CAIRN Commission Paper- Final Draft

Stochastic Frontier Approach of Measuring Agricultural Productivity and Efficiency: Accounting for Innovations

BY

Alexander B. Darku[1], Stavroula Malla[2] and Kien C. Tran[3]

[1 2 3] *Department of Economics, University of Lethbridge, Alberta, Canada*

December 13, 2012

CAIRN
AN ERCA RESEARCH NETWORK

Stochastic Frontier Approach of Measuring Agricultural Productivity and Efficiency: Accounting for Innovations

Alexander B. Darku[†]
Stavroula Malla[•]
and
Kien C. Tran[±]

ABSTRACT

It has been recognized that modeling agricultural productivity growth and inefficiency is a challenging task since there exists many methodologies and each one has its own merits and drawbacks. The main purpose and motivation of this paper is to offer some new modeling approaches to overcome the heterogeneity problem in measuring productivity and efficiency using stochastic frontier modeling. As an extension, the paper also shows how to incorporate endogeneity of the input in the production process by mean of auxiliary information. Regarding the estimation, the paper proposed various procedures that are fairly easy to implement, given the current existing computing power and readily automated software packages. Finally, the paper suggests a way to decompose the TFP growth and efficiency measurements for policy formulation purposes.

Final Draft

December 2012

[†] [•] [±] Department of Economics University of Lethbridge, Alberta Canada

** *A commission paper written for CAIRN Network.*

1. Introduction

Agricultural productivity growth in Canada is a topic of continuing interest to policy makers and researchers who aim to improve on economic sustainability, efficiency, living standards, and international competitiveness. Recent evidence suggests that agricultural productivity growth in Canada has significantly slowed down as well as lagging behind that of the U.S. and many OECD countries (see for example, Veeman and Gray (2009, 2010)).

Measuring productivity growth in agriculture sector is imperative but a challenging task. There exist many methodologies for measuring productivity growth and efficiency, and each method has its own advantages and disadvantages depending on the research question at hand. For example, Index Numbers is quite simple to implement but it is susceptible to the noise in the data as well as measurement errors. In addition, it does not allow for the measurements of technical change, scale effects or technical inefficiency. On the other hand, linear programming approach such as Data Envelopment Analysis (DEA) is more flexible in the sense that it is robust to the functional form of the production frontier and it can accommodates various production constraints and efficiency measurement. It is nevertheless, still susceptible to the noise in the data and outliers. Furthermore, as in Index Number approach, it not clear how heterogeneity and measurement errors can be handled in such framework. Recent work to make the DEA approach stochastic seems to be promising. In the approach, the original DEA is combined with regression analysis to overcome some of these deficiencies, see for example Simar and Wilson (2007). Similarly, by combining Index Numbers approach with some econometric techniques, one can overcome some of existence problems mentioned but at the expense of increasing computational complexity and imposition of more restrictive assumptions, see for example Steward (2006), Steward et al (2009). However, the combined approaches, albeit can overcome some of the deficiencies, introduces a new problem of how to incorporate the estimation errors from using the regression analysis. If a practitioner is fortunate enough to have in hand a relatively large data set, this estimation errors is negligible. However, the same cannot be said when the data set is relatively small. Finally, econometric approaches, such as parametric and nonparametric/semiparametric stochastic frontier, can provide much more flexible models that can accommodate heterogeneity and measurement errors problems. However, the cost of this flexibility is the imposition of additional distributional assumptions on the error structures and it is much more complicated to implement.

Generally speaking, there exist no such unified approaches that are computational simple, and at the same time, impose minimal restrictions on the model. Thus, in practice, applied researchers often need to decide what sacrifices they are willing to make in term of

computational simplicity and reliability of the results when choosing a particular approach or approaches in examining the total factor productivity growth. As Tsionas and Kumbhakar (2004) point out, the approach or approaches that incorporate inefficiency is more attractive for several reasons. First, it helps to identify which farmers/regions/provinces are inefficient and, if so, to what extent. By identifying the inefficient farmers/regions/provinces, policies designed to promote efficiency can be made more effective by directing the necessary aid to those who are in the greatest need of assistance. Second, after identifying the presence of inefficiency, it is natural to examine the factors responsible for inefficiency (e.g., Battese and Coelli 1995). Once some explanatory factors are found, programs can be designed to improve efficiency of the affected farmers/regions/provinces.

In most agriculture applications of production frontier models, frontier parameters are often assumed to be the same for all cross-section units and time. Such assumption might be questionable, especially in the application where the number of time observations is large. In such a case, it is reasonable to ask: Are the production frontier parameters constant across production units? Are they time-invariant? These questions, in the production frontier literature and especially in agriculture, have not been adequately addressed.

The main goal of this paper is to offer an alternative modeling approach to incorporate inefficiency and to allow for the production parameters to differ across cross-section unit as well as over time. Specifically, the production parameters are allowed to depend on a set of factors that are different across units and over time. This is particularly important avenue for agricultural productivity since factors such as innovations and/or past and current research and development (R&D) investments can have significant positive impact on the production parameters that can enhance total factor productivity.

The rest of the paper is organized as follows. Section 2 presents the general parametric and semiparametric stochastic frontier models that accommodates the production heterogeneity. A simple decomposition method of total factor productivity growth is discussed in Section 3. Section 4 provides various parametric and semiparametric estimation approaches of the models presented in section 2. Hypothesis testing for selecting the appropriate model is given in Section 5. Section 6 extends the model to include endogeneity and/or measurement errors of the inputs used in the production process. Section 7 briefly discusses data limitation. Section 8 concludes the paper.

2. Stochastic Production Frontier Models

In this section we will describe various ways to model heterogeneity in production frontier. We start with the most general model and then provide discussion of various special cases by imposing certain parameters restrictions on the frontier.

Consider the following stochastic production frontier model:

$$y_{it} = \alpha(z_{it}) + x_{it}^{'}\beta(z_{it}) + v_{it} - u_{it}, \quad i = 1,\ldots,n; t = 1,\ldots,T \tag{1}$$

where prime denotes transpose of a vector or matrix, $y_{it} = \ln Y_{it}$ is output of, x_{it} is a $k \times 1$ vector of logarithm of inputs of farm i at time t and other factors that can influence output. These include labor, capital, materials, fertilizers, time trend, etc., z_{it} is a $q \times 1$ vector of covariates that are not part of x_{it}, v_{it} is a random symmetric errors representing factors that are beyond the farmers control, $u_{it} \geq 0$ is a one-sided random errors representing technical inefficiency in which farmers are not implementing the best possible practice in the production process, $\delta(.) = (\alpha(.), \beta(.)^{'})^{'}$ is a $(k+1) \times 1$ vector of unspecified smooth function of z_{it}. Model (1) would be appropriate when, say, if one wants to allow for heterogeneity in the production function and the source of heterogeneity is known a priori. By way of example, let $x_{it} = (labor_{it}, capital_{it})$ and $z_{it-j} = R\&D_{it-j}$, then model (1) suggests that the labor and capital input coefficients may vary directly with the firm's past R&D investment input. Thus, both marginal productivity of labor and capital depend on the firm's R&D past values. As a result, returns to scale and total factor productivity growth may also be function of the past R&D. In a special case where $\beta(.) = \beta$, model (1) reduces to a partially linear stochastic frontier model discussed in Park, Sickle and Simar (1998). The partially linear stochastic frontier model assumes the slope coefficients β are invariant with respect to the past R&D, hence the past R&D variable can only shift the level of the production frontier. In this case, the past R&D variable is said to have a 'neutral' effects on the production frontier. In contrast, model (1) allows the past R&D to affect the stochastic frontier 'nonneutrally.'

Another example is when z_{it} is taken to be quality of capital (i.e., age of capital stock) or quality of labour (i.e., educational attainment), then model (1) takes into account of quality of inputs that is known to affect the factor productivity growth. Last but not least, one might be interested in examining how the input marginal productivities and total factor productivity growth/decay over certain periods of time without splitting the sample (in this case, $z_{it} = \tau = t / T$). Thus, model (1) provides a flexible way to model heterogeneity (including time

heterogeneity) among farmers through the input marginal productivities and that the source of heterogeneity is known.

Model (1) contains other interesting models as special cases. First, if we set $\alpha(z_{it}) = \alpha_0 + z_{it}^{'}\alpha_1$, and $\beta(.) = \beta$, then model (1) becomes:

$$y_{it} = \alpha_0 + z_{it}^{'}\alpha_1 + x_{it}^{'}\beta + v_{it} - u_{it} \tag{2}$$

In this case, if z_{it} is the past R&D investment, then model (2) considers past R&D values explicitly as part of the inputs in the production process, and it is assumed to neutrally shift the production function so that the source of heterogeneity is captured through the intercept. Second, if $\alpha(z_{it}) = \alpha$ and $\beta(z_{it}) = \beta_0 + z_{it}^{'}\beta_1$, then (1) becomes:

$$y_{it} = \alpha + x_{it}^{'}\beta_0 + (z_{it}x_{it})^{'}\beta_1 + v_{it} - u_{it} \tag{3}$$

and in this case, the source of heterogeneity is captured through the slopes by allowing say, the past R&D investment to interact with the input process of the production. Liam and Miller (2004) construct a model similar to (3) to examining the total factor productivity growth for OECD countries. They used the quality of capital and education as the sources of heterogeneity instead of past R&D investments. Finally, by allowing for the intercept and slopes to vary with z_{it}, so that

$$y_{it} = \mu + z_{it}^{'}\alpha + x_{it}^{'}\beta_0 + (z_{it}x_{it})^{'}\beta_1 + v_{it} - u_{it} \tag{4}$$

is a combination of model (2) and (3), and it is easy to see that model (4) nests model (2) and (3) as special cases. Consequently, standard F-statistics can be used to select the appropriate model from (2)-(4). However, choosing between model (1) and (4) is not obvious and it requires a different type of test statistic because model (1) is a semiparametric model. We will return to this discussion in the later section. For the rest of the paper, we will focus our attention on model (1) and (4) only.

3. Total Factor Productivity (TFP) Growth Decomposition

At its most basic form, TFP growth can be defined as the increase in the output not accounted for by the increase inputs that are used in the production process, and efficiency gains from using resources more efficiently by applying practices from the present stock of knowledge. In this sense, TFP growth is just a residual measure that accounts for the effect that factors other than input growth and efficiency have on output growth.

Conceptually, from the production frontier perspective, TFP growth can be decomposed into three components: scale effects, technological change, and changes in the degree of technical efficiency. Scale effects refer to the proportionate increase in output due to proportionate increase in all inputs in the production process, i.e., in the presence of increasing economies of scale suggests that the production of additional outputs will require a less proportionate increase in inputs use. Technological change relates to the technological progress including not only advances in physical technologies but also innovation in the overall knowledge base that lead to better decision making and planning. Improvements in the degree of technical inefficiency refer to the situation where resources can be used more efficiently by applying best practice from the present stock of knowledge.

The relative contributions of technical change, scale and inefficiency to productivity growth in Canadian agricultural sector can offer important insight into the impact of different policies on productivity growth. For instance, productivity growth composed largely of technical change points to the primacy of technology development and adoption in productivity advance, suggesting that research and development is an appropriate policy tool. On the other hand, if a scale effect is found to play a significant role, indicating that structural change in agricultural sector as the main driver of productivity growth, and therefore policy design to accelerate structural change may enhance productivity growth.

The TFP decomposition can be constructed as follows. For a given variable w, let $\dot{w}$ denotes the time derivative of w (i.e., $\partial w / \partial t$), then from the standard definition of TFP growth, differentiating (1) with respect to time, we obtain the following:

$$\dot{y}_{it} = \alpha_z(z_{it})\dot{z}_{it} + \dot{x}_{it}^{'}\beta(z_{it}) + x_{it}^{'}\beta_z(z_{it})\dot{z}_{it} + \dot{v}_{it} - \dot{u}_{it} \tag{5}$$

where $\alpha_z(.) = \dfrac{\partial \alpha(.)}{\partial z^{'}}$, $\beta_z(.) = \dfrac{\partial \beta(.)}{\partial z^{'}}$. Similarly, differentiating (4) with respect to time, yields

$$\dot{y}_{it} = \dot{z}_{it}^{'}\alpha + \dot{x}_{it}^{'}\beta_0 + (\dot{z}_{it}x_{it} + z_{it}\dot{x}_{it})^{'}\beta_1 + \dot{v}_{it} - \dot{u}_{it} \tag{6}$$

From (5) and (6), the TFP growth can be computed as:

$$TFP = \dot{y}_{it} - \alpha_z(z_{it})\dot{z}_{it} + \dot{x}_{it}^{'}\beta(z_{it}) + x_{it}^{'}\beta_z(z_{it})\dot{z}_{it} - \dot{u}_{it} \tag{7a}$$

$$TFP = \dot{y}_{it} - \dot{z}_{it}^{'}\alpha + \dot{x}_{it}^{'}\beta_0 + (\dot{z}_{it}x_{it} + z_{it}\dot{x}_{it})^{'}\beta_1 - \dot{u}_{it} \tag{7b}$$

Equation (7a) states that the TFP growth can be decomposed into output growth $\dot{y}_{it}$, input growth $\dot{x}_{it}^{'}\beta(z_{it})$, change in the exogenous factors that affecting the marginal productivity overtime, $\alpha_z(z_{it})\dot{z}_{it} + x_{it}^{'}\beta_z(z_{it})\dot{z}_{it}$ (that is, if z_{it} is to be interpreted as past R&D investment, then this term measures the changes in past R&D investment overtime that changes the inputs marginal productivity), and change in technical inefficiency, $\dot{u}_{it}$. Note that, since x_{it} contains a time trend, the input growth can further be decomposed into a scale effects and technical change. Similar interpretation of the TFP growth decomposition can be given for equation (7b). In practice, to compute the TFP growth decomposition, we usually use $\Delta \ln y_{it} = (\ln y_{it} - \ln y_{it-1})$ to approximate the time derivative variable $\dot{y}_{it}$, and similarly for other time derivative variables, and the unknown parameters and/or unknown functions are replaced by their consistent estimates. We now turn our attention to the estimation of the unknown parameters and/or the unknown functions.

4. Estimation

We will first provide a discussion on the estimation of model (4) since this model nests the other two models as special cases. Then we will generalize the estimation approach to model (1).

4.1. Global Maximum Likelihood (GML) Approach:

At the first glance, in the absence of the one-side error term u_{it} ($u_{it} = 0$. i.e., farmers are technically efficient), model (4) is nothing more than just a simple multiple regression model. Thus, in principle a simple ordinary least squared (OLS) approach can be used to estimate the parameters of the model. However, in the presence of u_{it} ($u_{it} > 0$, i.e., farmers are technically inefficient), more careful treatment is needed in using OLS approach because the composed errors $\epsilon_{it} = v_{it} - u_{it}$ do not have zero mean and hence OLS will lead to bias in the estimate of the intercept as well as inefficient estimation of the slope parameters. Fortunately, given the

structure of the model, the direction and the magnitude of the bias are known for this model, and hence a simple corrected OLS (COLS) procedure can be used to obtain the parameter estimates, see for example, Kumbhakar and Lovell (2000).

Alternatively, one could use the Maximum Likelihood (ML) approach to obtain consistent and efficient estimation of all the parameters in (4). To do this, one needs to further make explicit assumption regarding the distributional of v_{it} and u_{it}. Following standard practice, v_{it} is assumed to be *i.i.d.* $N(0,\sigma_v^2)$ and u_{it} is *i.i.d.* $\left|N(0,\sigma_u^2\right|$. Of course, other distributional assumptions on u_{it} such as exponential, truncated normal or Gamma can be used in place of the half-normal. However, previous works indicated that the ML estimates are not very sensitive to the distributional choice of u_{it} (see for example, Greene 2002).

Given the distributional assumptions of v_{it} and u_{it}, the conditional probability density function of $\varepsilon_{it} = v_{it} - u_{it}$ given x_{it} is given by:

$$f(\varepsilon_{it} \mid x_{it}) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(-\frac{\lambda\varepsilon_{it}}{\sigma}\right)$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_v$, $\phi(.)$ and $\Phi(.)$ are the probability density function (pdf) and cumulative distribution function (CDF) of a standard normal variable. In order to avoid nonnegativity restrictions on the variance parameters σ^2 and λ, we choose to re-parameterize these parameters as $\tilde{\sigma}^2 = \ln(\sigma^2)$ and $\tilde{\lambda} = \ln(\tilde{\lambda})$. Thus, the conditional pdf of y_{it} given x_{it} is:

$$g(y_{it} \mid x_{it}) = \frac{2}{\tilde{\sigma}}\phi\left(\frac{y_{it} - \mu - z_{it}^{'}\alpha - x_{it}^{'}\beta_0 - (z_{it}x_{it})^{'}\beta_1}{\tilde{\sigma}}\right)\Phi\left(-\frac{\tilde{\lambda}}{\tilde{\sigma}}(y_{it} - \mu - z_{it}^{'}\alpha - x_{it}^{'}\beta_0 - (z_{it}x_{it})^{'}\beta_1)\right)$$

and the conditional log-likelihood function for a sample of nT observations is given by:

$$\ln L(\theta) = -\frac{nT}{2}\ln 2\pi + \ln\tilde{\sigma}^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it} - \mu - z_{it}^{'}\alpha - x_{it}^{'}\beta_0 - (z_{it}x_{it})^{'}\beta_1 / \tilde{\sigma}^{2} +$$

$$\sum_{i=1}^{n}\sum_{i=1}^{T}\ln\Phi\left(-\frac{\tilde{\lambda}}{\tilde{\sigma}}(y_{it} - \mu - z_{it}^{'}\alpha - x_{it}^{'}\beta_0 - (z_{it}x_{it})^{'}\beta_1)\right) \qquad (8)$$

where $\theta = (\mu, \alpha, \beta_0, \beta_1, \tilde{\sigma}^2, \tilde{\lambda})$. Maximizing (8) with respect to θ yields the ML estimates. Note that, the log-likelihood function in (8) is highly non-linear and hence some sort of numerical algorithms such as David-Fletcher-Powell (DFP) or Gauss-Newton (GN), and starting values must be used to obtain the parameters estimates. However, given the current existing computing power and readily automated stochastic frontier estimation programs such as "FRONTIER" or "LIMDEP", the ML procedure is very simple to implement.

Once the parameters are estimated, the estimate of technical inefficiency term u_{it} can be constructed based on Jondrow et al. (1982) prediction formula:

$$\hat{u}_{it} = E(u_{it}|\varepsilon_{it}) = \frac{\hat{\sigma}\hat{\lambda}}{1+\hat{\lambda}^2}\left[\frac{\phi(\hat{\varepsilon}_{it}\hat{\lambda}/\hat{\sigma})}{\Phi(\hat{\varepsilon}_{it}\hat{\lambda}/\hat{\sigma})} - \frac{\hat{\varepsilon}_{it}\hat{\lambda}}{\hat{\sigma}}\right] \tag{9}$$

where $\hat{\varepsilon}_{it}$, $\hat{\sigma}$ and $\hat{\lambda}$ are the ML estimates of ε_{it}, σ and λ respectively. As usual in the frontier models, if the variables are measured in logs, a point estimate of the technical efficiency is then provided by $EFF_{it} = exp(-\hat{u}_{it}) \in [0,1]$.

4.2. Local Maximum Likelihood (LML) Approach

The parametric approach in modeling the marginal productivities as linear functions of the z_{it} as in model (4), although simple but very restrictive, and it is susceptible to functional form misspecification. Accordingly, it is generally more desirable to work with more general and flexible function to avoid misspecification in the functional form. To this end, we suggest to model the marginal productivities nonparametrically as in specification of model (1). That is, we assume the input coefficients are smooth function of the z_{it}, but these functions are unknown and need to be estimated.

For convenient and with slightly abuse of notation, we rewrite model (1) as:

$$y_{it} = x_{it}^{'}\beta(z_{it}) + v_{it} - u_{it} \tag{10}$$

where $x_{it} = (1, x_{it})$ and $\beta(z_{it}) = (\alpha(z_{it}), \beta(z_{it}))$. For the remainder of this section, we will use (10) as our semiparametric model instead of (1). For a given value of z and for z_{it} in the neighborhood of z, using Taylor series expansion, the vector of unknown function, $\beta(z_{it})$ in (10) can be approximated by local linear functions (or higher order polynomial, if one desires) as:

$$\beta(z_{it}) \cong \beta(z) + \beta'(z)(z_{it} - z) = a(z) + b'(z)(z_{it} - z) \tag{11}$$

where "prime" denotes the derivative vector of the function evaluated at z. One convenient feature of "local linear" method is that it automatically provides an estimator for the first-order derivatives of $\beta(z)$ which is needed in the computation of the TFP growth. Replacing $\beta(z_{it})$ in (10) with (11) we obtain:

$$y_{it} = x'_{it}a(z) + x'_{it}(z_{it} - z)'b(z) + v_{it} - u_{it} \tag{12}$$

In the absence of the technical inefficiency (i.e., $u_{it} = 0$), the parameters in (12) can be estimated using local linear least squares (LLLS) approach (see for example, Li and Racine (2007).) by minimizing:

$$S(a,b) = \sum_{i=1}^{n}\sum_{t=1}^{T} \left[y_{it} - x'_{it}a(z) - x'_{it}(z_{it} - z)'b(z) \right]^2 K\left(\frac{z_{it} - z}{h}\right) \tag{13}$$

where $K(.)$ is known as kernel function (or kernel weight) which assigns more weight to observation z_{it} that is closed to z, and less weight when z_{it} is far away from z, h is referred to as a smoothing parameter (or alternatively, a bandwidth or window width). Note that for a given value of z, the minimization problem in (13) is similar to the global parametric Generalized Least Squared (GLS) method and hence it is quite simple to compute. Let $\{\hat{a}(z), \hat{b}(z)\}$ be the solutions to (13), then it is easy to see that a consistent estimators of $\beta(z)$ and $\beta'(z)$ are given by $\hat{\alpha}(z)$ and $\hat{b}(z)$, respectively.

When the technical inefficiency term is presence, as in most cases when estimating stochastic frontier models, the simplest approach is to use a localized version of the global parametric COLS, which essentially is the same problem as in (13) except once the parameter estimates are obtained, one needs to correct for the bias in the estimate of the intercept term. However, this approach generally will produce inefficient parameter estimates, and hence it is not recommended. To obtain more efficient estimation, we can use the localized version of the MLE (see for example Tran and Tsionas (2011)). To do this, two minor modifications of the log-likelihood in (8) are needed. First, we need to localize the term ε_{it} so that $\varepsilon_{it}(z) = y_{it} - x'_{it}a(z) - x'_{it}(z_{it} - z)'b(z)$, and second, assign a kernel weight to the log-likelihood function. Thus, the conditional local log-likelihood function can be written as:

$$\ln L(\delta(z)) = -\frac{nT}{2}(\ln 2\pi + \ln \tilde{\sigma}^2)K\left(\frac{z_{it}-z}{h}\right) - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{(y_{it}-x_{it}^{'}a(z)-x_{it}^{'}(z_{it}-z)^{'}b(z))}{\tilde{\sigma}}\right)^2 K\left(\frac{z_{it}-z}{h}\right)$$

$$+\sum_{i=1}^{n}\sum_{t=1}^{T}\ln\Phi\left(-\frac{\tilde{\lambda}(y_{it}-x_{it}^{'}a(z)-x_{it}^{'}(z_{it}-z)^{'}b(z))}{\tilde{\sigma}}\right)K\left(\frac{z_{it}-z}{h}\right) \tag{14}$$

where $\delta(z) = \{a(z),b(z),\tilde{\sigma}^2,\tilde{\lambda}\}$. Note that since the variance parameters in (14) are constant whilst the production parameters are dependent of the localization of z, hence in order to obtain the local maximum likelihood estimates of all the parameters in (14), we recommend the use of the two-step (or preferably iterative) profile maximum likelihood algorithm of Tran and Tsionas (2011) which is described as follows.

Step 1: Use a linear parametric 'anchorage' model for $\beta(z_{it})$, that is, set $\beta(z_{it}) = z_{it}^{'}\alpha$ as in model (4) and obtain an initial ML estimate $\hat{\theta}^{(j)} = (\hat{\mu}^{(j)},\hat{\alpha}^{(j)},\hat{\beta}_0^{(j)},\hat{\beta}_1^{(j)},\hat{\tilde{\sigma}}^{2(j)},\hat{\tilde{\lambda}}^{(j)})$ where $j=0$ (i.e., maximize the conditional log-likelihood function (8)).

Step 2: For each z_{it} $(i=1,\cdots,n;t=1,\cdots T)$ in the sample, fix $(\tilde{\sigma}^2,\tilde{\lambda})$ at its initial estimates $(\hat{\tilde{\sigma}}^{2(j)},\hat{\tilde{\lambda}}^{(j)})$ from Step 1, and maximize the conditional *local* log-likelihood (14) with respect to $(a(z),b(z))$. Note that if the sample size nT is large, (14) could be performed on a random subsample R where $R \ll nT$ to reduce the computational burden.

Step 3: From step 2, fix $(a(z),b(z))$ at its estimates $(\hat{a}^{(j)}(z),\hat{b}^{(j)}(z))$, and maximize the conditional *global* log-likelihood function (8) to obtain $(\hat{\tilde{\sigma}}^{2(j+1)},\hat{\tilde{\lambda}}^{(j+1)})$.

Step 4: Using $(\hat{\tilde{\sigma}}^{2(j+1)},\hat{\tilde{\lambda}}^{(j+1)})$, repeat step 2 and step 3 until $\hat{\delta}(z) = \{\hat{a}(z),\hat{b}(z),\hat{\tilde{\sigma}}^2,\hat{\tilde{\lambda}}\}$ converges.

Step 5: Use the converged value $(\hat{a}(z),\hat{b}(z))$ from step 4 and compute $\hat{\beta}(z) = \hat{a}(z)$ and $\hat{\beta}^{'}(z)$ (the derivative vector $\beta^{'}(z)$.)

To implement the above algorithm (as well as the local linear estimators in (13)) in practice, one is required to specify the kernel weighting function and the bandwidth (or smoothing parameter). Typically in both theoretical and practical setting, nonparametric/semiparametric estimation has been established as relatively insensitive to choice of the kernel weighting function; and this insensitivity property carries with the profile MLE as

well. Thus a standard normal density is often used as a kernel function. However, the same cannot be said for the bandwidth selection. Different bandwidth can generate radically different impressions of the underlying function; hence for sound analysis and inference, an optimal principle must be adopted in selecting the bandwidth. In this paper we suggest an optimal bandwidth that is simple to use which is based on a simple cross-validation procedure. First, select h as:

$$h = h_{base} s_z (nT)^{-1/5}$$

where s_z is the vector of empirical standard deviations of the random variables vector z. So the bandwidth is adjusted for different scales of the variables and different sample sizes. Then the cross-validation criterion is evaluated for a grid of values for h_{base}.
The cross-validation proceeds as follows. For a given value of h_{base} compute

$$CV(h_{base}) = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} [y_{it} - \hat{m}^{(it)}(z) - \hat{u}_{it}^{(it)}]^2 \tag{15}$$

where $\hat{m}^{(it)} = x_{it}^{'(it)}(z_{it} - z)' \hat{\beta}(z)$ and $\hat{u}_{it}^{(it)}$ are the leave-one-out version of the local linear estimators described in (13) (i.e, use subsample where the observation (i,t) is deleted from the full sample.) The optimal value for h_{base} is then easily found by an appropriate grid search that minimizes (15). Note that if the sample size nT is large, maximization process of (14) and the evaluation in (15) can be computation intensive. To reduce this computational burden, a random subsample of say, r units, where $r \ll nT$ could be used.
Alternatively, one could use a "rule-of-thumb" bandwidth selection criterion such as $h_j = s_{z_j} (nT)^{-1/(4+q)}$ where s_{z_j} is the sample standard deviation of $\{z_{j,it}\}$, $j = 1, \cdots, q$. This selection approach is much simpler and requires little computation.

Once we have the model parameter estimates, the estimates of individual technical inefficiency over time can easily be obtained using (9) and these in turn, can be used to compute the TFP growth as in (7).

5. Testing Parametric versus Semiparametric Models

From our discussion in the previous section, even though model (1) is more general than model (4), one may want to estimate the parametric model (4), if it is, in fact the true model. In finite (or small) sample, it is usually more efficient to estimate the correct specified parametric model

than estimate a semiparametric model such as (1). However, if (1) is a correct specification and model (4) is not, the estimation results based on misspecification parametric model (4) will usually lead to inconsistent estimation results and hence will have serious consequences in computing the TFP growth. Therefore, in practice, it is of interest to test whether the parametric model (4) is an adequate representation of the data.

Let $\alpha^*(z_{it}) = \alpha_0 + z_{it}'\alpha_1$ and $\beta^*(z_{it}) = \beta_0 + z_{it}'\beta_1$, the null hypothesis that a parametric model (4) is a correct specification can be written as follows:

$$H_0 : \alpha(z_{it}) - \alpha^*(z_{it}) = 0 \text{ and } \beta(z_{it}) - \beta^*(z_{it}) = 0$$

almost everywhere. The alternative hypothesis is that model (1) is the correct specification but not model (4):

$$H_1 : \alpha(z_{it}) - \alpha^*(z_{it}) \neq 0 \text{ and } \beta(z_{it}) - \beta^*(z_{it}) \neq 0$$

on a set of positive measure. Li et al. (2002) proposed a semiparametric test similar to the above hypothesis and it can easily be extended to our case. Specifically, the proposed test statistic is given by:

$$\hat{\mathcal{I}}_{NT} = \frac{\hat{\sigma}_v}{(nT)^2 h^q} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j \neq i}^{n} \sum_{s \neq t}^{T} x_{it}' x_{js} \hat{v}_{it} \hat{v}_{js} K\left(\frac{z_{it} - z_{js}}{h}\right) \tag{16}$$

where $\hat{v}_{it} = y_{it} - \hat{\alpha}_0 - z_{it}'\hat{\alpha}_1 - x_{it}'\hat{\beta}_0 - (z_{it}x_{it})'\hat{\beta}_1 + \hat{u}_{it}$ is the residual obtained from the parametric model (4), $\hat{\sigma}_v^2 = \frac{1}{(nT)^2 h^q} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j \neq i}^{n} \sum_{s \neq t}^{T} \hat{v}_{it}^2 \hat{v}_{js}^2 (x_{it}' x_{js})^2 K^2(\frac{z_{it} - z_{js}}{h})$, and the same kernel function as in (14) can be used. Under certain regularity conditions, the test statistic $\hat{\mathcal{I}}_{NT} \sim N(0,1)$. Note that the test statistic (16) is actually a one-sided test in the sense that when $n \to \infty$, $\hat{\mathcal{I}}_{nT} \to \infty$ under the alternative hypothesis H_1. Thus in practice, one rejects H_0 if $\hat{\mathcal{I}}_{nT} > c_\alpha$ at the significant level α, where c_α is the upper α^{th} percentile from a standard normal distribution.

6. Extension

Insofar, the above discussion confines to the case of heterogeneity in the production process. Another important problem that often arises in practice, but usually has been ignored, is when

the inputs are endogenous or measured with errors. In the absence of the inefficiency, the solutions to the problem of measurement errors or endogeneity in standard regression model are well understood and well developed. However, in the context of stochastic frontier framework, this problem is more difficult to solve. The main reason is the difficult of treating the measurement errors or endogeneity with the inefficiency simultaneously. In this extension section, we provide some remedies to overcome this difficulty. Specifically, our discussion is based on some of the recent development on this particular problem by Kutlu (2010), Tran and Tsionas (2012).

To simplify our discussion as much as possible, we will confine it to the simplest case of parametric production frontier where there is one endogenous input (or one variable measures with error) only. Extension to the case where there are more than one endogenous input is straightforward and it does not affect the estimation procedure described below (see Tran and Tsionas (2012)).

Consider the following simpler version of stochastic production frontier model of (4) with one endogenous (or measurement error) input:

$$y_{it} = \mu + z_{it}\alpha + x_{it}\beta_0 + (z_{it}x_{it})\beta_1 + v_{it} - u_{it} \tag{17}$$

$$x_{it} = w_{it}\gamma + \epsilon_{it} \tag{18}$$

where x_{it} is an endogenous input so that it is correlated with the random error v_{it}, z_{it} is a sclar variable defines as before such as R&D or Innovations, w_{it} is of dimension $q(q \geq 1)$, and w_{it} is assumed to be strictly exogenous with respect to ϵ_{it}, v_{it} and u_{it}. As before, it is also assumed that $v_{it} \sim i.i.d.N(0,\sigma_v^2)$, $u_{it} \sim i.i.d.\left|N(0,\sigma_u^2)\right|$ and u_{it} independent of $x_{it}, z_{it}, w_{it}, v_{it}$ and ϵ_{it}. Model (17)-(18) suggests that there is auxiliary information available regarding the endogenous (or measurement error) input x_{it} and this information help to identify the production parameters and inefficiency.

Intuitively, the parameters in (17) and (18) can easily be estimated using the standard two-stage estimation approach where in the first stage, OLS is used to estimate γ by regressing x_{it} on w_{it} and obtain the predicted values $\hat{x}_{it} = w_{it}\hat{\gamma}$; and in the second stage, ML procedure is used to estimate the remaining parameters by replacing x_{it} by $\hat{x}_{it}$ in (17). However, there are two main problems with this approach. First, the estimated standard errors in the second stage estimation are incorrect and inconsistent, because it does not account for the estimation errors in the first stage. Second, the estimated parameters in the second stage are generally

inconsistent. To overcome these problems, Tran and Tsionas (2012) propose simple generalize method of moments (GMM) procedure that is fairly easy to implement.

To construct the GMM estimator, the error terms ε_{it} and v_{it} are assumed to satisfy the following:

$$\begin{pmatrix} \tilde{\epsilon}_{it} \\ v_{it} \end{pmatrix} \equiv \begin{pmatrix} \sigma_{\varepsilon}\epsilon_{it} \\ v_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix} \right) \tag{19}$$

where ρ is the correlation coefficient between v_{it} and ϵ_{it}. By Cholesky decomposition of the variance-covariance matrix of $(\tilde{\epsilon}_{it}\, v_{it})'$, we can write (3) as:

$$\begin{pmatrix} \tilde{\epsilon}_{it} \\ v_{it} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \sigma_v\rho & \sigma_v\sqrt{1-\rho^2} \end{bmatrix} \begin{pmatrix} \tilde{\epsilon}_{it} \\ \tilde{\omega}_{it} \end{pmatrix} \tag{20}$$

where $\tilde{\omega}_{it} \sim N(0,1)$ and independent of $\tilde{\epsilon}_{it}$. Using (20), we can write (17) and (18) as

$$y_{it} = \mu + z_{it}\alpha + x_{it}\beta_0 + (z_{it}x_{it})\beta_1 + \sigma_v\sigma_{\varepsilon}\rho\ x_{it} - w_{it}\gamma\ + \omega_{it} - u_{it} \tag{21}$$

where $\omega_{it} \sim N(0,(1-\rho^2)\sigma_v^2)$. Let $\xi_{it} = \omega_{it} - u_{it}$, $\sigma_s^2 = (1-\rho^2)\sigma_v^2 + \sigma_u^2$ and $\lambda = \sigma_u / \sigma_v\sqrt{1-\rho^2}$, then the probability density function of ξ_{it} is given by

$$f(\xi_{it}) = \frac{2}{\sigma_s}\phi\left(\frac{\xi_{it}}{\sigma_s}\right)\Phi\left(\frac{-\lambda\xi_{it}}{\sigma_s}\right), \qquad -\infty < \xi_{it} < \infty \tag{22}$$

whereas before, $\phi(.)$ and $\Phi(.)$ denote the standard normal PDF and CDF respectively. Let $y_i = (y_{i1},\cdots,y_{iT})'$, $x_i = (x_{i1},\cdots,x_{iT})'$, $z_i = (z_{i1},\cdots,z_{iT})'$, $w_i = (w_{i1}',\cdots,w_{iT}')'$ and denotes the $((q+5)\times 1)$ vector parameter θ as $\theta = (\alpha,\beta,\gamma,\eta,\lambda,\sigma_s^2)'$ where $\eta = -\sigma_s\rho / \sigma_{\varepsilon}(1+\lambda^2)(1-\rho^2)$, then for the sample observations (y_i, x_i, z_i, w_i), the joint log-likelihood function of y_i and x_i is given by:

$$\ln L(\theta) = \ln L_{y|x}(\theta) + \ln L_x \tag{23}$$

where

$$\begin{aligned}\ln L_{y|x}(\theta) = &\sum_{i=1}^{n} \ln f(y_i|x_i, z_i, w_i, \theta) \propto \\ &-\frac{nT}{2}\ln\sigma_s^2 + \frac{1}{\sigma_s}\sum_{i=1}^{n}\ln\Phi\Big[-\lambda(y_{it} - \mu - z_{it}\alpha - x_{it}\beta_0 - (z_{it}x_{it})\beta_1 - \eta(x_i - w_i\gamma))\Big] \quad (24a)\\ &-\frac{1}{2\sigma_s^2}\sum_{i=1}^{n}(y_i - \mu - z_{it}\alpha - x_{it}\beta_0 - (z_{it}x_{it})\beta_1 - \eta(x_i - w_i\gamma))^2\end{aligned}$$

and

$$\ln L_x = \sum_{i=1}^{n}\ln f(x_i) \propto -\frac{nT}{2}\ln(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}'\varepsilon_{it} \tag{24b}$$

Now given (18), from OLS procedure, the first-order condition is:

$$n^{-1}\sum_{i=1}^{n} w_i'(x_i - w_i\gamma) = n^{-1}\sum_{i=1}^{n} g_1(x_i, w_i, \gamma) = 0 \tag{25a}$$

where the definition of $g_1(.)$ should be apparent. Next, consider the log-likelihood function (24), and let $\partial \ln L(x_i, z_i, w_i, \theta) / \partial\theta$ denotes the $((q+5)\times 1)$ corresponding likelihood score vectors, Then we have

$$n^{-1}\sum_{i=1}^{n}\partial \ln f(x_i, z_i, w_i, \theta) / \partial\theta = n^{-1}\sum_{i=1}^{n} g_2(x_i, z_i, w_i, \theta) = 0 \tag{25b}$$

where again the definition of $g_2(.)$ should be obvious. Thus, equations (25a) and (25b) constitute a set of $(q+6)$ moment conditions that form the basis for our GMM estimator.
Define the moments vector $G(x_i, z_i, w_i, \gamma, \theta) = (g_1(x_i, w_i, \gamma)', g_2(x_i, z_i, w_i, \theta)')'$, then the joint GMM estimator takes the form

$$(\hat{\gamma}, \hat{\theta}) = \arg\min_{\gamma,\theta}\left(n^{-1}\sum_{i=1}^{n} G(x_i, z_i, w_i, \gamma, \theta)'\right)\left(n^{-1}\sum_{i=1}^{n} G(x_i, z_i, w_i, \gamma, \theta)\right) \tag{26}$$

The minimization problem in (26) can be performed using fully and readily automated GMM package which is available from all common software such as LIMDEP, SHAZAM, EVIEWS, etc., and no programming skill is required.

7. Data Limitation

Let us now turn our attention to the discussion of types of data use for the proposed approach. In this paper, the model and the estimation approaches are discussed in the context of panel data. It can be easily accommodate cross-section or time series data. However, it is well known that time series data is not recommended in stochastic frontier approach. Thus, in order to apply our suggested model, researcher needs to have either cross-section or panel data. For agriculture data, we would recommend that researcher would have at least provincial data over time (in case of panel) or regional data (in the case of cross-section).

8. Concluding Remarks

It has been recognized that modeling agricultural productivity growth and inefficiency is a challenging task since there exists many methodologies with their respective merits and drawbacks. Most approaches suffer from heterogeneity problem. The main purpose and motivation of this paper is to offer some new modeling approaches to overcome the heterogeneity problem in measuring productivity and efficiency using stochastic frontier modeling. To take full advantage of the proposed model for agricultural productivity studies, the paper suggests the use of regional or provincial level data over time, instead of aggregate time series data. As an extension, the paper also shows how to incorporate endogeneity of the inputs in the production process by means of auxiliary information.

Regarding the estimation, the paper proposed various procedures that are fairly easy to implement, given the current existing computing power and readily automated software packages. Finally, the paper suggests a way to measure technical efficiency and decompose the TFP growth into technical change, scale effects, and technical efficiency. The decomposition of TFP growth offers important insight into the impact of different policies to improve Canadian agricultural productivity growth. For instance productivity growth composed largely of technical

change points to the primacy of technology development and adoption in productivity advances, suggesting that research and development is an appropriate policy tool. On the other hand, if scale effect is found to play a significant role in agricultural productivity growth, indicating that structural change in agricultural sector is the main driver of productivity growth, then policies designed to accelerate structural changes may enhance productivity growth. Finally, if farms are found to be significantly inefficient, then resources can be directed to farmers' education in the application of best practice from the present stock of knowledge.

References

Battese, G.E and Coelli, T.J. (1995), "A Model for technical inefficiency effects in a stochastic frontier production function for panel data," *Empirical Economics*, 20, 325-332.

Greene, W. H. (2002), LIMDEP, Version 8.0: Econometric Modeling Guide. Vol. 2. Econometric Software, Plainview. N.Y., E24-3-E24-4.

Kumbhakar, S.C. and Lovell, C.A.K. (2002), *Stochastic Frontier Analysis*, Cambridge University Press, New York, N.Y.

Kutlu, L. (2010), "Battese-Coelli estimator with endogenous regressors," *Economics Letters*, 109, 79-81.

Limam, Y.R. and S.M. Miller (2004), "Explaining Economic Growth: Factor Accumulation, Total Factor Productivity Growth, and Production Efficiency Improvement," University of Connecticut Working Paper #20.

Park, B.U., Simar, L. and Sickle, R. (1998), "Stochastic panel frontiers: a semiparametric approach," *Journal of Econometrics*, 84, 121-152.

Simar, L and Wilson, P.W. (2007), "Estimation and inference in two-stage, semi-parametric models of production processes," *Journal of Econometrics*, 136, 31-64.

Steward, B. (2006), "Measures and causes of productivity growth in Prairie agriculture: 1940-2004," *M.S. Thesis*, University of Alberta.

Steward, B., Veeman, T.S. and Unterschultz (2009), "Crops and livestock productivity growth in the Prairie: The impacts of technical change and scale," *Canadian Journal of Agricultural Economics*, 57(3), 379-394.

Tran, K.C. and E.G. Tsionas (2011), "Profile Likelihood estimation of partially linear stochastic frontier models," *Working Paper,* Athens University of Economics and Business, Athens, Greece.

Tran, K.C. and E.G. Tsionas (2012), "GMM estimation of stochastic frontier model with endogenous regressors," *Economics Letters* (Forthcoming).

Tsionas, E.G. and Kumbhakar, S.C. (2004),"Markov switching stochastic frontier model," *Econometrics Journal*, 7, 398-425.

Veeman, T. S., & Gray, R., (2009), "Agricultural production and productivity in Canada," *Choices,* 24(4).

Veeman, T. S., & Gray, R., (2010), "The Shifting Patterns of Agricultural Production and Productivity in Canada," In J. M. Alston, B. A. Babcock, and P. G. Pardey (Eds.), *The Shifting Patterns of Agricultural Production and Productivity Worldwide*. Midwest Agribusiness Trade Research and Information Center, Iowa State University, Ames, Iowa.