Efficiency Analysis with Stochastic Frontier Models using Popular Statistical Softwares

Bao Hoang Nguyen[*]   Robin C. Sickles[†]   Valentin Zelenyuk [‡]

June 15, 2021

Abstract

Our chapter provides a brief introduction to the stochastic frontier paradigm – one of the most powerful techniques for performance analysis developed over the last few decades to address various research questions for many contexts with empirical applications in a wide variety of economic sectors such as banking, healthcare, agriculture, and so on. We also document the estimation routines used to implement the classical models as well as the recent developments in this research area for practitioners, especially those who are willing to use Stata, but also with tips on sources for R and Matlab users.

Keywords: Technical efficiency, Stochastic frontier analysis, Panel data, Semi-parametric, Stata, Matlab, R.

[*]School of Economics, University of Queensland, Brisbane, Qld 4072, Australia

[†]Economics Department, Rice University, Houston, TX 77251-1892, USA

[‡]School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Brisbane, Qld 4072, Australia

1   Introduction

In this chapter, we provide practitioners, who are interested in analysing the performance of production units, with a brief introduction to the stochastic frontier paradigm – one of the most powerful techniques for performance analysis developed in the last century.[1] Stochastic frontier analysis employs econometric models to estimate production frontiers and technical (in)efficiency with respect to these frontiers. Since its first introduction by Aigner et al. (1977) and Meeusen and van den Broeck (1977), stochastic frontier analysis has been applied to study the productivity and efficiency of production units in various economic sectors, such as banking (e.g.,Ferrier and Lovell (1990); Adams et al. (1999); Kumbhakar and Tsionas (2005); Malikov et al. (2016)), healthcare (e.g., Zuckerman et al. (1994); Rosko (2001); Greene (2004); Mutter et al. (2013); Comans et al. (2020)), agriculture (e.g., Battese and Coelli (1995); Battese and Broca (1997); Kumbhakar and Tsionas (2008)), to mention a few. Moreover, the methodology is also used to undertake cross-country studies on various important aspects of society such as the healthcare system (Greene, 2004) and taxation (Fenochietto and Pessino, 2013).

Our chapter also documents the estimation routines used to implement the classical models as well as the recent developments in this research area for practitioners, especially those who are willing to use Stata, but also with tips on where to find analogous programs for R and Matlab users.[2] Interested readers can find more comprehensive overviews in Sickles and Zelenyuk (2019, Chap. 11-16), Kumbhakar et al. (2021a) and Kumbhakar et al. (2021b).

The structure of this chapter is as follows. We start our discussion with the basic stochastic frontier model. We then extend our discussion to various generalisations of the stochastic frontier paradigm, including stochastic panel data models, stochastic frontier models with determinants of inefficiency, also referred to in the literature as "environmental factors", and the semi-parametric stochastic frontier models. To provide readers with an accessible toolkit to implement these methods, we also document available command-

---

[1]Another powerful technique for performance analysis is data envelopment analysis – the technique based on the mathematical linear programming method proposed by Farrell (1957) and popularised by Charnes et al. (1978).

[2]On this aspect, our chapter complements earlier surveys on empirical frontier application and productivity and efficiency analysis software, e.g., Daraio et al., 2019 and Daraio et al., 2020. Besides, the chapter also complements the previous contributions of Belotti et al. (2013) and Kumbhakar et al. (2015), who focused only on stochastic frontier analysis using Stata, by providing the sources on analogous implementations in Matlab and R. Moreover, we also include the discussion about the semi-parametric stochastic frontier models with ready-to-use Stata codes to implement the model proposed by Simar et al. (2017), which to the best of our knowledge have not been documented elsewhere before.

s/packages in popular statistical softwares. We focus on the implementation via Stata, and also provide brief comments on the sources on analogous implementations in Matlab and R. Besides, we also provide an empirical illustration for the methods discussed in this chapter.

2 Basic Stochastic Frontier Models

The stochastic frontier paradigm can be viewed as a generalization of the classical production function approach, where the optimal allocation in production is a testable restriction rather than a prior assumption usually assumed by the neoclassical production theory (Sickles and Zelenyuk, 2019).

The distinctive feature of the stochastic frontier paradigm (compared to the canonical average production model paradigm) is its non-symmetric two-component error, composed of a regular idiosyncratic disturbance and an additional one-sided non-negative error component.[3] The former accounts for factors such as measurement error, misspecification, and the randomness of the production process, whereas the latter aims to represent the technical inefficiency that reduces the actual output from its maximum feasible level.[4] Assumptions in the canonical model used in stochastic frontier analysis on the conditional independence of both error terms and the regressors as well as their independence from each other have been lifted over the years in a series of refinements of the basic model. We will discuss these in turn later in our chapter.

2.1 Aigner et al. (1977) Model

The canonical model of the stochastic frontier paradigm was proposed independently by Aigner et al. (1977) (hereafter ALS) and Meeusen and van den Broeck (1977). The ALS

---

[3]In the panel data context, which we will discuss in the next sections, the composed error can include four components.

[4]In this chapter, our discussion will follow the traditional exposition based on the production function. Similar exposition (with some adaptations) applies to other characterizations of the production side, such as cost function and revenue function. Meanwhile, more elaboration is needed if one is interested in measuring profit efficiency (see Färe et al. (2019), Sickles and Zelenyuk (2019, Chapter 2) and references therein).

model is formulated as [5]

$$\ln y_i = \ln f\left(x_i|\beta\right) + \varepsilon_i, \quad i = 1, \dots, n,$$
$$\varepsilon_i = v_i - u_i,$$
$$v_i \sim_{\text{iid}} \mathcal{N}\left(0, \sigma_v^2\right),$$
$$u_i \sim_{\text{iid}} \mathcal{N}^+\left(0, \sigma_u^2\right), \tag{1}$$

where $y_i \in \Re_+^1$ is the output, $x_i \in \Re_+^p$ is a vector of p inputs and β is a vector of the parameters corresponding to x_i.[6] The error term ε_i is composed of a normally distributed disturbance, v_i, representing the measurement and specification error, and a positive disturbance u_i (following the half-normal distribution), representing technical inefficiency.[7] Furthermore, v_i and u_i are assumed to be statistically independent from each other and from x_i. With the distribution assumptions on u_i and v_i, the likelihood function for the model is constructed and the model is then estimated using the maximum likelihood estimator.

Once the parameters of the model have been estimated, one can obtain the expected level of technical inefficiency by estimating

$$E\left[u\right] = \sqrt{2/\pi}\,\sigma_u, \tag{2}$$

and the expected level of efficiency by using the following approximation[8]

$$E\left[\exp\left(-u\right)\right] \approx 1 - E\left[u\right]. \tag{3}$$

If one is interested in the estimates of individual (in)efficiency of a specific production unit, more elaboration is needed. The most popular approach in the literature is to follow

---

[5]The formulation here is a convenient representation of a production relationship, where actual output is decomposed into the maximum output (with noise) and inefficiency, i.e., $y_i = f\left(x_i|\beta\right)\exp\left(\varepsilon_i\right) = f\left(x_i|\beta\right)\exp\left(v_i\right)\exp\left(-u_i\right)$. After log-transformation, we have a linear relationship as shown in (1).

[6]Multiple outputs also can be considered. For example, this can be done by employing a distance function instead of the production function or by looking at the estimation of the cost frontier or by converting outputs into polar coordinates (e.g., see Simar and Zelenyuk (2011)). One can also use dimension reduction techniques to reduce the dimension outputs or inputs into smaller dimensions, e.g., via Principle Component Analysis, or using economic or price-based aggregation (e.g., see related discussion in Zelenyuk (2020) and an application in Nguyen and Zelenyuk (2021)). The latter approach can be especially useful in the case of very large dimensions (sometimes called 'big wide data' cases), e.g., as is done for measuring the total output of countries (e.g., GDP), industries or firms (total revenue) or for some inputs (e.g. capital). Due to space limitation we will focus here on the single output case, as was also considered in ALS and many other studies.

[7]Other distributional assumptions such as exponential, truncated normal, gamma, etc., can be used for the inefficiency term (e.g., see Meeusen and van den Broeck, 1977; Greene, $1980a,b$; Stevenson, 1980; Greene, 1990; Almanidis and Sickles, 2012; Almanidis et al., 2014).

[8]The exact expression of the expected level of efficiency is given by $E\left[\exp\left(-u\right)\right] = 2\Phi\left(-\sigma_u\right)\exp\left(\frac{\sigma_u^2}{2}\right)$.

Jondrow et al. (1982) (hereafter JLMS), where the inefficiency of a production unit can be estimated or predicted using the expected value of u_i conditional on the realisation of the composed error of the model, i.e., $E\left(u_i|\varepsilon_i\right)$,[9] given by

$$E\left(u_i|\varepsilon_i\right) = \frac{\sigma_*\phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i}, \tag{4}$$

where

$$\mu_{*i} = \frac{-\sigma_u^2\varepsilon_i}{\sigma_v^2 + \sigma_u^2}, \tag{5}$$

and

$$\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}, \tag{6}$$

while $\phi\left(\cdot\right)$ and $\Phi\left(\cdot\right)$ are *pdf* and *cdf* of the standard normal distribution, respectively.[10] It is worth noting that while being originally developed for ALS, the JLMS-type procedure can be extended to predict (in)efficiency of a specific firm in the other models estimated by the maximum likelihood estimator (see more discussion in Kumbhakar, 1987).

2.2 Implementation of ALS Model

There are several options to estimate the basic stochastic frontier model in Stata. One can use the official Stata command `frontier` or utilise the command `sfcross` written by Belotti et al. (2013)[11] or even set up the likelihood function using the `sfmodel` command then estimating the model using the official Stata routine for the maximum likelihood, `ml max`, as described in the handbook of Kumbhakar et al. (2015).[12] These commands generate equivalent results for the basic stochastic frontier models and differ only in the formatting and listing of results and the options available for the different treatments of

---

[9]It is worth noting here that although this estimator is unbiased, it is an inconsistent estimator of individual inefficiency (see more discussion in Jondrow et al., 1982).

[10]One also can estimate the efficiency of a production unit by using the relationship $E\left[\exp\left(-u_i\right)|\varepsilon_i\right] \approx 1 - E\left[u_i|\varepsilon_i\right]$ or utilising the exact expression $E\left[\exp\left(-u_i\right)|\varepsilon_i\right] = \exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right)\frac{\Phi\left(\frac{\mu_{*i}}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}$ (Battese and Coelli, 1988).

[11]The `sfcross` command (and the `sfpanel` command that we will discuss later for the panel data context) can be installed by executing the following command lines in Stata: `net install sfcross, all from (http://www.econometrics.it/stata)` and `net install sfpanel, all from (http://www.econometrics.it/stata)`.

[12]The `sfmodel` and other user-written commands provided in the handbook of Kumbhakar et al. (2015) can be installed in Stata by executing the following command lines: `net install sfbook_install, from (https://sites.google.com/site/sfbook2014/home/install/) replace` and `sfbook_install`, see more details in Kumbhakar et al. (2015) and its website, `https://sites.google.com/site/sfbook2014/`.

Table 1: Illustration for implementation of the Aigner et al. (1977) model

```
****************************************************************************
*** Illustration for implementation of the Aigner et al. (1977) model **
***************************Partial Stata Codes***************************
****************************************************************************
/* Note that output, inputs are in log forms and stored in global Stata
   variables $y, $xlist, respectively */
/* Implementation using the standard Stata commands */
frontier $y $xlist, distribution(hnormal)
predict ineff_ALS_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_ALS_1, te /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfcross $y $xlist, distribution(hnormal)
predict ineff_ALS_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_ALS_2, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Kumbhakar et al. (2015) */
sfmodel $y, prod frontier($xlist) distribution(h)
ml max
sf_predict, jlms(ineff_ALS_3) /* Predict inefficiency, i.e., E(u|e) */
sf_predict, bc(eff_ALS_3) /* Predict efficiency, i.e., E(exp(-u)|e) */
```

error distributions for the one-sided efficiency term and the inclusion of environmental factors.

As we progress in our chapter we consider a richer set of generalisations of the canonical stochastic frontier paradigm. Also, user-written commands provide us with more flexibility to estimate models that are not available with the current official Stata commands. Moreover, the user-written commands by Belotti et al. (2013) and Kumbhakar et al. (2015) also equip us with options to provide and refine the initial values for the maximum likelihood estimation, which can be very useful when dealing with complex likelihood functions.

After estimating the models, the estimates of technical inefficiency and efficiency can be obtained by using the postestimation routine `predict` (for the models estimated in the Stata version 16 by the official Stata command and the command written by Belotti et al. (2013)) or `sf_predict` (for the models estimated by the command written by Kumbhakar et al. (2015)). As an illustration, a snippet of Stata codes for implementing the ALS model is provided in Table 1.

R software also has several packages to implement the estimation of the basic stochastic frontier model. For example, one can use the package `frontier` written by Coelli and Henningsen (2020)[13] or utilise the function `sfa` in the package `Benchmarking` written by Bogetoft and Otto (2019).

In order to estimate the basic stochastic frontier model, Matlab users need to set up the likelihood function and then utilise the optimisation routines, such as `fminunc` to optimise the likelihood function. Sickles and Zelenyuk (2019) provide a suite of Matlab codes to estimate a variety of stochastic frontier models on the website that accompanies their book.[14][15] Although they do not include the ALS model, one can easily adapt their codes to obtain the estimates for this basic stochastic frontier model.

3   Early generation of stochastic panel data models

The basic stochastic frontier model discussed in the previous section is formulated in the cross-sectional setting and suffers from a number of drawbacks. As discussed in Schmidt and Sickles (1984), the three main disadvantages of the basic cross-sectional stochastic frontier model are: (i) there does not exist a consistent estimator of individual efficiency, (ii) the parametric distributional assumptions are usually required for the two error components (inefficiency and noise) to estimate the model and to predict the overall and individual (in)efficiency, and (iii) the assumption that inefficiency is independent of regressors is usually not plausible.

Over the past four decades, substantial efforts have been made to address these drawbacks of the cross-sectional stochastic frontier model. Among those, particular interest hinges on exploiting the advantages of panel data structure. Schmidt and Sickles (1984) were among the first who provided a general framework to extend the cross-sectional stochastic frontier model to the panel data setting, which also encompasses the Pitt and Lee (1981) full parametric random effect model.

---

[13]The package `frontier` uses the Fortran source codes of Frontier 4.1 originally developed by Tim Coelli (see more details in the manual of the package available at `https://cran.r-project.org/web/packages/frontier/frontier.pdf`).

[14]The website can be found at `https://sites.google.com/site/productivityefficiency/home`.

[15]The Matlab codes accompanying Sickles and Zelenyuk (2019) are also converted to R codes by Sickles et al. (2020), which can be accessed via the link provided on the book website or directly via `https://sites.google.com/site/productivityinr`.

3.1 Schmidt and Sickles (1984) Model

The model in Schmidt and Sickles (1984) can be formulated as follow

$$y_{it} = \beta_0 + x'_{it}\beta + v_{it} - u_i, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{7}$$

where $y_{it} \in \Re^1_+$ is the output, $x_{it} \in \Re^p_+$ is a vector of p inputs of production unit i in time t. v_{it} is the regular disturbance, while the unobserved individual heterogeneity, u_i, represents technical inefficiency. Model (7) can be rewritten as

$$y_{it} = \beta_0^* + x'_{it}\beta + v_{it} - u_i^* = c_i + x'_{it}\beta + v_{it}, \tag{8}$$

where $\beta_0^* = \beta_0 - E(u_i)$, $u_i^* = u_i - E(u_i)$, $E(u_i) \geq 0$, $c_i = \beta_0^* - u_i^* = \beta_0 - u_i$.

Model (8) turns out to be a usual panel data model and can be estimated using the standard estimation methods in the panel data literature, such as the within estimator (i.e., in the fixed effect framework), the generalised least-square estimator (i.e., in the random effect framework), and the Hausman-Taylor estimator. After estimating the model, one can obtain the estimate $\hat{c}_i$ of c_i and follow Schmidt and Sickles (1984) to construct a consistent estimator of technical inefficiency

$$\hat{u}_i = \max(\hat{c}_i) - \hat{c}_i \geq 0, \quad i = 1, \ldots, n. \tag{9}$$

The estimated inefficiency in (9) is measured with respect to the best practice production unit in the sample, which is implicitly assumed to be 100% efficient.

3.2 Implementation of Schmidt and Sickles (1984) Model

One can estimate the Schmidt and Sickles (1984) model using standard routines in Stata. Specifically, the official Stata command `xtreg` can be utilised to estimate the standard panel data model in (8) and the postestimation command `predict` can be used to obtain the estimate $\hat{c}_i$ of c_i. It is then straightforward to code formula (9) into Stata to get the estimates of technical inefficiency. Alternatively, one can use the command `sfpanel` written by Belotti et al. (2013) with the option `model(fe)` or `model(regls)` to estimate Schmidt and Sickles (1984) model in a fixed or random effects framework, respectively. As an illustration, a snippet of Stata codes for implementing the Schmidt and Sickles (1984) model (in the fixed effects framework) is provided in Table 2.

It is worth noting that model (8) and the individual inefficiency in (9) are estimated without any parametric assumptions on the distributions of composed errors. Alternatively, one can impose parametric assumptions on the distributions of the error components in model(7), e.g., a half-normal distribution for u_i and a normal distribution for

Table 2: Illustration for implementation of the Schmidt and Sickles (1984) model in the fixed effects framework

```
*************************************************************************
**********************Illustration for implementation ******************
********************of the Schmidt and Sickles (1984) model ************
**************************Partial Stata Codes***************************
*************************************************************************
/* The illustration here is for the fixed effects framework */
/* Note that output, inputs are in log forms and stored in global Stata
    variables $y, $xlist, respectively */
/* Implementation using the standard Stata commands */
xtreg $y $xlist, fe /* Need to declare data to be panel before using
    xtreg command*/
predict ci, u /* Obtain the estimate of ci */
quietly summarize ci
gen ineff_SS_1 = r(max) - ci /* Predict inefficiency*/
gen eff_SS_1 = exp(-ineff_SS_1) /* Predict efficiency*/
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(fe)
predict ineff_SS_2, u /* Predict inefficiency*/
gen eff_SS_2 = exp(-ineff_SS_2) /* Predict efficiency*/
```

Table 3: Illustration for implementation of the Pitt and Lee (1981) model in the random effects framework

```
*****************************************************************************
*** Illustration for implementation of the Pitt and Lee (1981) model ***
***************************Partial Stata Codes*************************
*****************************************************************************
/* Note that output, inputs are in log forms and stored in global Stata
   variables $y, $xlist, respectively */
/* Implementation using the standard Stata commands */
predict ineff_PL_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_PL_1, te /* Predict inefficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(pl81)
predict ineff_PL_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_PL_2, bc /* Predict inefficiency, i.e., E(exp(-u)|e) */
```

v_{it} as discussed in Pitt and Lee (1981) and Schmidt and Sickles (1984). The model then can be estimated using the maximum likelihood estimator and the individual technical efficiency can be obtained by employing the JLMS procedure (extended to the panel data setting by Kumbhakar, 1987). This model is estimated in Stata using the user-written command sfpanel from Belotti et al. (2013) with the option model(pl81). Alternatively, if one assumes that u_i follows a truncated normal distribution, i.e., $u_i \sim \mathcal{N}^+ \left(\mu, \sigma_u^2 \right)$, then the official Stata command xtfrontier with the option ti can be utilised. A snippet of Stata codes for implementing the Pitt and Lee (1981) model is provided in Table 3.

Estimation of the Schmidt and Sickles (1984) model also can be implemented in Matlab and R using the codes provided by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the download links in footnote 14 and footnote 15).

3.3 Cornwell et al. (1990) Model

The technical inefficiency estimated within the Schmidt and Sickles (1984) framework is time-invariant, which may be an unrealistic restriction in many applied settings, especially in a long panel. To allow for time varying inefficiency in the Schmidt and Sickles (1984) framework, one can follow the suggestion in Cornwell et al. (1990) to replace c_i by, e.g., c_{it}, where c_{it} is a quadratic function of time trend t with the parameters (coefficients)

being firm-specific, specifically

$$c_{it} = \theta_{0i} + \theta_{1i}t + \theta_{2i}t^2. \tag{10}$$

The parameters in (10) can be estimated by regressing the residual from the model (7) for production unit i on a constant, time, and time-squared (see more discussion in Cornwell et al., 1990). The fitted value from this model provides us with a consistent estimate (for large N) of c_{it}, denoted as $\hat{c}_{it}$. The individual technical inefficiency of production unit i at time t then can be estimated using an analogous procedure to Schmidt and Sickles (1984), specifically[16]

$$\hat{u}_{it} = \hat{c}_t - \hat{c}_{it}, \tag{11}$$

where

$$\hat{c}_t = \max_j (\hat{c}_{jt}), \quad t = 1, \ldots, T. \tag{12}$$

3.4 Implementation of Cornwell et al. (1990) Model

Estimation of the Cornwell et al. (1990) model can be implemented using standard Stata routines in a set of procedures similar to those we discussed for the Schmidt and Sickles (1984) model. Alternatively, one can utilise the user-written command `sfpanel` from Belotti et al. (2013) with the option `model(fecss)`. A snippet of codes for implementing the Cornwell et al. (1990) model using `sfpanel` command is provided in Table 4.

Being similar to the Schmidt and Sickles (1984) model, one can estimate the Cornwell et al. (1990) model in Matlab and R using the codes provided by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the download links in footnote 14 and footnote 15).

3.5 Kumbhakar (1990) and Battese and Coelli (1992) Models

If one is willing to impose distributional assumptions on the inefficiency component (as well as on the random disturbance term), the maximum likelihood estimation can be utilised to estimate time-varying efficiency models. Kumbhakar (1990) and Battese and Coelli (1992) appear to be the most popular models of this type. In the Kumbhakar (1990) model, time-varying inefficiency is modelled as

$$\begin{aligned} u_{it} &= \left(1 + \exp\left(at + bt^2\right)\right)^{-1} \tau_i, \\ \tau_i &\sim_{\text{iid}} \mathcal{N}^+ \left(0, \sigma_\tau^2\right), \end{aligned} \tag{13}$$

---

[16]Cornwell et al. (1990) outlined estimators for a general model in which any set of regressors could be drivers of efficiency change, if efficiency was interpreted as firm-specific heterogeneity. These regressors could be time varying. Thus the Cornwell et al. (1990) model was the first study about which we are aware to address the issue of environmental variables influencing efficiency levels.

Table 4: Illustration for implementation of the Cornwell et al. (1990) model

```
****************************************************************************
* Illustration for implementation of the Cornwell et al. (1990) model **
**************************Partial Stata Codes************************
****************************************************************************
/* Note that output, inputs are in log forms and stored in global Stata
   variables $y, $xlist, respectively */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(fecss)
predict ineff_CSS, u /* Predict inefficiency*/
gen eff_CSS = exp(-ineff_CSS) /* Predict efficiency*/
```

while in Battese and Coelli (1992), time-varying inefficiency is specified as

$$
\begin{aligned}
u_{it} &= \{\exp\left[-\eta\left(t - T\right)\right]\}\,\tau_i, \\
\tau_i &\sim_{\text{iid}} \mathcal{N}^+\left(\mu, \sigma_\tau^2\right),
\end{aligned}
\tag{14}
$$

where a,b, and η are parameters to be estimated, and in both models, the random disturbance follows a normal distribution, i.e, $v_{it} \sim_{\text{iid}} \mathcal{N}\left(0, \sigma_v^2\right)$.

Being similar to the Cornwell et al. (1990) model, the Kumbhakar (1990) and Battese and Coelli (1992) models extend the Pitt and Lee (1981) model by allowing the mean of inefficiency to vary over time, but they are more parsimonious in the sense that temporal patterns only depend or one or two parameters. The Cornwell et al. (1990) model, however, has an advantage in that it allows temporal patterns to vary across production units. Moreover, as discussed above, estimation of the Cornwell et al. (1990) model does not require parametric assumptions for inefficiency term.

3.6 Implementation of Kumbhakar (1990) and Battese and Coelli (1992) Models

The Battese and Coelli (1992) model, also known as a "time decay" model, can be estimated using Stata commands in its version 16 platform as well as by using user-written commands. Specifically, the estimation can be implemented by using the `xtfrontier` command with the option `tvd` or the command `sfpanel` from Belotti et al. (2013) with the option `model(bc92)`. The official Stata command `xtfrontier` cannot carry out estimation of the Kumbhakar (1990) model, which is available using the option `model(kumb90)` with the command `sfpanel` from Belotti et al. (2013). A snippet of Stata codes for

Table 5: Illustration for implementation of the Kumbhakar (1990) and Battese and Coelli (1992) models

```
***************************************************************************
********* Illustration for implementation of the Kumbhakar (1990)  ******
************** and Battese and Coelli (1992) models *********************
********************** Partial Stata Codes ***************************
***************************************************************************
/* Note that output, inputs are in log forms and stored in global Stata
   variables $y, $xlist, respectively */
/* Implementation using the standard Stata commands */
xtfrontier $y $xlist, tvd /* the Battese and Coelli (1992) model */
predict ineff_BC_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_BC_1, te /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the commands from Belotti et al. (2013) */
sfpanel $y $xlist, model(bc92) /* the Battese and Coelli (1992) model */
predict ineff_BC_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_BC_2, bc /* Predict efficiency, i.e., E(exp(-u)|e) */

sfpanel $y $xlist, model(kumb90) /* the Kumbhakar (1990) model */
predict ineff_K, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_K, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
```

implementing Kumbhakar (1990) and Battese and Coelli (1992) models is provided in Table 5.

The estimation of the Battese and Coelli (1992) model can be implemented in R software by using the package `frontier` written by Coelli and Henningsen (2020). Alternatively, R users and Matlab users can utilise the codes prepared by Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the download links in footnote 14 and footnote 15).

4   Recent Advances of Stochastic Panel Data Models

The stochastic panel data models discussed so far have a major drawback in that technical inefficiency is not distinguishable from the unobserved individual heterogeneity, and thus technical inefficiency confounds with all time-invariant unobserved individual effects. Various approaches have been proposed in the literature to mitigate this (and other) issues. Here, we will focus on a few, namely Greene (2005a,b), Chen et al. (2014), Colombi

et al. (2014), Kumbhakar et al. (2014) and Belotti and Ilardi (2018).

4.1   Greene (2005*a,b*) Models

Greene (2005*a,b*) proposed a stochastic panel data model in which unobserved individual heterogeneity separates from (transitory) technical efficiency. The model is formulated as

$$
\begin{aligned}
y_{it} &= c_i + x'_{it}\beta + \varepsilon_{it}, \\
\varepsilon_{it} &= v_{it} - u_{it},
\end{aligned}
\tag{15}
$$

where v_{it} as before is a regular disturbance, while $u_{it} \geq 0$ is the source of inefficiency.

Estimation of the model in (15) is challenging, especially in the fixed effects framework. The two main challenges are: (i) the estimation of parameters may be inconsistent due to the incidental parameters problem, and (ii) there does not exist a closed-form expression of the likelihood function of the within or first-difference transformation of the composed error if one follows standard procedures. Greene (2005*a*) proposed to use the maximum likelihood dummy variable estimator to estimate the model in the fixed effects framework and provided simulation evidence showing that the incidental parameters problem is not serious for relatively large T.[17]

4.2   Implementation of Greene (2005*a,b*) Models

One can implement estimation of the Greene (2005*a,b*) models in Stata by using the user-written command `sfpanel` from Belotti et al. (2013) with the option `model(tfe)` in the fixed effects framework and with the option `model(tre)` in the random effects framework.

Recently, Chen et al. (2014) derived a closed-form expression for the likelihood function of the within and first difference transformation of the model by exploiting the properties of the closed-skew normal distribution class. The model in (15) then can be estimated consistently in the fixed effects framework using the marginal maximum likelihood estimator. Belotti and Ilardi (2018) further extend the work of Chen et al. (2014) by considering the simulated marginal maximum likelihood estimator.

Chen et al. (2014) and Belotti and Ilardi (2018) estimators can be implemented in Stata using the command `sftfe` written by Belotti and Ilardi (2018) with the options `estimator(within)` and `estimator(mmsle)`, respectively.[18] A snippet of Stata codes for implementing the Greene (2005*a,b*) models is provided in Table 6.

---

[17]Greene (2005*a*) also utilised the simulated maximum likelihood estimator to estimate the model in the random effect frameworks

[18]The `sftfe` command can be installed by executing the following command line in Stata: `net install sftfe, all from (http://www.econometrics.it/stata)`.

Table 6: Illustration for implementation of the Greene (2005a,b) models

```
******************************************************************************
***    Illustration for implementation of the Green (2005a,b) models    ***
***************************Partial Stata Codes**************************
******************************************************************************
/* The illustration here is for the fixed effect framework */
/* Note that output, inputs are in log forms and stored in global Stata
    variables $y, $xlist, respectively */
/* Implementation using  the maximum likelihood dummy variable estimator
     (the commands from Belotti et al. (2013)) */
sfpanel $y $xlist, distribution(hnormal) model(tfe)
predict ineff_G_1, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_1, bc /* Predict efficiency, i.e., E(exp(-u)|e) */
/* Implementation using the marginal maximum likelihood estimator (the
    commands from Belotti and Ilardi (2018)) */
sftfe $y $xlist,  distribution(hnormal) estimator(within)
predict ineff_G_2, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_2, jlms /* Predict efficiency */
/* Implementation using the simulated marginal maximum likelihood
    estimator (the commands from Belotti and Ilardi (2018)) */
sftfe $y $xlist,  distribution(hnormal) estimator(mmsle)
predict ineff_G_3, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_G_3, jlms /* Predict efficiency */
```

To the best of our knowledge, routines to estimate the Greene (2005a,b) model are not yet available in R and Matlab in a public domain.

4.3 Colombi et al. (2014) and Kumbhakar et al. (2014) Models

The model specified in (15), although distinguishing between unobserved individual heterogeneity and technical inefficiency, only considers transitory inefficiency. Kumbhakar et al. (2014) and Colombi et al. (2014) further extend the model by decomposing the inefficiency into transitory and persistent components whose formulation is

$$
\begin{aligned}
y_{it} &= \beta_0 + x_{it}'\beta + c_i - \eta_i + v_{it} - u_{it}, \\
c_i &\sim_{\mathrm{iid}} \mathcal{N}\left(0, \sigma_c^2\right), \\
\eta_i &\sim_{\mathrm{iid}} \mathcal{N}^+\left(0, \sigma_\eta^2\right), \\
v_{it} &\sim_{\mathrm{iid}} \mathcal{N}\left(0, \sigma_v^2\right), \\
u_{it} &\sim_{\mathrm{iid}} \mathcal{N}^+\left(0, \sigma_u^2\right),
\end{aligned}
\tag{16}
$$

where c_i represents the unobserved individual heterogeneity, η_i represents the persistent inefficiency, u_{it} represents transitory inefficiency, and v_{it} is the regular disturbance. The model in (16) can be estimated using a single-stage maximum likelihood method (Colombi et al., 2014) or a multi-step procedure (Kumbhakar et al., 2014). The multi-step procedure although being inefficient relative to the single-stage maximum likelihood estimation, it is simpler and easier to implement. For the multi-step procedure, the model in (16) can be rewritten as

$$
y_{it} = \beta_0^* + x_{it}'\beta + \alpha_i + \varepsilon_{it},
\tag{17}
$$

where

$$
\beta_0^* = \beta_0 - E\left[\eta_i\right] - E\left[u_{it}\right],
\tag{18}
$$

$$
\alpha_i = c_i - \eta_i + E\left(\eta_i\right),
\tag{19}
$$

$$
\varepsilon_{it} = v_{it} - u_{it} + E\left[u_{it}\right].
\tag{20}
$$

The model in (17) turns out to be a standard panel data model and can be estimated by the usual panel data estimation methods. After estimating (17), one can obtain the predicted values of α_i and ε_{it}, $\hat{\alpha}_i$ and $\hat{\varepsilon}_{it}$, and then the persistent and transitory inefficiency components are estimated by applying standard stochastic frontier techniques to (19) and (20) with α_i and ε_{it} replaced by $\hat{\alpha}_i$ and $\hat{\varepsilon}_{it}$.

Table 7: Illustration for implementation of the Kumbhakar et al. (2014) model

```
****************************************************************************
*Illustration for implementation of the Kumbhakar et al. (2014) models *
*************************Partial Stata Codes*************************
****************************************************************************
/* The illustration here is for the random effect framework */
/* Note that output, inputs are in log forms and stored in global Stata
    variables $y, $xlist, respectively */
/* Implementation using the standard Stata commands */
xtreg $y $xlist, re
predict alp, u /* Obtain estimates of alpha */
predict esl, e /* Obtain estimates of the composed error */
/* Estimate equation (19) using the basic stochastic frontier model to
    obtain persistent (in)efficiency */
frontier alp, distribution(hnormal)
predict ineff_pers, u /* Predict persistent inefficiency, E(u|e) */
predict eff_pers, te /* Predict persistent efficiency, E(exp(-u)|e) */
/* Estimate equation (19) using the basic stochastic frontier model to
    obtain transitory (in)efficiency */
frontier esl, distribution(hnormal)
predict ineff_trans, u /* Predict transitory inefficiency, E(u|e) */
predict eff_trans, te /* Predict transitory efficiency, E(exp(-u)|e) */
```

4.4 Implementation of Colombi et al. (2014) and Kumbhakar et al. (2014) Models

The multi-step procedure to estimate the model specified in Colombi et al. (2014) and Kumbhakar et al. (2014) can be implemented in Stata using the command for panel data estimation, `xtreg`, together with the routines for basic stochastic frontier model estimation, e.g., `frontier` or `sfcross`. A snippet of Stata codes for implementing the Kumbhakar et al. (2014) model is provided in Table 7.

Similarly, R users can utilise panel data estimation routines (e.g., `plm` function) combined with commands for estimation of the basic stochastic frontier model discussed above (e.g, `sfa` or `frontier`) to estimate the Colombi et al. (2014) and Kumbhakar et al. (2014) models.

The implementation of the procedure in Matlab requires more effort since it is not easy (as in Stata or R) to perform panel data regression with this platform. With Matlab, one

needs to write their own code or download and install the panel data toolbox, e.g. the one written by Álvarez et al. (2017), to estimate a panel data regression model.

5 Stochastic Frontier Models with Determinants of Inefficiency

An interesting generalisation of stochastic frontier paradigm is extending the models to examine the impact of exogenous determinants on technical inefficiency. It is usually done by parameterising the parameters of inefficiency distribution, i.e., the pre-truncated mean or/and variance, as a function of exogenous variables. The approaches are applicable in both cross-sectional and panel data settings, and since it can be easily extended to panel data settings, here we focus our discussion on the cross-sectional context.

5.1 Popular Models

Cornwell et al. (1990) was the first to develop a model in which determinants of efficiency could be included in the stochastic frontier formulation. However, due to the linear way in which the determinants of efficiency were included in the regression model their fixed effect estimator could not point identify both a covariate's effect on efficiency and its effect on the level of production. Kumbhakar et al. (1991) addressed this identification problem by specifying the efficiency determinants as a nonlinear function, parameterising the pretruncated mean of inefficiency as a function of exogenous variable, specifically[19]

$$
\begin{aligned}
u_i &\sim \mathcal{N}^+ \left(\mu_i, \sigma_u^2 \right), \\
\mu_i &= z_i' \delta,
\end{aligned}
\tag{21}
$$

where $z_i \in \Re^k$ is a vector of k exogenous variables (including constant term) and δ is a vector of the parameters to be estimated. Alternatively, Caudill et al. (1995) proposed specifying the variance of the inefficiency distribution as

$$
\begin{aligned}
u_i &\sim \mathcal{N}^+ \left(0, \sigma_{ui}^2 \right), \\
\sigma_{ui}^2 &= \exp \left(z_i' \delta \right).
\end{aligned}
\tag{22}
$$

One can also at the same time parameterise both the pre-truncated mean and variance of inefficiency as a function of exogenous variables, i.e, combining (21) and (22), as in Wang (2002). These parametric stochastic frontier models are typically estimated using

---

[19]This model specification was cast in the panel data context and popularised by Battese and Coelli (1995).

the maximum likelihood estimator in much the same way as the basic stochastic frontier model.

Wang and Schmidt (2002) suggested a different specification for modelling the determinants of inefficiency based on a scaling property,[20] specifically

$$u_i \sim g\left(z_i|\delta\right)u_i^*, \tag{23}$$

where $g(\cdot)$ is a positive function of the exogenous variables (the scaling function) and u_i^* is a positive random variable. With this specification, the distribution of inefficiency is the same for all production units, i.e., governed by u_i^*, while the scale of the inefficiency distribution changes across production units depending on z_i. The scaling property was further explored in Alvarez et al. (2006). Among others, they provided a nice economic interpretation for the scaling property in that u_i^* represents the baseline (in)efficiency of a production unit capturing things like the natural skills of its managers, meanwhile the scaling function allows (or prevents) the production unit to exploit these natural skills through other variables, z_i, such as the experience and education of the managers, or the environment in which the production unit operates. Moreover, Alvarez et al. (2006) also devoted their attention to testing the hypothesis of the scaling property.

5.2 Implementation of Stochastic Frontier Models with Determinants of Inefficiency

Most of the parametric models discussed in this section can be easily implemented using Stata since the estimation routines for the basic stochastic frontier model in Stata also provide options to specify the pretruncated mean and/or variance of inefficiency as a function of the exogenous variables.

In particular, the Cornwell et al. (1990) estimator of course can be implemented using standard panel techniques and linear projections. The model specified in (23) can be estimated using nonlinear least squares without imposing any parametric assumption on the distribution of u_i^* or by the maximum likelihood based on the parametric distribution of the composed error. The maximum likelihood approach can be implemented in Stata by setting up the likelihood using the `sfmodel` command from Kumbhakar et al. (2015) with the option `hscale(·)` and the log likelihood can be maximized using the standard Stata routine `ml max`.

A snippet of Stata codes for implementing stochastic frontier models with determinants of inefficiency is provided in Table 8.

---

[20]It is worth mentioning here that although being popularised by Wang and Schmidt (2002), Simar et al. (1994) appear to be the first who analysed the scaling property in detail.

Table 8: Illustration for implementation of stochastic frontier models with determinants of inefficiency

```
****************************************************************************
**** Illustration for implementation of stochastic frontier models *****
*************** with determinants of inefficiency *********************
*********************** Partial Stata Codes ***************************
****************************************************************************
/* Note that output, inputs, and exogenous variables are stored in
   global Stata variables $y, $xlist, $zlist, respectively. Output and
   inputs are in log forms */
/* Implementation for the Kumbhakar et al. (1991) model (using the
   commands from Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(tnormal) emean($zlist)
predict ineff_KGM, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_KGM, bc /* Predict inefficiency, i.e., E(exp(-u)|e) */
/* Implementation for the Caudill et al. (1995)  model (using the
   commands from Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(hnormal) usigma($zlist)
predict ineff_CFG, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_CFG, bc /* Predict inefficiency, i.e., E(exp(-u)|e) */
/* Implementation for the Wang (2002) model (using the commands from
   Belotti et al. (2013) ) */
sfcross $y $xlist, distribution(tnormal) emean($zlist) usigma($zlist)
predict ineff_W, u /* Predict inefficiency, i.e., E(u|e) */
predict eff_W, bc /* Predict inefficiency, i.e., E(exp(-u)|e) */
/* Implementation for the Wang and Schmidt (2002) model (using the
   commands from Kumbhakar et al. (2015) ) */
sfmodel $y, prod dist(t) frontier($xlist) scaling hscale($zlist) tau cu
ml max
sf_predict, jlms(ineff_WH) /* Predict inefficiency, i.e., E(u|e) */
sf_predict, bc(eff_WH) /* Predict inefficiency, i.e., E(exp(-u)|e) */
```

6 Semi-parametric Stochastic Frontier Models

Another generalisation of the stochastic frontier paradigm is to relax parametric assumptions imposed on the functional form of the production frontier and to some extent the parametric assumption on the distribution of inefficiency.

6.1 The variety of models

Banker and Maindiratta (1992) appear to be among the first attempting to estimate stochastic frontier models semiparametrically. They proposed a framework combining stochastic and deterministic frontier (i.e., data envelopment analysis) approaches and developed techniques for the maximum likelihood estimation with nonparametric characterisation of classes of monotone and concave production frontiers. Other early attempts belong to Fan et al. (1996) and Kneip and Simar (1996), who suggested using nonparametric kernel regression methods in the framework of parametric maximum likelihood estimation. Specifically, Fan et al. (1996) proposed a multi-stage semiparametric likelihood estimation approach, in which the Nadaraya-Watson nonparametric estimator is employed in the first stage to estimate the average production relationship and a full parametric maximum likelihood estimator is used in the next stage to back out the conditional mean of inefficiency, which is utilised in the last stage to identify the frontier. Kneip and Simar (1996) followed a similar strategy with Fan et al. (1996) but in a panel data setting.

Semi-parametric panel frontiers were also considered in a series of papers by Park et al. (1998, 2003, 2007) wherein firm inefficiency effects are endogenous. They constructed the semiparametric efficiency bounds and the corresponding semiparametric efficient estimators for such models under differing assumptions about the form of endogeneity, the serial dependence of the idiosyncratic error, and possible dynamic structures for the panel data model. They used kernel smoothers in these modeling efforts as did Adams et al. (1997, 1999), and Adams and Sickles (2007). Current Stata software for these models is in the development stage in Badunenko et al. (2021), while existing Matlab and R codes for these semi-nonparametric panel frontier models can be found on the website that accompanies Sickles and Zelenyuk (2019) (see the links in footnote 14 and footnote 15). Model averaging methods utilised in Sickles (2005), Duygun et al. (2017) and Isaksson et al. (2020) also can be found on that website and are currently being used in developing consensus productivity growth estimates for the United Nations Industrial Development Organization (UNIDO).

Another approach to estimate stochastic frontier models semiparametrically was pro-

posed by Kumbhakar et al. (2007), who suggested employing the local likelihood estimation. The key distinction between this approach and the parametric likelihood approach is that the estimation is localised in the sense that individual contribution to the likelihood is determined by the kernel-based weights instead of the equal weights. Kneip et al. (2015) extended the work of Kumbhakar et al. (2007) by relaxing the parametric assumption on the distribution of inefficiency, while Park et al. (2015) suggested an alternative parameterisation of the local likelihood and outlined a framework for allowing categorical variables in the local likelihood context.

Semi-parametric methods have also been introduced into the stochastic frontier paradigm to deal with specifications of inefficiency. Cornwell et al. (1990) utilised a second-order Taylor series in a time trend to model time varying inefficiency while Lee and Schmidt (1993) specified the time-varying and cross-sectionally varying inefficiency using a one-factor multiplicative model. Extensions to mixed models and more general factor models were pursued by Ahn et al. (2007, 2013), Kneip et al. (2004), Kneip and Sickles (2011), and Kneip et al. (2012). The latter model is programmed in Matlab and R on the software website for Sickles and Zelenyuk (2019) and Sickles et al. (2020) (see the links in footnote 14 and footnote 15) and its coding in Stata is in process in Badunenko et al. (2021). Finally, the Kneip and Sickles (2011) general cross-sectional and time-varying factor model is available in the R package from Oualid Bada and discussed at length in Bada and Liebl (2014).[21]

6.2   Simar et al. (2017) Model

Recently, Simar et al. (2017) suggested using the local least squares method as an alternative for the local likelihood approach to estimate the stochastic frontier models. The local least squares approach is much simpler to compute and easier to implement compared to the local likelihood, and we will focus our discussion here on this approach.

The model in Simar et al. (2017) can be formulated as follows

$$y_i = m\left(x_i, \ z_i\right) + v_i - u_i, \quad i = 1, \ldots, n, \tag{24}$$

where $m\left(x_i, z_i\right)$ is the production frontier, $y_i \in \Re_+^1$ is the output, $x_i \in \Re_+^p$ is a vector of inputs, and $z_i \in \Re^k$ is a vector of k variables that can influence the production process. v_i is statistical noise, which is assumed to have zero mean, i.e., $E\left(v_i|x_i, z_i\right) = 0$, and positive finite variance, i.e., $VAR\left(v_i|x_i, z_i\right) \in (0, \infty)$. Meanwhile, u_i is the inefficiency term following a one-sided distribution, with a positive mean, i.e., $E\left(u_i|x_i, z_i\right) = \mu_u\left(x_i, z_i\right) \in (0, \infty)$

---

[21]Software instructions and downloadable codes are accessible at `https://www.jstatsoft.org/article/view/v059i06`.

and positive finite variance, i.e., $VAR\left(u_i|x_i, z_i\right) \in (0, \infty)$. As in other stochastic frontier models, u_i and v_i are also assumed to be independent, conditionally on (x_i, z_i).

Now, let us define

$$\varepsilon_i^* = v_i - u_i + \mu_u\left(x_i, z_i\right), \tag{25}$$

and

$$r_1\left(x_i, z_i\right) = m\left(x_i, z_i\right) - \mu_u\left(x_i, z_i\right). \tag{26}$$

We can rewrite (24) as

$$y_i = r_1\left(x_i, z_i\right) + \varepsilon_i^*. \tag{27}$$

Since $E\left(\varepsilon_i^*|x_i, z_i\right) = 0$, we can use standard nonparametric methods (e.g., local polynomial least squares) to estimate $r_1\left(x_i, z_i\right)$. In order to estimate the individual inefficiency, we also need to make a parametric assumption on the distribution of inefficiency, e.g.,

$$u_i|x_i, z_i \sim \mathcal{N}^+\left(0, \sigma_u^2\left(x_i, z_i\right)\right). \tag{28}$$

With the distributional assumption, the conditional mean of inefficiency can be estimated using the following relationships

$$\sigma_u^3\left(x_i, z_i\right) = \sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi - 4}\right) r_3\left(x_i, z_i\right), \tag{29}$$

and

$$\mu_u\left(x_i, z_i\right) = \sqrt{\frac{2}{\pi}}\sigma_u\left(x_i, z_i\right), \tag{30}$$

where $r_3\left(x_i, z_i\right) = E\left(\left(\varepsilon_i^*\right)^3|x_i, z_i\right)$ is the third moment of the composed error. Specifically, the residuals from the nonparametric estimation of the model in (27), $\widehat{\varepsilon}_i^*$, can be utilised to obtain the nonparametrical estimates of the third moment of the composed error, $\widehat{r}_3\left(x_i, z_i\right)$. The estimates of technical inefficiency then can be obtained by plugging the $\widehat{r}_3\left(x_i, z_i\right)$ into (29) and (30).[22]

6.3 Implementation of Simar et al. (2017) Model

Estimation of the Simar et al. (2017) model can be implemented using the standard Stata routines with a bit of additional programming. The key command is `npregress` which helps to perform the local least square estimation in Stata environment. As an

---

[22]The distributional assumptions on u_i and v_i allow obtaining a generalised version of JLMS type estimates, although more interesting in the semi/non-parametric context are the estimates of $E(u_i|x_i = x, z_i = z)$, which can be done for any values of interest for (x, z). The elasticities of $E(u_i|x_i = x, z_i = z)$ can also be obtained, which can be done without any parametric assumptions on distributions, just by assuming that u_i comes from a one-parameter scale family (see Section 4 in Simar et al., 2017, for more details).

illustration, we provide here in Table 9 a part of a Stata do file that implements the procedure discussed in the previous subsection to estimate the Simar et al. (2017) model.

Similarly, one can implement the estimation of the Simar et al. (2017) model in R with the local least squares estimation being carried out by the `np` package with a bit of additional programming similar to the one we presented here (and as was done by Parmeter and Zelenyuk (2019)). The implementation of the model in Matlab requires more effort since one needs to write his/her own codes for the local least square estimation (as was done by Simar et al. (2017)). Preparation of user-friendly packages in R and Matlab is currently in progress.

Table 9: Illustration for implementation of the Simar et al. (2017) model

```
****************************************************************************
*** Illustration for implementation of the Simar et al. (2017) model****
**************************Partial Stata Codes************************
****************************************************************************
/* Estimate the model in (27) using local linear estimator and store
   predicted value in variable r1hat. Note that output, inputs, and
   exogenous variables are stored in global Stata variables $y, $xlist,
   $zlist, respectively. Output and inputs are in log forms */
/* Note that the default options in npregress command is to use
   Epanechnikov kernel and select bandwidth by cross-validation, i.e.,
   by minimizing the integrated mean squared error of the prediction. */
npregress kernel $y $xlist $zlist, estimator(linear) predict(r1hat)
   noderivatives
/* Obtain the residual and the residual cubed from  estimation of the
   model in equation (27)*/
gen ehat = $y - r1hat
gen ehat3 = ehat^3
/* Estimate the third moment of the composed error using local linear
   estimator and store predicted value in variable r3hat */
npregress kernel ehat3 $xlist $zlist, estimator(linear) predict(r3hat)
   noderivatives
/* Calculate sigma u hat cubed using equation (29) */
gen sigmauhat3 = sqrt(_pi/2)*(_pi/(_pi-4))*r3hat
/* Calculate sigma u hat. Note that following Simar et al. (2017), we
   set negative values of sigma u hat equal zero */
gen sigmauhat = max(sigmauhat3^(1/3),0)
/* Calculate estimated values of inefficiency using equation (30)*/
gen muhat = sqrt(2/_pi)*sigmauhat
```

7 Empirical Illustration

In this section, we provide a small empirical illustration of the models discussed in the previous sections, including the basic stochastic frontier model, the stochastic panel data models, and the semi-parametric stochastic frontier model.[23][24] For this purpose, we use the data set about rice producers in the Philippines, which was also utilised for the similar purpose and popularised in the literature by Coelli et al. (2005).[25][26]

Specifically, the data set includes the information about 43 rice producers in the Tarlac region of the Philippines in a period of 8 years from 1990 to 1997. We extract from the data set the information on one output and three inputs including the area planted, labour used and fertiliser used. The output is measured in tonnes of freshly threshed rice, while the inputs are measured in hectares, man-days of family and hired labour, and kilograms of active ingredients, respectively (see more details about the description of the data in Coelli et al., 2005).

For this empirical illustration, we deliberately apply all the models to the data and focus our discussion on the estimated inefficiency to reflect the differences in results across the models. Moreover, for all the models that require a functional form for the production relationship, we assume a linear in log production function, i.e., the Cobb-Douglas production function.[27] The Stata codes for implementing this analysis are provided in the Appendix.

The summary statistics of the estimated inefficiency are provided in Table 10, and their histograms are shown in Figure 1.[28] Meanwhile, the variations of the estimated inefficiency across the years are shown in Figure 2.

At first glance, we can see that the means of estimated inefficiency vary significantly across the models, ranging from 0.20 (the Battese and Coelli (1992) model) to 0.45 (the Kumbhakar et al. (2014) model).[29] This is understandable since each model depends on

---

[23]For the results to some extent to be comparable, we deliberately do not include in this empirical illustration the stochastic frontier models with determinants of inefficiency.

[24]Also, due to the computational difficulty in optimising the likelihood function, the result from Kumbhakar (1990) is not available for the dataset used in this empirical illustration.

[25]Downloaded from `http://www.uq.edu.au/economics/cepa/crob2005/software/CROB2005.zip`.

[26]For an illustration with this data with various DEA models see, e.g., Simar and Zelenyuk (2020).

[27]To estimate the cross-sectional models, e.g., Aigner et al. (1977) and Simar et al. (2017) models, we pool the data across years.

[28]The estimated distribution of estimated inefficiency from the Simar et al. (2017) model is showing some mass at zero (i.e., the phenomenon referred to as 'wrong skewness' in stochastic frontier analysis) because 79 out 344 observations have $\hat{\sigma}_u^3(x_i, z_i) < 0$ and their inefficiency is set to equal to 0.

[29]It is important to clarify here that for all the models, the means we refer to are averages of the estimates of individual inefficiencies.

Table 10: Summary statistics of the estimated inefficiency

Models	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
Aigner et al. (1977)	0.36	0.25	0.04	0.18	0.29	0.47	2.00
Schmidt and Sickles (1984) (fixed effects)	0.34	0.20	0.00	0.18	0.30	0.50	0.98
Schmidt and Sickles (1984) (random effects)	0.23	0.12	0.00	0.13	0.21	0.33	0.60
Pitt and Lee (1981)	0.21	0.14	0.05	0.10	0.16	0.31	0.70
Cornwell et al. (1990)	0.44	0.27	0.00	0.27	0.41	0.57	2.09
Battese and Coelli (1992)	0.20	0.15	0.04	0.08	0.14	0.28	0.92
Greene (2005a,b) (random effects)	0.33	0.24	0.03	0.17	0.27	0.43	1.89
Greene (2005a,b) (fixed effects)	0.35	0.23	0.02	0.18	0.30	0.47	1.87
Kumbhakar et al. (2014) (Total)	0.45	0.25	0.10	0.27	0.38	0.55	2.01
Kumbhakar et al. (2014) (Persistent)	0.15	0.10	0.02	0.06	0.13	0.24	0.49
Kumbhakar et al. (2014) (Transitory)	0.29	0.20	0.03	0.16	0.24	0.37	1.67
Simar et al. (2017)	0.25	0.14	0.00	0.17	0.30	0.35	0.45

different sets of assumptions. Moreover, it is important for practitioners to be aware of these differences and carefully justify the assumptions of the model of their choice before proceeding with their analysis. For example, with this data set, the difference in estimated inefficiency between the fixed effects and random effects frameworks is significant when all the unobserved individual heterogeneity is viewed as inefficiency (e.g, in the Schmidt and Sickles (1984) model), but the difference is minimal when inefficiency is distinguished from the unobserved individual heterogeneity (e.g., in the Greene (2005a,b) model).

Furthermore, recall that the temporal pattern of inefficiency is also specified differently in different models. As illustrated in Figure 2, the estimated inefficiency is constant over time in the Schmidt and Sickles (1984) and Pitt and Lee (1981) models, but follows a quadratic trend in the Cornwell et al. (1990) model and has a linear trend in Battese and Coelli (1992). Meanwhile, other models, such as Greene (2005a,b) and Kumbhakar et al. (2014), do not impose any temporal patterns on the time varying component of inefficiency.

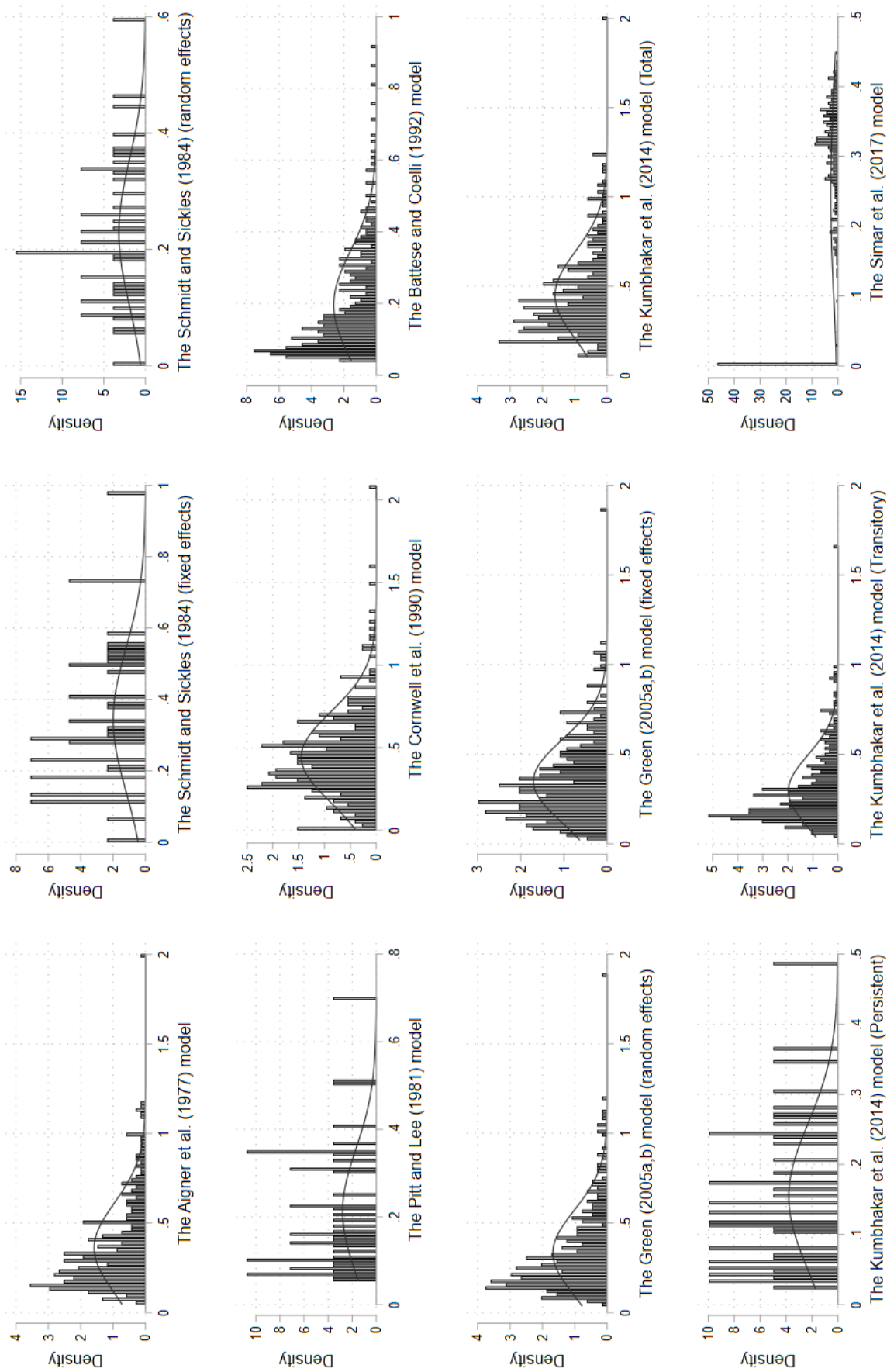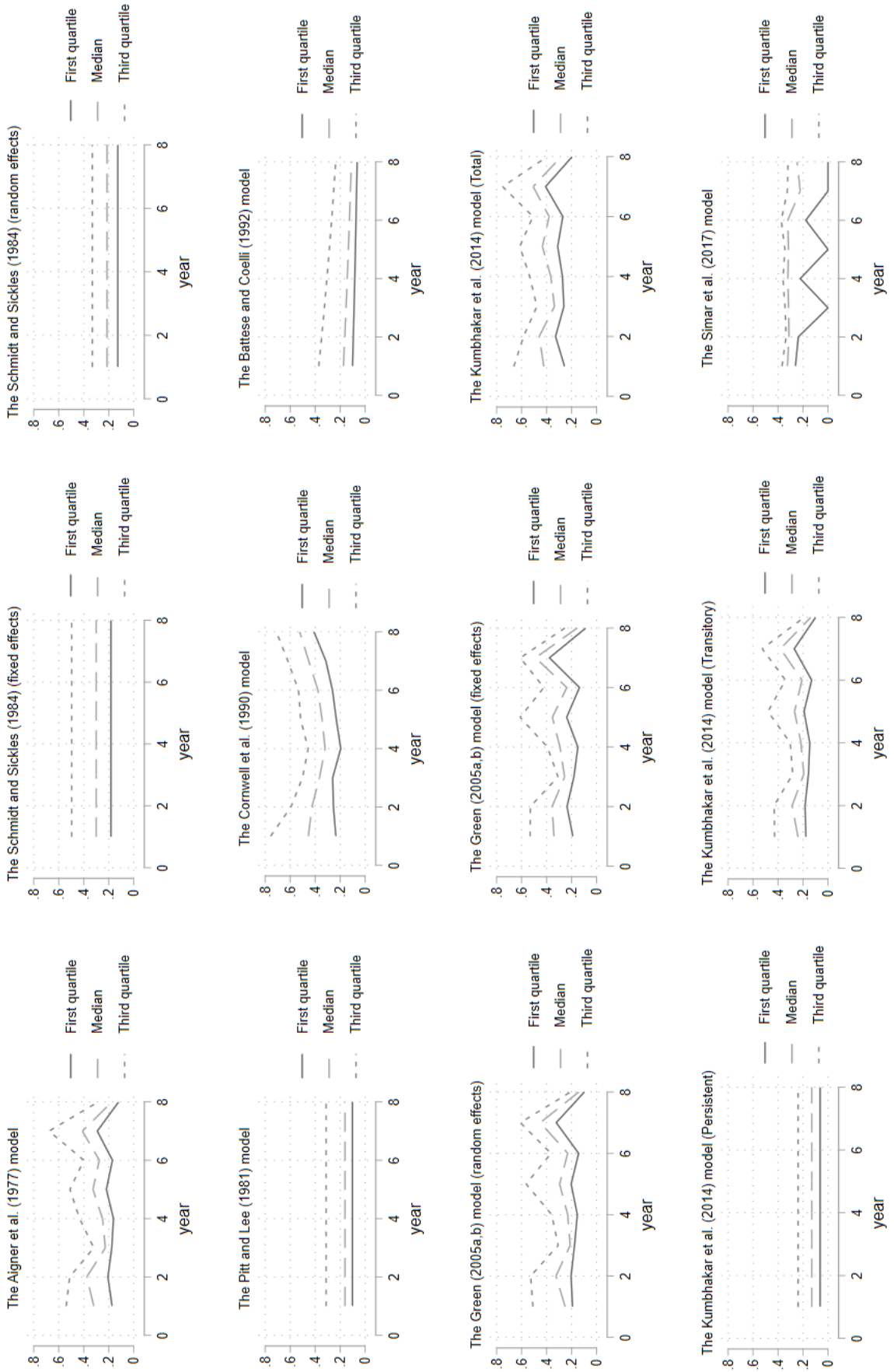Figure 1: Histograms of the estimated inefficiency

28

Figure 2: The estimated inefficiency across years

8 Concluding Remarks

This chapter discussed a variety of stochastic frontier models to estimate the technical efficiency of production units. Our chapter also documented the estimation routines used to implement these methods for practitioners, especially those who are willing to use Stata, but also with tips on where to find analogous programs for R and Matlab users.

Although many recent developments in the field were covered in this chapter, it was still a relatively brief introduction to the stochastic frontier paradigm with some other generalisations remaining untouched, such as Bayesian stochastic frontier[30], stochastic metafrontier[31], spillovers and spatial frontiers[32], and endogeneity[33]. We refer interested readers to more extensive resources (e.g., Sickles and Zelenyuk, 2019; Kumbhakar et al., 2021a,b) for more detailed discussions of these and other topics.

Finally, many other important developments in the field are still in progress, thus we encourage readers to check for updates as well as contribute themselves to such developments and discoveries.

Acknowledgments

---

[30]E.g., see Van den Broeck et al., 1994; Griffin and Steel, 2004, 2007; Liu et al., 2017.

[31]E.g., see Battese et al., 2004; O'Donnell et al., 2008; Huang et al., 2014.

[32]E.g., see Glass et al., 2016; Orea and Álvarez, 2019.

[33]E.g., see Amsler et al., 2016; Kutlu, 2010; Karakaplan and Kutlu, 2015, 2017; Karakaplan, 2017.

Appendix

Listing 1: Stata codes for the empirical illustration

```stata
set more off
clear all
log using "SFABookChapter_Empirical", text replace
// Load data
import delimited "rice.txt", delimiter(space) varnames(1) ///
        encoding(ISO-8859-2)
// Generate variables
foreach X of varlist prod area labor npk {
        generate l`X' = ln(`X')
        }
global y lprod
global xlist llabor larea lnpk  /*Cobb-Douglas function*/
global id fmercode
global t yeardum
xtset $id $t
***********************************************************************
/* the Aigner et al. (1977) model */
***********************************************************************
sfcross $y $xlist, distribution(hnormal)
estimates store ALS
predict ineff_ALS, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_ALS "The Aigner et al. (1977) model"
***********************************************************************
/* the  Schmidt and Sickles (1984)model */
***************************Partial Stata Codes********************
/* the fixed effect framework */
sfpanel $y $xlist, model(fe)
estimates store SS_fe
predict ineff_SS_fe, u /* Predict inefficiency*/
label variable ineff_SS_fe ///
        "The Schmidt and Sickles (1984)(fixed effects)"
/* the random effect framework */
sfpanel $y $xlist, model(regls)
estimates store SS_re
predict ineff_SS_re, u /* Predict inefficiency*/
label variable ineff_SS_re ///
        "The Schmidt and Sickles (1984) (random effects)"
```

```
*****************************************************************
/* the Pitt and Lee (1981) model */
*****************************************************************
sfpanel $y $xlist, model(pl81)
estimates store PL
predict ineff_PL, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_PL "The Pitt and Lee (1981) model"
*****************************************************************
/* the Cornwell et al. (1990) model */
*****************************************************************
sfpanel $y $xlist, model(fecss)
estimates store CSS
predict ineff_CSS, u /* Predict inefficiency*/
label variable ineff_CSS "The Cornwell et al. (1990) model"
*****************************************************************
/* the Battese and Coelli (1992) model */
*****************************************************************
sfpanel $y $xlist, model(bc92)
estimates store BC
predict ineff_BC, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_BC "The Battese and Coelli (1992) model"
*****************************************************************
/* the Green (2005a,b) models */
*****************************************************************
/* the random effect framework */
sfpanel $y $xlist, distribution(hnormal) model(tre)
estimates store G_tre
predict ineff_G_tre, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_G_tre "The Green (2005a,b) model (random effects)"
/* the fixed effect framework */
/* Implementation using the marginal maximum likelihood estimator */
sftfe $y $xlist,  distribution(hnormal) estimator(within)
estimates store G_mmle
predict ineff_G_mmle, u /* Predict inefficiency, i.e., E(u|e) */
label variable ineff_G_mmle "The Green (2005a,b) model (fixed effects)"
*****************************************************************
/* the Kumbhakar et al. (2014) models */
*****************************************************************
xtreg $y $xlist, re
estimates store KLH
predict alp, u /* Obtain estimates of alpha */
predict esl, e /* Obtain estimates of the composed error */
/* Estimate equation (19) using the basic stochastic frontier model to
    obtain persistent (in)efficiency */
```

```
gen constant = 1 /* Generate contant to use in sfcross */
sfcross alp constant , distribution ( hnormal ) noconstant
predict ineff_KHL_pers , u /* Predict inefficiency , i.e., E(u|e) */
label variable ineff_KHL_pers ///
        "The Kumbhakar et al. (2014) model ( Persistent )"
/* Estimate equation (19) using the basic stochastic frontier model to
    obtain transitory (in)efficiency */
sfcross esl constant , distribution ( hnormal ) noconstant
predict ineff_KHL_trans , u /* Predict inefficiency , i.e., E(u|e) */
label variable ineff_KHL_trans ///
        "The Kumbhakar et al. (2014) model ( Transitory )"
/* Calculate the total inefficiency */
gen ineff_KLH = ineff_KHL_pers + ineff_KHL_trans
label variable ineff_KLH "The Kumbhakar et al. (2014) model ( Total )"
****************************************************************************
/* the Simar et al. (2017) */
****************************************************************************
/* Estimate model in (27) using local linear estimator and store
    predicted value in variable r1hat. Note that output , inputs , and
    exogenous variables are stored in global Stata variables $y , $xlist ,
    $zlist , respectively */
/* Note that the default options in npregress command is to use
    Epanechnikov kernel and select bandwidth by cross - validation , i.e.,
    by minimizing the integrated mean squared error of the prediction. */
npregress kernel $y $xlist , estimator ( linear ) predict ( r1hat ) ///
        noderivatives
/* Obtain the residual and the residual cubed from the estimation of
    model in equation (27)*/
gen ehat = $y - r1hat
gen ehat3 = ehat^3
/* Estimate the third momment of the composed error using local linear
    estimator and store predicted value in variable r3hat */
npregress kernel ehat3 $xlist , estimator ( linear ) predict ( r3hat ) ///
        noderivatives
/* Calculate sigma u hat cubed using equation (29) */
gen sigmauhat3 = sqrt ( _pi /2) *( _pi /( _pi -4) ) * r3hat
/* Calculate sigma u hat. Note that following Simar et al. (2017) , we
    set negative values of sigma u hat equal zero */
gen sigmauhat = max ( sigmauhat3 ^(1/3) ,0)
/* Calculate estimated values of inefficiency using equation (30)*/
gen ineff_SKVZ = sqrt (2/ _pi ) * sigmauhat
label variable ineff_SKVZ "The Simar et al. (2017) model"
```

```stata
*****************************************************************************
/* Summarising and exporting the results */
*****************************************************************************
/* The estimated coefficients of the frontiers */
esttab ALS SS_fe SS_re PL CSS BC G_tre G_mmle KLH
esttab ALS SS_fe SS_re PL CSS BC G_tre G_mmle KLH  ///
        using coefficients.csv, replace
/* The estimated inefficiency */
global myvars ineff_ALS ineff_SS_fe ineff_SS_re ineff_PL ineff_CSS ///
        ineff_BC ineff_G_tre ineff_G_mmle ineff_KLH ineff_KHL_pers ///
        ineff_KHL_trans ineff_SKVZ
estpost summarize $myvars, detail
esttab using inefficiency.csv, ///
        cells("count mean sd min p25 p50 p75 max") replace
*****************************************************************************
/* Histograms of esitmated inefficiency */
*****************************************************************************
foreach X of varlist $myvars {
        histogram `X', bin(100) normal `kden'
        graph save `X'.gph, replace
        }
graph combine ineff_ALS.gph ineff_SS_fe.gph ineff_SS_re.gph ///
        ineff_PL.gph ineff_CSS.gph ineff_BC.gph ineff_G_tre.gph ///
        ineff_G_mmle.gph ineff_KLH.gph ineff_KHL_pers.gph ///
        ineff_KHL_trans.gph ineff_SKVZ.gph, col(3) scale(1)
graph export histogramineff.png, replace
*****************************************************************************
/* Plot esitmated inefficiency across years */
*****************************************************************************
sort yeardum
label variable yeardum "year"
foreach X of varlist $myvars {
        by yeardum, sort: egen `X'_Q1 = pctile(`X'), p(25)
        by yeardum, sort: egen `X'_Q2 = pctile(`X'), p(50)
        by yeardum, sort: egen `X'_Q3 = pctile(`X'), p(75)
        label variable `X'_Q1 "First quartile"
        label variable `X'_Q3 "Third quartile"
        label variable `X'_Q2 "Median"
        local labeltext : variable label `X'
        graph two line `X'_Q1 `X'_Q2 `X'_Q3 yeardum, ///
                title(`labeltext', size(small))
        graph save `X'_trend.gph, replace
        }
graph combine ineff_ALS_trend.gph ineff_SS_fe_trend.gph ///
```

```stata
        ineff_SS_re_trend.gph ineff_PL_trend.gph ///
        ineff_CSS_trend.gph ineff_BC_trend.gph ///
        ineff_G_tre_trend.gph ineff_G_mmle_trend.gph ///
        ineff_KLH_trend.gph ineff_KHL_pers_trend.gph ///
        ineff_KHL_trans_trend.gph ineff_SKVZ_trend.gph, ///
        col(3) scale(1) xcommon ycommon
graph export allineff_trend.png, replace
log close
```

References

Adams, R. M., Berger, A. N. and Sickles, R. C. (1997), Computation and inference in semiparametric efficient estimation, *in* H. Amman, B. Rustem and A. Whinston, eds, 'Computational Approaches to Economic Problems', Springer US, Boston, MA, pp. 57–70.

Adams, R. M., Berger, A. N. and Sickles, R. C. (1999), 'Semiparametric approaches to stochastic panel frontiers with applications in the banking industry', *Journal of Business & Economic Statistics* **17**(3), 349–358.

Adams, R. M. and Sickles, R. C. (2007), 'Semiparametric efficient distribution free estimation of panel models', *Communications in Statistics–Theory and Methods* **36**(13), 2425–2442.

Ahn, S. C., Lee, Y. H. and Schmidt, P. (2007), 'Stochastic frontier models with multiple time-varying individual effects', *Journal of Productivity Analysis* **27**(1), 1–12.

Ahn, S. C., Lee, Y. H. and Schmidt, P. (2013), 'Panel data models with multiple time-varying individual effects', *Journal of Econometrics* **174**(1), 1–14.

Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977), 'Formulation and astimation of stochastic frontier production function models', *Journal of Econometrics* **64**(6), 1263–1297.

Almanidis, P., Qian, J. and Sickles, R. C. (2014), Stochastic frontier models with bounded inefficiency, *in* R. C. Sickles and W. C. Horrace, eds, 'Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications', Springer New York, pp. 47–81.

Almanidis, P. and Sickles, R. C. (2012), The skewness problem in stochastic frontier models: Fact or fiction?, *in* I. V. Keilegom and P. Wilson, eds, 'Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift in Honor of Léopold Simar', Springer Publishing: New York.

Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006), 'Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics', *Journal of Productivity Analysis* **25**(3), 201–212.

Álvarez, I. C., Barbero, J., Zofío, J. L. et al. (2017), 'A panel data toolbox for matlab', *Journal of Statistical Software* **76**(6), 1–27.

Amsler, C., Prokhorov, A. and Schmidt, P. (2016), 'Endogeneity in stochastic frontier models', *Journal of Econometrics* **190**(2), 280–288.

Bada, O. and Liebl, D. (2014), 'phtt: Panel data analysis with heterogeneous time trends in R', *Journal of Statistical Software* **59**(6).

Badunenko, O., Sickles, R. C. and Zelenyuk, V. (2021), Productivity and efficiency analysis in Stata. monograph in progress, Rice University.

Banker, R. D. and Maindiratta, A. (1992), 'Maximum likelihood estimation of monotone and concave production frontiers', *Journal of Productivity Analysis* **3**(4), 401–415.

Battese, G. E. and Broca, S. S. (1997), 'Functional forms of stochastic frontier production functions and models for technical inefficiency effects: a comparative study for wheat farmers in pakistan', *Journal of Productivity Analysis* **8**(4), 395–414.

Battese, G. E. and Coelli, T. J. (1988), 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data', *Journal of Econometrics* **38**(3), 387–399.

Battese, G. E. and Coelli, T. J. (1992), 'Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India', *Journal of Productivity Analysis* **3**(1-2), 153–169.

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**(2), 325–332.

Battese, G. E., Rao, D. P. and O'Donnell, C. J. (2004), 'A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies', *Journal of productivity analysis* **21**(1), 91–103.

Belotti, F., Daidone, S., Ilardi, G. and Atella, V. (2013), 'Stochastic frontier analysis using stata', *The Stata Journal* **13**(4), 719–758.

Belotti, F. and Ilardi, G. (2018), 'Consistent inference in fixed-effects stochastic frontier models', *Journal of Econometrics* **202**(2), 161–177.

Bogetoft, P. and Otto, L. (2019), *Benchmarking: Benchmark and Frontier Analysis Using DEA and SFA*, R package version 0.28.

Caudill, S. B., Ford, J. M. and Gropper, D. M. (1995), 'Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity', *Journal of Business & Economic Statistics* **13**(1), 105–111.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978), 'Measuring the efficiency of decision making units', *European Journal of Operational Research* **2**(6), 429–444.

Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014), 'Consistent estimation of the fixed effects stochastic frontier model', *Journal of Econometrics* **181**(2), 65–76.

Coelli, T. and Henningsen, A. (2020), *Package 'frontier'*, R package version 1.1-8.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G. E. (2005), *An introduction to efficiency and productivity analysis*, Springer Science & Business Media.

Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014), 'Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency', *Journal of Productivity Analysis* **42**(2), 123–136.

Comans, T., Nguyen, K.-H., Zelenyuk, V., Gray, L., Tran, A., Nguyen, B. H., Wang, Z. and Moretto, N. (2020), The cost of residential aged care, Technical Report Research Paper 9, The Royal Commission into Aged Care Quality and Safety.

Cornwell, C., Schmidt, P. and Sickles, R. C. (1990), 'Production frontiers with cross-sectional and time-series variation in efficiency levels', *Journal of Econometrics* **46**(1-2), 185–200.

Daraio, C., Kerstens, K. H., Nepomuceno, T. C. C. and Sickles, R. (2019), 'Productivity and efficiency analysis software: an exploratory bibliographical survey of the options', *Journal of Economic Surveys* **33**(1), 85–100.

Daraio, C., Kerstens, K., Nepomuceno, T. and Sickles, R. C. (2020), 'Empirical surveys of frontier applications: a meta-review', *International Transactions in Operational Research* **27**(2), 709–738.

Duygun, M., Hao, J., Isaksson, A. and Sickles, R. C. (2017), 'World productivity growth: A model averaging approach', *Pacific Economic Review* **22**(4), 587–619.

Fan, Y., Li, Q. and Weersink, A. (1996), 'Semiparametric estimation of stochastic production frontier models', *Journal of Business & Economic Statistics* **14**(4), 460–468.

Färe, R., He, X., Li, S. and Zelenyuk, V. (2019), 'A unifying framework for farrell profit efficiency measurement', *Operations Research* **67**(1), 183–197.

Farrell, M. J. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society. Series A (General)* **120**(3), 253–290.

Fenochietto, R. and Pessino, C. (2013), 'Understanding countries' tax effort', *IMF Working Paper* (WP/13/244).

Ferrier, G. D. and Lovell, C. K. (1990), 'Measuring cost efficiency in banking: Econometric and linear programming evidence', *Journal of Econometrics* **46**(1-2), 229–245.

Glass, A., Kenjegalieva, K. and Sickles, R. C. (2016), 'Spatial autoregressive and spatial Durbin stochastic frontier models for panel data', *Journal of Econometrics* **190**, 289–300.

Greene, W. (2004), 'Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems', *Health Economics* **13**(10), 959–980.

Greene, W. (2005*a*), 'Fixed and random effects in stochastic frontier models', *Journal of productivity analysis* **23**(1), 7–32.

Greene, W. (2005*b*), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics* **126**(2), 269–303.

Greene, W. H. (1980*a*), 'Maximum likelihood estimation of econometric frontier functions', *Journal of Econometrics* **13**(1), 27–56.

Greene, W. H. (1980*b*), 'On the estimation of a flexible frontier production model', *Journal of Econometrics* **13**(1), 101–115.

Greene, W. H. (1990), 'A gamma-distributed stochastic frontier model', *Journal of Econometrics* **46**(1), 141–163.

Griffin, J. E. and Steel, M. F. (2004), 'Semiparametric Bayesian inference for stochastic frontier models', *Journal of Econometrics* **123**(1), 121–152.

Griffin, J. E. and Steel, M. F. (2007), 'Bayesian stochastic frontier analysis using winbugs', *Journal of Productivity Analysis* **27**(3), 163–176.

Huang, C. J., Huang, T.-H. and Liu, N.-H. (2014), 'A new approach to estimating the metafrontier production function based on a stochastic frontier framework', *Journal of productivity Analysis* **42**(3), 241–254.

Isaksson, A., Shang, C. and Sickles, R. C. (2020), 'Nonstructural analysis of productivity growth for the industrialized countries: a jackknife model averaging approach', *Econometric Reviews* pp. 1–38.

Jondrow, J., Lovell, C. K., Materov, I. S. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**(2-3), 233–238.

Karakaplan, M. and Kutlu, L. (2015), 'Handling endogeneity in stochastic frontier analysis', *Available at SSRN 2607276* .

Karakaplan, M. U. (2017), 'Fitting endogenous stochastic frontier models in stata', *The Stata Journal* **17**(1), 39–55.

Karakaplan, M. U. and Kutlu, L. (2017), 'Endogeneity in panel stochastic frontier models: an application to the Japanese cotton spinning industry', *Applied Economics* **49**(59), 5935–5939.

Kneip, A. and Sickles, R. C. (2011), Panel data, factor models, and the Solow residual, *in* I. Van Keilegom and P. W. Wilson, eds, 'Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift for Léopold Simar', Physica-Verlag HD, Heidelberg, pp. 83–114.

Kneip, A., Sickles, R. C. and Song, W. (2004), Functional data analysis and mixed effect models, *in* J. Antoch, ed., 'COMPSTAT 2004 — Proceedings in Computational Statistics', Physica-Verlag HD, Heidelberg, pp. 315–326.

Kneip, A., Sickles, R. C. and Song, W. (2012), 'A new panel data treatment for heterogeneity in time trends', *Econometric Theory* pp. 590–628.

Kneip, A. and Simar, L. (1996), 'A general framework for frontier estimation with panel data', *Journal of Productivity Analysis* **7**(2), 187–212.

Kneip, A., Simar, L. and Van Keilegom, I. (2015), 'Frontier estimation in the presence of measurement error with unknown variance', *Journal of Econometrics* **184**(2), 379–393.

Kumbhakar, S. C. (1987), 'The specification of technical and allocative inefficiency in stochastic production and profit frontiers', *Journal of Econometrics* **34**(3), 335–348.

Kumbhakar, S. C. (1990), 'Production frontiers, panel data, and time-varying technical inefficiency', *Journal of Econometrics* **46**(1-2), 201–211.

Kumbhakar, S. C., Ghosh, S. and McGuckin, J. T. (1991), 'A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms', *Journal of Business & Economic Statistics* **9**(3), 279–286.

Kumbhakar, S. C., Lien, G. and Hardaker, J. B. (2014), 'Technical efficiency in competing panel data models: a study of Norwegian grain farming', *Journal of Productivity Analysis* **41**(2), 321–337.

Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. G. (2007), 'Nonparametric stochastic frontiers: A local maximum likelihood approach', *Journal of Econometrics* **137**(1), 1–27.

Kumbhakar, S. C., Parmeter, C. F. and Zelenyuk, V. (2021*a*), Stochastic frontier analysis: Foundations and advances I, *in* S. C. Ray, R. Chambers and S. Kumbhakar, eds, 'Handbook of Production Economics', Springer Singapore, Singapore, pp. 1–40.

Kumbhakar, S. C., Parmeter, C. F. and Zelenyuk, V. (2021*b*), Stochastic frontier analysis: Foundations and advances II, *in* S. C. Ray, R. Chambers and S. Kumbhakar, eds, 'Handbook of Production Economics', Springer Singapore, Singapore, pp. 1–38.

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

Kumbhakar, S. C. and Tsionas, E. G. (2008), 'Estimation of input-oriented technical efficiency using a nonhomogeneous stochastic production frontier model', *Agricultural Economics* **38**(1), 99–108.

Kumbhakar, S. C., Wang, H. and Horncastle, A. P. (2015), *A practitioner's guide to stochastic frontier analysis using Stata*, Cambridge University Press.

Kutlu, L. (2010), 'Battese-Coelli estimator with endogenous regressors', *Economics Letters* **109**(2), 79–81.

Lee, Y. H. and Schmidt, P. (1993), A production frontier model with flexible temporal variation in technical efficiency, *in* H. O. Fried, S. S. Schmidt and C. K. Lovell, eds, 'The measurement of productive efficiency: Techniques and applications', Oxford University Press New York, pp. 237–255.

Liu, J., Sickles, R. C. and Tsionas, E. G. (2017), 'Bayesian treatments for panel data stochastic frontier models with time varying heterogeneity', *Econometrics* **5**, 1–21.

Malikov, E., Kumbhakar, S. C. and Tsionas, M. G. (2016), 'A cost system approach to the stochastic directional technology distance function with undesirable outputs: The case of US banks in 2001–2010', *Journal of Applied Econometrics* **31**(7), 1407–1429.

Meeusen, W. and van den Broeck, J. (1977), 'Efficiency estimation from Cobb-Douglas production functions with composed error', *International Economic Review* **18**(2), 435–444.

Mutter, R. L., Greene, W. H., Spector, W., Rosko, M. D. and Mukamel, D. B. (2013), 'Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes', *Journal of Productivity Analysis* **39**(2), 101–110.

Nguyen, B. H. and Zelenyuk, V. (2021), Aggregation of outputs and inputs for DEA analysis of hospital efficiency: Economics, operations research and data science perspectives, *in* J. Zhu and V. Charles, eds, 'Data-Enabled Analytics: DEA for Big Data', Springer. forthcoming.

O'Donnell, C. J., Rao, D. P. and Battese, G. E. (2008), 'Metafrontier frameworks for the study of firm-level efficiencies and technology ratios', *Empirical economics* **34**(2), 231–255.

Orea, L. and Álvarez, I. C. (2019), 'A new stochastic frontier model with cross-sectional effects in both noise and inefficiency terms', *Journal of Econometrics* **213**(2), 556–577.

Park, B. U., Sickles, R. C. and Simar, L. (1998), 'Stochastic panel frontiers: A semiparametric approach', *Journal of Econometrics* **84**(2), 273–301.

Park, B. U., Sickles, R. C. and Simar, L. (2003), 'Semiparametric-efficient estimation of ar (1) panel data models', *Journal of Econometrics* **117**(2), 279–309.

Park, B. U., Sickles, R. C. and Simar, L. (2007), 'Semiparametric efficient estimation of dynamic panel data models', *Journal of Econometrics* **136**(1), 281–301.

Park, B. U., Simar, L. and Zelenyuk, V. (2015), 'Categorical data in local maximum likelihood: theory and applications to productivity analysis', *Journal of Productivity Analysis* **43**(2), 199–214.

Parmeter, C. F. and Zelenyuk, V. (2019), 'Combining the virtues of stochastic frontier and data envelopment analysis', *Operations Research* **67**(6), 1628–1658.

Pitt, M. M. and Lee, L.-F. (1981), 'The measurement and sources of technical inefficiency in the Indonesian weaving industry', *Journal of Development Economics* **9**(1), 43–64.

Rosko, M. D. (2001), 'Cost efficiency of US hospitals: a stochastic frontier approach', *Health Economics* **10**(6), 539–551.

Schmidt, P. and Sickles, R. C. (1984), 'Production frontiers and panel data', *Journal of Business & Economic Statistics* **2**(4), 367–374.

Sickles, R. C. (2005), 'Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings', *Journal of Econometrics* **126**(2), 305–334.

Sickles, R. C., Song, W. and Zelenyuk, V. (2020), Econometric analysis of productivity: Theory and implementation in R, *in* H. D. Vinod and C. Rao, eds, 'Financial, Macro and Micro Econometrics Using R', Vol. 42 of *Handbook of Statistics*, Elsevier, pp. 267 – 297.

Sickles, R. and Zelenyuk, V. (2019), *Measurement of Productivity and Efficiency*, Cambridge: Cambridge University Press.

Simar, L., Lovell, C. K. and van den Eeckaut, P. (1994), 'Stochastic frontiers incorporating exogenous influences on efficiency', *Discussion papers* **9403**.

Simar, L., Van Keilegom, I. and Zelenyuk, V. (2017), 'Nonparametric least squares methods for stochastic frontier models', *Journal of Productivity Analysis* **47**(3), 189–204.

Simar, L. and Zelenyuk, V. (2011), 'Stochastic FDH/DEA estimators for frontier analysis', *Journal of Productivity Analysis* **36**(1), 1–20.

Simar, L. and Zelenyuk, V. (2020), 'Improving finite sample approximation by central limit theorems for estimates from data envelopment analysis', *European Journal of Operational Research* **284**(3), 1002–1015.

Stevenson, R. E. (1980), 'Likelihood functions for generalized stochastic frontier estimation', *Journal of Econometrics* **13**(1), 57–66.

Van den Broeck, J., Koop, G., Osiewalski, J. and Steel, M. F. (1994), 'Stochastic frontier models: A Bayesian perspective', *Journal of Econometrics* **61**(2), 273–303.

Wang, H. J. (2002), 'Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model', *Journal of Productivity Analysis* **18**(3), 241–253.

Wang, H. J. and Schmidt, P. (2002), 'One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels', *Journal of Productivity Analysis* **157**, 286–296.

Zelenyuk, V. (2020), 'Aggregation of inputs and outputs prior to data envelopment analysis under big data', *European Journal of Operational Research* **282**(1), 172–187.

Zuckerman, S., Hadley, J. and Lezzoni, L. (1994), 'Measuring hospital efficiency with frontier cost functions', *Journal of Health Economics* **13**(3), 255–280.