}u_{li}\right) = \sigma_{sl}$ and $\mathbb{E}\left(u_{si}u_{lt}\right) = 0$ for any $s$, $l$, and $i \neq t$. The goal is to obtain consistent and more efficient estimates of the parameters, $\beta_s$, and the coefficient function, $\theta_s(\cdot)$ at an arbitrary point $z \in \mathbb{R}^{p_s}$.

## 3.1 Estimation

Following Robinson (1988) and taking conditional expectation of Eq. (3.1) given $z_{si}$, we have

$$\mathbb{E}\left(y_{si} \mid z_{si}\right) = \theta_s(z_{si}) + \mathbb{E}\left(x_{si} \mid z_{si}\right)' \beta_s. \tag{3.2}$$

[4]For example, if we have three inputs and one output, we can write the parameter vector of the translog profit function, (2.13), as $(\beta_1^\pi, \beta_2^\pi, \beta_3^\pi, \beta_{11}^\pi, \beta_{12}^\pi, \beta_{13}^\pi, \beta_{22}^\pi, \beta_{23}^\pi, \beta_{33}^\pi)'$, then the right-hand-side variable vector of the share equations, (2.14), can be written as: $(1, 0, 0, \ln \widetilde{\mathcal{W}}_{1i}, \ln \widetilde{\mathcal{W}}_{2i}, \ln \widetilde{\mathcal{W}}_{3i}, 0, 0, 0)'$ for $j = 1$, $(0, 1, 0, 0, \ln \widetilde{\mathcal{W}}_{1i}, 0, \ln \widetilde{\mathcal{W}}_{2i}, \ln \widetilde{\mathcal{W}}_{3i}, 0)'$ for $j = 2$, and $(0, 0, 1, 0, 0, \ln \widetilde{\mathcal{W}}_{1i}, 0, \ln \widetilde{\mathcal{W}}_{2i}, \ln \widetilde{\mathcal{W}}_{3i})'$ for $j = 3$.

Eq. (3.2) provides a moment condition for estimating $\theta_s(\cdot)$, i.e., $\theta_s(z_{si}) = \mathbb{E}\left(y_{si} \mid z_{si}\right) - \mathbb{E}\left(x_{si} \mid z_{si}\right)' \beta_s$. Subtracting (3.2) from (3.1) gives

$$y^*_{si} = x^{*'}_{si}\beta_s + u_{si}, \tag{3.3}$$

where $y^*_{si} \equiv y_{si} - \mathbb{E}\left(y_{si} \mid z_{si}\right)$ and $x^*_{si} \equiv x_{si} - \mathbb{E}\left(x_{si} \mid z_{si}\right)$. This system of equations can be viewed as Zellner (1962) SUR model. Let $g_{sy}(z) \equiv \mathbb{E}\left(y_{si} \mid z_{si} = z\right)$ and $g_{sx}(z) \equiv \mathbb{E}\left(x_{si} \mid z_{si} = z\right)$. Note that $y^*_{si}$ and $x^*_{si}$ are not observable due to unknown conditional means $g_{sy}(z_{si})$ and $g_{sx}(z_{si})$, respectively. To estimate them, we hereby use the simple Nadaraya-Watson estimators

$$\widehat{g}_{sy}(z) = \frac{\sum_{i=1}^n K_s\left(\frac{z_{si}-z}{h_s}\right) y_{si}}{\sum_{i=1}^n K_s\left(\frac{z_{si}-z}{h_s}\right)} \quad \text{and} \quad \widehat{g}_{sx}(z) = \frac{\sum_{i=1}^n K_s\left(\frac{z_{si}-z}{h_s}\right) x_{si}}{\sum_{i=1}^n K_s\left(\frac{z_{si}-z}{h_s}\right)}, \tag{3.4}$$

where $K_s(\cdot)$ is a product kernel function and $h_s$ is the associated bandwidth. Estimates of $y^*_{si}$ and $x^*_{si}$ follow naturally as $\widehat{y}^*_{si} = y_{si} - \widehat{g}_{sy}(z_{si})$ and $\widehat{x}^*_{si} = x_{si} - \widehat{g}_{sx}(z_{si})$, respectively.

We first ignore cross-equation correlation and estimate $\beta_s$ equation by equation using the ordinary least squares (OLS). The zero conditional mean, $\mathbb{E}\left(u_{si} \mid z_{si}, x_{si}\right) = 0$, leads to Robinson's single-equation estimator for $\beta_s$ using the estimated $\widehat{y}^*_{si}$ and $\widehat{x}^*_{si}$, i.e.,

$$\widehat{\beta}_s = \left(\sum_{i=1}^n \widehat{x}^*_{si}\widehat{x}^{*'}_{si}\right)^{-1} \sum_{i=1}^n \widehat{x}^*_{si}\widehat{y}^*_{si}.$$

With $\widehat{\beta}_s$, we have a pilot estimator for $\theta_s(z)$ as $\widehat{\theta}_s(z) = \widehat{g}_{sy}(z) - \widehat{g}_{sx}(z)'\widehat{\beta}_s$. That is, no additional regression is required to recover the nonparametric component. We call $\widehat{\theta}_s(z)$ Robinson single-equation nonparametric estimator. Let $\widehat{u}_{si} = \widehat{y}^*_{si} - \widehat{x}^{*'}_{si}\widehat{\beta}_s$, and the contemporaneous cross-equation correlation $\sigma_{sl}$ is then estimated by $\widehat{\sigma}_{sl} = \frac{1}{n}\sum_{i=1}^n \widehat{u}_{si}\widehat{u}_{li}$ for any $s, l = 1, \ldots, m$.

To estimate $\beta_s$ using the SUR system, we stack Eq. (3.3) by observation for all the $m$ equations such that

$$y^*_i = x^{*\prime}_i\beta + u_i,$$

where $y_i^* \equiv (y_{1i}^*, y_{2i}^*, \ldots, y_{mi}^*)'$, $u_i \equiv (u_{1i}, u_{2i}, \ldots, u_{mi})'$, $x_i^* = \begin{pmatrix} x_{1i}^* & 0 & \cdots & 0 \\ 0 & x_{2i}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{mi}^* \end{pmatrix}$ is a block diagonal matrix of dimension $q \times m$ with $q \equiv \sum_{s=1}^{m} q_s$, and $\beta = (\beta_1', \cdots, \beta_m')'$ is a $q$-vector of coefficients. Under a stricter zero conditional mean assumption, i.e., $\mathbb{E}\left(u_i \mid z_i, x_i\right) = 0$, and by a spherical transformation, a moment condition for $\beta$ is

$$\beta = \left(\mathbb{E}\left(x_i^* \Sigma_m^{-1} x_i^{*\prime}\right)\right)^{-1} \mathbb{E}\left(x_i^* \Sigma_m^{-1} y_i^*\right),$$

where $\Sigma_m \equiv \mathrm{Var}\left(u_i\right) = \{\sigma_{sl}\}_{s=1,l=1}^{m,m}$ is the $m \times m$ variance-covariance matrix with a typical element $\sigma_{sl}$.

If $\{x_i^*, y_i^*\}_{i=1}^n$ were observed, the well-known GLS and Zellner (1962) SUR estimators can be constructed by

$$\beta_{\mathrm{gls}} = \left(\sum_{i=1}^{n} x_i^* \Sigma_m^{-1} x_i^{*\prime}\right)^{-1} \left(\sum_{i=1}^{n} x_i^* \Sigma_m^{-1} y_i^*\right), \tag{3.5}$$

and

$$\beta_{\mathrm{sur}} = \left(\sum_{i=1}^{n} x_i^* \widehat{\Sigma}_m^{-1} x_i^{*\prime}\right)^{-1} \left(\sum_{i=1}^{n} x_i^* \widehat{\Sigma}_m^{-1} y_i^*\right),$$

respectively, where $\widehat{\Sigma}_m$ is an estimate for $\Sigma_m$ constructed using $\widehat{\sigma}_{sl}$ in the first step. Replacing $x_i^*$ and $y_i^*$ with their corresponding nonparametric estimates $\widehat{x}_i^*$ and $\widehat{y}_i^*$, we have a feasible SUR estimator for $\beta$ as

$$\begin{aligned} \widehat{\beta}_{\mathrm{sur}} &= \left(\sum_{i=1}^{n} \widehat{x}_i^* \widehat{\Sigma}_m^{-1} \widehat{x}_i^{*\prime}\right)^{-1} \left(\sum_{i=1}^{n} \widehat{x}_i^* \widehat{\Sigma}_m^{-1} \widehat{y}_i^*\right) \\ &= \left(\widehat{x}^{*\prime} \left(I_n \otimes \widehat{\Sigma}_m^{-1}\right) \widehat{x}^*\right)^{-1} \left(\widehat{x}^{*\prime} \left(I_n \otimes \widehat{\Sigma}_m^{-1}\right) \widehat{y}^*\right), \end{aligned}$$

where $\widehat{x}^* = (\widehat{x}_1^*, \widehat{x}_2^*, \cdots, \widehat{x}_n^*)'$ is an $mn \times q$ stacked data matrix and $\widehat{y}^* = (\widehat{y}_1^{*\prime}, \widehat{y}_2^{*\prime}, \cdots, \widehat{y}_n^{*\prime})'$ is an $mn$-vector. To follow the conventional method of stacking data by equation, rather than by observation, $\widehat{\beta}_{\text{sur}}$ can be equivalently written as

$$\widehat{\beta}_{\text{sur}} = \left(\widehat{X}^{*\prime}\left(\widehat{\Sigma}_m^{-1}\otimes I_n\right)\widehat{X}^{*}\right)^{-1}\left(\widehat{X}^{*\prime}\left(\widehat{\Sigma}_m^{-1}\otimes I_n\right)\widehat{Y}^{*}\right), \tag{3.6}$$

where $\widehat{X}^{*} = \begin{pmatrix} \widehat{X}_1^{*} & 0 & \cdots & 0 \\ 0 & \widehat{X}_2^{*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{X}_m^{*} \end{pmatrix}$ is an $mn \times q$ stacked block diagonal matrix with a typical $n \times q_s$ element $\widehat{X}_s^{*} = (\widehat{x}_{s1}^{*}, \cdots, \widehat{x}_{sn}^{*})'$ and $\widehat{Y}^{*} = (\widehat{Y}_1^{*\prime}, \widehat{Y}_2^{*\prime}, \cdots, \widehat{Y}_m^{*\prime})'$ is an $mn$-vector with a typical $n \times 1$ element $\widehat{Y}_s^{*} = (\widehat{y}_{s1}^{*}, \cdots, \widehat{y}_{sn}^{*})'$. With $\widehat{\beta}_{\text{sur}}$, a natural estimator for the nonparametric component, $\theta_s(z)$, can be obtained as $\widetilde{\theta}_s(z) = \widehat{g}_{sy}(z) - \widehat{g}_{sx}(z)'\widehat{\beta}_{s,\text{sur}}$, where $\widehat{\beta}_{s,\text{sur}}$ is the corresponding $q_s \times 1$ vector component in $\widehat{\beta}_{\text{sur}}$ for the $s^{\text{th}}$ equation. We call $\widetilde{\theta}_s(z)$ two-step nonparametric estimator.

## 3.2 Asymptotic properties of $\widehat{\beta}_{\text{sur}}$ and $\widetilde{\theta}_s(\cdot)$

Let $\beta_{s,\text{sur}}$ and $\beta_{s,\text{gls}}$ be defined in a similar manner as $\widehat{\beta}_{s,\text{sur}}$. From Zellner (1962), we know that as long as $(\widehat{\Sigma}_m^{-1} - \Sigma_m^{-1})$ has a parametric $\sqrt{n}$ convergence rate, $\beta_{s,\text{sur}}$ is more efficient than the OLS estimator which only uses information based on the $s^{\text{th}}$ equation, and $\beta_{\text{sur}}$ has an asymptotic normal distribution as $\sqrt{n}\,(\beta_{\text{sur}} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\beta)$, where $\mathcal{V}_\beta \equiv (\mathbb{E}\,(x_i^{*}\Sigma_m^{-1}x_i^{*\prime}))^{-1}$.[5] We find that using estimated $x_i^{*}$ and $y_i^{*}$ in $\beta_{\text{sur}}$ does not affect its asymptotic property as long as we choose suitable bandwidths.[6] We establish the asymptotic equivalence between $\widehat{\beta}_{\text{sur}}$ and $\beta_{\text{sur}}$ by showing that $\sqrt{n}\left(\widehat{\beta}_{\text{sur}} - \beta_{\text{sur}}\right) = o_p(1)$ in Appendix A. To do that, we first lay out a set of assumptions. In what follows, $C > 0$ denotes a generic constant and $C^r$ denotes the class of functions such that i) each of its elements is $r$-times partially continuously differentiable, and ii) all their partial derivatives up to order $r$ are uniformly bounded.

**Assumption A1.** a) The random sequence $\{(x_i', z_i', u_i')'\}_{i=1}^{n}$ is an independent and identically distributed (IID) process and their relationship can be described by (3.1). b) The $m$-

[5] The asymptotic distribution of $\beta_{s,\text{sur}}$ can be obtained easily from the joint one of $\beta_{\text{sur}}$. It is normal with asymptotic covariance being the corresponding $s^{\text{th}}$ $q_s \times q_s$ diagonal block of $\mathcal{V}_\beta$.

[6] We need to undersmooth in the first step so that the nonparametric bias term becomes negligible when it comes to estimating the parametric coefficients $\beta$.

vector $u_i$ has zero conditional mean and constant conditional covariance, i.e., $\mathbb{E}(u_i \mid z_i, x_i) = 0$ and $\mathbb{E}(u_i^2 \mid z_i, x_i) = \Sigma_m = \{\sigma_{sl}\}_{s,l=1}^{m,m}$; $\Sigma_m$ is positive definite. c) $\sigma_{sx,k}^2 \equiv \mathbb{E}\left(x_{si,k}^{*2} \mid z_i\right) \leq C$ for any $k = 1, \cdots, q_s$ and $\sigma_{sy}^2 \equiv \mathbb{E}(y_{si}^{*2} \mid z_i) \leq C$.

**Assumption A2.** The kernel $K_s$ satisfies $K_s(z) = \prod_{i=1}^{p_s} k_s(z_i)$ where $z = (z_1, \cdots, z_{p_s})$. $k_s$ is symmetric about zero and satisfies: a) $\int k_s(z)\,\mathrm{d}z = 1$; b) $k_s$ is a kernel of order $r_s$, i.e., $\int k_s(z) z^i\,\mathrm{d}z = 0$ for $i = 1, \cdots, r_s - 1$, and $\int |k_s(z) z^{r_s}|\,\mathrm{d}z \leq C$.

**Assumption A3.** a) The density function of $z_{si}$, $f_{sz}(\cdot)$, is uniformly bounded away from zero and infinity. b) The functions $f_{sz}(\cdot)$, $g_{sx}(\cdot)$, and $g_{sy}(\cdot)$ all belong to $C^{r_s+1}$.

**Assumption A4.** Denote $L_{sn} \equiv \left(\frac{\log n}{n h_s^{p_s}}\right)^{1/2} + h_s^{r_s}$ where $h_s = n^{-\frac{1}{2r_s+p_s}} \to 0$ as $n \to \infty$ with $r_s > p_s/2$.

In Assumption A2, we use a high-order kernel, and together with appropriate smoothness of the nonparametric functions in A3 b), we achieve bias reduction for the nonparametric estimators. With a suitable bandwidth choice, those biases become negligible when the asymptotic normality of $\widehat{\beta}_{\text{sur}}$ is obtained.[7] For practical uses, such high-order kernels can be easily constructed using methods, for example, in Geng et al. (2020, Eq. 10). Theorem 1 below establishes the square-root $n$ asymptotic normality of $\widehat{\beta}_{\text{sur}}$ with proof in Appendix A.

**Theorem 1.** *Under Assumptions A1–A4, we have*

$$\sqrt{n}\left(\widehat{\beta}_{\text{sur}} - \beta\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\beta),$$

*where* $\mathcal{V}_\beta = \left(\mathbb{E}\left(x_i^* \Sigma_m^{-1} x_i^{*\prime}\right)\right)^{-1}$.

**Remark 1.** *The efficiency gain of $\widehat{\beta}_{\text{sur}}$ that takes advantage of cross-equation correlation relative to one that does not, for example, the Robinson (1988) estimator, is easy to establish. Given that $\widehat{\beta}_s$ is the single-equation linear estimator, by Robinson (1988) its asymptotic covariance matrix is $\mathcal{V}_{\beta,\mathrm{R}} \equiv \left(\mathbb{E}\left(x_i^* x_i^{*\prime}\right)\right)^{-1} \mathbb{E}\left(x_i^* \Sigma_m x_i^{*\prime}\right) \left(\mathbb{E}\left(x_i^* x_i^{*\prime}\right)\right)^{-1}$. By the matrix form of Cauchy-Schwarz inequality, it is straightforward that $\mathbb{E}\left(x_i^* x_i^{*\prime}\right) \left(\mathbb{E}\left(x_i^* \Sigma_m^{-1} x_i^{*\prime}\right)\right)^{-1} \mathbb{E}\left(x_i^* x_i^{*\prime}\right) \leq \mathbb{E}\left(x_i^* \Sigma_m x_i^{*\prime}\right)$ which leads to the desired result $\mathcal{V}_\beta \leq \mathcal{V}_{\beta,\mathrm{R}}$.*

[7] It can be shown that if the dimension of $z_{si}$ is smaller than four, an ordinary kernel of order two would suffice.

**Remark 2.** *Estimator for the covariance matrix $\mathcal{V}_\beta$ is critical for statistical inference and hypothesis testing. A consistent estimator can be easily constructed by $\widehat{\mathcal{V}}_\beta \equiv \left(\sum_{i=1}^n \widehat{x}_i^* \widehat{\Sigma}_m^{-1} \widehat{x}_i^{*\prime}\right)^{-1}$ where the square root of the diagonal elements give the corresponding standard errors of each element in $\widehat{\beta}_{\text{sur}}$.*

The next theorem establishes the asymptotic normality of the nonparametric estimator $\widetilde{\theta}_s(\cdot)$ with proof in Appendix A.

**Theorem 2.** *Let $\mu_{k_s,r_s} \equiv \int k_s(z) z^{r_s}\, \mathrm{d}z$, $\mathrm{D}_j^k f(z) \equiv \frac{\partial^k}{\partial_j^k} f(z)$, and $\mathrm{D}_j^0 f(z) \equiv f(z)$ for any $k \geq 1$ and $1 \leq j \leq k$. Under Assumption A1–A4 and assuming that $\mathbb{E}\left(|u_{si}|^{2+\delta} \mid z_{si}, x_{si}\right) \leq C$ for some $\delta > 0$, we have*

$$\sqrt{nh_s^{p_s}} \left(\widetilde{\theta}_s(z) - \theta_s(z) - b_{s,1}(z)\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{V}_{s,1}(z)\right),$$

*where $b_{s,1}(z) \equiv h_s^{p_s} f_{sz}^{-1}(z) \mu_{k_s,r_s} \sum_{k=1}^{r_s} \frac{1}{k!(r_s-k)} \sum_{j=1}^{p_s} \mathrm{D}_j^k \theta_s(z) \mathrm{D}_j^{r_s-k} f_{sz}(z) + o_p(h_s^{p_s})$ and $\mathcal{V}_{s,1}(z) \equiv f_{sz}^{-1}(z) \sigma_{ss} \int K_s^2(\gamma)\, \mathrm{d}\gamma$.*

**Remark 3.** *Given the order of the bias, it follows immediately from Theorem 2 that $\widetilde{\theta}_s(z)$ is consistent for $\theta_s(z)$.*

**Remark 4.** *It is clear from the proof of Theorem 2 that $\widetilde{\theta}_s(z)$ is asymptotically equivalent to Robinson's nonparametric estimator $\widehat{\theta}_s(z)$, and they have the same asymptotic distribution. Let $\ddot{\theta}_s(z, \ddot{\beta}_s) \equiv \widehat{g}_{sy}(z) - \widehat{g}_{sx}(z)' \ddot{\beta}_s$ denote a general nonparametric estimator for $\theta_s(z)$ which is constructed using a linear estimator $\ddot{\beta}_s$. As long as $\ddot{\beta}_s$ has a convergence rate faster than the nonparametric rate, i.e., $\sqrt{nh_s^{p_s}}(\ddot{\beta}_s - \beta_s) = o_p(1)$, $\ddot{\theta}_s(z, \ddot{\beta}_s)$ would have the asymptotic normality given in Theorem 2. This indicates that cross-equation correlation might not be effectively used in our current setup for improving estimation efficiency of the nonparametric component. From the simulations given in Section 4, we show that it is indeed the case in terms of finite sample performance.*

### 3.3 More efficient nonparametric estimation

In this section, we propose a potentially more efficient nonparametric estimator by taking cross-equation correlation into account. The idea is basically in the same vein as in Su et al.

(2013) who extended Martins-Filho & Yao (2009) and improved efficiency in estimating a nonparametric model with a general parametric error covariance. In our case, after a spherical transformation of the error vector for the whole system of equations, we construct a new regressand for the nonparametric regression taking the linear part as given. We show that an estimator that accounts for cross-equation correlation is efficient relative to those estimators that do not, e.g., $\widehat{\theta}_s(z)$ or $\widetilde{\theta}_s(z)$.[8]

We first stack Eq. (3.3) for all the $n$ observations and rearrange, giving

$$Y_s - X_s\beta_s = \Theta_s(Z_s) + U_s,$$

where $Y_s = (y_{s1}, \cdots, y_{sn})'$, $X_s = (x_{s1}, \cdots, x_{sn})'$, $Z_s$ and $U_s$ are defined analogously as $Y_s$, and $\Theta_s(Z_s) = (\theta_s(z_{s1}), \cdots, \theta_s(z_{sn}))'$. Then stacking all the $m$ equations, we have

$$Y - X\beta = \Theta(Z) + U, \tag{3.7}$$

where $Y$ and $X$ are defined analogously as $\widehat{Y}^*$ and $\widehat{X}^*$ in Eq. (3.6), respectively, $Z = (Z_1, \cdots, Z_m)$, $U = (U_1', \cdots, U_m')'$, and $\Theta(Z) = (\Theta_1(Z_1)', \cdots, \Theta_m(Z_m))'$. It is easy to see that $\Sigma \equiv \mathbb{E}(UU') = \Sigma_m \otimes I_n$. Given that $\Sigma_m$ is positive definite, there exists a unique lower triangular matrix $P_m$ such that $\Sigma_m = P_m P_m'$, the Cholesky decomposition of $\Sigma_m$. Then we have $\Sigma = PP'$ where $P = P_m \otimes I_n$. Denote $V = P^{-1}$ and $V_m = P_m^{-1}$, then $V = V_m \otimes I_n$. Let $\mathcal{E} = (\mathcal{E}_1', \cdots, \mathcal{E}_m')' \equiv VU$ such that

$$\begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_m \end{pmatrix} = \begin{pmatrix} v_{11}I_n & 0_n & \cdots & 0_n \\ v_{21}I_n & v_{22}I_n & \cdots & 0_n \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1}I_n & v_{m2}I_n & \cdots & v_{mm}I_n \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{pmatrix},$$

where $\mathcal{E}_s = \sum_{l=1}^{s} v_{sl}U_l$ with a typical element $\varepsilon_{si} = \sum_{l=1}^{s} v_{sl}u_{li}$. $\mathcal{E}$ is spherical as $\mathbb{E}(\mathcal{E}\mathcal{E}') =$

[8] It is well known that the efficiency of the GLS estimator for fully linear models is invariant to the covariance decomposition method since the estimator only depends on the inverse of the covariance matrix rather than on its square root; see, e.g., Eq. (3.5). More interestingly, we find that how to decompose the error covariance matrix might play a role in the extent to which efficiency of the nonparametric estimation improves; see Remark 6.

$\mathbb{E}(VUU'V) = VPP'V' = I_{mn}$, and $\mathbb{E}(\mathcal{E}_s\mathcal{E}_s') = I_n$ follows immediately. Denote $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_m)' \equiv Y - X\beta = \Theta(Z) + U$ where $\mathbf{Y}_s = Y_s - X_s\beta_s$ with a typical element $\mathbf{y}_{si} = y_{si} - x_{si}'\beta_s$. Now taking $\beta$ as given and following Su et al. (2013), we define a new regressand vector for the $s^{\text{th}}$ equation, i.e., $\mathbf{Y}_s^* \equiv v_{ss}\Theta_s(Z_s) + \mathcal{E}_s$,[9] and conduct a local-linear nonparametric regression of a feasible regressand $\widehat{\mathbf{y}}_{si}^*/\widehat{v}_{ss}$ on $z_{si}$[10] which leads to our nonparametric SUR estimator

$$\widetilde{\theta}_{s,\text{sur}}(z) = e_s(0_{p_s})' \left[M_{s,z}'\mathcal{K}_{s,z}M_{s,z}\right]^{-1} M_{s,z}'\mathcal{K}_{s,z}\frac{1}{\widehat{v}_{ss}}\widehat{\mathbf{Y}}_s^*, \tag{3.8}$$

where $\widehat{\mathbf{Y}}_s^* = (\widehat{\mathbf{y}}_{s1}^*, \cdots, \widehat{\mathbf{y}}_{sn}^*)' \equiv \widehat{v}_{ss}\widehat{\Theta}_s(Z_s) + \widehat{\mathcal{E}}_s$, $\widehat{\mathcal{E}}_s \equiv \sum_{l=1}^{s} \widehat{v}_{sl}\widehat{U}_l$ with $\widehat{U}_s = (\widehat{u}_{s1}, \cdots, \widehat{u}_{sn})'$ for $s = 1, \cdots, m$, $\widehat{\Theta}_s(Z_s) = \left(\widehat{\theta}_s(z_{s1}), \cdots, \widehat{\theta}_s(z_{sn})\right)'$, $\widehat{v}_{sl}$ is a typical $(s,l)^{\text{th}}$ element of $\widehat{V}_m = \widehat{P}_m^{-1}$ with $\widehat{P}_m$ being the Cholesky factor of $\widehat{\Sigma}_m$ for any $s, l = 1, \cdots, m$, $e_s(x) \equiv (1, x')' : \mathbb{R}^{p_s} \to \mathbb{R}^{p_s+1}$, $0_{p_s}$ is a $p_s \times 1$ vector of zeros, $M_{s,z} \equiv (M_{s1}(z), \cdots, M_{sn}(z))'$ with $M_{si}(z) \equiv \left(1, \frac{1}{h_{2s}}(z_{si} - z)\right)'$ and $h_{2s}$ the associated bandwidth, and $\mathcal{K}_{s,z} \equiv \text{diag}\{K_{si}(z)\}_{i=1}^n$ with $K_{si}(z) \equiv \frac{1}{h_{2s}^{p_s}}K_s\left(\frac{1}{h_{2s}}(z_{si} - z)\right)$. Denote the bandwidth in the pilot estimator $\widehat{\theta}_s(z)$ by $h_{1s}$. The asymptotic normality of $\widetilde{\theta}_{s,\text{sur}}(z)$ is given in the following theorem. It can be established in a similar manner as in Su et al. (2013), and therefore is omitted to save space.

**Theorem 3.** *Under Assumptions A1–A3, and if for any $s, l = 1, \cdots, m$, $h_{1s}/h_{2l} \to 0$, $nh_{1s}^{p_s} \to \infty$ and $nh_{2s}^{p_s+2r_s} \to C \in [0, \infty]$ as $n \to \infty$, we have*

$$\sqrt{nh_{2s}^{p_s}}\left(\widetilde{\theta}_{s,\text{sur}}(z) - \theta_s(z) - b_{s,2}(z)\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{V}_{s,2}(z)\right), \tag{3.9}$$

*where $b_{s,2}(z) \equiv \frac{1}{r_s}h_{2s}^{r_s}\mu_{k_s,r_s}\sum_{\iota=1}^{p_s} \mathrm{D}_{\iota\iota}^{r_s}\theta_s(z) + o_p(h_{2s}^{r_s})$ and $\mathcal{V}_{s,2}(z) \equiv f_{sz}^{-1}(z)v_{ss}^{-2}\int K_s^2(\gamma)\,\mathrm{d}\gamma$.*

**Remark 5.** *The relative efficiency gain of $\widetilde{\theta}_{s,\text{sur}}(z)$ to $\widehat{\theta}_s(z)$ or $\widetilde{\theta}_s(z)$ can be easily verified by comparing $\mathcal{V}_{s,2}$ and $\mathcal{V}_{s,1}$. Denote a typical element in $P_m$ by $p_{sl}$ for $s, l = 1, \cdots, m$. Given $\Sigma_m = P_mP_m'$, we have $\sigma_{ss} = \sum_{l=1}^{s} p_{sl}^2 \geq p_{ss}^2 = v_{ss}^{-2}$ which gives the desired result $\mathcal{V}_{s,2} \leq \mathcal{V}_{s,1}$.*

**Remark 6.** *Note that the construction of this second-stage nonparametric SUR estimator*

[9] This originates from a spherical transformation of Eq. (3.7) where $\mathbf{Y}^* = (\mathbf{Y}_1^{*\prime}, \cdots, \mathbf{Y}_m^{*\prime})' \equiv H \cdot \Theta(Z) + VU = V\mathbf{Y} + (H - V) \cdot \Theta(Z)$ and $H$ is a diagonal matrix with the same diagonal elements as $V$; see Su et al. (2013) for more details.

[10] It can be easily shown that a local-polynomial kernel estimation of equation $\mathbf{y}_{si}^* = v_{ss}\theta_s(z_{si}) + \varepsilon_{si}$ can be achieved by treating $\mathbf{y}_{si}^*/v_{ss}$ as the regressand.

*$\widetilde{\theta}_{s,\mathrm{sur}}(z)$ in Eq. (3.8) depends on the choice of square root for $\Sigma_m$ via elements of $\widehat{V}$, the inverse of the square root of $\widehat{\Sigma}_m$. The other square root for $\Sigma_m$ that is induced by the Spectral decomposition ($\Sigma_m = B_m B_m$ where $B_m = B_m'$) does also work in our context. An interesting finding here is that the asymptotic variance of $\widetilde{\theta}_{s,\mathrm{sur}}(z)$ and thus its efficiency depend on such choice via $v_{ss}$ in $\mathcal{V}_{s,2}(z)$. Conventionally, researchers either simply overlook this dependency and thus distinctions between these two methods inducing the square root (see Martins-Filho & Yao 2009; Su et al. 2013); or simply choose one of them without discussing further the impact of such choice (see Linton et al. 2004). A more detailed investigation into it seems necessary and we leave it as future work.*

### 3.4 Bandwidth selection

In (3.4) we assume the same bandwidth $h_s$ when estimating $g_{sy}(\cdot)$ and $g_{sx}(\cdot)$ simply for ease of theoretical proof. In practice, we choose the data-driven least squares cross-validation (LSCV) method, rather than specifying them in an *ad hoc* manner. That is, for each $s$, we select bandwidths $h_{sy}$ to minimize

$$\sum_{i=1}^{n} \left( y_{si} - \widehat{g}_{sy}^{(-i)}(z_{si}) \right)^2, \tag{3.10}$$

where $\widehat{g}_{sy}^{(-i)}(z_{si}) = \sum_{j \neq i} K_s \left( \frac{z_{sj} - z_{si}}{h_{sy}} \right) y_{sj} \big/ \sum_{j \neq i} K_s \left( \frac{z_{sj} - z_{si}}{h_{sy}} \right)$ is the leave-one-out kernel conditional mean. This is obtained by evaluating $g_{sy}(\cdot)$ at $z_{si}$ using all the information except that about $i$. The $j^{\mathrm{th}}$ element of the bandwidths $h_{sx}$ is selected similarly by replacing $y_{si}$ and $h_{sy}$ in (3.10) with the $j^{\mathrm{th}}$ element of $x_{si}$ and $h_{sx}$, respectively, for $j = 1, \cdots, q_s$.

The bandwidth vector, $h_{2s}$, of the nonparametric SUR estimation in Section 3.3 can be obtained in a similar manner. That is, for each $s$, we minimize

$$\sum_{i=1}^{n} \left( \widehat{\mathbf{y}}_{si}^{*} / \widehat{v}_{ss} - \widetilde{\theta}_{s,\mathrm{sur}}^{(-i)}(z_{si}) \right)^2, \tag{3.11}$$

where $\widetilde{\theta}_{s,\mathrm{sur}}^{(-i)}(z_{si}) = e_s(0_{p_s})' \left( \sum_{j \neq i}^{n} M_{sj}(z_{si}) M_{sj}(z_{si})' K_s \left( \frac{z_{sj} - z_{si}}{h_{2s}} \right) \right)^{-1}$ $\sum_{j \neq i}^{n} M_{sj}(z_{si}) (\widehat{\mathbf{y}}_{sj}^{*} / \widehat{v}_{ss}) K_s \left( \frac{z_{sj} - z_{si}}{h_{2s}} \right)$ is the leave-one-out estimator of $\theta_s(z_{si})$.

### 3.5 Test for parametric functional form

If the functional form of $\theta_s(z_{si})$ in Eq. (3.1) were known, a standard parametric SUR model would give more efficient estimates than the partially linear counterpart. The parametric SUR model can be written as

$$y_{si} = \theta_s(z_{si}, \delta_s) + x'_{si}\beta^o_s + u^o_{si},$$

where $\theta_s(z_{si}, \delta_s) = z'_{si}\delta_s$ is a linear parametric function with a $p_s$-vector of parameters, $\delta_s$. In the case that this parametric model is misspecified, any parametric estimator based on it would be inconsistent. Therefore, it would be interesting to test the parametric SUR against its partially linear counterpart.

To do this, we follow Cai et al. (2000) goodness-of-fit test procedure and compare the RSS's from the parametric and the semiparametric fittings. The null hypothesis is $H_0$ : $\theta_s(z_{si}) = z'_{si}\delta_s$ for any $s$. Let $\{\widehat{u}^o_{si}\}^{m,n}_{s=1,i=1}$ be the residuals from the standard parametric SUR regression (Zellner 1962) where $\widehat{u}^o_{si} = y_{si} - z'_{si}\widehat{\delta}_s - x'_{si}\widehat{\beta}^o_s$. The RSS of the parametric SUR is calculated as $RSS_0 = \frac{1}{mn}\widehat{u}^{o\prime}\widehat{u}^o$ where $\widehat{u}^o = (\widehat{u}^{o\prime}_1, \widehat{u}^{o\prime}_2, \cdots, \widehat{u}^{o\prime}_n)'$ and $\widehat{u}^o_i = (\widehat{u}^o_{1i}, \widehat{u}^o_{2i}, \cdots, \widehat{u}^o_{mi})'$. Given that the residuals of the partially linear SUR are $\widehat{u}_i = \widehat{y}^*_i - \widehat{x}^{*\prime}_i\widehat{\beta}_{\text{sur}}$, the RSS of the partially linear SUR is calculated as $RSS_1 = \frac{1}{mn}\widehat{u}'\widehat{u}$ where $\widehat{u} = (\widehat{u}'_1, \widehat{u}'_2, \cdots, \widehat{u}'_n)'$.

The test statistic can then be constructed by $T_n = RSS_0/RSS_1 - 1$. If the parametric model is close enough to its partially linear counterpart, then $T_n$ will be close to zero; otherwise it will be greater than zero. Therefore, the test is one-sided. In what follows, we propose a cluster bootstrap procedure to obtain the $p$-value of the test and determine whether to reject the null hypothesis or not with the following steps.

Step 1: Re-center the residuals from the partially linear SUR model and obtain $\{\widehat{u}_i - \bar{\widehat{u}}\}^n_{i=1}$, where $\bar{\widehat{u}} = \frac{1}{n}\sum^n_{i=1}\widehat{u}_i$, $\forall s$.[11]

Step 2: Generate bootstrapped residuals, $u^b_i$, by re-sampling the re-centered residuals with replacement, $\forall i$.[12]

[11] We follow Cai et al. (2000) and bootstrap the residuals from the partially linear SUR model because its residuals are consistent under both the null and the alternative hypotheses.

[12] For the cluster bootstrap, each $i$ is viewed as a cluster and the re-sampling is applied to all the $m$ elements of the cluster. The number of clusters $n$ goes to infinity, while the size of each cluster $m$ is finite. See Cameron & Trivedi (2005, Chapter 11) for more details about the bootstrap methods.

Step 3: Generate $y_i^b = z_i'\widehat{\delta} + x_i'\widehat{\beta}^o + u_i^b$. That is, the bootstrapped residuals, $u_i^b$, are added to the fitted values under the null to generate the bootstrapped dependent variable.

Step 4: Use the bootstrapped sample $\{y_i^b, x_i, z_i\}_{i=1}^n$ to calculate the bootstrapped test statistic $T_n^b$.

Step 5: Repeat Step 2–4 for a large number of times, say, $B = 399$, and the $p$-value is calculated as the mean frequency of the event $\{T_n^b > T_n\}$, i.e., $\frac{1}{B}\sum_{b=1}^B \mathbf{I}(T_n^b > T_n)$, where $\mathbf{I}(\cdot)$ is an indicator function that equals one if its argument is true. The null hypothesis can be rejected if the $p$-value is less than a pre-specified level of significance.

# 4 Simulations

In Section 3, we show that our SUR estimator for the linear part is asymptotically efficient relative to Robinson (1988) estimator, and that our nonparametric SUR estimator is asymptotically more efficient than the Robinson single-equation and two-step nonparametric estimators by taking into account cross-equation correlation. In this section, we conduct some simulations and compare the finite sample performances of these estimators for a two-equation system with a data generating process (DGP):

$$
\begin{aligned}
y_{1i} &= \theta_1(z_{1i}) + \beta_1 x_{1i} + u_{1i}, \\
y_{2i} &= \theta_2(z_{2i}) + \beta_2 x_{2i} + u_{2i},
\end{aligned} \tag{4.1}
$$

where $i = 1, \ldots, n$. $z_{1i}$ and $z_{2i}$ are both generated as i.i.d. $\mathcal{U}[0, 2]$. We set $\theta_1(z_{1i}) = \sin(z_{1i})$, $\theta_2(z_{2i}) = \cos(z_{2i})$, $\beta_1 = 1$, and $\beta_2 = 2$. $x_{si} = \varrho z_{si} + e_{si}$, where $\varrho = 0.6$ captures the degree of correlation between $x_{si}$ and $z_{si}$, and $e_{si} \sim$ i.i.d. $\mathcal{N}(1, 0.5^2)$, $\forall s = 1, 2$. $(u_{1i}, u_{2i})' \sim$ i.i.d. multivariate normal $\mathcal{N}(0, \Omega)$ with $\Omega = \{\sigma_{sl}\}_{s,l=1}^{2,2}$. Here we let $\sigma_{11} = \sigma_{22} = 1$ and consider cross-equation correlation such that $\sigma_{12} = \sigma_{21} = 0.6$.[13] Sample size $n$ is set at $100, 200, 400, 800$, and $1600$. Bandwidths in all cases are selected via the LSCV method as discussed in Section 3.4.

For a parametric component estimator, we report its bias, variance (var), and mean

[13] We also consider the no cross-equation correlation case, i.e., $\sigma_{12} = \sigma_{21} = 0$, and find that for all sample sizes, the parametric and nonparametric component estimators have similar MSEs and AMSEs, respectively, and therefore results are omitted to save space, but are available upon request.

squared error (MSE) over $R = 1000$ Monte Carlo replications. For a nonparametric component estimator, we report its average bias (Abias), average variance (Avariance or Avar), and average MSE (AMSE) over the $R$ replications. Specifically,

$$\text{Abias} = \frac{1}{R}\sum_{\gamma=1}^{R}\left(\text{bias}_{\theta_s}^{(\gamma)}\right) = \frac{1}{R}\sum_{\gamma=1}^{R}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_s^{(\gamma)}(z_{si}) - \theta_s^{(\gamma)}(z_{si})\right)\right),$$

where $\widehat{\theta}_s^{(\gamma)}(z_{si})$ and $\theta_s^{(\gamma)}(z_{si})$ are the estimated and true nonparametric components for the $\gamma^{\text{th}}$ replication, respectively,

$$\text{AMSE} = \frac{1}{R}\sum_{\gamma=1}^{R}\left(\text{MSE}_{\theta_s}^{(\gamma)}\right) = \frac{1}{R}\sum_{\gamma=1}^{R}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_s^{(\gamma)}(z_{si}) - \theta_s^{(\gamma)}(z_{si})\right)^2\right),$$

and

$$\text{Avariance} = \frac{1}{R}\sum_{\gamma=1}^{R}\left(\text{MSE}_{\theta_s}^{(\gamma)} - \left(\text{bias}_{\theta_s}^{(\gamma)}\right)^2\right).$$

It is clear from Table 1 that, when cross-equation correlation exists, the parametric component estimates of the partially linear SUR model have smaller variances and MSEs than those of Robinson's, and these relative efficiency gains are quite stable and robust for all sample sizes considered. With an increase in the sample size, the variances and MSEs drop significantly. Most of the time, the parametric component estimates of the partially linear SUR model have smaller biases than those of Robinson's. For the nonparametric part, the superiority of the two-step nonparametric estimator over Robinson's in terms of Avariance is not obvious as sample size increases. This result echoes with the previous finding in Remark 4 that cross-equation correlation information might not be effectively used for the two-step nonparametric estimator. However, with the extra step of the nonparametric SUR estimation that effectively takes into account the cross-equation correlation, although there is an increase in Abias, the reduction in Avariance is large enough such that the AMSEs become much smaller than those of the Robinson single-equation and two-step nonparametric estimators for all sample sizes considered.

Table 1: Finite Sample Performance *with* Cross-Equation Correlation ($\boldsymbol{\sigma_{12} = 0.6}$)

| | $\beta_1$ | | | $\beta_2$ | | | $\theta_1(\cdot)$ | | | $\theta_2(\cdot)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Bias | Var | MSE | Bias | Var | MSE | Abias | Avar | AMSE | Abias | Avar | AMSE |
| **Robinson's single-equation** — $\widehat{\beta}_s, \widehat{\theta}_s(\cdot)$ | | | | | | | | | | | | |
| 100 | 0.0025 | 0.0469 | 0.0469 | 0.0087 | 0.0464 | 0.0465 | 0.0000 | 0.0438 | 0.1754 | -0.0078 | 0.0435 | 0.1727 |
| 200 | 0.0063 | 0.0198 | 0.0198 | -0.0045 | 0.0211 | 0.0211 | -0.0014 | 0.0230 | 0.0785 | 0.0148 | 0.0242 | 0.0822 |
| 400 | 0.0015 | 0.0094 | 0.0094 | 0.0045 | 0.0100 | 0.0100 | 0.0010 | 0.0122 | 0.0393 | -0.0034 | 0.0124 | 0.0413 |
| 800 | 0.0028 | 0.0051 | 0.0051 | 0.0052 | 0.0052 | 0.0052 | -0.0031 | 0.0073 | 0.0214 | -0.0052 | 0.0073 | 0.0220 |
| 1600 | 0.0010 | 0.0025 | 0.0025 | 0.0002 | 0.0024 | 0.0024 | -0.0005 | 0.0041 | 0.0113 | 0.0029 | 0.0040 | 0.0111 |
| **Partially linear SUR** — $\widehat{\beta}_{s,\mathrm{sur}}, \widetilde{\theta}_s(\cdot)$ | | | | | | | | | | | | |
| 100 | 0.0046 | 0.0298 | 0.0298 | 0.0097 | 0.0309 | 0.0310 | -0.0033 | 0.0421 | 0.1292 | -0.0093 | 0.0414 | 0.1306 |
| 200 | 0.0050 | 0.0128 | 0.0128 | -0.0017 | 0.0140 | 0.0140 | 0.0007 | 0.0223 | 0.0590 | 0.0104 | 0.0234 | 0.0634 |
| 400 | -0.0011 | 0.0063 | 0.0063 | 0.0032 | 0.0066 | 0.0066 | 0.0052 | 0.0118 | 0.0306 | -0.0011 | 0.0121 | 0.0322 |
| 800 | 0.0022 | 0.0034 | 0.0034 | 0.0039 | 0.0033 | 0.0034 | -0.0021 | 0.0072 | 0.0170 | -0.0031 | 0.0071 | 0.0169 |
| 1600 | 0.0012 | 0.0016 | 0.0016 | 0.0020 | 0.0017 | 0.0017 | -0.0008 | 0.0040 | 0.0088 | 0.0001 | 0.0040 | 0.0090 |
| **Nonparametric SUR estimator** — $\widetilde{\theta}_{s,\mathrm{sur}}(\cdot)$ | | | | | | | | | | | | |
| 100 | - | - | - | - | - | - | -0.0240 | 0.0376 | 0.1261 | -0.0215 | 0.0241 | 0.1142 |
| 200 | - | - | - | - | - | - | -0.0218 | 0.0207 | 0.0586 | -0.0031 | 0.0130 | 0.0535 |
| 400 | - | - | - | - | - | - | -0.0167 | 0.0110 | 0.0304 | -0.0154 | 0.0068 | 0.0276 |
| 800 | - | - | - | - | - | - | -0.0228 | 0.0057 | 0.0162 | -0.0168 | 0.0036 | 0.0139 |
| 1600 | - | - | - | - | - | - | -0.0171 | 0.0030 | 0.0081 | -0.0117 | 0.0020 | 0.0073 |

**Note**: (1) This table compares the effects of sample size on the partially linear SUR and Robinson equation-by-equation models when there exists cross-equation correlation (i.e., $\sigma_{12} = 0.6$). (2) For the partially linear SUR model, the parametric components are estimated with the feasible SUR estimator, and the nonparametric components are estimated with the two-step nonparametric estimator.

# 5 Empirical example

As an empirical example, we use a firm-level Italian banking data set obtained from Bankscope.[14] This data set covers information about 739 banks over 10 years from 1996 to 2005. The observations are indexed by $i = 1, \dots, 2989$. In this paper, we use the intermediation approach (Sealey & Lindley 1977) and consider capital, labor, and purchased funds as the three banking inputs.[15] They are used to produce the three outputs: loans, other earning assets, and off-balance sheet accounts. The environmental factors include the three types of banking risks—1) credit risk, measured by loan loss provisions in log; 2) solvency risk, measured by equity capital amount in log; and 3) liquidity risk, measured by amount of liquidity assets in log—along with a time trend. Note that an increase in one of these risk measures (e.g., loan loss provisions in log) indicates a decrease in the corresponding risk level (e.g., credit risk level). These environmental variables may affect the total cost of banks in unknown fashions. Table B1 in Appendix B contains summary statistics of all the variables used.[16] For ease of comparison, the three different model specifications are estimated using the same data: 1) the partially linear SUR model, i.e., (2.6) and (2.7); 2) the partially linear single-equation model, i.e., (2.6) only; and 3) the linear SUR model, i.e., (2.6) and (2.7) in which $\theta(\cdot)$ in (2.6) is replaced by a linear parametric function of $z$. Results from all the models are reported in Table B2, Table 2, and Figures 1–3.

Table B2 in Appendix B reports the translog cost function parameter estimates for input prices and outputs. These parameters are not input price or output elasticities, and therefore have few economic meanings. However, it can be seen that most parameter estimates are the same in signs or similar in magnitudes across the three models. Most standard errors under the partially linear SUR model are smaller than their partially linear single-equation and linear SUR counterparts. The partially linear SUR and single-equation models contain exactly the same number of parameters—this is because the share equations of the translog cost system are the first-order partial derivatives of the translog cost function with respect to input prices. However, the former utilizes more information from the additional share

[14] See *bankscope2.bvdep.com* for details.

[15] Purchased funds such as customer deposits are viewed as inputs which are used to produce loans and other assets. See Hughes & Mester (1993) for more details about treating deposits as inputs in estimating a cost function.

[16] See Resti (1997) for a brief introduction to the Italian banking system.

equations for estimation, and therefore is expected to produce more efficient estimates.[17] The partially linear and linear SUR models have the same number of equations, including the share equations. However, the translog cost function of the partially linear model, Eq. (2.6), is more flexibly specified than its linear parametric counterpart. This indicates that the partially linear SUR has better goodness-of-fit, and therefore smaller estimated error variance than the linear SUR.

Table 2 reports the mean and quartile values (Q1–Q3) of estimated elasticities and RTS. The estimated input price elasticities are the estimated cost shares of inputs—they should be close to the actual cost shares ($S_1$–$S_3$) reported in Table B1. It can be seen that the partially linear SUR and linear SUR yield similar input price elasticity estimates, and the mean estimates of these two models are close to the actual means of capital, labor, and purchased funds shares, respectively, in Table B1. In particular, purchased funds (i.e., interest expenses) have the largest share of cost in producing the three outputs—the mean of the estimated $\partial \ln C/\partial \ln W_3$ is 0.4364 from the partially linear SUR model. This result is unsurprising as banks utilize portfolios of risk-free and risky-assets to generate income. In contrast, the mean estimated cost share of capital is only 0.0777 from the partially linear single-equation regression. Without exploiting the share equations, the seemingly flexible single-equation regression fails to predict the capital share.

All the three models generate similar output elasticities, which can be viewed as the marginal costs of producing the outputs, except that the elasticity estimates of other earning assets ($O_2$) under the linear SUR are slightly smaller than those under the other two models. It is obvious that for all the models, it is the most costly to produce loans: for an average bank under the partially linear SUR model, as the loan output increases by 1%, the total cost increases by 0.4780%, *ceteris paribus*. To reduce the non-performing loans (NPLs) and increase loan quality, extra tasks must be undertaken, such as additional credit monitoring and screening of the borrowers, working out new contract terms on problem loans, selling off the NPLs, etc.—all these activities would involve extra costs/efforts on the part of bank managers (Berger & DeYoung 1997).

RTS is one of the most discussed topics in banking. For example, Mester (1996) found

[17] A caution is that too many share equations might inflate the estimated error variance in practice.

Table 2: Summary Statistics of Elasticities and RTS

| | Input price elasticity | | | Output elasticity | | | RTS |
|---|---|---|---|---|---|---|---|
| | $\frac{\partial \ln C}{\partial \ln W_1}$ | $\frac{\partial \ln C}{\partial \ln W_2}$ | $\frac{\partial \ln C}{\partial \ln W_3}$ | $\frac{\partial \ln C}{\partial \ln O_1}$ | $\frac{\partial \ln C}{\partial \ln O_2}$ | $\frac{\partial \ln C}{\partial \ln O_3}$ | |
| **Partially linear SUR** | | | | | | | |
| Mean | 0.2492 | 0.3144 | 0.4364 | 0.4780 | 0.4127 | 0.0383 | 1.0843 |
| Q1(25%) | 0.2210 | 0.2708 | 0.3855 | 0.4563 | 0.3861 | 0.0185 | 1.0178 |
| Q2(50%) | 0.2489 | 0.3146 | 0.4376 | 0.4777 | 0.4062 | 0.0349 | 1.0916 |
| Q3(75%) | 0.2779 | 0.3603 | 0.4872 | 0.5006 | 0.4324 | 0.0541 | 1.1521 |
| **Partially linear single-equation** | | | | | | | |
| Mean | 0.0777 | 0.5533 | 0.3690 | 0.4788 | 0.3997 | 0.0428 | 1.0995 |
| Q1(25%) | 0.0527 | 0.5056 | 0.3175 | 0.4526 | 0.3609 | 0.0180 | 1.0098 |
| Q2(50%) | 0.0752 | 0.5556 | 0.3687 | 0.4749 | 0.3894 | 0.0385 | 1.1096 |
| Q3(75%) | 0.0997 | 0.6050 | 0.4187 | 0.5026 | 0.4310 | 0.0642 | 1.1888 |
| **Linear SUR** | | | | | | | |
| Mean | 0.2467 | 0.3159 | 0.4373 | 0.4681 | 0.2939 | 0.0497 | 1.2330 |
| Q1(25%) | 0.2168 | 0.2735 | 0.3789 | 0.4165 | 0.2247 | 0.0019 | 1.2150 |
| Q2(50%) | 0.2468 | 0.3164 | 0.4370 | 0.4746 | 0.2927 | 0.0495 | 1.2330 |
| Q3(75%) | 0.2786 | 0.3615 | 0.4935 | 0.5289 | 0.3633 | 0.0952 | 1.2500 |

**Note**: (1) The numeriare input price elasticity is calculated as $\partial \ln C/\partial \ln W_1 = 1 - \sum_{j=2}^{J} \partial \ln C/\partial \ln W_j$. (2) The RTS is calculated as $\left(\sum_{p=1}^{P} \partial \ln C/\partial \ln O_p\right)^{-1}$.

increasing RTS by estimating a single translog cost function using a sample of 214 U.S. banks from 1991 to 1992. Hughes & Mester (1998) had the similar finding by estimating a translog cost system, including the share equations, with a sample of 286 U.S. banks from 1989 to 1990. Our finding in this paper—RTS is greater than unity—is in line with the previous two papers. This indicates that banks could benefit from expansion via, e.g., mergers and acquisitions. However, the RTS estimates under the linear SUR is overly large as they substantially deviate from unity. The testing procedure described in Section 3.5 suggests that the null of linear SUR is rejected at the 1% level and the alternative of the partially linear SUR is preferred, given that the bootstrapped $p$-value is zero to the fourth decimal place with 399 replications.

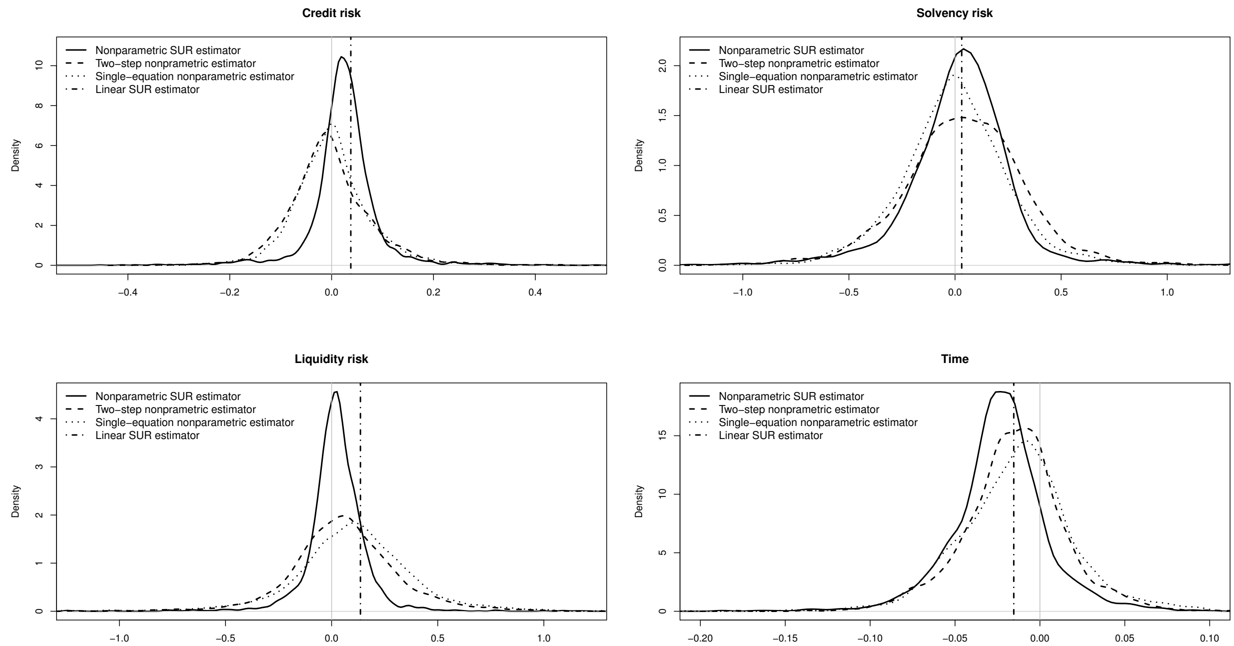
In all the three models, the productivity parameter, $\theta$, depends on the environmental factors $z$. It is easy to see that the marginal effects of $z$ on $\theta$ are also the marginal effects of $z$ on total cost of production. For the partially linear models, the $z$ variables shift the cost function in fully flexible manners, given that $\theta$ is an unknown nonparametric function of $z$, whereas for the parametric model, they only affect banks' costs in linear manners. In this application, the $z$ variables include the three banking risks and time. It is important to take account of risk-taking behavior of banks when investigating their performances (Hughes & Mester 1998). Since risk exposure provides a bank with higher profit while risk management that prevents the bank from failure requires additional costs, it would be desirable to examine the marginal impact of risks on a bank's performance, especially when a cost function is employed to represent its technology. Meanwhile, the marginal effects of time on $\theta$ measure technical change. The local-linear nonparametric SUR estimator provided in Section 3.3 facilitates the calculation of these marginal effects.

To easily compare and contrast the marginal effect estimates, Figure 1 plots the kernel density functions for each of the estimators, along with a grey vertical line at zero to highlight the heterogeneity of the marginal effects. It is obvious that the marginal effects from the linear SUR are all constants—in particular, an increase in the measurements of risks (i.e., a decrease in the risks), *ceteris paribus*, causes the total cost to increase, as the degenerate marginal effects of all the risk variables are positive. This is because risk management requires extra costs, including additional non-interest expenses in administering the loan portfolios and managing financial capital and liquidity assets. However, the linear SUR ignores the heterogeneity in these marginal effects. For banks that are exposed to lower levels of risks, the managers might have less motivation to spend extra time and efforts in controlling the risks, by skimping on the resources contributed to risk management (Hughes & Mester 1993), leading to cost reduction. Indeed, the correlation coefficients between $\partial\theta/\partial z$ and the corresponding $z$ are all significantly negative at the 5% level, which indicates that low level of risks (i.e., large values of $z$'s) is associated with negative $\partial\theta/\partial z$.[18] This heterogeneity can be captured by the flexible partially linear models as shown in Figure 1. For the partially linear SUR model, both the two-step nonparametric estimator and nonparametric SUR estimator are employed. It is obvious that for all the plots in Figure 1, the nonparametric SUR estimates have much smaller variations than the two-step and single-equation nonparametric counterparts. It can also be seen that for the marginal effects of credit risk, the two-step and single-equation nonparametric estimators generate similar estimates. However, it seems that the single-equation one underestimates (overestimates) the marginal effects of the solvency (liquidity) risk as most parts of its related density function are to the left (right) of those of the density functions of the two-step nonparametric and nonparametric SUR estimates, respectively. It is important for policy makers to correctly understand these heterogeneous marginal effects, and therefore make appropriate regulations for the banks to reduce the probability of bank failure with sound risk management while not profoundly harming profitability.

Finally, technical progress is found for the Italian banking system during the period 1996–2005: the marginal effect of time on total cost is -0.0154 from the linear SUR, indicating that the total cost decreases by 1.54% per annum, *ceteris paribus*. This marginal effect estimate is close to the mean counterparts from the two-step and single-equation nonparametric estimators, which generate similar technical change estimates. According to Berger (2003), technical progress in banking stems from improvements in information technology and the use of financial engineering models to improve credit and portfolio analysis. Altunbas et al.

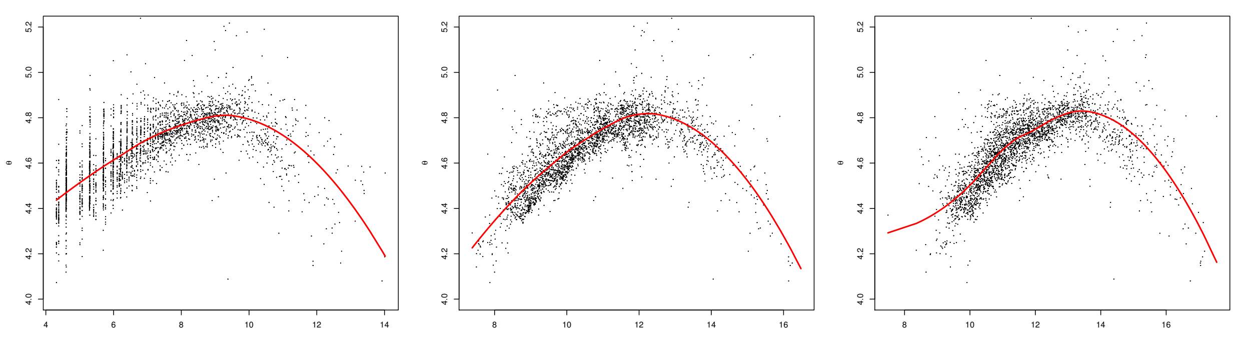
[18]To see this more clearly, in Figure 2 we first show a scatter plot of estimated $\theta$ with the nonparametric SUR estimator against each risk variable, and then fit a locally weighted scatter plot smoothing (LOWESS) regression line on each scatter plot. All the three fitted lines are inverse-U shaped. This means that the slopes, $\partial\theta/\partial z$, decrease as the $z$'s increase, and therefore supports the negative correlation between the risk measures and their marginal effects on the cost of production.

Figure 1: Kernel Density Plots of Marginal Effects of $z$ on $\theta(\cdot)$




**Note**: The kernel density functions of the marginal effects of each environmental variable, $z$, on the cost of production, $\theta(\cdot)$, are plotted. For example, the upper left panel reports the marginal effects of credit risk on cost of production. In this panel, kernel density functions from the four estimators are plotted for ease of comparison. In addition, a grey vertical line at zero is drawn to highlight the heterogeneity of the marginal effects.

Figure 2: LOWESS Plots of $\theta(\cdot)$ on $z$




**Note**: The locally weighted scatter plot smoothing (LOWESS) regression lines of $\theta(\cdot)$ estimated with the nonparametric SUR estimator on the risk variables are plotted. For example, in the first panel, we show a scatter plot of estimated $\theta(\cdot)$ on credit risk, before fitting a LOWESS regression line on the scatter plot.

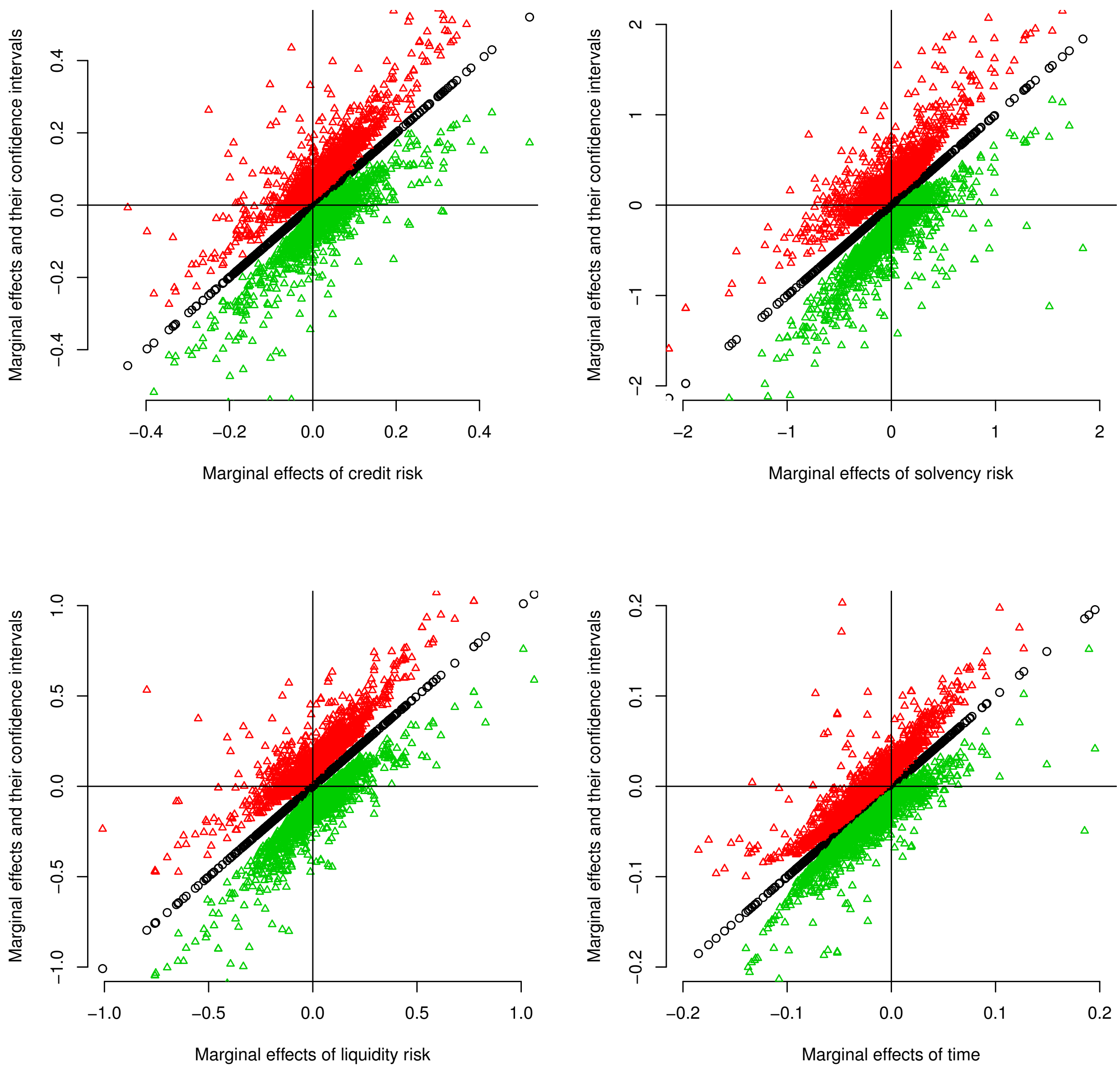
(1999) found similar evidence of cost diminution per annum due to technical progress using a sample of European banks from 1989 to 1996. In fact, most of the nonparametric estimates are to the left of the grey vertical line at zero in Figure 1, and therefore indicate technical progress. Furthermore, most parts of the density function of the nonparametric SUR estimates are to the left of those of the density functions of the two-step and single-equation nonparametric estimates, indicating that the evidence of technical progress is stronger with the nonparametric SUR estimator. The linear SUR fails to capture any possible technical regress in banking that could happen during financial crisis. However, all the nonparametric estimates generate technical regress for some observations—most of the technical regress happened from 1997 to 1998 during the Asian financial crisis, which triggered a worldwide economic downturn due to financial integration.

To see whether a marginal effect estimate from the nonparametric SUR estimator is statistically significant at a conventional level, we conduct a cluster wild bootstrap to obtain the standard error of each of the marginal effect estimate. Specifically, we first re-center the residuals from the partially linear SUR model and obtain $\{\widehat{u}_i - \bar{\widehat{u}}\}_{i=1}^n$, where $\bar{\widehat{u}} = \frac{1}{n}\sum_{i=1}^n \widehat{u}_i$, $\forall s$. Then, generate bootstrapped residuals, $u_i^{bb}$, where $u_i^{bb} = [(1-\sqrt{5})/2](\widehat{u}_i - \bar{\widehat{u}})$ with probability $(1+\sqrt{5})/(2\sqrt{5})$ and $u_i^{bb} = [(1+\sqrt{5})/2](\widehat{u}_i - \bar{\widehat{u}})$ with probability $(\sqrt{5}-1)/(2\sqrt{5})$, $\forall i$. Next, generate the bootstrapped dependent variable by adding the bootstrapped residuals to the fitted values of the partially linear SUR model, i.e., $y_i^{bb} = \widetilde{\theta}_{sur}(z_i) + x_i'\widehat{\beta}_{sur} + u_i^{bb}$, where $y_i^{bb}$, $u_i^{bb}$, and $x_i$ are defined analogously as $y_i^*$, $u_i$, and $x_i^*$, respectively, and $\widetilde{\theta}_{sur}(z_i) \equiv (\widetilde{\theta}_{1,\mathrm{sur}}(z_i), \widetilde{\theta}_{2,\mathrm{sur}}(z_i), \ldots, \widetilde{\theta}_{m,\mathrm{sur}}(z_i))'$.[19] Finally, we use the bootstrapped sample $\{y_i^{bb}, x_i, z_i\}_{i=1}^n$ to estimate the marginal effects, and call them the bootstrapped marginal effect estimates. Repeat the preceding steps 399 times, and then for each observation, we would have a sample of 400 marginal effect estimates, including the original marginal effect estimate. The standard error of the estimate for each observation is calculated as the standard deviation using the original and bootstrapped marginal effect estimates. For each observation, the confidence interval is constructed by adding and subtracting twice the standard error from the original marginal effect estimate.

To report all the marginal effect estimates and their statistical significance in a concise

[19] For the translog cost system, the nonparametric component is in the cost function only. Therefore, in this particular example with a cost function and two share equations, $\widetilde{\theta}_{2,\mathrm{sur}}(z_i) = \widetilde{\theta}_{3,\mathrm{sur}}(z_i) = 0$.

Figure 3: Marginal Effect Estimates and Their Confidence Intervals




**Note**: The 45 degree plots of the marginal effect estimates of each environmental variable, $z$, on the cost of production, $\theta(\cdot)$, are presented. For example, the upper left panel reports the marginal effect estimates of credit risk on cost of production (as circles), along with their confidence intervals (as triangles). The vertical line at zero is drawn to highlight the heterogeneity of the marginal effects, and the horizontal line at zero indicates statistical significance.

manner, we use the 45 degree plot suggested by Zhang et al. (2012) and Henderson et al. (2012) as shown in Figure 3. Particularly, in the upper left panel of Figure 3, we first plot marginal effect estimates of credit risk against themselves—this would generate a 45 degree scatter plot, and then plot the upper and lower confidence bounds against marginal effect estimates of credit risk, respectively. If an upper (lower) confidence bound is below (above) the horizontal line at zero, then the marginal effect estimate for this observation is significantly different from zero at the 5% level. If the horizontal line at zero passes through a confidence interval, then the marginal effect estimate for this observation would be statistically insignificant. Results show that marginal effects of credit risk, solvency risk, and liquidity risk are significantly positive (negative) for 29.98% (4.55%), 17.16% (9.00%), and 19.97% (7.09%) of the observations, respectively. This confirms the heterogeneity of risk management across banks. There are 61.86% (4.12%) of the observations that exhibit significant technical progress (regress). This shows that a majority of the banks enjoyed technical progress over most years while technical regress occurred occasionally amid adverse financial environment.

# 6 Discussion: heteroscedasticity in estimating production system

The cost/profit function is correlated with the share equations as implied by Shephard/Hotelling's Lemma, and the share equations are correlated with each other due to input substitutability (or complementarity). However, production decisions and degrees of input substitutability might vary across firms because firms use different levels of inputs or face different production environments (e.g., firm age, size, etc.), leading to heterogeneity of technology. Therefore, contemporaneous correlations among equations can vary across observations (Chavas & Segerson 1987, Mandy & Martins-Filho 1993).

To model the observation-specific contemporaneous correlations, let $\Omega = \{\omega_{sl}\}_{s=1,l=1}^{m,m}$ be the $m \times m$ variance-covariance matrix with a typical element $\omega_{sl}$, where

$$
\omega_{sl} = \begin{pmatrix} \sigma_{sl}^1 & 0 & \cdots & 0 \\ 0 & \sigma_{sl}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{sl}^n \end{pmatrix},
$$

and $\sigma_{sl}^i$ denotes contemporaneous correlations between the $s^{\text{th}}$ and $l^{\text{th}}$ equation for the $i^{\text{th}}$ observation, $\forall s, l = 1, \ldots, m$ and $i = 1, \ldots, n$. To estimate these correlations, define $\sigma_{sl}^i = \mathbb{E}\left(u_{si}u_{li} \mid x_{si}, x_{li}, z_{si}, z_{li}\right)$. First, obtain $\widehat{u}_{si}$, $\forall s = 1, \ldots, m$, from the equation-by-equation OLS estimation of (3.3). Then,

1. For $s \neq l$, estimate $\widehat{u}_{si}\widehat{u}_{li} = \widetilde{g}(x_{si}, x_{li}, z_{si}, z_{li}) + \widetilde{e}_{si}$, where $\widetilde{g}(\cdot)$ is an unknown nonparametric function and $\mathbb{E}\left(\widetilde{e}_{si} \mid x_{si}, x_{li}, z_{si}, z_{li}\right) = 0$. The estimated cross-equation scedastic function, $\widehat{\sigma}_{sl}^i$, is $\widehat{\widetilde{g}}(x_{si}, x_{li}, z_{si}, z_{li})$, and captures between-equation heteroscedasticity.

2. For $s = l$, estimate $\widehat{u}_{si}^2 = \bar{g}(x_{si}, z_{si}) + \bar{e}_{si}$, where $\bar{g}(\cdot)$ is another unknown nonparametric function and $\mathbb{E}\left(\bar{e}_{si} \mid x_{si}, z_{si}\right) = 0$, subject to the observation-specific inequality constraints that $\widehat{\bar{g}}(\cdot) \geq 0$, using the constraint weighted bootstrapping procedure (Hall & Huang 2001, Du et al. 2013). The estimated scedastic function, $\widehat{\sigma}_{ss}^i$, is $\widehat{\bar{g}}(x_{si}, z_{si}) \geq 0$, and captures within-equation heteroscedasticity.

The between- and within-equation heteroscedasticity in the context of the partially linear SUR model can be detected by testing if the gradients of $\widehat{\widetilde{g}}(\cdot)$ and $\widehat{\bar{g}}(\cdot)$ are jointly significant at a conventional level, respectively.[20] Let $\widehat{\Omega} = \{\widehat{\omega}_{sl}\}_{s=1,l=1}^{m,m}$, and the heteroscedasticity-corrected SUR estimator can be written as

$$
\widehat{\beta}_{\text{sur}}^c = \left(\widehat{X}^{*\prime}\widehat{\Omega}^{-1}\widehat{X}^*\right)^{-1}\left(\widehat{X}^{*\prime}\widehat{\Omega}^{-1}\widehat{Y}^*\right),
$$

where $\widehat{X}^*$ is an $mn \times q$ stacked block diagonal matrix and $\widehat{Y}^*$ is an $mn$-vector as defined in Section 3.

[20] See Racine (1997) for a consistent significance testing procedure for nonparametric regression.

# 7 Conclusions

This paper extends Robinson (1988) partially linear single-equation regression model to a partially linear SUR model in a straightforward manner. The parametric component is estimated via a two-step feasible SUR estimation procedure, and the resulting SUR estimator achieves root-$n$ convergence and asymptotic normality. The nonparametric component is more efficiently estimated with a nonparametric SUR estimator based on the Cholesky decomposition. A model specification test for alternative parametric functional forms is proposed, and conducted using a novel cluster bootstrap procedure.

To establish a connection between the partially linear SUR model and production theory, we propose a partially linear translog cost system that consists of a partially linear translog cost function and a set of share equations derived via Shephard's Lemma. Given that the share equations have the same parameters as the cost function, it would be more efficient to estimate the cost system than estimating a single cost function. The nonparametric component of the cost function is interpreted as a productivity parameter which is an unknown nonparametric function of $z$.

A firm-level Italian banking data set obtained from Bankscope is used to estimate the translog cost system. We find that the partially linear SUR model produces more efficient and reasonable estimates than the partially linear single-equation and linear SUR counterparts. The model specification test suggests that the partially linear SUR is preferred to its linear counterpart. Furthermore, the partially linear SUR model estimated with the nonparametric SUR estimator better captures marginal effects of banking risks and time on cost of production than the other competing estimators considered in this paper.

Finally, the model can be further extended to a panel data SUR with fixed effects possibly via the least squares dummy variable approach. Besides, non-contemporaneous correlations among equations can possibly be modeled without changing the framework of the GLS estimator.

# References


Altunbas, Y., Goddard, J. & Molyneux, P. (1999), 'Technical change in banking', *Economics Letters* **64**(2), 215–221.

Baležentis, T., Liobikienė, G., Štreimikienė, D. & Sun, K. (2020), 'The impact of income inequality on consumption-based greenhouse gas emissions at the global level: A partially linear approach', *Journal of Environmental Management* **267**, 110635.

Berger, A. (2003), 'The economic effects of technological progress: Evidence from the banking industry', *Journal of Money, Credit and Banking* **35**(2), 141–76.

Berger, A. N. & DeYoung, R. (1997), 'Problem loans and cost efficiency in commercial banks', *Journal of Banking and Finance* **21**(6), 849–870.

Bhaumik, S. K., Kumbhakar, S. C. & Sun, K. (2015), 'A note on a semiparametric approach to estimating financing constraints in firms', *The European Journal of Finance* **21**(12), 992–1004.

Blundell, R., Duncan, A. & Pendakur, K. (1998), 'Semiparametric estimation and consumer demand', *Journal of Applied Econometrics* **13**(5), 435–461.

Bontemps, C., Simioni, M. & Surry, Y. (2006), Semiparametric hedonic price models: Assessing the effects of agricultural nonpoint source pollution, Economics working paper archive (Toulouse), French Institute for Agronomy Research (INRA), Economics Laboratory in Toulouse (ESR Toulouse).

Cai, Z., Fan, J. & Yao, Q. (2000), 'Functional-coefficient regression models for nonlinear time series', *Journal of the American Statistical Association* **95**(451), 941–956.

Cameron, A. & Trivedi, K. (2005), *Microeconometrics: Methods and Application*, Cambridge University Press, chapter 11, pp. 376–377.

Chavas, J.-P. & Segerson, K. (1987), 'Stochastic specification and estimation of share equation systems', *Journal of Econometrics* **35**(2), 337–358.

Du, P., Parmeter, C. & Racine, J. (2013), 'Nonparametric kernel regression with multiple predictors and multiple shape constraints', *Statistica Sinica* **23**(3), 1347–1371.

Geng, X., Martins-Filho, C. & Yao, F. (2020), 'Estimation of a partially linear additive model with generated covariates', *Journal of Statistical Planning and Inference* **208**, 94–118.

Hall, P. & Huang, H. (2001), 'Nonparametric kernel regression subject to monotonicity constraints', *The Annals of Statistics* **29**(3), 624–647.

Henderson, D. J., Kumbhakar, S. C., Li, Q. & Parmeter, C. F. (2015), 'Smooth coefficient estimation of a seemingly unrelated regression', *Journal of Econometrics* **189**(1), 148–162.

Henderson, D. J., Kumbhakar, S. C. & Parmeter, C. F. (2012), 'A simple method to visualize results in nonlinear regression models', *Economics Letters* **117**(3), 578–581.

Hoeffding, W. (1961), The strong law of large numbers for U-statistics, Technical report, Institute of Statistics mimeo series 302, Department of Statistics, North Carolina State University.

Hughes, J. P. & Mester, L. J. (1993), 'A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine', *Journal of Productivity Analysis* **4**(3), 293–315.

Hughes, J. P. & Mester, L. J. (1998), 'Bank capitalization and cost: Evidence of scale economies in risk management and signaling', *The Review of Economics and Statistics* **80**(2), 314–325.

Kumbhakar, S. (1991), 'The measurement and decomposition of cost-inefficiency: The translog cost system', *Oxford Economic Papers* **43**(4), 667–683.

Kumbhakar, S. C. (2001), 'Estimation of profit functions when profit is not maximum', *American Journal of Agricultural Economics* **83**(1), 1–19.

Kumbhakar, S. & Tsionas, E. (2005), 'Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach', *Journal of Econometrics* **126**(2), 355–384.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Linton, O. B., Mammen, E., Lin, X. & Carroll, R. J. (2004), 'Correlation and marginal longitudinal kernel nonparametric regression', *Proceedings of the Second Seattle Symposium in Biostatistics* **179**.

Mandy, D. M. & Martins-Filho, C. (1993), 'Seemingly unrelated regressions under additive heteroscedasticity: Theory and share equation applications', *Journal of Econometrics* **58**(3), 315–346.

Martins-Filho, C. & Yao, F. (2006), 'A note on the use of v and u statistics in nonparametric models of regression', *Annals of the Institute of Statistical Mathematics* **58**, 389C406.

Martins-Filho, C. & Yao, F. (2009), 'Nonparametric regression estimation with general parametric error covariance', *Journal of Multivariate Analysis* **100**(3), 309–333.

Mester, L. J. (1996), 'A study of bank efficiency taking into account risk-preferences', *Journal of Banking and Finance* **20**(6), 1025–1045.

Millimet, D., List, J. & Stengos, T. (2003), 'The environmental Kuznets curve: Real progress or misspecified models?', *Review of Economics and Statistics* **85**(4), 1038–1047.

Racine, J. (1997), 'Consistent significance testing for nonparametric regression', *Journal of Business & Economic Statistics* **15**(3), 369–378.

Resti, A. (1997), 'Evaluating the cost-efficiency of the Italian banking system: What can be learned from the joint application of parametric and non-parametric techniques', *Journal of Banking and Finance* **21**(2), 221–250.

Robinson, P. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Sealey, C. W. & Lindley, J. T. (1977), 'Inputs, outputs, and a theory of production and cost at depository financial institutions', *The Journal of Finance* **32**(4), 1251–1266.

Su, L., Ullah, A. & Wang, Y. (2013), 'Nonparametric regression estimation with general parametric error covariance: A more efficient two-step estimator', *Empirical Economics* **45**, 1009–1024.

Xu, Q., You, J. & Zhou, B. (2008), 'Seemingly unrelated nonparametric models with positive correlation and constrained error variances', *Economics Letters* **99**(2), 223–227.

Yao, F. & Martins-Filho, C. (2015), 'An asymptotic characterization of finite degree U-statistics with sample size-dependent kernels: Applications to nonparametric estimators and test statistics', *Communications in Statistics - Theory and Methods* **44**(15), 3251–3265.

You, J. & Zhou, X. (2010), 'Statistical inference on seemingly unrelated varying coefficient partially linear models', *Statistica Neerlandica* **64**(2), 227–253.

You, J. & Zhou, X. (2014), 'Asymptotic theory in fixed effects panel data seemingly unrelated partially linear regression models', *Econometric Theory* **30**(2), 407–435.

Zellner, A. (1962), 'An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias', *Journal of the American Statistical Association* **57**(298), 348–368.

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

# Appendix A Proof of theorems

*Proof of Theorem 1.* From Zellner (1962), we have $\sqrt{n}\,(\beta_{\text{sur}} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\beta)$ where $\mathcal{V}_\beta = \left(\mathbb{E}\left(x_i^* \Sigma_m^{-1} x_i^{*\prime}\right)\right)^{-1}$ provided that $\widehat{\Sigma}_m^{-1}$ has a parametric convergence rate to $\Sigma_m^{-1}$. This condition for a semiparametric model is easy to verify given that the unconditional error variance estimator for the nonparametric regression has a parametric convergence rate; see, e.g., Theorem 2 of Martins-Filho & Yao (2006). We complete the proof by showing that $\sqrt{n}\left(\widehat{\beta}_{\text{sur}} - \beta_{\text{sur}}\right) = o_p(1)$.

Let $A_n \equiv \frac{1}{n}\sum_{i=1}^n x_i^* \widehat{\Sigma}_m^{-1} x_i^{*\prime}$, $B_n \equiv \frac{1}{n}\sum_{i=1}^n x_i^* \widehat{\Sigma}_m^{-1} y_i^*$, $\widehat{A}_n \equiv \frac{1}{n}\sum_{i=1}^n \widehat{x}_i^* \widehat{\Sigma}_m^{-1} \widehat{x}_i^{*\prime}$, and $\widehat{B}_n \equiv \frac{1}{n}\sum_{i=1}^n \widehat{x}_i^* \widehat{\Sigma}_m^{-1} \widehat{y}_i^*$. We have $\widehat{\beta}_{\text{sur}} = \widehat{A}_n^{-1}\widehat{B}_n$ and $\beta_{\text{sur}} = A_n^{-1}B_n$. By Theorem 2.6 in Li & Racine (2007) together with the bias reduction method using a high-order kernel,[21] under Assumption A1–A4, for a compact subset $\mathcal{G}_z$, we have $\sup_{z\in\mathcal{G}_z}|\widehat{g}_{sx}(z) - g_{sx}(z)| = O_p(L_{sn})$ where $L_{sn} \equiv \left(\frac{\log n}{nh_s^{p_s}}\right)^{\frac{1}{2}} + h_s^{r_s}$. In what follows, we show that $\widehat{A}_n = A_n + o_p\left(n^{-1/2}\right)$ and then $\widehat{B}_n = B_n + o_p(n^{-1/2})$ follows in a similar manner. Given that both $A_n$ and $B_n$ are $O_p(1)$, we have $\widehat{A}_n^{-1}\widehat{B}_n = A_n^{-1}B_n + o_p(n^{-1/2})$, which completes the proof.

Let $d_i \equiv \widehat{x}_i^* - x_i^*$, then

$$\widehat{A}_n - A_n = \frac{1}{n}\sum_{i=1}^n d_i \widehat{\Sigma}_m^{-1} x_i^{*\prime} + \frac{1}{n}\sum_{i=1}^n x_i^* \widehat{\Sigma}_m^{-1} d_i' + \frac{1}{n}\sum_{i=1}^n d_i \widehat{\Sigma}_m^{-1} d_i' \equiv A_{1n} + A_{2n} + A_{3n}.$$

Given that $d_i = O_p(L_n)$ uniformly where $L_n \equiv \sum_s^m L_{sn}$, we have $A_{3n} = O_p(L_n^2) = o(n^{-1/2})$ by Assumption A4.

As $A_{2n} = A_{1n}'$, we only examine the order of $A_{1n}$ here. Let $\widehat{\Sigma}_m^{-1} \equiv \{\widehat{v}_{sl}\}_{s,l=1}^{m,m}$ and a typical $(s,l)^{\text{th}}$ block of $A_{1n}$ be $a_{sl} \equiv \widehat{v}_{sl}\frac{1}{n}\sum_{i=1}^n (\widehat{x}_{si}^* - x_{si}^*)x_{li}^{*\prime}$ of dimension $q_s \times q_l$. Then the $(k,j)^{\text{th}}$ element of $a_{sl}$ is

$$a_{sl,kj} \equiv -\widehat{v}_{sl}\frac{1}{n}\sum_{i=1}^n \left(\widehat{g}_{sx,k}(z_{si}) - g_{sx,k}(z_{si})\right) x_{li,j}^*,$$

where $\widehat{g}_{sx,k}(z)$ denotes the $k^{\text{th}}$ element of the vector $\widehat{g}_{sx}(z)$ and $g_{sx,k}(z)$ and $x_{li,j}^*$ are defined

[21] See also Theorem 2 of Geng et al. (2020).

similarly. By Eq. (3.4), we have

$$\widehat{g}_{sx,k}(z_{si}) - g_{sx,k}(z_{si}) = \frac{1}{\widehat{f}_{sz}(z_{si})} \frac{1}{nh_s^{p_s}} \sum_{t=1}^{n} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right)\left[x_{st,k}^{*} + (g_{sx,k}(z_{st}) - g_{sx,k}(z_{si}))\right],$$

where $\widehat{f}_{sz}(z) \equiv (nh_s^{p_s})^{-1}\sum_{t=1}^{n} K_s\left(\frac{z_{st}-z}{h_s}\right)$ is the Rosenblatt density estimator for the density function $f_{sz}(z)$ of $z_{si}$. As $\widehat{f}_{sz}(z)$ has the same uniform convergence rate of $O_p(L_{sn})$ to $f_{sz}(z)$ and $f_{sz}(z)$ is uniformly bounded away from zero, we have

$$\widehat{g}_{sx,k}(z_{si}) - g_{sx,k}(z_{si}) = \frac{1}{f_{sz}(z_{si})} \frac{1}{nh_s^{p_s}} \sum_{t=1}^{n} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right)\left[x_{st,k}^{*} + (g_{sx,k}(z_{st}) - g_{sx,k}(z_{si}))\right] + O_p(L_{sn}^2).$$

Given that $\frac{1}{n}\sum_{i=1}^{n}|x_{li,j}^{*}| = O_p(1)$, we have $a_{sl,kj} = -\widehat{v}_{sl}\left(Q_{1n} + Q_{2n}\right) + o_p(n^{-1/2})$, where

$$Q_{1n} \equiv \frac{1}{n^2 h_s^{p_s}} \sum_{i=1}^{n}\sum_{t=1}^{n} \frac{1}{f_{sz}(z_{si})} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right) x_{st,k}^{*} x_{li,j}^{*},$$

and

$$Q_{2n} \equiv \frac{1}{n^2 h_s^{p_s}} \sum_{i=1}^{n}\sum_{t=1}^{n} \frac{1}{f_{sz}(z_{si})} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right) (g_{sx,k}(z_{st}) - g_{sx,k}(z_{si})) x_{li,j}^{*}.$$

For $Q_{1n}$, let $\psi_{nit} \equiv (h_s^{p_s} f_{sz}(z_{si}))^{-1} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right) x_{st,k}^{*} x_{li,j}^{*}$ and we have $Q_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit}$ $= \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} + \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{1n} + E_{2n}$. Given that $\mathbb{E}\left(E_{1n}\right) = 0$ and $\mathrm{Var}\left(E_{1n}\right) = n^{-3}\mathbb{E}\left(\psi_{nii}\right) = O(n^{-3}h_s^{-2p_s}) = o(n^{-1})$, by Chebyshev's Inequality we have $E_{1n} = o_p(n^{-1/2})$. Let $U_n \equiv \binom{n}{k}^{-1}\sum_{\substack{i=1\\ i<t}}^{n}\sum_{t=1}^{n}\phi_{nit}$ be a $U$-Statistic of degree 2 where $\phi_{nit} \equiv \psi_{nit} + \psi_{nti}$ and we have $|E_{2n}| \leq C|U_n|$. The order of the $U$-Statistic can be analyzed by examining the orders of each component in its Hoeffding's $H$-decomposition in Hoeffding (1961). Let $w_i \equiv (z_{si}', z_{li}', x_{si,k}, x_{li,j})'$. In our case, we have $U_n = \theta_n + 2H_n^{(1)} + H_n^{(2)}$, where $\theta_n \equiv \mathbb{E}\left(\phi_{nit}\right) = 0$ as $\mathbb{E}\left(x_{st,k}^{*} \mid z_{st}\right) = 0$, $H_n^{(1)} \equiv \frac{1}{n}\sum_{i=1}^{n}\phi_{1n}(w_i) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\phi_{nit} \mid w_i\right) = 0$ as $\mathbb{E}\left(x_{st,k}^{*} \mid z_{st}\right) = \mathbb{E}\left(x_{lt,j}^{*} \mid z_t\right) = 0$, and $H_n^{(2)} = U_n$. By Theorem 1 in Yao & Martins-Filho (2015), the order of $H_n^{(2)}$ is determined by $n$ and the leading variance $\sigma_{2n}^2 \equiv \mathrm{Var}\left(\phi_n\right)$, specifically, $H_n^{(2)} = O_p\left((\sigma_{2n}^2/n^2)^{1/2}\right)$. It is straightforward that $\sigma_{2n}^2 = O_p(h^{-p_s})$ given $\mathbb{E}\left(x_{st,k}^{*2} \mid z_t\right) \leq C$ , which gives that $H_n^{(2)} = O_p\left((nh^{p_s})^{-1/2}n^{-1/2}\right) = o_p\left(n^{-1/2}\right)$. Therefore, we have $Q_{1n} = o_p\left(n^{-1/2}\right)$.

Analysis about the order of $Q_{2n}$ is similar. We use the same set of notations for the $U$-

Statistic and let $\psi_{nit} \equiv (h_s^{p_s} f_{sz}(z_{si}))^{-1} K_s\left(\frac{z_{st}-z_{si}}{h_s}\right)(g_{sx,k}(z_{st})-g_{sx,k}(z_{si}))x^*_{li,j}$. As $\psi_{nii}=0$, we have $Q_{2n} = \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{3n}$ and $|E_{3n}| \leq C|U_n|$ where $U_n \equiv \binom{n}{k}^{-1}\sum_{\substack{i=1\\ i<t}}^{n}\sum_{t=1}^{n}\phi_{nit} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ is a $U$-Statistic of degree 2 with $\phi_{nit} \equiv \psi_{nit}+\psi_{nti}$. Similar to the arguments for $Q_{1n}$, here let $w_i \equiv (z'_{si}, z'_{li}, x_{li,j})'$ and we have $\theta_n = 0$ and $\mathbb{E}\left(\psi_{nti} \mid w_i\right) = 0$ as $\mathbb{E}\left(x^*_{li,j} \mid z_i\right) = 0$. Then we have

$$
\begin{aligned}
\phi_{1n} &\equiv \mathbb{E}\left(\phi_{nit} \mid w_i\right) = \mathbb{E}\left(\psi_{nit} \mid w_i\right) \\
&= (h_s^{p_s} f_{sz}(z_{si}))^{-1} x^*_{li,j}\mathbb{E}\left(K_s\left(\frac{z_{st}-z_{si}}{h_s}\right)(g_{sx,k}(z_{st})-g_{sx,k}(z_{si})) \mid z_{si}\right),
\end{aligned}
$$

and $\sigma^2_{1n} \equiv \mathrm{Var}\left(\phi_{1n}\right) \leq \mathbb{E}\left(\phi^2_{1n}\right) = O_p\left(h_s^{2r_s}\right) = o_p(1)$ by Assumption A3 b) with a high-order kernel. For $\sigma^2_{2n}$, we have

$$
\begin{aligned}
\sigma^2_{2n} &= \mathrm{Var}\left(\phi_{nit}\right) \leq C\mathbb{E}\left(\psi^2_{nit}\right) \\
&= C\sigma^2_{lx,j}\mathbb{E}\left((h_s^{p_s} f_{sz}(z_{si}))^{-2} K_s^2\left(\frac{z_{st}-z_{si}}{h_s}\right)(g_{sx,k}(z_{st})-g_{sx,k}(z_{si}))^2\right) \\
&= O_p(h_s^{2-p_s}).
\end{aligned}
$$

By Theorem 1 in Yao & Martins-Filho (2015), we have $H_n^{(1)} = O_p\left((\sigma^2_{1n}/n)^{1/2}\right) = o_p(n^{-1/2})$ and $H_n^{(2)} = O_p\left((\sigma^2_{2n}/n^2)^{1/2}\right) = O_p\left(n^{-1/2}h_s(nh_s^{p_s})^{-1/2}\right) = o_p(n^{-1/2})$. Therefore, we have $Q_{2n} = o_p\left(n^{-1/2}\right)$ and the proof is complete. □

*Proof of Theorem 2.* By Theorem 1, we have $\widehat{\beta}_{s,\mathrm{sur}} - \beta_s = O_p\left(n^{-1/2}\right)$. Therefore,

$$
\begin{aligned}
\sqrt{nh_s^{p_s}}\left(\widetilde{\theta}_s(z) - \theta_s(z)\right) &= \sqrt{nh_s^{p_s}}\left(\widehat{g}_{sy}(z) - g_{sy}(z) - (\widehat{g}_{sx}(z) - g_{sx}(z))'\beta_s\right) \\
&\quad - \sqrt{nh_s^{p_s}}\left(\widehat{g}_{sx}(z) - g_{sx}(z)\right)'\left(\widehat{\beta}_{s,\mathrm{sur}} - \beta_s\right) - \sqrt{nh_s^{p_s}}g_{sx}(z)'\left(\widehat{\beta}_{s,\mathrm{sur}} - \beta_s\right),
\end{aligned}
$$

where the last two terms are $o_p(1)$ given the consistency of $\widehat{g}_{sx}(z)$ and $h_s = o(1)$. For the first term, its asymptotic properties can be analyzed in a similar manner as the proof of Theorem 4 by Geng et al. (2020). Actually the analysis is simpler here as there is no endogeneity issue and we do not need to use the so-called “instrument” function for identification of an additive structure of the nonparametric part. Following their Step 1 and 2 in the proof (we do not have the term in Step 3), we have

$$\sqrt{nh_s^{p_s}}\left(\widehat{g}_{sy}(z)-g_{sy}(z)-b_{sy}(z)\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_{sy}\right),$$

where $b_{sy}(z) \equiv h_s^{p_s} f_{sz}^{-1}(z)\mu_{k_s,r_s}\sum_{k=1}^{r_s}\frac{1}{k!(r_s-k)}\sum_{j=1}^{p_s}\mathrm{D}_j^k g_{sy}(z)\mathrm{D}_j^{r_s-k}f_{sz}(z)+o_p(h_s^{p_s})$ and $\mathcal{V}_{sy} = f_{sz}^{-1}(z)\sigma_{sy}^2\int K_s^2(\gamma)\,\mathrm{d}\gamma$. Extending this analysis to the term $\sqrt{nh_s^{p_s}}\left(\left(\widehat{g}_{sx}(z)-g_{sx}(z)\right)'\beta_s\right)$ and given that $\theta_s(z)=g_{sy}(z)-g_{sx}(z)'\beta_s$, we have

$$\sqrt{nh_s^{p_s}}\left(\widehat{g}_{sy}(z)-g_{sy}(z)-\left(\widehat{g}_{sx}(z)-g_{sx}(z)\right)'\beta_s-b_{s,1}(z)\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_{s,1}\right),$$

where $b_{s,1}(z) \equiv h_s^{p_s} f_{sz}^{-1}(z)\mu_{k_s,r_s}\sum_{k=1}^{r_s}\frac{1}{k!(r_s-k)}\sum_{j=1}^{p_s}\mathrm{D}_j^k \theta_s(z)\mathrm{D}_j^{r_s-k}f_{sz}(z)+o_p(h_s^{p_s})$ and $\mathcal{V}_{s,1} = f_{sz}^{-1}(z)\sigma_{ss}\int K_s^2(\gamma)\,\mathrm{d}\gamma$. The proof is complete. $\square$

# Appendix B Data summary and translog parameters

Table B1: Summary Statistics of the Variables

| Symbol | Variable name | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| $C$ | Total cost (thousands USD) | 148,745.76 | 581,477.33 | 637.05 | 9,795,565.82 |
| $O_1$ | Loans (thousands USD) | 1,586,346.49 | 6,297,792.74 | 5,501.32 | 103,512,670.90 |
| $O_2$ | Other earning assets (thousands USD) | 1,044,387.57 | 4,490,291.04 | 5,005.51 | 90,949,857.57 |
| $O_3$ | Off balance sheet (thousands USD) | 785,074.86 | 4,290,790.06 | 74.60 | 76,676,128.04 |
| $W_1$ | Capital price | 0.93 | 0.43 | 0.00 | 2.41 |
| $W_2$ | Labor price | 0.02 | 0.00 | 0.00 | 0.03 |
| $W_3$ | Purchased funds price | 0.04 | 0.01 | 0.01 | 0.07 |
| $S_1$ | Capital share | 0.26 | 0.07 | 0.05 | 0.83 |
| $S_2$ | Labor share | 0.31 | 0.07 | 0.02 | 0.51 |
| $S_3$ | Purchased funds share | 0.43 | 0.10 | 0.09 | 0.88 |
| **Environmental variables** ($z$) | | | | | |
| Credit risk | Loan loss provisions in log | 7.18 | 1.92 | 4.31 | 14.02 |
| Solvency risk | Equity capital amount in log | 10.72 | 1.59 | 7.38 | 16.48 |
| Liquidity risk | Amount of liquidity assets in log | 11.68 | 1.58 | 7.49 | 17.55 |
| Time ($t$) | Time trend | 4.97 | 2.32 | 1.00 | 10.00 |

**Note**: (1) The number of obs. is 2,989. (2) Time $t$ is calculated as ($year - 1995$), where $year$ goes from 1996 to 2005.

Table B2: Translog Cost Function Parameter Estimates

| | Partially linear SUR | | Partially linear single-equation | | Linear SUR | |
|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| $\beta_{22}$ | 0.1630 | 0.0019 | 0.1506 | 0.0182 | 0.1569 | 0.0021 |
| $\beta_{23}$ | -0.1305 | 0.0015 | -0.1282 | 0.0110 | -0.1256 | 0.0016 |
| $\rho_{21}$ | 0.0275 | 0.0014 | -0.0052 | 0.0049 | 0.0325 | 0.0016 |
| $\rho_{22}$ | -0.0124 | 0.0012 | 0.0130 | 0.0064 | -0.0162 | 0.0015 |
| $\rho_{23}$ | -0.0115 | 0.0009 | -0.0256 | 0.0023 | -0.0122 | 0.0011 |
| $\beta_{33}$ | 0.1640 | 0.0020 | 0.1576 | 0.0139 | 0.1680 | 0.0019 |
| $\rho_{31}$ | -0.0163 | 0.0011 | -0.0163 | 0.0017 | -0.0501 | 0.0019 |
| $\rho_{32}$ | 0.0018 | 0.0009 | 0.0042 | 0.0047 | 0.0327 | 0.0017 |
| $\rho_{33}$ | 0.0166 | 0.0011 | 0.0190 | 0.0019 | 0.0193 | 0.0012 |
| $\alpha_{11}$ | 0.0353 | 0.0069 | 0.0445 | 0.0067 | 0.1626 | 0.0103 |
| $\alpha_{12}$ | -0.0220 | 0.0041 | -0.0201 | 0.0039 | -0.1009 | 0.0072 |
| $\alpha_{13}$ | -0.0028 | 0.0013 | -0.0089 | 0.0017 | -0.0568 | 0.0049 |
| $\alpha_{22}$ | 0.0334 | 0.0138 | 0.0414 | 0.0127 | 0.1628 | 0.0084 |
| $\alpha_{23}$ | 0.0082 | 0.0015 | 0.0074 | 0.0014 | -0.0489 | 0.0052 |
| $\alpha_{33}$ | 0.0082 | 0.0023 | 0.0169 | 0.0023 | 0.0882 | 0.0038 |
| $\beta_{2}$ | 0.4720 | 0.0086 | 0.9194 | 0.0835 | 0.4567 | 0.0107 |
| $\beta_{3}$ | 0.4478 | 0.0110 | 0.3145 | 0.0666 | 0.4978 | 0.0113 |
| $\alpha_{1}$ | 0.3925 | 0.0720 | 0.1943 | 0.0692 | 0.2510 | 0.0488 |
| $\alpha_{2}$ | 0.1531 | 0.1589 | 0.1373 | 0.1467 | 0.1565 | 0.0488 |
| $\alpha_{3}$ | -0.1051 | 0.0210 | -0.1566 | 0.0207 | 0.4010 | 0.0305 |

**Note**: The translog parameter estimates of the input prices and outputs, along with their standard errors, are reported for the three models.