
AGRICULTURAL POLICY
WORKING PAPER SERIES

WP2020-09

Climate variability and farm inefficiency: a spatial
stochastic frontier analysis of Senegalese agriculture

K. Christophe Adjin
Christian H.C.A. Henning
Institute for Agricultural Economics
University of Kiel

The Agricultural Working Paper Series is published by the
Chair of Agricultural Policy at the University of Kiel.
The authors take the full responsibility for the content.

K. Christophe Adjin
Christian H.C.A. Henning

Climate variability and farm inefficiency: a spatial stochastic frontier analysis of
Senegalese agriculture

Institute for Agricultural Economics
University of Kiel

Kiel, 2020

WP2020-09

<http://www.agrarpol.uni-kiel.de/de/publikationen/working-papers-of-agricultural-policy>

About the authors:

K. Christophe Adjin is a doctoral researcher at the Chair of agricultural policy at the Institute for Agricultural Economics, University of Kiel (Germany).

Christian H.C.A. Henning is professor and Chair of agricultural policy at the Institute for Agricultural Economics, University of Kiel (Germany)

Corresponding author: cadjin@ae.uni-kiel.de

Abstract

This paper aimed to analyse Senegalese farmers' technical efficiency in the context of climate variability and spatial heterogeneity. To achieve this, firstly using simulated data, we evaluated the newly developed spatial stochastic frontier estimation technique based on skew-normal distributions. Secondly, using cross-sectional survey data we conducted an empirical analysis for 4423 Senegalese farm households. Simulation results show that the estimation approach used is appropriate and produces consistent results with large sample sizes, although it might suffer from a "starting values" problem. Empirical findings reveal that agricultural production in Senegal mostly depends on the allocated area and it is highly affected by climatic factors such as rainfall and temperature. Moreover, within a radius of 4 km the technical efficiency of farms appears to be significantly affected by unobserved spatial features. Furthermore, this farm technical efficiency can on average be increased by 20%, when accounting for spatial heterogeneity.

Keywords: Climate variability, Farm efficiency, Spatial heterogeneity, Senegal.

JEL classification: Q54, C21, D24.

1 Introduction

In Senegal, agriculture remains an important sector of the economy, and it accounts for approximately 32% of the country's total employment, more than 16% of the national GDP (World Bank, 2019) and 21%¹ of total exports (Republique du Senegal, 2018). However, the sector is challenged by many factors, and among those, climate variability. Senegalese agriculture is mostly rain-fed and less than 1% of agricultural land is under irrigation. The sector is therefore highly vulnerable to rainfall variability (D'Alessandro *et al.*, 2015). During the period 1941-2000, rainfall in the country has been characterized by a great annual irregularity and the shortening of the duration of the rainy season. Also, from year to year, the amount of precipitation substantially varies both temporally and spatially (Ba, 2006). According to Jalloh *et al.* (2013), the country will experience in the coming decades an increase in rainfall variability and droughts due to climate change, especially in the eastern part. The impacts of such erratic weather and climate shifts in Senegal will lead to productivity losses in the agricultural sector (Jalloh *et al.*, 2013). Moreover, such climate conditions also pose serious challenges for the cereals production sub-sector, which constitutes primary sources of food for rural populations. According to Jalloh *et al.* (2013) changing climate conditions combined with population growth, could lead to a 30% reduction in per capita cereals production by 2025.

The presence of erratic weather and climate conditions implies that Senegalese farmers would have to integrate not only conventional production factors but also climatic risks factors (such as precipitation, temperature, etc.) into their production systems. For instance, the study of Ba (2006) has shown that close droughts due to climate change has led to a transformation of not only some productions systems but also the decline or disappearance of some crops. Globally, climate change and its impact on agricultural productivity have gained increased attention during recent decades among research communities (e.g. Hughes *et al.* (2011); Lachaud *et al.* (2017)). As observed by Mulwa and Kabubo-Mariara (2017), despite the increasing number of climate change and variability studies and efficiency studies, there is a lack of literature linking climate change and variability to farm-level efficiency in Africa. Notable exceptions are the papers of Sherlund *et al.* (2002) in Cote d'Ivoire, Mulwa and Kabubo-Mariara (2017) in Kenya and Oyekale (2012) in Nigeria. This study, therefore aims to analyse agricultural technical efficiency in climate variability conditions in the Senegalese context, and it contributes to the literature in several

¹Calculation is based on values of exportation figures in Republique du Senegal (2018) and includes exports of fishes products, groundnut products, cotton and cotton fabrics.

ways.

First, few studies have analysed the farm efficiencies of Senegalese agriculture. Furthermore, previous studies have focused on groundnut farming (Thiam and Bravo-Ureta, 2003), rice sub-sector (Diagne *et al.*, 2013; Ngom *et al.*, 2016; Seck, 2017), vegetable sub-sector (Dedehouanou, 2014), or some selected crops (Okuyama *et al.*, 2017). Our analysis wants to take the agricultural sector as a whole. Moreover, although several studies have shown the significant effect that climatic factors can have on farmers technical efficiencies, none of these previous studies has really integrated them into their modelling. We, therefore, incorporated in our designated SFA models climatic factors namely rainfall, temperature, and their anomalies. Furthermore, to take into account spatiality and understand how its unobserved features affect generally farm production and technical efficiency, we took advantage of a new approach developed in spatial econometrics and apply it to the Senegalese context. Finally, we extended de Graaff (2020) simulation works in the spatial stochastic frontier field, primarily by conducting simulations for the spatial lag models, and secondly by generating and varying (via spatial correlation coefficients) the spatial weight matrices used in the simulation analysis.

Simulation results confirm that the estimation method designed by de Graaff (2020) based on the skew-normal distribution of errors in the stochastic production frontier is appropriate and can lead to consistent estimates. Although the Maximum Likelihood estimation procedure is sensitive to the starting values, notably the sign of inefficiency dispersion coefficient. Empirical results revealed that farmers technical efficiency in Senegal is affected by both climatic features and spatiality. These results have important policy implications when designing adaptation strategies to climate change.

The paper is divided into six sections, including this introduction. The second section presents a brief literature on spatial stochastic frontier models. The third section describes the approach used and the estimation technique. Section 4 presents the simulation results, while section 5 presents and discusses the empirical results. Finally, section 6 presents the conclusions and explores some policy implications.

2 Literature on spatial stochastic frontier

Recently, in the field of productivity and efficiency analysis, analysts have tried to incorporate spatial interactions into frontier models, resulting in some novel spatial

frontier models. Based on this literature and assuming some spatial interactions between decision-making units, one could distinguish four main types of models designed to introduce spatial interactions into the stochastic frontier approach. These models are (i) the spatial auto-regressive stochastic frontier (SAR), (ii) the spatial lag on the exogenous variable stochastic frontier (SLX), (iii) the spatial error stochastic frontier (SEM) and (iv) the spatial inefficiency stochastic frontier (SIM).

The SAR model integrates the spatial dependence into the dependent variable, rendering this latter auto-regressive. It has been widely used by modellers in many fields and contexts. Regarding the efficiency analysis field, although few studies are known, this specification is the one of most interest to researchers. Barrios and Lavado (2010) used the SAR-type stochastic frontier model accounting for spatial externalities. They applied it to data sets from the Philippines to demonstrate the importance of spatial components, and to derive unbiased estimations of technical efficiency. To evaluate the impact of programs that aim to improve the productivity of firms, this SAR-type model was also used by Affuso (2010) in Tanzania to analyse the efficiencies of matched subsamples of treated and non-treated farmers. Using this spatial autoregressive model, Glass *et al.* (2013) extended the decomposition of the standard factor productivity growth to incorporate direct and indirect impacts of units, introducing at the same time the concept of efficiency spillovers. The authors applied their specification to a panel data of 40 European countries from 1995 to 2008. Glass *et al.* (2014) applied the same approach in a cost frontier model to highlight the importance of efficiency spillovers in the U.S. manufacturing sector. Han *et al.* (2016) by allowing endogenous interaction effects in the frontier model, used data from 21 OECD countries from 1960 to 2001, to derive the spillover effects of public capital stock. Similarly, Ramajo and Hewings (2018) used a SAR model of stochastic frontier approach to estimate the technical efficiency of 120 European Union regions over the period 1995-2007. Their study revealed a strong geographic pattern of regional efficiency showing productivity convergence of European regions during the same period.

The spatial lag of exogenous variable, like the spatial lag model of X (SLX) defined by LeSage and Pace (2009), characterize a link between an output of a particular decision-making unit and the inputs of its neighbours. This model intends to explain the spatial influence of neighbours' input values (or the indirect flow of resources) on a given unit's output. In the context of efficiency analysis, Adetutu *et al.* (2015) applied such a model to account for local spatial dependence and to shift the production frontier. The spatial error stochastic frontier is similar to the spatial error model of LeSage and Pace (2009) and unlike the previous models, it is

designed to bring out the spatial heterogeneity between production units. Druska and Horrace (2004) used such specification, by including spatial auto-regressive disturbances into the classical SFA. Using generalized moments method on a panel data, they estimated time-invariant inefficiencies and concluded to the existence of spatial correlations into data and this latter affects the magnitude and variability of the production function and the estimated technical efficiencies.

The spatial inefficiency stochastic frontier model is an auto-regressive specification concerning the inefficiency error term of the composite error. This specification illustrates the spatial correlation between the levels of efficiency of neighbouring units. To account for the possible unknown geographical variation of the outputs of farms in Brazil, Schmidt *et al.* (2009) included in their model, a latent spatial structure in the inefficiency error term. Their findings showed that standards models induce significantly different inefficiencies across units. Tonini and Pede (2011) also incorporated spatial dependency in the inefficiency term to measure total factor productivity of European agriculture from 1993 – 2006 in 29 countries. They found that not allowing for spatial dependency underestimates the cumulated technical inefficiency changes. Areal *et al.* (2012) used such specification in a Bayesian setting to show spatial dependence in technical efficiency in a panel data of dairy farms in England and Wales. More recently Pede *et al.* (2018) and Skevas (2020) followed the approach of Areal *et al.* (2012) and applied it respectively to irrigated and rain-fed agroecosystem rice farming in the Philippines and to Dutch dairy farms. Fusco and Vidoli (2013) applied the same model to a cross-sectional data of wine industries in Italy and demonstrated the uniform and strong spatial dependence between neighbouring units. Tsionas and Michaelides (2016) also used this same stochastic frontier model with the decomposition of inefficiency into an idiosyncratic and a spatial spillover component. They applied their method to a production data of Italian regions over the period 1970 - 1993. Carvalho (2018) proposed a spatial Bayesian random effects stochastic frontier model that allows for unobserved heterogeneity and spillovers between firms' efficiencies. de Graaff (2020) estimated the SIM model after assuming a skew-normal distribution for the composite error term.

From these four main models, several combinations could be made to form different types of spatial stochastic frontiers. Pavlyuk (2013), for instance, developed a full model with all types of spatial interaction. However, its estimation is challenging due to identification and computation issues. After some restrictions on spatial parameters, Pavlyuk (2011, 2013) used data sets of the regional tourism markets in the Baltic States and the European airports to demonstrate significant spatial dependencies. Orea and Álvarez (2019) developed a stochastic frontier model that

allows for cross-sectional spatial correlation in both the noise and inefficiency terms, and that can be estimated by maximum likelihood and non-linear least squares. Glass *et al.* (2016) considered a spatial Durbin stochastic frontier model which combines the SAR and SLX specifications. They applied it respectively to European countries and demonstrated the asymmetry between efficiency spillovers to and from European countries.

3 Spatial stochastic production frontier approach

3.1 The model

From the seminal works of Aigner *et al.* (1977) and Meeusen and van Den Broeck (1977), the standard stochastic production frontier is defined as:

$$y_i = f(X_i, \beta) \exp(v_i - u_i), \quad (1)$$

where y_i is the observed output of the farm unit i ($i = 1, 2, \dots, N$), X_i is a vector of inputs, $f(X_i, \beta)$ is the production function, with β as the parameters to be estimated; v_i is a two-sided stochastic term that accounts for statistical noise, u_i is a non-negative stochastic term representing farm inefficiency. In this model, the possible production y_i , is bounded above by the stochastic quantity $f(X_i, \beta) \exp(v_i)$ that consists of a deterministic part $f(X_i, \beta)$ common to all farms and a farm-specific part $\exp(v_i)$ that captures the effect of random shocks. Errors v_i are assumed to be independently and identically distributed as a normal distribution $N^+(0, \sigma_v^2)$, independent of u_i , which are assumed to be non-negative with either a half-normal distribution (Aigner *et al.*, 1977), a truncated normal distribution (Stevenson, 1980), an exponential distribution (Meeusen and van Den Broeck, 1977) or a gamma distribution (Greene, 1990). Assuming a Cobb–Douglas production function for the output production y_i , a logarithmic specification of the function 1 gives in vector notation:

$$\ln(y) = \ln(X) + v - u. \quad (2)$$

Following previous literature, spatial interactions can be incorporated into this standard SFA in two main ways, through the dependent variable resulting in the so-called spatial auto-regressive stochastic frontier model (SAR-SFA), and through the errors giving the spatial error stochastic frontier model (SEM-SFA). The SAR-SFA model

can be written as:

$$\ln(y) = \rho W \ln(y) + \ln(X) \beta + v - u, \quad (3)$$

where ρ is the spatial lag parameter and W is the spatial weight matrix, and y , X , β , v , and u are as defined previously. Furthermore, the SEM-SFA can be specified as:

$$\ln(y) = \ln(X) \beta + \epsilon, \quad \text{with } \epsilon = v - u = \lambda W \epsilon + \tilde{\epsilon}, \quad (4)$$

where λ is the spatial error lag parameter, and W , y , X , β , v , and u are as defined previously.

3.2 Estimation of parameters

To estimate parameters of the standard SFA with the maximum likelihood technique, one usually assumes that v follows a normal distribution ($v \sim N(0, \sigma_v^2)$) and u a half-normal distribution ($u \sim N^+(0, \sigma_u^2)$). Therefore, the marginal density function of ϵ is given by:

$$f(\epsilon) = \frac{2}{\sigma} \phi\left(\frac{\epsilon}{\sigma}\right) \Phi\left(-\frac{\epsilon\gamma}{\sigma}\right), \quad (5)$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\gamma = \frac{\sigma_u}{\sigma_v}$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\phi(\cdot)$ is the standard normal probability density functions. From equation 5 the log likelihood function for N farms as proposed by (Aigner *et al.*, 1977) is:

$$\ln L = -\frac{N}{2} \ln\left(\frac{\pi\sigma^2}{2}\right) + \sum_i \ln \Phi\left(-\frac{\epsilon_i\gamma}{\sigma}\right) - \frac{1}{2\sigma^2} \sum_i \epsilon_i^2. \quad (6)$$

Once the error in the designed stochastic frontier exhibits a more complex structure, estimations of the parameters become cumbersome. Therefore, for the estimation of the spatial models, we adopted a skew-normal distribution approach proposed by de Graaff (2020) which enables us to straightforward estimate models 3 and 4 using the maximum likelihood technique. In the skew-normal approach, the composite error ϵ can be rewritten as a sum of a normal and a truncated normal distributions (de Graaff, 2020) which gives:

$$\epsilon = \delta |\mu| + \sqrt{1 - \delta^2} \nu, \quad (7)$$

where μ and ν are independent variables $N(0, 1)$ and $\delta \in (-1, 1)$. The stochastic variable ϵ is generated by means of convolution, however ϵ can also be obtained by conditioning:

$$\epsilon = (\nu | \mu > 0), \quad (8)$$

where (μ, ν) follows a bivariate normal distribution with δ as a correlation coefficient. It can be shown that both equations 7 and 8 lead to the same skew-normal density function:

$$\epsilon Z \sim SN(\alpha) = 2\phi(x)\Phi(\alpha x), \quad (9)$$

where α is the skewness parameter that determines the shape of the density function². As shown by de Graaff (2020), if $\epsilon \sim SN(\alpha)$ and $\ln(y) = \ln(X)\beta$, then the affine transformation of $\ln(y) \sim SN(\ln(X)\beta, \sigma^2, \alpha)$ can be expressed as:

$$\epsilon \sim 2\phi(\ln(y) - \ln(X)\beta; \sigma^2)\Phi(\alpha(\ln(y) - \ln(X)\beta)). \quad (10)$$

In this case, $\ln(X)\beta$, σ^2 , and α could be defined as a location parameter, a scale parameter and a skewness parameter respectively. The relation between equations 2 and 10 can be defined by stating $\ln(y) - \ln(X)\beta = \pi(v | u > 0) = \epsilon$, where

$$\epsilon = \begin{bmatrix} \mu \\ \nu \end{bmatrix} \sim N(0, \Omega^*), \quad \Omega^* = \begin{bmatrix} 1 & \delta' \\ \delta' & \sigma^2 \end{bmatrix}, \quad (11)$$

and where $\alpha = \delta^2\sqrt{1 - \delta^2}$, $\delta = \sigma_u$, and $\sqrt{1 - \delta^2}\sigma_\epsilon = \sigma_v$. The latter equality denotes the intrinsic relation between u and v which is implicit in specification 2. It is important to note that specification 2 only holds when $\delta < 0$. From density equation 10, the log likelihood for N observations can be specified as (de Graaff, 2020):

$$\ln L = -\frac{N}{2}\ln\pi - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\epsilon'\epsilon + \sum_i \ln(2\Phi(\alpha\epsilon_i)), \quad (12)$$

where ϵ_i is the i th observation of the vector ϵ . As argued by de Graaff (2020), a skew-normal distribution allows the use of a single error term instead of a composite one. This process has several advantages when working with multivariate distributions, and the interpretation of the parameters seems as well more intuitive (using scale, location, and skewness parameters). However, a disadvantage is the need to use a re-parametrization of the parameters in order to estimate them properly. The log likelihood function specified in equation 12 can be straightforwardly adapted to the spatial lag stochastic frontier and spatial error stochastic frontier model (de Graaff, 2020). The log likelihood for the spatial lag stochastic frontier for

²See de Graaff (2020); Azzalini (1985, 2005); Azzalini and Valle (1996); Azzalini and Capitanio (1999); Arellano-Valle and Azzalini (2006) and Arellano-Valle and Azzalini (2008) for more literature on the skew-normal distributions.

N units is expressed as:

$$\ln L = -\frac{N}{2} \ln \pi - \frac{N}{2} \ln (\sigma^2) + \ln |I - \rho W| - \frac{1}{2} \epsilon' \epsilon + \sum_i \ln (2\Phi(\alpha \epsilon_i)), \quad (13)$$

where $\epsilon = \frac{1}{\sigma} [\ln(y) - \ln(X)\beta]$ and I the identity matrix. The log likelihood for a stochastic frontier model with spatial dependence in the error term can be defined as:

$$\ln L = -\frac{N}{2} \ln \pi - \frac{N}{2} \ln (\sigma^2) + \ln |I - \lambda W| - \frac{1}{2} \epsilon' \epsilon + \sum_i \ln (2\Phi(\alpha \epsilon_i)), \quad (14)$$

where $\epsilon = \frac{1}{\sigma} [I - \lambda W] [\ln(y) - \ln(X)\beta]$, and I , N , and w are defined as previously.

3.3 Measurement of technical efficiency

After estimating the likelihood functions in equations 13 and 14, the obtained parameters $\hat{\epsilon}$, $\hat{\delta}$ using $\hat{\alpha}$ and $\hat{\sigma}$ can therefore be used to draw simulations from $u | \epsilon$ and derive the expectation for each farm. Following de Graaff (2020), we will use the generic formula for $u | \epsilon$ defined by Dominguez-Molina *et al.* (2003) as a normal distribution with mean and variance equal to:

$$\text{Mean} = \frac{1}{\frac{\delta^2}{\sigma^2} + 1} \frac{\delta}{\sigma^2} \epsilon, \quad \text{Variance} = \frac{1}{\frac{\delta^2}{\sigma^2} + 1}.$$

4 Monte Carlo simulation

4.1 Simulation procedure

Following a similar approach as de Graaff (2020), we set up a simulation procedure with a base cross-section data generating process defined as:

$$\begin{aligned} Y &= 1 + \ln(A) + \epsilon, \\ \text{where } \epsilon &= \delta | u | + \sqrt{1 - \delta^2} v \\ \text{and } \ln(A) &\sim U(5, 14) \\ u &\sim N(0, 0.3) \\ v &\sim N(0, 0.3). \end{aligned} \quad (15)$$

Here, v and u are drawn using a normal distribution, the production input is drawn using a uniform distribution with associated coefficient β_1 equals to 1, and the

constant term (denoted β_0) is also set to 1. From this base specification we randomly generated data sets by considering alternative scenarios (see table 1). The experiments were then performed by varying the number of observations N (250 and 1000), and the values of the coefficient of correlation δ (-0.2, -0.5, and -0.8). However, contrary to de Graaff (2020) we also vary the values of the spatial dependence coefficients λ and ρ (0.2, 0.5, and 0.8), by considering a randomly generated spatial weight matrix of the type, inverse geographic distance, using latitudes and longitudes drawn from a uniform distribution ($U(0, 20)$). Note that the weights matrices are all row-standardized, and their diagonals set to 0. For each scenario, 1,000 replications are made.

Table 1: Monte Carlo simulation scenarios

Variables	Description	Base scenario	Alternative scenarios
N	Number of observations	250	1000
λ, ρ	Spatial correlation	0.2	0.5, 0.8
δ	Inefficiency coefficient	-0.2	-0.5, -0.8

For each generated data, we estimate the values of parameters β_0 (constant term), β_1 , σ , α , δ , the likelihood value, the mean technical efficiency. We also compute the true values of α using the estimated δ , the true likelihood value, and the true mean technical efficiency. This process is replicated 1,000 times for all the scenarios defined in Table 1. For each scenario, we compute the mean value and its standard deviation. We also calculate the bias and root mean squared error (RMSE) for all estimated values using the following formula:

$$\text{Bias}_k = \frac{1}{1000} \sum_{r=1}^{1000} (E_{kr} - T_{kr}), \quad \text{RMSE}_k = \sqrt{\frac{1}{1000} \sum_{r=1}^{1000} (E_{kr} - T_{kr})^2},$$

where Bias_k is the bias in scenario k , RMSE_k is the root mean squared error in scenario k , E_{kr} is the estimated value in replication r of scenario k , and T_{kr} is the true value in replication r of scenario k . Moreover, for a better analysis, we also plot for each scenario box-plots for estimated parameters, which better exhibit the distribution of the parameters (medians, ranges, biases with the true data generating process values).

4.2 Simulation results

Tables 6, 7, and 8 present the simulation results for different setups of λ (0.2, 0.5, 0.8) respectively (case of spatial error SFA). Furthermore, figures 5, 6, and 7 in appendix show the distributions of parameters β_0 , β_1 , δ , σ and λ . These results exhibits some patterns. Estimated coefficients β_0 and β_1 shows relatively very low biases and RSME for all scenarios. For instance, the biases for β_1 are close to 0. Also, β_0 and β_1 distributions show that they converge to their true values when the sample size is high. For parameter λ , with low true values, biases and standard deviations are high. Parameter σ exhibits low biases whether $N=250$ or $N=1000$, and when values of δ are -0.2 or -0.5. Biases seems to increase with the increase (absolute term) of values of δ . In some cases, estimated values of σ are totally out of their range as shown by figure 5 for $\delta=0.8$. Regarding parameter δ , biases diminishes with the increase in sample size in all scenarios. However, box-plots show relatively skewed distributions for δ with wide range in most scenarios (when $\delta=-0.2, -0.5$), denoting the difficult precision in the estimation of this parameter when the sample size is small and the absolute value of true δ is low. As α and δ are related, the same pattern is observed for α . In conclusion, as also observed by de Graaff (2020), values of the estimated parameters converge to their true values when the sample size becomes larger (as shown by the box-plots, a wider range in the distribution of estimated values is globally observed when $N=250$, compared to $N=1000$).

Tables 9, 10, and 11 reports respectively the simulation results for the different setups of ρ (0.2, 0.5, 0.8) (case of spatial auto-regressive SFA), and figures 8, 9, and 10 in the appendix plot in this specific case the distributions of parameters β_0 , β_1 , δ , σ and ρ respectively. One can observe that parameter β_0 shows high biases and RMSE in most scenarios, especially when true $\rho = 0.8$. But the estimated coefficient β_1 shows very low biases and RMSE in all scenarios. In the case of estimated parameter α , the RMSE are high for all scenarios. In addition, the biases are high when $\delta = -0.2$. Similar pattern can be observed for the estimated values of δ .

These simulation works show that the suggested estimation technique can provide consistent results, mostly in the case of the spatial error stochastic frontier model. However, it is worth noting that for the efficiency parameter, it seems important to use a starting value with a negative sign which is the expected sign of the coefficient. Indeed, during simulations, we observed that when the starting value of the inefficiency parameter δ has a positive value, the estimated coefficient gives values close to zero. We conducted several simulation rounds and observed similar

results, however, we did not report the results here. This is the only issue that we observed in the simulation, and even when using the positive starting values, this did not affect the estimation of other coefficients. We believe this problem might be related to the re-parametrization in the equations: centered versus non-centered parameterizations, which has been widely discussed in literature³. The starting values used for the simulation work and empirical analysis are defined following principles of stochastic frontier analysis and they are presented in table 2.

Table 2: Starting values for maximum likelihood estimation

Parameters	Estimated models		
	SFA	SAR-SFA	SEM-SFA
β_0	β_0^{OLS}	β_0^{SAR}	β_0^{SEM}
β_1	β_1^{OLS}	β_1^{SAR}	β_1^{SEM}
σ	σ^{OLS}	σ^{SAR}	σ^{SEM}
λ		λ^{SEM}	λ^{SEM}
ρ		ρ^{SEM}	ρ^{SEM}
δ	$-\sigma^{OLS}\sqrt{1-\pi/2}$	$-\sigma^{SAR}\sqrt{1-\pi/2}$	$-\sigma^{SEM}\sqrt{1-\pi/2}$

5 Empirical application

5.1 Model specification

This section describes the specifications of the production frontiers and of the weight matrices used. We specify a base production function (model 1), using a Cobb-Douglas functional form between the output and the inputs used⁴. Therefore the estimated parameters of the conventional inputs can be interpreted as partial elasticities of production. From this base model, we specified three variant models: in the first variant (model 2), we incorporated climate variables directly into the non-stochastic component of the production function. Meanwhile in the second variant (model 3), in addition to the climate variables, we included some other environmental variables. Our general model is therefore specified as follows:

$$\ln(y_i) = \theta_0 + \sum_{k=1}^k \theta_k \ln(x_{ik}) + \sum_{l=1}^l \gamma_l D_{il} + \sum_{m=1}^m \Delta_m Z_{im} + \varepsilon_i \quad (16)$$

³The re-parametrization problem has been mentioned to us via e-mails by de Graaff (2020) himself, however, we did not find any solution to address it in the present work.

⁴The Cobb-Douglas functional has convenient properties. As argued by O'Donnell (2016), it generally satisfies non-negativity and monotony, unlike Translog, a commonly used alternative.

where y_i represents the output of the i^{th} farmer, x_{ik} denote vectors of the production inputs k ; D_{il} represent climatic variable l ; Z_{im} represent other non-stochastic environmental variables m ; θ_0 , θ_k , γ_l , and Δ_m are parameters to be estimated and ε_i is the error term. The output here is the household total crop production. The three inputs are land, labour and operating costs. The climatic variables are rainfall, temperature, rainfall and temperature anomalies. The other non-stochastic environmental variables include the components of soil, the farmer education, the use of improved seeds, and membership in farmers organizations. In this empirical analysis, we estimated this model using a step-wise approach. First, we assumed no spatial interaction and estimated the model as a linear model (OLS). Secondly, after testing for the presence of spatial dependence in the production function (we discussed this later), we incorporated spatial interaction via the spatial weight matrix W and estimated linear spatial models (Spatial lag and Spatial error). The spatial lag is of the form $\ln(y_i) = \rho W \ln(y_i) \theta_0 + \sum_{k=1}^k \theta_k \ln(x_{ik}) + \sum_{l=1}^l \gamma_l D_{il} + \sum_{m=1}^m \Delta_m Z_{im} + \varepsilon_i$ and the spatial error is of the form: $\ln(y_i) = \theta_0 + \sum_{k=1}^k \theta_k \ln(x_{ik}) + \sum_{l=1}^l \gamma_l D_{il} + \sum_{m=1}^m \Delta_m Z_{im} + \varepsilon_i$, with $\varepsilon_i = \lambda W \varepsilon_i + \tilde{\varepsilon}_i$, with ρ and λ as parameters to be estimated. Finally, assuming that ε follows a skew-normal distribution, we estimated a standard SFA and then spatial frontier models (spatial SFA lag and Spatial SFA error), to derive farmers technical efficiency scores robust to the presence of spatial dependence. The performances of the estimated models are compared to select the "best" fitting model based on various tests.

Prior to estimating the spatial models, we specify the weight matrix for the Senegalese agricultural sector. The spatial weight matrix W is a symmetric matrix, where its elements w_{ij} express proximity of a household i with a household j . In common practice, to enable an interpretation of model coefficients, W is row standardized so that the sum of the row elements equals to one. In addition, the diagonal elements w_{ii} are set to zero, in order to prevent the effect of the i^{th} household from directly predicting itself. Many specifications of weight matrices have been used in the literature, and specifying the weight matrix is arbitrary. However, prior knowledge of the study population and economic theory can help to guide in the specification of these matrices. Following Areal *et al.* (2012) and Pede *et al.* (2018), we specify a spatial weight matrix of a power form with: $w_{ij} = \exp(-d_{ij}^2/s^2)$, where d_{ij} is the euclidean distance between households i and j , s is the cut-off distance. We choose a cut-off distance of 4 km based on a previous study in this thesis. It is assumed therefore that beyond this cut-off distance there is no spatial effect.

Moreover, before implementing the spatial models, we checked for spatial interaction by computing the Moran index using the dependent variable and the residuals

of the OLS of each model⁵. Using the same variables, we plotted Moran scatter plots to visually show the spatial interaction. Furthermore, we conducted a series of tests including the standard Lagrange multiplier (LM)(Anselin, 1988) test and its robust counterparts (Anselin *et al.*, 1996)⁶.

5.2 Data source and variables

The data used for the empirical analysis was derived from a survey conducted in Senegal, which randomly sampled 4480 households that mainly produce rainfed cereals. The survey was implemented under the Agricultural Policy Support Project (Projet d’Appui aux Politiques Agricoles, PAPA)⁷, which is an initiative of the Government of Senegal funded by USAID-Senegal as part of the ”Feed The Future” initiative, and implemented for a period of 3 years (2015 - 2018) by the Senegalese Ministry of Agriculture and Rural Facilities with technical support from the International Food Policy Research Institute (IFPRI). The data, which covers the main agricultural season of 2016/2017, contains information on crop production and different inputs used. After the data cleaning and removing observations with no information on crop production, the final sample comprises of 4245 households located in all six Senegalese agro-ecological zones. Climate data were retrieved, using the surveyed households location coordinates, from publicly available databases of the Climate Hazards Center of the University of California (<https://www.chc.ucsb.edu/data>). The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)⁸ for the period 1981-2016 was used for rainfall variables, and the monthly Climate Hazards Group InfraRed Temperature with Station (CHIRTSmax)⁹ for the period

⁵Moran’s I statistic for outcome is computed as: $I = (\sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})) / (\sum_i (Y_i - \bar{Y})^2)$, where w_{ij} is the spatial weight between households i and j ; Y_i is the outcome of household i ; and \bar{Y} is the mean of the outcome. The range of Moran’s I is $(-1, 1)$, with 1 indicating perfect spatial similarity (or positive spatial correlation), 0 indicating no spatial correlation, and -1 indicating perfect dispersion (or negative correlation). If we observe a significant spatial autocorrelation based on Moran’s I statistic, spatial regressions models should be used to correct for the spatial autocorrelation errors.

⁶The standard two LM tests are: $LM_{error} = [e'W_e / (e'e/N)]^2 / [tr(W^2 + W'W)]$ and $LM_{lag} = [e'Wy / (e'e/N)]^2 / D$. Robust LM tests are defined as: $RLM_{error} = [e'W_e / (e'e/N)]^2 / [tr(W^2 + W'W)]$ and $RLM_{lag} = [e'Wy / (e'e/N)]^2 / D$, where e denotes the estimated residual from the non-spatial model; N is the number of farmers; and W are defined as previously.

⁷Official website of the project is <http://www.papa.gouv.sn/>.

⁸CHIRPS data is a 35+ years quasi-global rainfall data set. Spanning 50°S-50°N (and all longitudes) and ranging from 1981 to near-present, CHIRPS incorporates the in-house climatology data CHPclim, 0.05° resolution satellite imagery, and in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring (Funk *et al.*, 2015).

⁹CHIRTSmax is a global 2-m maximum temperature (Tmax) product that directly combines satellite and station-based estimates of Tmax to produce routinely updated data to support the monitoring of temperature extremes. The CHIRTSmax development process integrated a long-

1983-2016 was used for temperature variables. Table 3 presents the definition and summary statistics of the variables used in the analysis.

Table 3: Description of variables

Variables	Description and measurement	Mean	SD	Max	Min
Production	Total crop production (1000 FCFA)	467.684	764.653	14375	0.180
Land	Total area of land cultivated (Ha)	4.411	4.527	50.000	0.001
Family labor	Total family labor (Adult equivalent)	3.228	2.258	20.100	0.150
Operating costs	Total Operating costs (1000 FCFA)	55.970	118.610	3418.500	1.500
Rainfall	Annual rainfall (m)	0.672	0.292	1.484	0.187
Temperature	Average maximum temperature (Celsius degree)	35.942	1.314	38.43	30.90
Rainfall anomaly	Rainfall anomaly 1981-2015	0.033	0.050	0.170	-0.186
Temperature anomaly	Average maximum temperature Anomaly 1983-2015	0.020	0.006	0.038	0.008
Clay	Percentage of clay (%)	19.97	7.235	40.00	3.00
Silt	Percentage of silt (%)	13.47	5.735	31.00	2.00
Improved seed	Use of improved seed (1=yes, 0=no)	0.259	0.4380	1	0
Education	Formal education (1=yes, 0=no)	0.370	0.483	1	0
FBO Membership	Membership in Farmer Based Organization (1=yes, 0=no)	0.088	0.283	1	0
N	Number of Observations	4423	4423	4423	4423

SD: Standard Deviation

The dependent variable used in the models is the total crop production, which is expressed in Franc CFA¹⁰, and it includes all farm crop production outputs valued at the market prices. Farmers produce 32 crops which include major crops, namely cereals (rice, maize, sorghum, millet), groundnut, and cotton. The first input is the land, which is the sum of all land area dedicated to crop production during the 2016/2017 growing season. Variable labour, the second input, is the quantity of total labour (in adult equivalent). The total of operating costs in FCFA, which is the last input, includes the costs of seeds, fertilizers, non-family labour, and other costs such as transport, maintenance, etc. Climate variables include rainfall, rainfall anomaly, temperature and temperature anomaly. Based on study of Ba (2006), the climate in Senegal is made up of two seasons: the rainy season which lasts two to four months (exceptionally five months) depending on the region and the very long dry season. Also, these two seasons are more complex to define because they are difficult to delineate in time and space. The spatial and temporal features of the rain are very variable from one region to another (Ba, 2006). The spatial features of the

term climatology with satellite information and available station data. The result is a monthly estimate of the daily maximum temperature for the 1983-2016 time period (Funk *et al.*, 2019).

¹⁰Local currency, 1 FCFA = 0,0017 USD as at 8 May 2020.

Senegalese climate justify once again the spatial approach used in this paper. We, therefore, consider rainfall as the annual precipitation received by each household. Temperature is the yearly average maximum temperature of the household location expressed in degree Celsius. We also computed the anomalies of temperature and rainfall. We follow Lachaud *et al.* (2017) and computed anomalies as the deviation of the 2016 annual rainfall and temperature observation from the long-term mean (1981–2015 for rainfall and 1983-2015 for temperature)¹¹. As argued by Lachaud *et al.* (2017) and Barrios *et al.* (2010), there are some advantages associated with the use of anomalies. First, factors like the station location and elevation are less critical, and agricultural decisions are based on expected weather behaviour, therefore production might be affected by any weather deviation from the expectations. In the sample, the mean monthly maximum temperature is about 36 degrees Celsius, while the mean annual rainfall received by a household in the year 2016 is 672 mm. The long-term average annual rainfall over the period 1981-2015 is 651 mm and the long-term average maximum temperature over the period 1983-2015 is 35 degree Celsius. The other variables included in the models that can produce cross-sectional variation in the production frontier between farmers are their education, use of improved seeds, percentages of clay and silt in soils, and membership in farmers organizations.

5.3 Empirical results

Diagnostics for spatial interdependence

Table 4 reports the results of the diagnostic tests for spatial interaction. It shows that the Moran’s I statistics are all positive and highly significant ($p < 0.01$). For instance, the Moran I using the residuals are 0.377, 0.328, and 0.310 for models 1, 2 and 3 respectively, indicating that there is a strong positive spatial correlation between farmers crop production, as well as a strong spatial dependence. Farmers with relatively high production seem to live close to other farmers with high production, and farmers with relatively low production tend to live near farmers with low production. These results are corroborated by the Moran plots (Figures 3 and 4 in the appendix), which show how observations outcomes are strongly and positively correlated to their neighbours’ outcomes (one can observe the clustering of values in the upper right quadrant and lower left quadrant, suggesting the positive spatial autocorrelation). These results suggest that spatial correlation should be considered

¹¹Anomaly = (current year value - long term mean) / long term mean

in our analysis. After estimating the OLS, we computed the standard Lagrange Multiplier tests (Anselin, 1988) and their robust counterparts (Anselin *et al.*, 1996). The null hypotheses were mostly rejected at $p < 0.01$, indicating that spatial interaction should be incorporated into our models. At this step, tests results are non-conclusive. However, when comparing tests statistics, the results are mostly in favour of the spatial error models. The test statistics of the error models are much higher than that of the lag models, we therefore continue the analysis with the spatial error models.

Table 4: Moran Index and Lagrange Multiplier tests

	Model 1	Model 2	Model 3
Moran I Statistic (outcome)		0.418***	
Moran I Statistic (residuals)	0.377***	0.328***	0.299***
Standard LM Error	2049.7***	1550.6***	1285.8***
Robust LM Error	1374.4***	1036.3***	735.57***
Standard LM Lag	716.72***	536.69***	550.7***
Robust LM Lag	41.426***	22.414***	0.494

Notes: *** $p < 0.01$. LM: Lagrange Multiplier

Production frontier

Tables 12, 13, and 14 present the results of estimated models 1, 2, and 3 respectively, with different estimation techniques. At this stage of the analysis, the preferred models are the spatial error models for the base model (SEM 1) as well as for the variants (SEM 2 and SEM 3). We still need to decide between these linear spatial error models (SEM 1, SEM 2, and SEM 3) and the spatial SFA error models (SEM-SFA 1, SEM-SFA 2, and SEM-SFA 3) which are spatial frontier analysis. Based on the likelihood ratio tests (LR), the spatial SFA error specifications are preferred. The results of the LR tests comparing the various models are presented in table 5. Furthermore, based on the same likelihood ratio test, the preferred spatial SFA error model is the one which includes climatic and other variables (SEM-SFA 3). This model suggests that including climatic variables and other variables bring additional information to the model. We, therefore, continue the discussion with this model. The maximum likelihood estimates of the SEM-SFA 3, which includes climatic and other variables is reported in column (6) of table 14. Results

show as expected¹² that the model produces statistical significant (at 1% level) and positive partial production elasticities. These results suggest that in Senegal, land and labour are the inputs that make the highest contribution to agricultural production. Land contribution to agricultural production is about 86% while labour, which is mainly made of the family labour, contributes about 9%. Similar results have also been reported in other parts of the developing world, for instance in Latin America and Caribbean by Lachaud *et al.* (2017). Meanwhile, the total operating costs which includes the values of fertilizers, seeds, hired labour and other cost such as transport and maintenance only plays a minor role in the farm crop production, with a contribution of about 2%. When comparing these results with the estimates of other models (OLS 1, OLS 2, OLS 3, SFA 1, SFA 2, SFA 3, SEM-SFA 1, etc.), one can observe similar patterns: statistical significant and positive partial production elasticities, with a slight difference in magnitudes. Neither the standard frontier analysis nor the incorporation of spatial interaction changed the results of the frontier estimates, although the likelihood value improves significantly in the spatial models.

Table 5: Likelihood ratios tests

	SEM 2	SEM 3	SEM-SFA 1	SEM-SFA 2	SEM-SFA 3
SEM 1	88.393***	152.177***	3074.76***		
SEM 2		63.783***		3068.269***	
SEM 3					3072.986***
SEM-SFA 1				81.90***	150.40***
SEM-SFA 2					68.50***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Climatic and spatial effects

Climatic variables exhibit statistically significant coefficients with the exception of average maximum temperature. The rainfall parameter is positive, and as expected higher precipitation is beneficial for farm crop production. The variables representing rainfall and temperature anomalies are both negative (-3.193 and -16.967) and significant at 1% level. These results would mean that the deviation of 2016 annual rainfall and temperature from the long-term trend has negatively affected farm crop production.

The spatial error parameter λ (with value of 0.525) of the preferred model is

¹²Satisfaction of regularity conditions from production economic theory, i.e., partial output elasticities should be non-negative and less than 1(Lachaud *et al.*, 2017).

positive and statistically different from zero. In comparison, across all alternative models, this parameter ranges from 0.523 and 0.588, and its values are statistically significant. At the same time, the spatial lag parameter ρ , though smaller than λ , also exhibits positive and statistically significant values (range from 0.310 to 0.346). The results suggest the presence of a high spatiality arising from unobservable spatial factors in the agricultural production sector of Senegal. This implies that computing technical efficiency scores using non-spatial production frontiers would have led to biased estimates.

The standard deviation of the error term (σ) for the preferred model is 0.843 and significant at 1% level. This value is very close in terms of magnitude to the estimated ones in all alternative models, which are also significant. The highest σ can be observed for the standard SFA in specification 1 (SFA 1) and the lowest for the spatial lag SFA in specification 2 (SAR-SFA 2). The dispersion parameter for the inefficiency (α) of the preferred model is negative and statistically significant. This result suggests that most of the farmers are producing below the production frontier. However, it is easy to notice that this parameter is not always statistically significant in the alternative specifications, and its magnitude also varies significantly. For instance, in the base model, the value of this parameter α is statistically significant for the standard SFA (SFA 1) but in absolute terms, lower than that of the spatial error SFA (SEM-SFA 1). In specifications 2 and 3, the same pattern is observed, however, the magnitude of the α parameter is much lower than that of the spatial error SFA (SEM-SFA 1, 2, and 3). One explanation is that, first the difference in absolute terms in the inefficiency parameter could be due to the effect of climatic variables which may have acted as a "corrective" factor, hence suggesting that not incorporating the climatic variables would simply lead to missing variable bias. Secondly, these differences are due to the existence of spatial features in farm production. Once again, not having included them in our models would have led to biased estimates. When comparing the "instability" of the α parameter to the "stability" of the σ parameter, one can conclude that climate variability and spatial features are very important factors determining the inefficiency of farm households in Senegal. The results also reveal that components of soils (clay, silt), use of improved seed, formal education of household head and membership in farmer organizations are important factors in improving farm production.

Technical Efficiency

Figure 1 shows the kernel distributions of technical efficiency estimates for all model specifications (both non-spatial and spatial models). The average value of technical efficiency scores for the preferred model is 0.792. This result suggests that there is still a room to improve farm efficiency by at least 20%. The mean efficiency estimated for all specifications varies between 0.752 and 0.792. The magnitudes of the means of efficiency scores seem consistent through estimations methods and models. However, their distributions do not follow the same patterns. The distribution of the efficiency scores of the spatial SFA error in all specifications have a more flattened shape and is very distinct from the other models. This can be a consequence of the high absolute value of α observed previously. Moreover, we calculated the percentage difference between the average technical efficiency for non-spatial SFA models and for the different models that account for spatial dependence. Results show that accounting for spatial dependence in the analysis leads to slight increases in the estimates of technical efficiency in all cases. For instance, in specifications 3, the increase of technical efficiency is only about 4% in comparison to estimate of a non-spatial SFA. This result would indicate that the effects of incorporating spatial dependence in the analysis is more observable through the distribution of farm technical efficiency scores (shown by the flattened curve), and suggesting that some households might have a certain level of technical efficiency mostly because of their geographic location. Therefore, we mapped the individual change in technical efficiency scores to explore any specificity of the Senegalese regions, and found that there are no peculiarities between localities in terms of efficiency change. Farmers who have high changes in efficiency could be found in any part of the country. Figure 2 mapped first the change in technical efficiency due to spatial heterogeneity (change between models SEM-SFA 3 and SFA 3) and secondly the change in efficiency due to spatial heterogeneity and climatic variability (change between models SEM-SFA 3 and SFA 1). These results might suggest that climate variability is persistent and affects Senegalese farmers efficiency irrespective of where they are located, and they should use adequate farming techniques and technologies in order to have an appropriate level of productivity. Finally, we conducted the same analysis using this time the translog functional form and the results are presented in tables 15, 16, and 17 and figure 11 in the appendix. Results show that the obtained estimates of parameters α , σ , and ρ and the kernel distributions of technical efficiencies are similar to those of the Cobb-Douglas functional form.

6 Conclusion

The main objective of this paper was to combine spatial econometrics techniques and a stochastic frontier approach, to analyse the effects of spatial dependence and climate variability on farmers' technical inefficiency. The paper contributes to the growing literature of efficiency analysis and to the existing knowledge of the Senegalese agricultural sector by incorporating spatial features in the production functions. For this purpose, we use a recently designed estimation technique for cross-sectional data in spatial econometrics, to conduct simulations and empirical analyses.

Simulation results show that the maximum likelihood technique estimation suggested by de Graaff (2020) and based on the skew-normal distribution of errors in the stochastic production frontier provides consistent results. However, it is worth noting that for the efficiency parameter, it seems important to use a starting value with a negative sign which is the expected sign of the coefficient. Empirical results reveal that Senegalese agriculture is more dependent on land area and dedicated labour. Moreover, results show that farms efficiency is highly affected by both climatic features and spatial heterogeneity, and not accounting for them might lead to biased results for the efficiency distribution. Particularly, we found that rainfall and temperature anomalies negatively affect the production frontier, and farm technical efficiency on average increases when estimations control for spatial heterogeneity and climate variability. Also, we did not find conclusive evidence that could have led to the choice of spatial dependence modelling in the form of a spatial autoregressive-type. Furthermore, findings also reveal that the changes in technical efficiency score could be observed in any part of the country, implying that the effects of climate variability and unobserved spatial features are not specific to any particular region, but common in the entire country.

In terms of policy implications, these results imply that farmers need to adapt to climatic effects by using appropriate and very localized technologies or farming practices that take into account the specific characteristics of their locations. Policy-makers should encourage the design and dissemination of agricultural technologies that are very adaptable to specific conditions of farmers. We also observed that membership in farmer organizations as a measure of social capital improves the production frontier. Farmers organizations and other social groups could complement the efforts of extension services, and be a good entry point of introducing climate adaptation and other farming techniques.

From a research perspective, future studies should pursue the simulations work and investigate more the empirical performances of the skew-normal approach. Such studies should explore the problem of starting values that we have encountered, and the "centered versus non-centered parameterizations" issues. These studies are also encouraged to apply the approach to a broad range of sectors (besides agriculture) in developing countries where most studies are needed and where spatiality is mostly neglected in the analysis.

Table 6: Simulation results for $\lambda = 0.2$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	1.029	0.137	0.029	0.140	1.000	0.946	0.132	-0.054	0.142	1.000	0.917	0.120	-0.083	0.146
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.319	0.037	0.013	0.039	0.346	0.301	0.039	-0.045	0.059	0.500	0.287	0.040	-0.213	0.217
	α	-0.204	-0.444	0.514	-0.240	0.567	-0.577	-0.523	0.558	0.054	0.560	-1.333	-1.158	0.739	0.175	0.759
	δ	-0.200	-0.312	0.326	-0.112	0.344	-0.500	-0.355	0.335	0.145	0.365	-0.800	-0.635	0.338	0.165	0.376
	λ	0.200	-0.251	1.248	-0.451	1.327	0.200	-0.259	1.365	-0.459	1.440	0.200	-0.241	1.336	-0.441	1.407
	TE	0.871	0.838	0.061	-0.033	0.071	0.869	0.854	0.059	-0.015	0.065	0.782	0.896	0.074	0.114	0.138
	LL	36.349	38.807	11.614	2.457	2.952	49.740	57.270	11.613	7.530	8.244	44.363	102.191	11.595	57.828	58.386
1000	β_0	1.000	1.026	0.104	0.026	0.107	1.000	0.959	0.109	-0.041	0.116	1.000	0.974	0.107	-0.026	0.111
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.313	0.024	0.007	0.025	0.346	0.298	0.025	-0.048	0.054	0.500	0.296	0.019	-0.204	0.204
	α	-0.204	-0.356	0.360	-0.152	0.390	-0.577	-0.487	0.391	0.090	0.401	-1.333	-1.291	0.336	0.042	0.339
	δ	-0.200	-0.287	0.265	-0.087	0.279	-0.500	-0.380	0.275	0.120	0.300	-0.800	-0.768	0.154	0.032	0.158
	λ	0.200	-0.029	0.903	-0.229	0.931	0.200	0.004	0.799	-0.196	0.822	0.200	0.041	0.752	-0.159	0.768
	TE	0.871	0.849	0.049	-0.022	0.055	0.869	0.871	0.045	0.002	0.047	0.782	0.914	0.048	0.132	0.140
	LL	145.069	147.823	22.293	2.754	3.287	199.056	221.970	22.462	22.914	23.669	177.702	401.685	23.275	223.983	224.536

Notes: TE and LL represent respectively technical efficiency and log likelihood value

Table 7: Simulation results for $\lambda = 0.5$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	1.022	0.179	0.022	0.180	1.000	0.874	0.169	-0.126	0.211	1.000	0.784	0.164	-0.216	0.271
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.319	0.037	0.012	0.039	0.346	0.299	0.039	-0.047	0.061	0.500	0.279	0.044	-0.221	0.225
	α	-0.204	-0.429	0.515	-0.224	0.561	-0.577	-0.475	0.556	0.103	0.565	-1.333	-0.998	0.813	0.336	0.879
	δ	-0.200	-0.300	0.325	-0.100	0.340	-0.500	-0.320	0.337	0.180	0.382	-0.800	-0.541	0.386	0.259	0.464
	λ	0.500	0.399	0.841	-0.101	0.847	0.500	0.400	0.871	-0.100	0.876	0.500	0.377	1.099	-0.123	1.105
	TE	0.871	0.834	0.061	-0.037	0.073	0.869	0.842	0.061	-0.027	0.070	0.782	0.874	0.074	0.092	0.120
	LL	35.913	38.455	11.619	2.542	3.049	49.303	56.897	11.607	7.594	8.308	43.926	101.683	11.609	57.756	58.319
1000	β_0	1.000	1.023	0.146	0.023	0.148	1.000	0.916	0.158	-0.084	0.179	1.000	0.909	0.175	-0.091	0.197
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.313	0.024	0.007	0.025	0.346	0.299	0.025	-0.048	0.054	0.500	0.295	0.021	-0.205	0.206
	α	-0.204	-0.354	0.364	-0.150	0.393	-0.577	-0.490	0.392	0.087	0.401	-1.333	-1.266	0.380	0.068	0.385
	δ	-0.200	-0.286	0.269	-0.086	0.283	-0.500	-0.382	0.277	0.118	0.301	-0.800	-0.752	0.188	0.048	0.194
	λ	0.500	0.567	0.720	0.067	0.723	0.500	0.587	0.692	0.087	0.697	0.500	0.612	0.681	0.112	0.689
	TE	0.871	0.845	0.052	-0.025	0.059	0.869	0.868	0.048	-0.000	0.050	0.782	0.909	0.051	0.126	0.137
	LL	144.531	147.350	22.290	2.818	3.351	198.519	221.493	22.469	22.974	23.732	177.164	401.120	23.289	223.956	224.510

Notes: TE and LL represent respectively technical efficiency and log likelihood value

Table 8: Simulation results for $\lambda = 0.8$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	0.914	0.252	-0.086	0.266	1.000	0.538	0.246	-0.462	0.523	1.000	0.181	0.210	-0.819	0.845
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.317	0.037	0.011	0.039	0.346	0.295	0.039	-0.051	0.064	0.500	0.263	0.045	-0.237	0.241
	α	-0.204	-0.384	0.516	-0.180	0.547	-0.577	-0.379	0.551	0.199	0.586	-1.333	-0.682	0.820	0.651	1.047
	δ	-0.200	-0.270	0.327	-0.070	0.334	-0.500	-0.255	0.335	0.245	0.415	-0.800	-0.365	0.401	0.435	0.592
	λ	0.800	0.982	0.757	0.182	0.778	0.800	0.987	0.760	0.187	0.782	0.800	0.993	0.789	0.193	0.812
	TE	0.871	0.830	0.076	-0.041	0.087	0.869	0.829	0.081	-0.040	0.092	0.782	0.839	0.074	0.057	0.096
	LL	34.765	37.640	11.635	2.875	3.386	48.156	56.053	11.611	7.897	8.603	42.779	100.466	11.660	57.687	58.259
1000	β_0	1.000	0.935	0.219	-0.065	0.229	1.000	0.637	0.247	-0.363	0.439	1.000	0.470	0.299	-0.530	0.608
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.313	0.024	0.007	0.025	0.346	0.299	0.025	-0.048	0.054	0.500	0.293	0.025	-0.207	0.209
	α	-0.204	-0.350	0.363	-0.146	0.391	-0.577	-0.486	0.391	0.091	0.401	-1.333	-1.214	0.455	0.120	0.470
	δ	-0.200	-0.283	0.269	-0.083	0.281	-0.500	-0.379	0.276	0.121	0.301	-0.800	-0.719	0.237	0.081	0.251
	λ	0.800	1.158	0.674	0.358	0.763	0.800	1.173	0.666	0.373	0.763	0.800	1.194	0.672	0.394	0.778
	TE	0.871	0.850	0.064	-0.020	0.068	0.869	0.878	0.061	0.009	0.064	0.782	0.918	0.060	0.136	0.149
	LL	143.191	146.310	22.293	3.118	3.646	197.179	220.444	22.467	23.265	24.018	175.824	399.923	23.275	224.098	224.652

Notes: TE and LL represent respectively technical efficiency and log likelihood

Table 9: Simulation results for $\rho = 0.2$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	1.161	1.011	0.161	1.023	1.000	1.089	0.935	0.089	0.938	1.000	1.069	0.760	0.069	0.763
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.320	0.037	0.014	0.040	0.346	0.303	0.038	-0.043	0.058	0.500	0.293	0.036	-0.207	0.210
	α	-0.204	-0.465	0.510	-0.261	0.572	-0.577	-0.557	0.559	0.020	0.559	-1.333	-1.300	0.613	0.034	0.614
	δ	-0.200	-0.329	0.320	-0.129	0.344	-0.500	-0.378	0.335	0.122	0.357	-0.800	-0.718	0.240	0.082	0.254
	ρ	0.200	0.191	0.076	-0.009	0.077	0.200	0.192	0.071	-0.008	0.071	0.200	0.193	0.058	-0.007	0.059
	TE	0.871	0.845	0.054	-0.026	0.061	0.869	0.855	0.057	-0.014	0.062	0.782	0.911	0.032	0.128	0.134
	LL	36.349	38.799	11.596	2.449	2.928	49.740	57.276	11.600	7.536	8.243	44.363	102.302	11.605	57.939	58.497
1000	β_0	1.000	1.100	0.855	0.100	0.860	1.000	1.033	0.789	0.033	0.789	1.000	1.020	0.638	0.020	0.638
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.313	0.024	0.007	0.025	0.346	0.300	0.025	-0.047	0.053	0.500	0.297	0.018	-0.203	0.204
	α	-0.204	-0.369	0.352	-0.165	0.388	-0.577	-0.517	0.374	0.061	0.378	-1.333	-1.248	0.496	0.085	0.503
	δ	-0.200	-0.300	0.257	-0.100	0.276	-0.500	-0.406	0.260	0.094	0.277	-0.800	-0.739	0.270	0.061	0.276
	ρ	0.200	0.195	0.065	-0.005	0.065	0.200	0.196	0.060	-0.004	0.060	0.200	0.198	0.050	-0.002	0.050
	TE	0.871	0.855	0.047	-0.145	0.153	0.869	0.875	0.045	-0.125	0.133	0.782	0.897	0.123	-0.103	0.161
	LL	145.069	147.837	22.331	2.767	3.335	199.056	221.995	22.530	22.939	23.715	177.702	401.779	23.247	224.077	224.626

Notes: TE and LL represent respectively technical efficiency and log likelihood value

Table 10: Simulation results for $\rho = 0.5$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	1.292	1.504	0.292	1.531	1.000	1.207	1.392	0.207	1.407	1.000	1.160	1.137	0.160	1.147
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.319	0.037	0.012	0.039	0.346	0.303	0.039	-0.044	0.058	0.500	0.293	0.036	-0.207	0.210
	α	-0.204	-0.436	0.518	-0.231	0.567	-0.577	-0.543	0.564	0.035	0.565	-1.333	-1.305	0.605	0.029	0.605
	δ	-0.200	-0.304	0.323	-0.104	0.339	-0.500	-0.366	0.338	0.134	0.363	-0.800	-0.723	0.230	0.077	0.243
	ρ	0.500	0.488	0.071	-0.012	0.072	0.500	0.489	0.066	-0.011	0.067	0.500	0.491	0.055	-0.009	0.055
	TE	0.871	0.839	0.053	-0.032	0.063	0.869	0.853	0.055	-0.016	0.062	0.782	0.912	0.030	0.129	0.134
	LL	35.913	38.361	11.598	2.449	2.930	49.303	56.840	11.604	7.537	8.245	43.926	101.869	11.609	57.943	58.500
1000	β_0	1.000	1.175	1.288	0.175	1.299	1.000	1.096	1.189	0.096	1.193	1.000	1.065	0.966	0.065	0.968
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.311	0.024	0.005	0.024	0.346	0.298	0.025	-0.049	0.055	0.500	0.298	0.016	-0.202	0.203
	α	-0.204	-0.308	0.371	-0.104	0.385	-0.577	-0.473	0.395	0.104	0.409	-1.333	-1.328	0.242	0.006	0.242
	δ	-0.200	-0.245	0.275	-0.045	0.279	-0.500	-0.368	0.278	0.132	0.307	-0.800	-0.789	0.062	0.011	0.063
	ρ	0.500	0.492	0.061	-0.008	0.062	0.500	0.494	0.057	-0.006	0.057	0.500	0.497	0.047	-0.003	0.047
	TE	0.871	0.831	0.059	-0.039	0.072	0.869	0.868	0.047	-0.001	0.049	0.782	0.919	0.007	0.137	0.137
	LL	144.531	147.283	22.335	2.752	3.321	198.519	221.447	22.526	22.929	23.704	177.164	401.281	23.255	224.117	224.666

Notes: TE and LL represent respectively technical efficiency and log likelihood value

Table 11: Simulation results for $\rho = 0.8$

N	Parameters	$\delta = -0.2$					$\delta = -0.5$					$\delta = -0.8$				
		True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE	True	Mean	SD	Bias	RMSE
250	β_0	1.000	2.225	3.250	1.225	3.472	1.000	2.026	3.032	1.026	3.200	1.000	1.751	2.513	0.751	2.622
	β_1	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.007	0.000	0.007	1.000	1.000	0.006	0.000	0.006
	σ	0.306	0.319	0.037	0.013	0.039	0.346	0.303	0.039	-0.044	0.058	0.500	0.293	0.036	-0.207	0.210
	α	-0.204	-0.437	0.518	-0.233	0.568	-0.577	-0.541	0.564	0.037	0.565	-1.333	-1.305	0.604	0.029	0.605
	δ	-0.200	-0.305	0.324	-0.105	0.341	-0.500	-0.365	0.337	0.135	0.363	-0.800	-0.723	0.229	0.077	0.242
	ρ	0.800	0.777	0.062	-0.023	0.066	0.800	0.780	0.058	-0.020	0.062	0.800	0.785	0.049	-0.015	0.051
	TE	0.871	0.839	0.054	-0.032	0.064	0.869	0.853	0.056	-0.016	0.062	0.782	0.912	0.029	0.130	0.134
	LL	34.765	37.224	11.603	2.458	2.945	48.156	55.702	11.610	7.546	8.257	42.779	100.732	11.618	57.953	58.511
1000	β_0	1.000	1.827	2.843	0.827	2.959	1.000	1.645	2.640	0.645	2.716	1.000	1.430	2.178	0.430	2.220
	β_1	1.000	1.000	0.004	0.000	0.004	1.000	1.000	0.003	0.000	0.003	1.000	1.000	0.003	0.000	0.003
	σ	0.306	0.311	0.024	0.005	0.025	0.346	0.298	0.025	-0.049	0.055	0.500	0.298	0.016	-0.202	0.203
	α	-0.204	-0.293	0.383	-0.089	0.393	-0.577	-0.460	0.406	0.117	0.423	-1.333	-1.328	0.242	0.005	0.242
	δ	-0.200	-0.229	0.286	-0.029	0.287	-0.500	-0.356	0.289	0.144	0.323	-0.800	-0.789	0.063	0.011	0.064
	ρ	0.800	0.784	0.054	-0.016	0.056	0.800	0.787	0.051	-0.013	0.052	0.800	0.792	0.042	-0.008	0.043
	TE	0.871	0.818	0.064	-0.052	0.083	0.869	0.856	0.061	-0.012	0.065	0.782	0.919	0.007	0.136	0.137
	LL	143.191	145.948	22.321	2.757	3.323	197.179	220.101	22.528	22.923	23.698	175.824	399.936	23.256	224.112	224.661

Notes: TE and LL represent respectively technical efficiency and log likelihood value

Table 12: Production frontier estimates, model without climatic and other environmental variables (Model 1)

	OLS 1 (1)	SAR 1 (2)	SEM 1 (3)	SFA 1 (4)	SAR-SFA 1 (5)	SEM-SFA 1 (6)
Constant	11.268 (0.029)***	7.151 (0.170)***	11.244 (0.037)***	11.720 (0.093)***	7.486 (0.238)***	12.287 (0.124)***
Land	0.772 (0.012)***	0.627 (0.014)***	0.853 (0.014)***	0.765 (0.013)***	0.624 (0.014)***	0.848 (0.015)***
Family labor	0.073 (0.018)***	0.108 (0.017)***	0.095 (0.017)***	0.073 (0.018)***	0.108 (0.017)***	0.096 (0.017)***
Operating costs	0.032 (0.003)***	0.027 (0.003)***	0.022 (0.003)***	0.033 (0.003)***	0.028 (0.003)***	0.023 (0.003)***
ρ		0.346 (0.014)***			0.345 (0.014)***	
λ			0.587 (0.014)***			0.588 (0.014)***
σ				0.989 (0.043)***	0.883 (0.056)***	0.855 (0.026)***
α				-0.700 (0.163)***	-0.529 (0.261)**	-0.812 (0.109)***
Mean Efficiency				0.784 (0.154)	0.777 (0.122)	0.791 (0.174)
Log Likelihood	-5708.521	-5442.834	-5110.123	-4174.473	-3909.837	-3572.743
N	4423	4423	4423	4423	4423	4423

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

Table 13: Production frontier estimates, model with climatic variables (Model 2)

	OLS 2 (1)	SAR 2 (2)	SEM 2 (3)	SFA 2 (4)	SAR-SFA 2 (5)	SEM-SFA 2 (6)
Constant	10.512 (0.428)***	7.806 (0.428)***	10.822 (0.796)***	10.531 (0.428)***	7.816 (0.301)***	11.928 (0.458)***
Land	0.803 (0.013)***	0.679 (0.015)***	0.860 (0.014)***	0.803 (0.013)***	0.679 (0.013)***	0.858 (0.015)***
Family labor	0.052 (0.018)***	0.082 (0.017)***	0.087 (0.017)***	0.052 (0.018)***	0.082 (0.017)***	0.088 (0.017)***
Operating costs	0.033 (0.003)***	0.028 (0.003)***	0.023 (0.003)***	0.033 (0.003)***	0.028 (0.003)***	0.023 (0.003)***
Rainfall	0.824 (0.060)***	0.652 (0.057)***	0.801 (0.112)***	0.825 (0.045)***	0.652 (0.047)***	0.766 (0.089)***
Temperature	0.020 (0.011)*	-0.009 (0.011)	0.010 (0.021)	0.020 (0.009)**	-0.009 (0.008)	0.002 (0.012)
Rainfall Anomaly	-3.483 (0.262)***	-2.525 (0.252)***	-3.354 (0.478)***	-3.482 (0.320)***	-2.525 (0.151)***	-3.313 (0.424)***
Temperature Anomaly	-20.649 (3.199)***	-14.041 (3.037)***	-18.113 (5.987)***	-20.664 (0.176)***	-14.044 (0.909)***	-17.919 (0.615)***
ρ		0.310 (0.014)***			0.310 (0.014)***	
λ			0.547 (0.015)***			0.550 (0.015)***
σ				0.846 (0.011)***	0.800 (0.009)***	0.821 (0.030)***
α				-0.026 (0.431)	-0.015 (0.464)	-0.668 (0.135)***
Mean Efficiency				0.754 (0.091)	0.752 (0.091)	0.789 (0.166)
Log Likelihood	-5536.755	-5328.830	-5065.926	-4003.861	-3795.935	-3531.792
N.	4423	4423	4423	4423	4423	4423

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

Table 14: Production frontier estimates, model with climatic and other environmental variables (Model 3)

	OLS 3 (1)	SAR 3 (2)	SEM 3 (3)	SFA 3 (4)	SAR-SFA 3 (5)	SEM-SFA 3 (6)
Constant	10.314 (0.422)***	7.598 (0.419)***	10.449 (0.757)***	10.322 (1.227)***	7.895 (0.298)***	11.737 (0.282)***
Land	0.851 (0.013)***	0.731 (0.015)***	0.867 (0.014)***	0.851 (0.013)***	0.730 (0.015)***	0.863 (0.014)***
Family labor	0.051 (0.018)***	0.080 (0.017)***	0.087 (0.017)***	0.051 (0.018)***	0.080 (0.017)***	0.087 (0.017)***
Operating costs	0.024 (0.003)***	0.019 (0.003)***	0.018 (0.003)***	0.024 (0.003)***	0.019 (0.003)***	0.019 (0.003)***
Rainfall	0.173 (0.084)**	-0.007 (0.080)	0.273 (0.144)*	0.173 (0.063)***	-0.015 (0.062)	0.198 (0.109)*
Temperature	0.013 (0.011)	-0.016 (0.011)	0.011 (0.020)	0.013 (0.013)	-0.017 (0.008)**	-0.000 (0.008)
Rainfall Anomaly	-3.389 (0.260)***	-2.501 (0.249)***	-3.321 (0.456)***	-3.390 (0.201)***	-2.491 (0.207)***	-3.249 (0.360)***
Temperature Anomaly	-6.465 (3.423)*	-1.104 (3.236)	-7.897 (6.033)	-6.460 (0.405)***	-1.047 (0.439)**	-7.088 (0.081)***
Clay soil	0.016 (0.004)***	0.023 (0.003)***	0.013 (0.005)***	0.016 (0.004)***	0.023 (0.004)***	0.014 (0.005)***
Silt soil	0.016 (0.004)***	0.009 (0.004)**	0.013 (0.006)**	0.016 (0.004)***	0.009 (0.004)**	0.014 (0.006)**
Improved seed	0.107 (0.031)***	0.099 (0.030)***	0.109 (0.030)***	0.107 (0.032)***	0.100 (0.030)***	0.114 (0.030)***
Education	0.060 (0.027)**	0.051 (0.025)**	0.066 (0.025)***	0.060 (0.027)**	0.051 (0.025)**	0.068 (0.026)***
FBO Membership	0.324 (0.047)***	0.259 (0.044)***	0.145 (0.045)***	0.324 (0.047)***	0.260 (0.044)***	0.148 (0.045)***
ρ		0.310 (0.014)***			0.310 (0.014)***	
λ			0.523 (0.016)***			0.525 (0.016)***
σ				0.827 (0.013)***	0.822 (0.069)***	0.843 (0.025)***
α				-0.013 (1.433)	-0.425 (0.388)	-0.788 (0.106)***
Mean Efficiency				0.752 (0.089)	0.782 (0.139)	0.792 (0.172)
Log Likelihood	-5436.144	-5222.687	-5034.034	-3903.249	-3689.896	-3497.542
N.	4423	4423	4423	4423	4423	4423

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

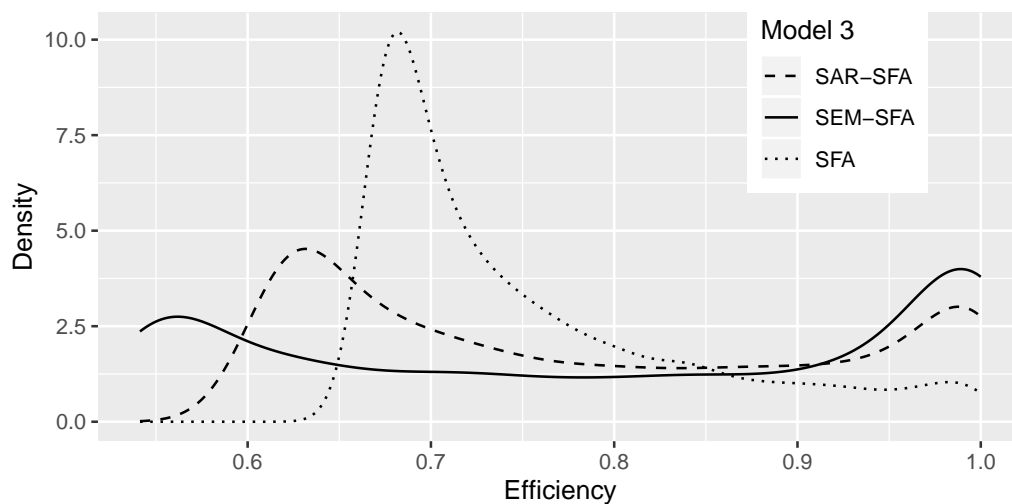
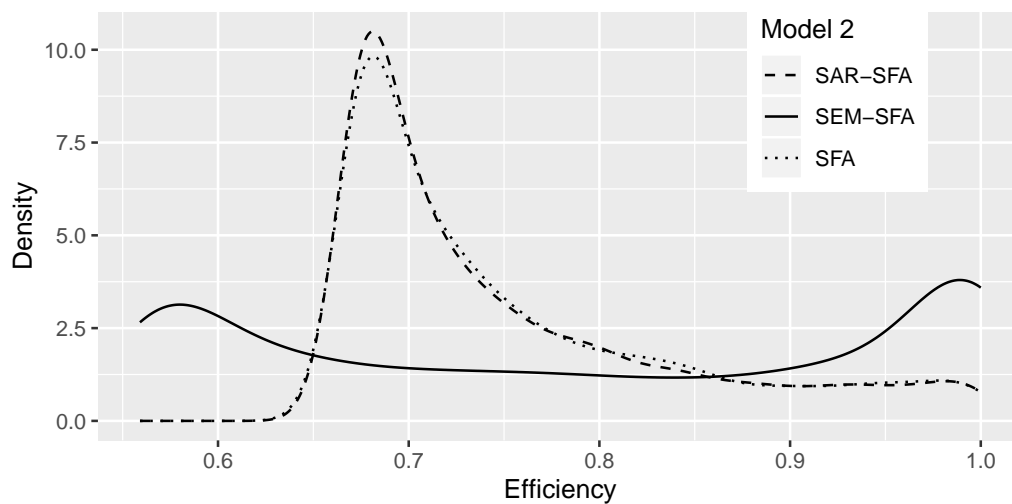
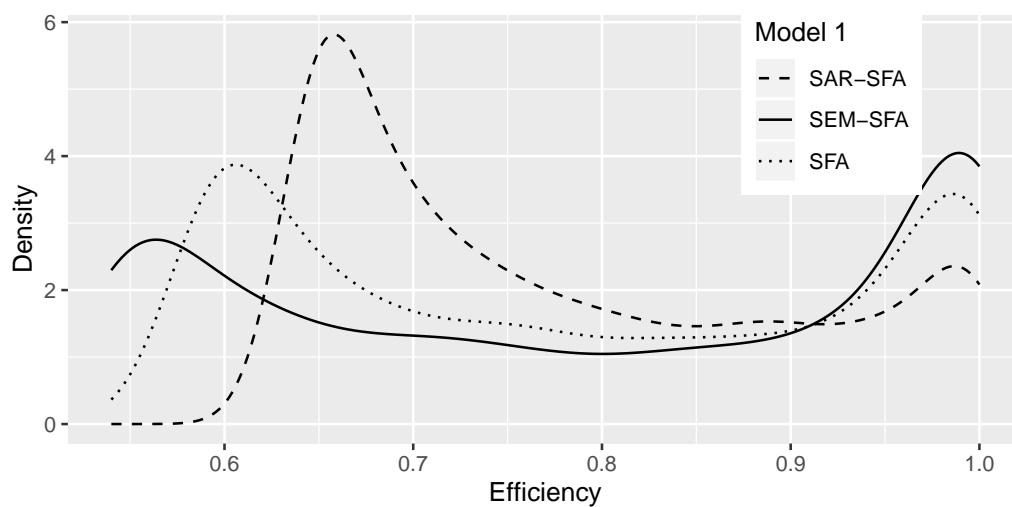
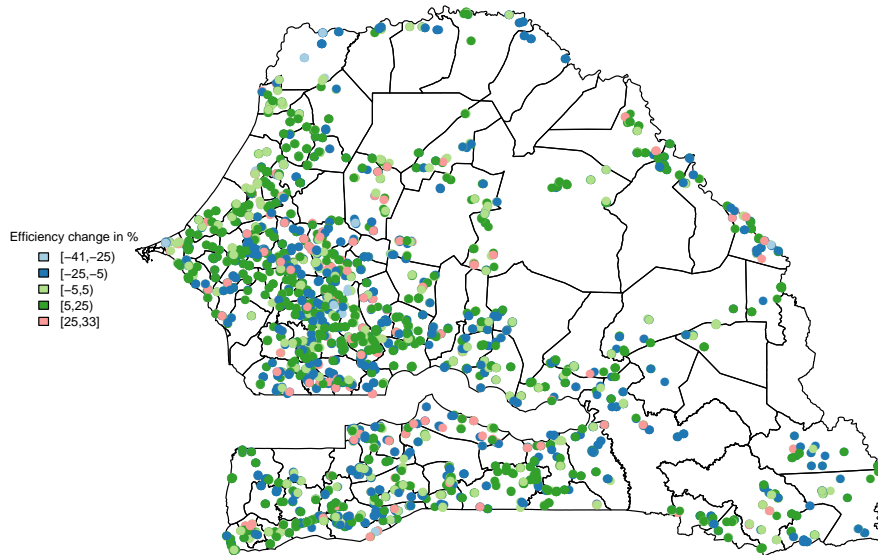
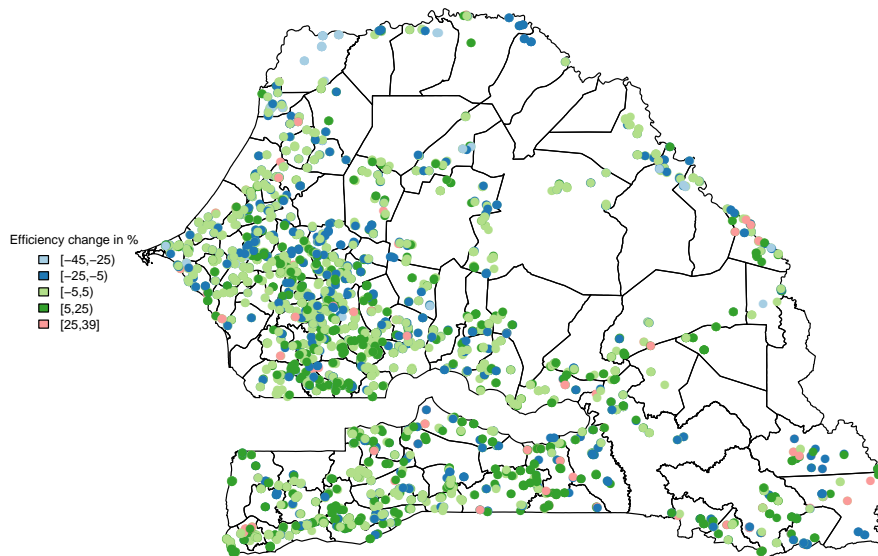


Figure 1: Kernel distributions of technical efficiency scores for models 1, 2 and 3



Due to spatial heterogeneity



Due to spatial heterogeneity and climate variability

Figure 2: Percentage change in individual efficiency scores

References

- Adetutu, M., Glass, A. J., Kenjegalieva, K. and Sickles, R. C. (2015) The effects of efficiency and TFP growth on pollution in Europe: a multistage spatial analysis, *Journal of Productivity Analysis*, **43**, 307–326.
- Affuso, E. (2010) Spatial autoregressive stochastic frontier analysis: An application to an impact evaluation study, *Available at SSRN 1740382*.
- Aigner, D., Lovell, C. and Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, **6**, 21 – 37.
- Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, NATO Asi Series. Series E, Applied Sciences, Springer Netherlands.
- Anselin, L., Bera, A. K., Florax, R. and Yoon, M. J. (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, **26**, 77–104.
- Areal, F. J., Balcombe, K. and Tiffin, R. (2012) Integrating spatial dependence into stochastic frontier analysis, *Australian Journal of Agricultural and Resource Economics*, **56**, 521–541.
- Arellano-Valle, R. B. and Azzalini, A. (2006) On the unification of families of skew-normal distributions, *Scandinavian Journal of Statistics*, **33**, 561–574.
- Arellano-Valle, R. B. and Azzalini, A. (2008) The centred parametrization for the multivariate skew-normal distribution, *Journal of Multivariate Analysis*, **99**, 1362 – 1382, special Issue: Multivariate Distributions, Inference and Applications in Memory of Norman L. Johnson.
- Azzalini, A. (1985) A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.
- Azzalini, A. (2005) The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics*, **32**, 159–188.
- Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**, 579–602.
- Azzalini, A. and Valle, A. D. (1996) The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.
- Ba, B. (2006) *Etude géographique de l’agriculture en Afrique noire: analyse des productions céréalières et des systèmes alimentaires au Sénégal*, Ph.D. thesis, University of Geneva.
- Barrios, E. B. and Lavado, R. F. (2010) Spatial stochastic frontier models, Tech. rep., PIDS Discussion Paper Series.
- Barrios, S., Bertinelli, L. and Strobl, E. (2010) Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy, *The Review of Economics and Statistics*, **92**, 350–366.
- Carvalho, A. (2018) Efficiency spillovers in Bayesian stochastic frontier models: application to electricity distribution in New Zealand, *Spatial Economic Analysis*, **13**, 171–190.
- D’Alessandro, S., Fall, A. A., Grey, G., Simpkin, S. and Wane, A. (2015) Senegal. Agricultural sector risk assessment, techreport 96296-SN, World Bank.
- de Graaff, T. (2020) On the estimation of spatial stochastic frontier models: an alternative skew-normal approach, *The Annals of Regional Science*, **64**, 267–285.
- Dedehouanou, F. (2014) Diversification towards high value export production, farm household productivity and efficiency in Senegal, in *17th Annual Conference on Global Economic Analysis, Dakar, Senegal*, Purdue University, West Lafayette, Global Trade Analysis Project (GTAP).
- Diagne, M., Demont, M., Seck, P. A. and Diaw, A. (2013) Self-sufficiency policy and

- irrigated rice productivity in the Senegal River Valley, *Food Security*, **5**, 55–68.
- Dominguez-Molina, J. A., González-Farías, G. and Ramos-Quiroga, R. (2003) Skew-normality in stochastic frontier analysis.
- Druska, V. and Horrace, W. C. (2004) Generalized Moments Estimation for Spatial Panel Data: Indonesian Rice Farming, *American Journal of Agricultural Economics*, **86**, 185–198.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. *et al.* (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Scientific data*, **2**, 1–21.
- Funk, C., Peterson, P., Peterson, S., Shukla, S., Davenport, F., Michaelsen, J., Knapp, K. R., Landsfeld, M., Husak, G., Harrison, L., Rowland, J., Budde, M., Meiburg, A., Dinku, T., Pedreros, D. and Mata, N. (2019) A high-resolution 1983–2016 tmax climate data record based on infrared temperatures and stations by the climate hazard center, *Journal of Climate*, **32**, 5639–5658.
- Fusco, E. and Vidoli, F. (2013) Spatial stochastic frontier models: controlling spatial global and local heterogeneity, *International Review of Applied Economics*, **27**, 679–694.
- Glass, A., Kenjegalieva, K. and Paez-Farrell, J. (2013) Productivity growth decomposition using a spatial autoregressive frontier model, *Economics Letters*, **119**, 291 – 295.
- Glass, A., Kenjegalieva, K. and Sickles, R. C. (2014) Estimating efficiency spillovers with state level evidence for manufacturing in the us, *Economics Letters*, **123**, 154 – 159.
- Glass, A. J., Kenjegalieva, K. and Sickles, R. C. (2016) A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers, *Journal of Econometrics*, **190**, 289 – 300, endogeneity Problems in Econometrics.
- Greene, W. H. (1990) A gamma-distributed stochastic frontier model, *Journal of Econometrics*, **46**, 141 – 163.
- Han, J., Ryu, D., Sickles, R. *et al.* (2016) How to measure spillover effects of public capital stock: a spatial autoregressive stochastic frontier model, Tech. rep.
- Hughes, N., Lawson, K., Davidson, A., Jackson, T. and Sheng, Y. (2011) Productivity pathways: climate-adjusted production frontiers for the Australian broadacre cropping industry, in *Australian Agricultural and Resource Economics Society Conference, 2011 Conference (55th), February 8–11, 2011, Melbourne, Australia*, 2011, 422-2016-26937, p. 33.
- Jalloh, A., Nelson, G., Thomas, T., Zougmore, R. and Roy-Macauley, H. (2013) *West African Agriculture and Climate Change: A Comprehensive Analysis*, Climate change in Africa, International Food Policy Research Institute.
- Lachaud, M. A., Bravo-Ureta, B. E. and Ludena, C. E. (2017) Agricultural productivity in latin america and the caribbean in the presence of unobserved heterogeneity and climatic effects, *Climatic Change*, **143**, 445–460.
- LeSage, J. and Pace, R. (2009) *Introduction to Spatial Econometrics*, Statistics, textbooks and monographs, CRC Press.
- Meeusen, W. and van Den Broeck, J. (1977) Efficiency estimation from cobb-douglas production functions with composed error, *International Economic Review*, **18**, 435–444.
- Mulwa, R. and Kabubo-Mariara, J. (2017) Productive efficiency and its determinants in a changing climate: A monotonic translog stochastic frontier analysis, Tech. rep., Environment for Development Initiative.
- Ngom, C. A. B., Sarr, F. and Fall, A. A. (2016) Mesure de l’efficacité technique des riziculteurs du bassin du fleuve sénégal, *Économie rurale. Agricultures, alimentations,*

- territoires*, pp. 91–105.
- O'Donnell, C. (2016) Using information about technologies, markets and firm behaviour to decompose a proper productivity index, *Journal of Econometrics*, **190**, 328 – 340.
- Okuyama, Y., Maruyama, A., Takagaki, M. and Kikuchi, M. (2017) Technical efficiency and production potential of selected cereal crops in senegal, *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, **118**, 187–197.
- 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**, 556 – 577.
- Oyekale, A. S. (2012) Impact of climate change on cocoa agriculture and technical efficiency of cocoa farmers in south-west nigeria, *Journal of Human Ecology*, **40**, 143–148.
- Pavlyuk, D. (2011) Application of the spatial stochastic frontier model for analysis of a regional tourism sector, *Transport and Telecommunication*, **12**, 28–38.
- Pavlyuk, D. (2013) Distinguishing between spatial heterogeneity and inefficiency: Spatial stochastic frontier analysis of european airports, *Transport and Telecommunication*, **14**, 29–38.
- Pede, V. O., Areal, F. J., Singbo, A., McKinley, J. and Kajisa, K. (2018) Spatial dependency and technical efficiency: an application of a Bayesian stochastic frontier model to irrigated and rainfed rice farmers in Bohol, Philippines, *Agricultural Economics*, **49**, 301–312.
- Ramajo, J. and Hewings, G. J. (2018) Modeling regional productive performance using a spatial stochastic frontier approach: New evidence for europe (1995-2007), *Regional Studies*, **52**, 1372–1387.
- Republique du Senegal (2018) Note d'Analyse du Commerce Extérieur, Edition 2017, Tech. rep., Agence Nationale de la Statistique et de la Démographie, ANSD.
- Schmidt, A., Moreira, A., Helfand, S. and Fonseca, T. (2009) Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency, *Journal of Productivity Analysis*, **31**, 101–112.
- Seck, A. (2017) Fertiliser subsidy and agricultural productivity in senegal, *The World Economy*, **40**, 1989–2006.
- Sherlund, S. M., Barrett, C. B. and Adesina, A. A. (2002) Smallholder technical efficiency controlling for environmental production conditions, *Journal of Development Economics*, **69**, 85 – 101.
- Skevas, I. (2020) Inference in the spatial autoregressive efficiency model with an application to dutch dairy farms, *European Journal of Operational Research*, **283**, 356 – 364.
- Stevenson, R. E. (1980) Likelihood functions for generalized stochastic frontier estimation, *Journal of Econometrics*, **13**, 57 – 66.
- Thiam, A. and Bravo-Ureta, B. (2003) Technical efficiency measures for a sample of senegalese groundnut producers using pooled cross-section time series data., *International Arachis Newsletter*, pp. 36–39.
- Tonini, A. and Pede, V. (2011) A generalized maximum entropy stochastic frontier measuring productivity accounting for spatial dependency, *Entropy*, **13**, 1916–1927.
- Tsonas, E. G. and Michaelides, P. G. (2016) A spatial stochastic frontier model with spillovers: Evidence for italian regions, *Scottish Journal of Political Economy*, **63**, 243–257.
- World Bank (2019) World bank data, world Bank Data Base.

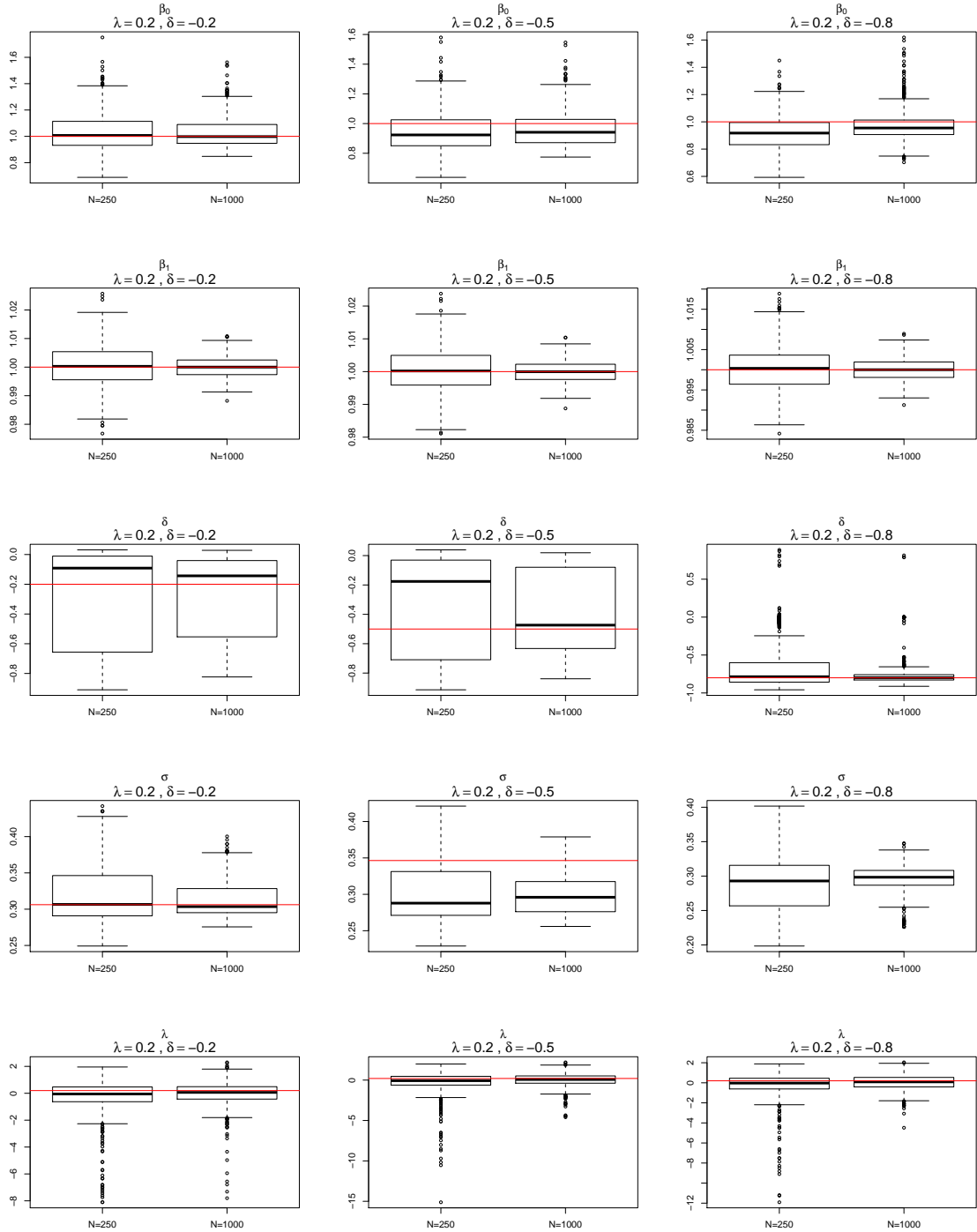


Figure 5: Box plots of β_0 , β_1 , δ , σ and λ for true $\lambda=0.2$

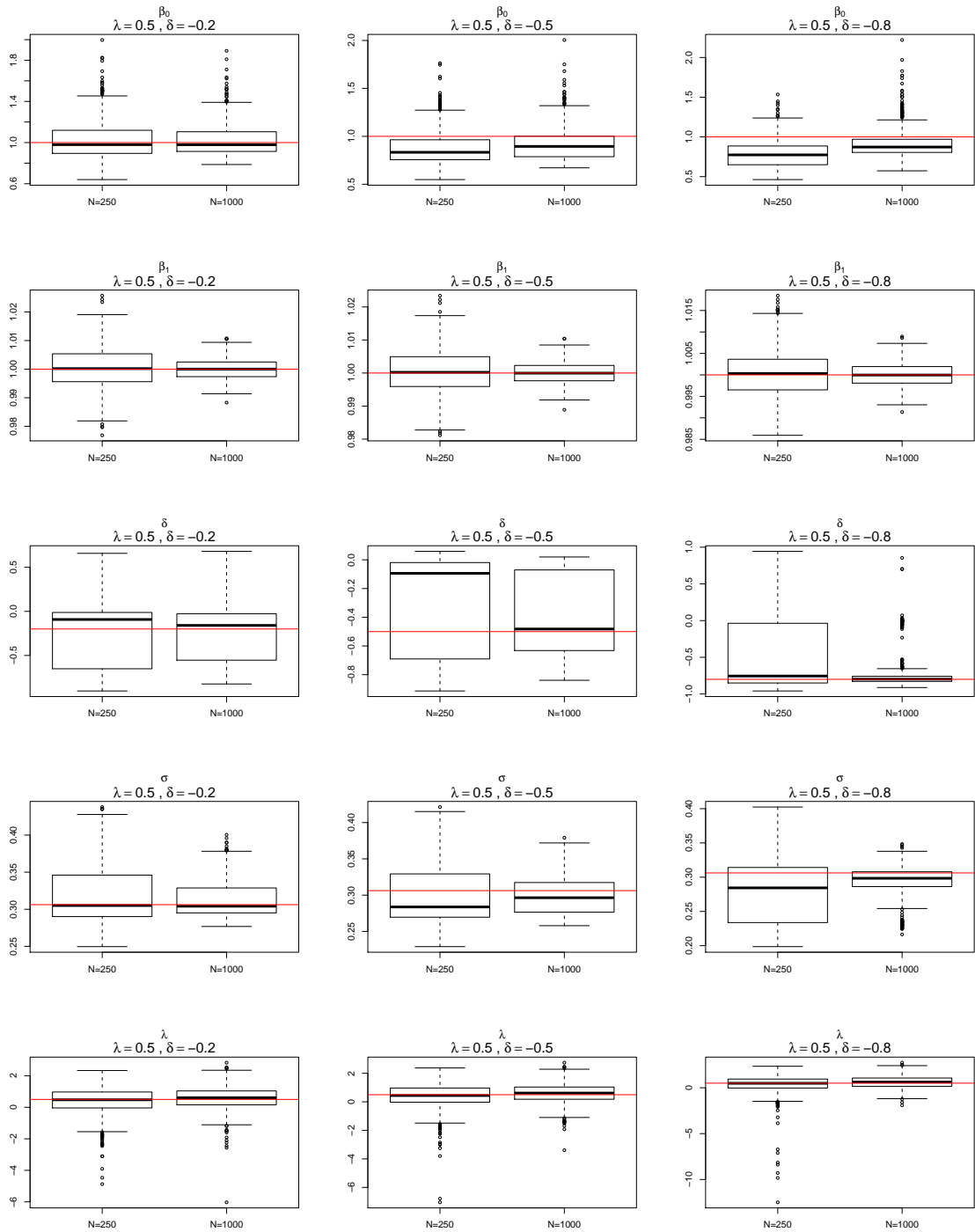


Figure 6: Box plots of β_0 , β_1 , δ , σ and λ for true $\lambda=0.5$

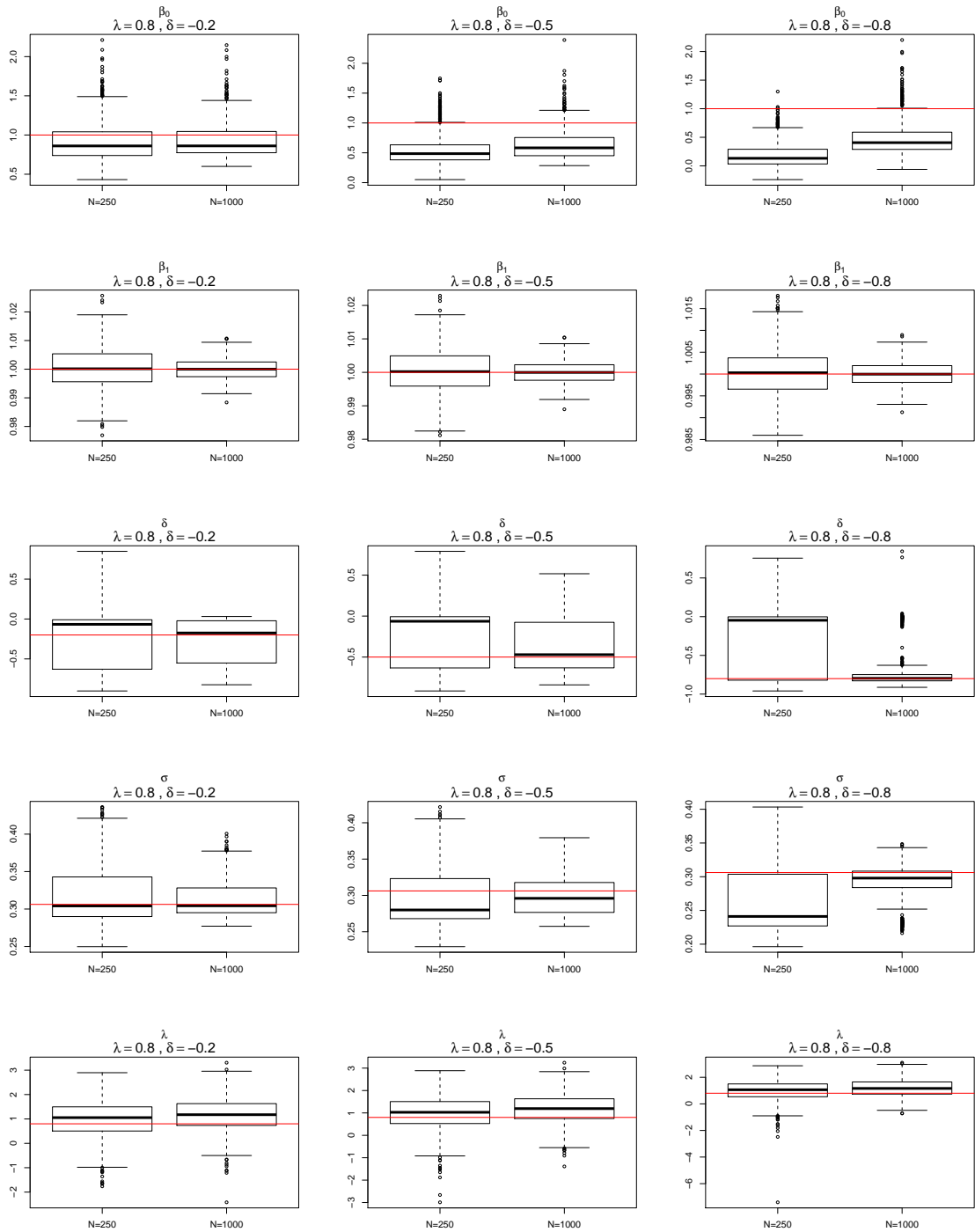


Figure 7: Box plots of β_0 , β_1 , δ , σ and λ for true $\lambda=0.8$

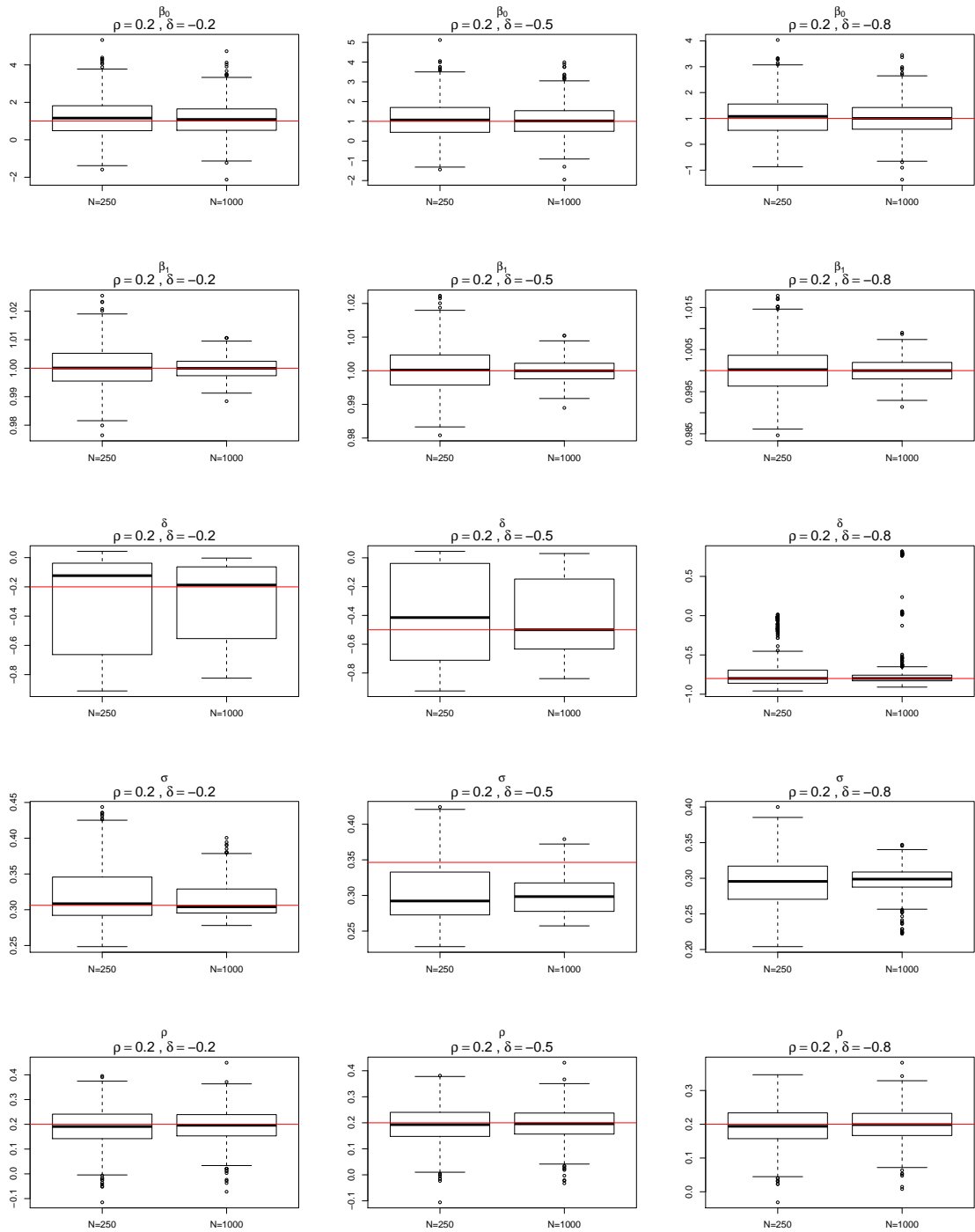


Figure 8: Box plots of β_0 , β_1 , δ , σ and ρ for true $\rho=0.2$

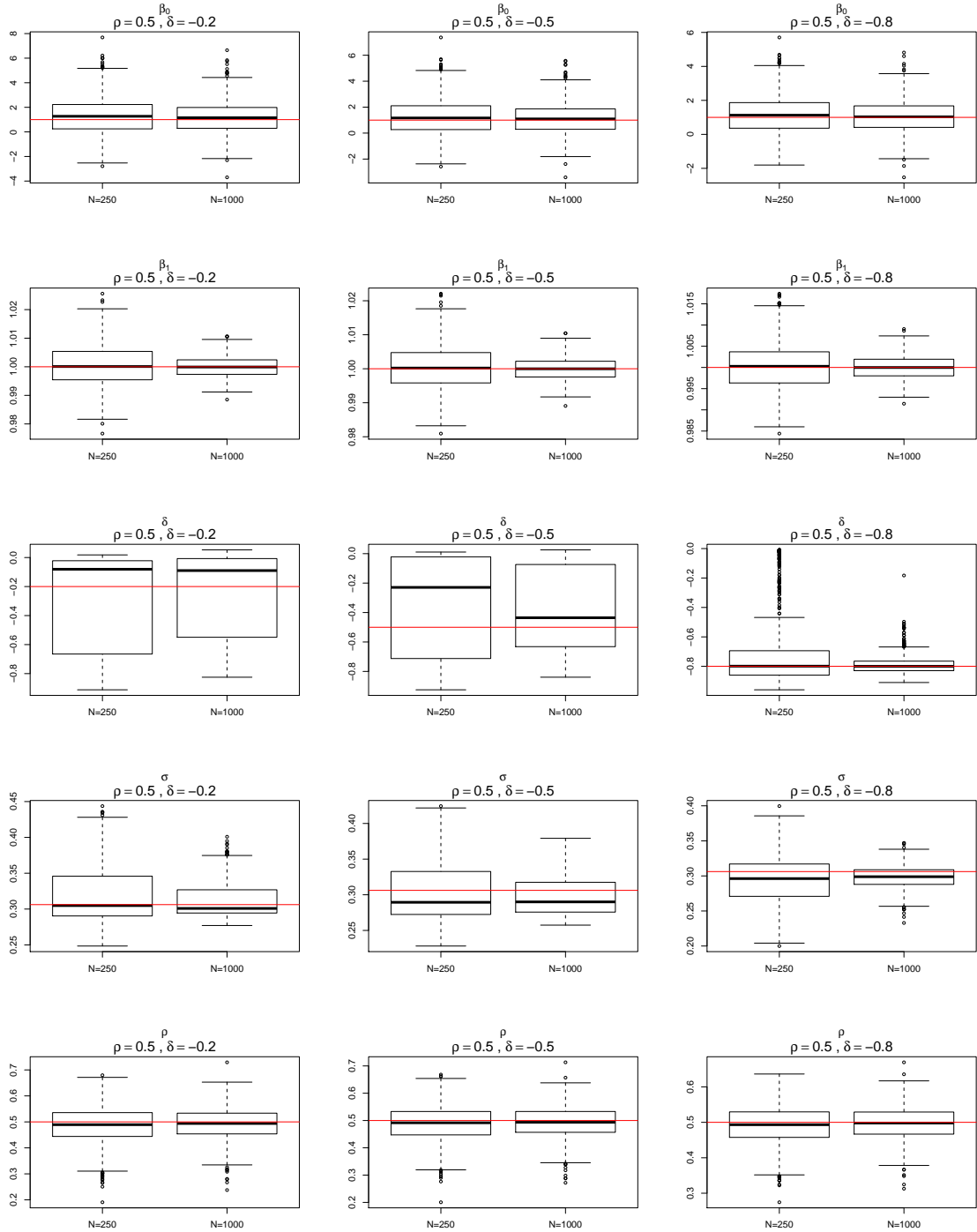


Figure 9: Box plots of β_0 , β_1 , δ , σ and ρ for true $\rho=0.5$

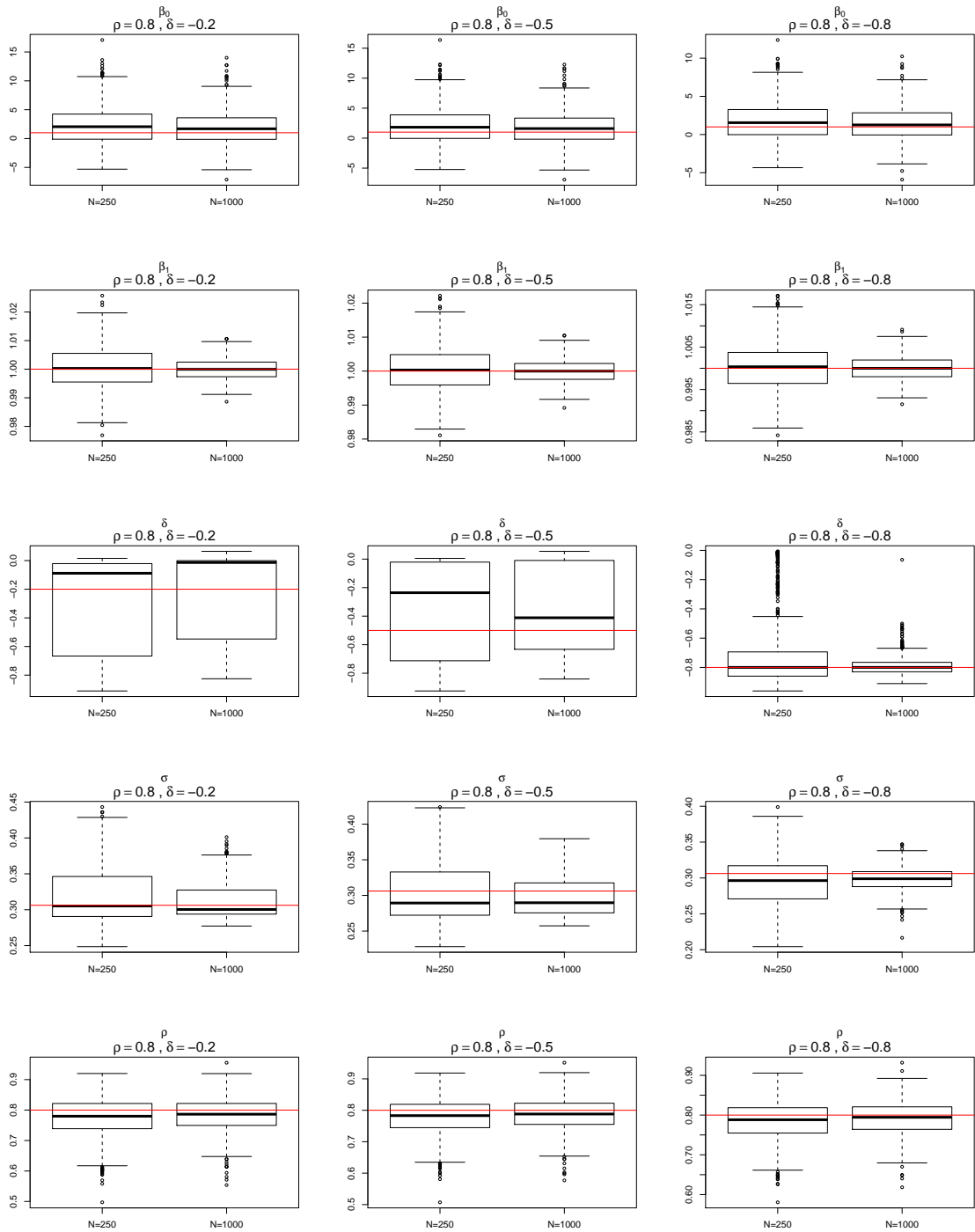


Figure 10: Box plots of β_0 , β_1 , δ , σ and ρ for true $\rho=0.8$

Table 15: Translog production frontier estimates, model without climatic and other environmental variables (Model 1)

	OLS	SAR	SEM	SFA	SAR-SFA	SEM-SFA
Constant	11.335 (0.040)***	7.435 (0.171)***	11.313 (0.044)***	11.918 (0.059)***	7.934 (0.202)***	11.313 (0.479)***
Land	0.774 (0.027)***	0.626 (0.027)***	0.786 (0.026)***	0.767 (0.027)***	0.623 (0.026)***	0.786 (0.026)***
Family labor	-0.024 (0.039)	0.002 (0.037)	0.008 (0.034)	-0.031 (0.039)	-0.003 (0.037)	0.008 (0.035)
Operating costs	-0.167 (0.014)***	-0.146 (0.013)***	-0.105 (0.012)***	-0.170 (0.014)***	-0.148 (0.013)***	-0.105 (0.013)***
Land x Land	-0.040 (0.010)***	-0.044 (0.009)***	-0.015 (0.009)	-0.043 (0.010)***	-0.044 (0.009)***	-0.015 (0.009)
Family labor x Family labor	0.110 (0.034)***	0.131 (0.031)***	0.101 (0.029)***	0.111 (0.033)***	0.131 (0.031)***	0.101 (0.029)***
Operating costs x Operating costs	0.040 (0.002)***	0.035 (0.002)***	0.025 (0.002)***	0.041 (0.002)***	0.035 (0.002)***	0.025 (0.002)***
Land x Family labor	0.068 (0.016)***	0.054 (0.015)***	0.042 (0.014)***	0.070 (0.016)***	0.055 (0.015)***	0.042 (0.014)***
Land x Operating costs	-0.018 (0.003)***	-0.013 (0.003)***	-0.005 (0.003)*	-0.019 (0.003)***	-0.014 (0.003)***	-0.005 (0.003)*
Family labor x Operating costs	-0.006 (0.004)	-0.005 (0.004)	-0.004 (0.003)	-0.005 (0.004)	-0.004 (0.004)	-0.004 (0.003)
ρ		0.328 (0.014)***			0.325 (0.015)***	
λ			0.570 (0.014)***			0.570 (0.014)***
σ				1.029 (0.028)***	0.919 (0.030)***	0.729 (0.008)***
α				-1.007 (0.104)***	-0.807 (0.120)***	-0.000 (0.355)
Mean Efficiency				0.788 (0.167)	0.790 (0.167)	0.748 (0.088)
Log Likelihood	-5553.057	-5306.002	-5037.265	-4010.806	-3769.499	-3504.370
Num. obs.	4423	4423	4423	4423	4423	4423

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

Table 16: Translog production frontier estimates, model with climatic variables (Model 2)

	OLS	SAR	SEM	SFA	SAR-SFA	SEM-SFA
Constant	10.485 (0.417)***	7.938 (0.418)***	10.757 (0.755)***	10.512 (0.350)***	7.948	11.978 (0.188)***
Land	0.821 (0.027)***	0.692 (0.027)***	0.802 (0.026)***	0.821 (0.027)***	0.692 (0.026)***	0.799 (0.026)***
Family labor	-0.011 (0.038)	-0.002 (0.036)	0.006 (0.034)	-0.011 (0.038)	-0.002 (0.036)	0.005 (0.034)
Operating costs	-0.153 (0.013)***	-0.140 (0.013)***	-0.104 (0.012)***	-0.153 (0.013)***	-0.140 (0.013)***	-0.106 (0.012)***
Land x Land	-0.030 (0.009)***	-0.034 (0.009)***	-0.013 (0.009)	-0.030 (0.009)***	-0.034 (0.009)***	-0.014 (0.009)
Family labor x Family labor	0.084 (0.032)***	0.111 (0.031)***	0.096 (0.029)***	0.085 (0.032)***	0.111 (0.030)***	0.097 (0.029)***
Operating costs x Operating costs	0.039 (0.002)***	0.034 (0.002)***	0.025 (0.002)***	0.039 (0.002)***	0.034 (0.002)***	0.026 (0.002)***
Land x Family labor	0.065 (0.015)***	0.054 (0.015)***	0.043 (0.014)***	0.065 (0.015)***	0.054 (0.015)***	0.043 (0.014)***
Land x Operating costs	-0.019 (0.003)***	-0.015 (0.003)***	-0.006 (0.003)**	-0.019 (0.003)***	-0.015 (0.003)***	-0.006 (0.003)**
Family labor x Operating costs	-0.008 (0.004)**	-0.006 (0.003)*	-0.004 (0.003)	-0.008 (0.004)**	-0.006 (0.003)*	-0.004 (0.003)
Rainfall	0.851 (0.058)***	0.677 (0.056)***	0.809 (0.106)***	0.852 (0.048)***	0.677 (0.042)***	0.761 (0.080)***
Temperature	0.020 (0.011)*	-0.007 (0.010)	0.012 (0.020)	0.020 (0.010)**	-0.007 (0.007)	0.003 (0.006)
Rainfall Anomaly	-2.938 (0.256)***	-2.083 (0.247)***	-2.983 (0.455)***	-2.937 (0.290)***	-2.082	-2.930 (0.200)***
Temperature Anomaly	-19.511 (3.124)***	-13.217 (2.978)***	-17.773 (5.682)***	-19.572 (0.853)***	-13.223 (0.279)***	-17.238
ρ		0.293 (0.014)***			0.293 (0.014)***	
λ			0.527 (0.015)***			0.530 (0.016)***
σ				0.818 (0.012)***	0.777 (0.009)***	0.828 (0.025)***
α				-0.036 (0.459)	-0.017 (0.451)	-0.755 (0.110)***
Mean Efficiency				0.755 (0.092)	0.752 (0.090)	0.791 (0.173)
Log Likelihood	-5388.054	-5194.900	-4992.608	-3855.159	-3662.006	-3456.784
Num. obs.	4423	4423	4423	4423	4423	4423

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

Table 17: Translog production frontier estimates, model with climatic and other environmental variables (Model 3)

	OLS	SAR	SEM	SFA	SAR-SFA	SEM-SFA
Constant	10.337 (0.414)***	7.757 (0.412)***	10.435 (0.725)***	10.940 (0.294)***	8.263 (0.272)***	11.780 (0.402)***
Land	0.850 (0.027)***	0.726 (0.026)***	0.810 (0.026)***	0.848 (0.027)***	0.724 (0.026)***	0.807 (0.026)***
Family labor	-0.004 (0.038)	0.006 (0.036)	0.009 (0.034)	-0.007 (0.038)	0.003 (0.036)	0.007 (0.034)
Operating costs	-0.145 (0.013)***	-0.131 (0.013)***	-0.105 (0.012)***	-0.146 (0.013)***	-0.132 (0.013)***	-0.107 (0.012)***
Land x Land	-0.025 (0.009)***	-0.029 (0.009)***	-0.013 (0.009)	-0.025 (0.009)***	-0.029 (0.009)***	-0.014 (0.009)
Family labor x Family labor	0.080 (0.032)**	0.106 (0.030)***	0.094 (0.029)***	0.082 (0.032)**	0.107 (0.030)***	0.095 (0.029)***
Operating costs x Operating costs	0.035 (0.002)***	0.031 (0.002)***	0.025 (0.002)***	0.035 (0.002)***	0.031 (0.002)***	0.025 (0.002)***
Land x Family labor	0.055 (0.015)***	0.044 (0.014)***	0.041 (0.014)***	0.055 (0.015)***	0.044 (0.014)***	0.041 (0.014)***
Land x Operating costs	-0.016 (0.003)***	-0.012 (0.003)***	-0.005 (0.003)**	-0.016 (0.003)***	-0.012 (0.003)***	-0.006 (0.003)**
Family labor x Operating costs	-0.007 (0.004)**	-0.006 (0.003)*	-0.004 (0.003)	-0.007 (0.004)*	-0.005 (0.003)	-0.004 (0.003)
Rainfall	0.295 (0.083)***	0.101 (0.079)	0.335 (0.138)**	0.261 (0.063)***	0.076 (0.062)	0.253 (0.107)**
Temperature	0.014 (0.011)	-0.014 (0.010)	0.013 (0.019)	0.009 (0.008)	-0.017 (0.008)**	0.000 (0.011)
Rainfall Anomaly	-2.946 (0.255)***	-2.142 (0.245)***	-2.968 (0.438)***	-2.905 (0.229)***	-2.110 (0.317)***	-2.889 (0.274)***
Temperature Anomaly	-7.383 (3.352)**	-2.032 (3.181)	-8.240 (5.785)	-7.082 (0.339)***	-1.753 (0.277)***	-7.221
Clay soil	0.012 (0.004)***	0.019 (0.003)***	0.011 (0.005)**	0.013 (0.004)***	0.019 (0.003)***	0.011 (0.005)**
Silt soil	0.015 (0.004)***	0.009 (0.004)**	0.013 (0.006)**	0.015 (0.004)***	0.009 (0.004)**	0.014 (0.006)**
Improved seed	0.122 (0.031)***	0.111 (0.029)***	0.117 (0.029)***	0.125 (0.031)***	0.115 (0.029)***	0.122 (0.029)***
Education	0.051 (0.026)*	0.042 (0.025)*	0.058 (0.025)**	0.051 (0.026)**	0.043 (0.024)*	0.060 (0.025)**
FBO Membership	0.224 (0.046)***	0.177 (0.043)***	0.120 (0.045)***	0.225 (0.046)***	0.178 (0.044)***	0.121 (0.045)***
ρ		0.296 (0.014)***			0.296 (0.014)***	
λ			0.507 (0.016)***			0.508 (0.016)***
σ				0.906 (0.032)***	0.850 (0.032)***	0.842 (0.023)***
α				-0.703 (0.133)***	-0.661 (0.140)***	-0.840 (0.099)***
Mean Efficiency				0.788 (0.160)	0.788 (0.162)	0.794 (0.176)
Log Likelihood	-5316.782	-5116.775	-4963.964	-3782.425	-3582.761	-3425.641
Num. obs.	4423	4423	4423	4423	4423	4423

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses

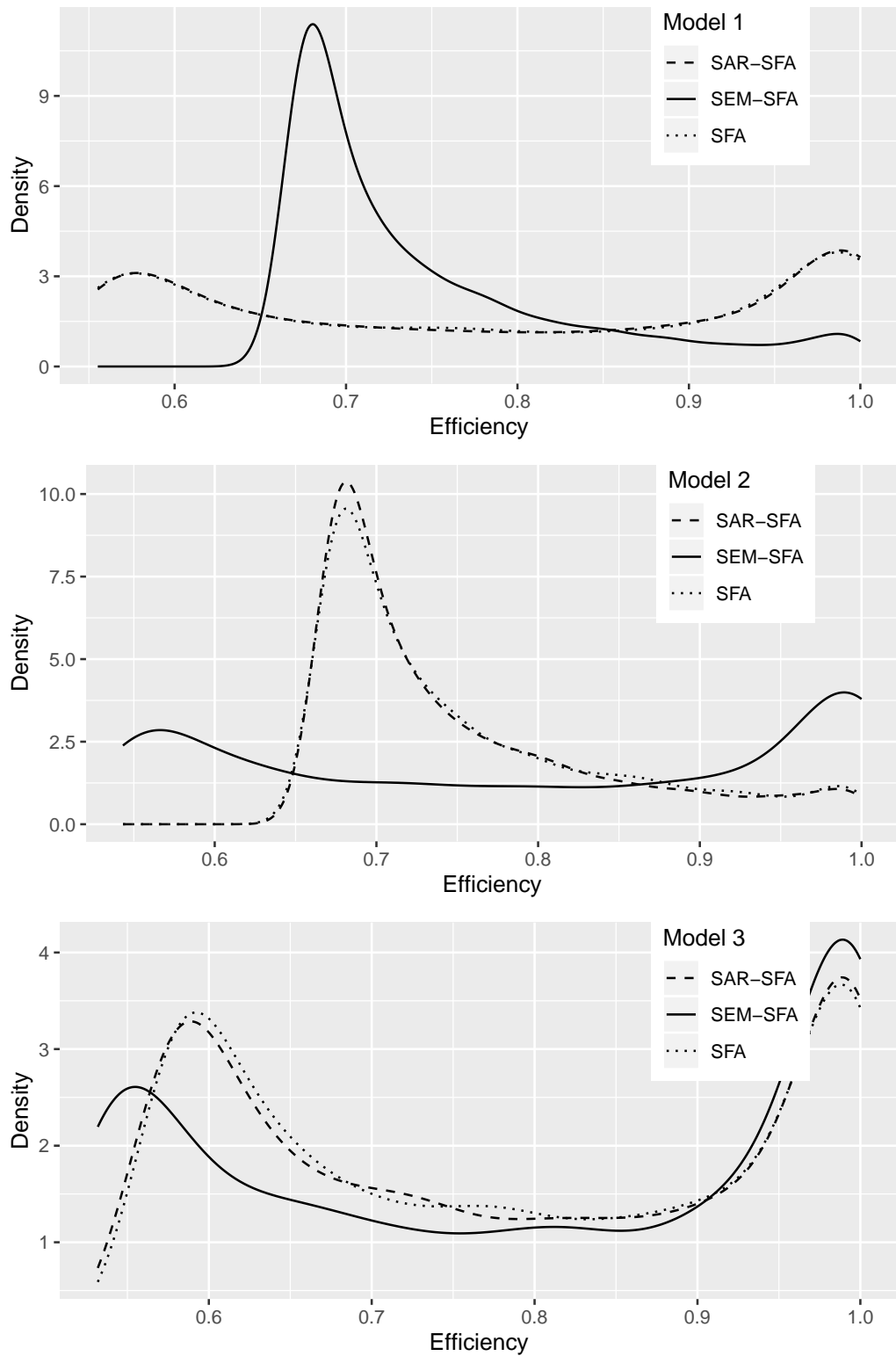


Figure 11: Kernel distributions of technical efficiency scores for models 1, 2 and 3 with Translog

```

# Packages
library(igraph)
library(spdep)
library(maxLik)
library(GMCM)

# Simulations functions
simul_spsfa <- function ( repetitions = NULL, seed = NULL,
                          nobS = NULL, Wmatr = NULL,
                          swdim = NULL, swneig = NULL, swp = NULL,
                          geolat = NULL, geolon = NULL,
                          minX = NULL, maxX = NULL, dgp = NULL, bcoef = NULL,
                          fake_rho = NULL, fake_lambda = NULL,
                          fake_sdv = NULL, fake_sdu = NULL, fake_delta = NULL,
                          approach = NULL, estim = NULL, efficiency = NULL,
                          method = NULL, control=NULL) {

# Set the seed
set.seed ( seed )

# Matrices to simulations results

Nobs <- matrix ( NA, nrow = repetitions, ncol=1)
Wmatrice <- matrix ( NA, nrow = repetitions, ncol = 1)
minXvar <- matrix ( NA, nrow = repetitions, ncol = 1)
maxXvar <- matrix ( NA, nrow = repetitions, ncol = 1)
dataGP <- matrix ( NA, nrow = repetitions, ncol = 1)
rho_true <- matrix ( NA, nrow = repetitions, ncol = 1)
lambda_true <- matrix ( NA, nrow = repetitions, ncol = 1)
splug_true <- matrix ( NA, nrow = repetitions, ncol = 1)

beta_true <- matrix ( NA, nrow = repetitions, ncol = 2)
sdv_true <- matrix ( NA, nrow = repetitions, ncol = 1)
sdu_true <- matrix ( NA, nrow = repetitions, ncol = 1)
alpha_true <- matrix ( NA, nrow = repetitions, ncol = 1)
sigma_true <- matrix ( NA, nrow = repetitions, ncol = 1)
delta_true <- matrix ( NA, nrow = repetitions, ncol = 1)
ditribution <- matrix ( NA, nrow = repetitions, ncol = 1)
estimation <- matrix ( NA, nrow = repetitions, ncol = 1)
U_true <- matrix ( NA, nrow = repetitions, ncol = nobS)
mTE_true <- matrix ( NA, nrow = repetitions, ncol = 1)
Loglik_true <- matrix ( NA, nrow = repetitions, ncol = 1)

beta_est <- matrix ( NA, nrow = repetitions, ncol = 2)
alpha_est <- matrix ( NA, nrow = repetitions, ncol = 1)
sigma_est <- matrix ( NA, nrow = repetitions, ncol = 1)
delta_est <- matrix ( NA, nrow = repetitions, ncol = 1)
splug_est <- matrix ( NA, nrow = repetitions, ncol = 1)
U_est <- matrix ( NA, nrow = repetitions, ncol = nobS)
mTE_est <- matrix ( NA, nrow = repetitions, ncol = 1)
Loglik_est <- matrix ( NA, nrow = repetitions, ncol = 1)

timespent <- matrix ( NA, nrow = repetitions, ncol = 1)

```

```

# Data generating process function
func_dgp_est <- function ( repetitions = NULL, seed = NULL,
                           nobs = NULL, Wmatr = NULL,
                           swdim = NULL, swneig = NULL, swp = NULL,
                           geolat = NULL, geolon = NULL,
                           minX = NULL, maxX = NULL, dgp = NULL, bcoef = NULL,
                           fake_rho = NULL, fake_lambda = NULL,
                           fake_sdv = NULL, fake_sdu = NULL, fake_delta = NULL,
                           approach = NULL, estim = NULL, efficiency = NULL,
                           method = NULL, control=NULL) {

  rm ( list=ls () )

  if (Wmatr == "sw"){ # Small World graph weight matrix
    Wmat <- sample_smallworld ( dim = swdim, size = nobs,
                               nei = swneig, p = swp)
    Wmatw <- as.matrix ( as_adjacency_matrix ( Wmat, type = "both" ) )

  } else { # Geographic distance weight matrix
    Dgeo <- data.frame ( lat = runif(nobs, 0, geolat),
                        lon = runif ( nobs, 0, geolon))
    Dmat <- dist ( cbind ( Dgeo$lon, Dgeo$lat ),
                  method = 'euclidean', upper=TRUE )
    Wmatw <- ( as.matrix ( (1 / Dmat), upper=TRUE, diag = TRUE ) )
    Wmatw[!is.finite(Wmatw)] <- 0
    diag ( Wmatw ) <- 0
  }
  # Row standardized
  Wmatws <- apply ( Wmatw, 2 , function(x) x/rowSums(Wmatw))
  Wmatlw <- mat2listw (Wmatw, style="W")

  # Inverse of matrix (I-rho*W)
  I_rhow <- invIrW ( Wmatws, fake_rho, method = "solve", feasible = NULL)
  I_lambdaw <- invIrW ( Wmatws, fake_lambda, method = "solve", feasible = NULL)

  # Creation of variables
  # Data is constant plus a uniform covariate
  X <- cbind ( c (1), runif ( nobs, minX, maxX ) )

  # Inefficiency term, # u ~ Normal(0, 0.3)
  u <- rnorm (nobs, mean=0, sd = fake_sdu )

  # Noise # normal idiosyncratic error term
  v <- rnorm (nobs, mean=0, sd = fake_sdv)

  # True values
  fake_alpha <- ( fake_delta / (sqrt( 1 - fake_delta^2 )) )
  fake_sigma <- fake_sdv / (sqrt( 1 - fake_delta^2 ))

  if ((dgp == "sem")|| ( dgp == "sde")){
    fake_splag <- fake_lambda
  } else { fake_splag <- fake_rho }
}

```

```

if (approach == "halfn") { # normal - half-normal approach
  e <- v - abs(u)
} else { # skew-normal approach
  e <- fake_delta * abs(u) + sqrt ( 1 - fake_delta^2 ) * v }

# Dependent variable
if ( dgp == "linear" ) { # No spatial
  y <- X%%bcoef + e

} else if ( dgp == "sar" ) { # SAR:  $y = \rho * W * y + X * \beta + e$ 
  y <- I_rhow %% ( X%%bcoef + e )
  I <- diag(nobs)
  fake_Bmat <- I - fake_splag * Wmatws
  fake_detBmat <- determinant(fake_Bmat, logarithm = TRUE)$modulus
  fake_epsilon <- (1/fake_sigma) * (fake_Bmat%%y - X%%bcoef)

  fake_Loglik <- - (nobs / 2) * log (pi) -
    (nobs/2) * log( fake_sigma^2 ) +
    fake_detBmat - (1/2) * (t ( fake_epsilon ) %% fake_epsilon ) +
    sum ( log ( 2 * pnorm ( fake_alpha * fake_epsilon ) ) )

} else if ( dgp == "sdb" ) { # SDB:  $y = \rho * W * y + X * \beta + W * X * \theta + e$ 
  y <- I_rhow %% ( X[,1] * bcoef[1] + X[,2] * (bcoef[2]/2) +
    Wmatws %% X[,2] * (bcoef[2]/2) + e )

} else if ( dgp == "sem" ) { # SEM:  $y = X * \beta + [I - W * \lambda]^{-1} * e$ 
  y <- X%%bcoef + I_lambdaw %% e
  I <- diag(nobs)
  fake_Bmat <- I - fake_splag * Wmatws

  fake_epsilon <- (1/fake_sigma) * fake_Bmat %% (y - X%%bcoef)
  fake_detBmat <- determinant(fake_Bmat, logarithm = TRUE)$modulus

  fake_Loglik <- (-1) * ( nobs/2 ) * log ( pi ) -
    (nobs/2) * log ( fake_sigma^2 ) +
    fake_detBmat - ( 1/2 ) * t ( fake_epsilon ) %% fake_epsilon +
    sum ( log ( 2 * pnorm ( fake_alpha * fake_epsilon ) ) )

} else { # dgp == "sde" # SDE:  $y = X * \beta + W * X * \theta + [I - W * \lambda]^{-1} * e$ 
  y <- X[,1] * bcoef[1] + X[,2] * (bcoef[2]/2) +
    Wmatws %% X[,2] * (bcoef[2]/2) + I_lambdaw %% e
}

#Estimating specific inefficiency

if (dgp == "sem") {
  fake_residuals <- ( 1 / fake_sigma ) * fake_Bmat %% ( y - X %% bcoef )
} else if ( dgp == "sar" ) {
  fake_residuals <- ( 1 / fake_sigma ) * (fake_Bmat %% y - X %% bcoef )
} else {
  fake_residuals <- ( 1 / fake_sigma ) * (y - X %% bcoef )
}

```

```

fake_means_u <- ( 1 / ( ( fake_delta^2 / fake_sigma^2 ) + 1 ) ) * (
  fake_delta / fake_sigma^2 ) * fake_residuals
fake_var_u <- (1) / ( (fake_delta^2 / fake_sigma^2 ) + 1)
fake_U <- dnorm(fake_residuals, mean = fake_means_u, sd = fake_var_u )
fake_mTE <- mean ( exp ( -fake_U ) )

start.time <- Sys.time()

#Estimations
if ( estim == "skewn_sem" ) {
  #SEM, log likelihood function (deGraaf, 2019)

  SpatialFrontierErrorFun <- function(pars, X, Y, W){
    p <- length(pars)
    lambda <- pars[p]
    alpha <- pars[(p-1)]
    sigma <- pars[(p-2)]
    beta <- pars[(1:(p-3))]

    sigma <- sqrt(sigma^2)
    lambda <- 2*(exp(lambda)/(1+exp(lambda)))-1
    nObs <- dim(X)[1]
    B <- (diag(nObs) - lambda * W)
    Xb <- X%*%beta
    alpha <- -sqrt(alpha^2)
    delta <- alpha/(sqrt(1+alpha^2))
    z <- (1/sigma)*(B%*(Y - Xb))
    term1 <- -(nObs/2)*log(pi) -(nObs/2)*log(sigma^2) + log(det(B)) -
      (1/2)*(t(z)%*%z)
    term2 <- colSums(log(2*pnorm(alpha*z)))

    return(term1 + term2)
  }

  # Starting values (spdep package)
  q <- dim(X)[2]
  erlag <- errorsarlm (y ~ X[,2] , listw = Wmatlw)
  beta0 <- erlag$coefficients[1:(q)]
  delta0 <- -sqrt(erlag$s2*(1-2/pi))
  alpha0 <- delta0/(sqrt(1 - delta0^2))
  sigma0 <- sqrt(erlag$s2)
  splag0 <- erlag$lambda[1]
  startv <- as.vector (c(beta0, sigma0, alpha0, splag0))
  estim_max <- maxLik ( SpatialFrontierErrorFun,
    X=X, Y=y, W=Wmatws,
    start = startv,
    method = method, control= control)
} else if ( estim == "skewn_sar" ) {
  #SAR, log likelihood function (deGraaf, 2019)

  SpatialFrontierLagFun <- function(pars, Y, X, W){

```

```

p <- length(pars)
rho <- pars[p]
alpha <- pars[(p-1)]
sigma <- pars[(p-2)]
beta <- pars[(1:(p-3))]
sigma <- sqrt(sigma^2)
nObs <- length(Y)
A <- (diag(nObs) - rho * W)
Xb <- X%%beta
alpha <- -sqrt(alpha^2)
delta <- alpha/(sqrt(1+alpha^2))
z <- (1/sigma)*(A%%Y - Xb)

term1 <- -(nObs/2)*log(pi) -(nObs/2)*log(sigma^2) +
  log(det(A)) - (1/2)*(t(z)%%z)
term2 <- colSums(log(2*pnorm(alpha*z)))

return(term1 + term2)
}

## Spatial SAR Degraaf
q <- dim(X)[2]
sp_lag_a <- lagsarlm (y ~ X[,2], listw = Wmatlw)
beta0 <- sp_lag_a$coefficients[1:(q)]
delta0 <- -sqrt(sp_lag_a$s2*(1-2/pi))
alpha0 <- delta0/(sqrt(1 - delta0^2))
sigma0 <- sqrt(sp_lag_a$s2)
rho0 <- sp_lag_a$rho[1]
startv <- c(beta0, sigma0, alpha0, rho0)
estim_max <- maxLik ( SpatialFrontierLagFun,
  X=X, Y=y, W=Wmatws,
  start = startv,
  method = method, control= control)

} else {
  estim_max <- NA }

beta_hat <- estim_max$estimate[1:q]
sigma_hat <- estim_max$estimate[q+1]
alpha_hat <- estim_max$estimate[q+2]
splug_hat <- estim_max$estimate[q+3]
Loglik_hat <- estim_max$maximum

# Estimating specific inefficiency

delta_hat <- alpha_hat / ( sqrt ( 1 + alpha_hat^2 ) )
I <- diag ( nobs )
Bmat_hat <- I - splag_hat*Wmatws

if (estim == "skewn_sem") {
  residuals <- ( 1 / sigma_hat ) * Bmat_hat %% ( y - X %% beta_hat )
} else if ( estim == "skewn_sar" ) {
  residuals <- ( 1 / sigma_hat ) * (Bmat_hat %% y - X %% beta_hat )

```

```

} else {
  residuals <- ( 1 / sigma_hat ) * (y - X %*% beta_hat )
}

means_u <- ( 1 / ( ( delta_hat^2 / sigma_hat^2 ) + 1 ) ) * (
  delta_hat / sigma_hat^2 ) * residuals
var_u <- (1) / ( (delta_hat^2 / sigma_hat^2 ) + 1)
U_hat <- dnorm(residuals, mean = means_u, sd = var_u )
mTE_hat <- mean ( exp ( -U_hat ) )

# Return of results
return ( list (nobs=nobs, Wmatr=Wmatr,
  minX=minX, maxX=maxX, dgp=dgp,
  fake_rho=fake_rho, fake_lambda=fake_lambda,
  fake_splag=fake_splag, bcoef=bcoef,
  fake_sdv=fake_sdv, fake_sdu=fake_sdu,
  fake_sigma = fake_sigma,
  fake_delta = fake_delta, fake_alpha = fake_alpha,
  fake_Loglik = fake_Loglik,
  fake_U = fake_U, fake_mTE = fake_mTE,
  Loglik_hat= Loglik_hat,
  approach=approach, estim=estim,
  beta_hat=beta_hat, alpha_hat=alpha_hat,
  sigma_hat=sigma_hat,
  splag_hat=splag_hat, delta_hat=delta_hat,
  Bmat_hat=Bmat_hat, residuals=residuals,
  means_u=means_u,
  var_u=var_u, U_hat=U_hat, mTE_hat=mTE_hat ) )
}

for (k in 1:repetitions){

  # Run the function
  Results <- func_dgp_est ( repetitions = NULL, seed = NULL,
    nobs = NULL, Wmatr = NULL,
    swdim = NULL, swneig = NULL, swp = NULL,
    geolat = NULL, geolon = NULL,
    minX = NULL, maxX = NULL, dgp = NULL, bcoef = NULL,
    fake_rho = NULL, fake_lambda = NULL,
    fake_sdv = NULL, fake_sdu = NULL, fake_delta = NULL,
    approach = NULL, estim = NULL, efficiency = NULL,
    method = NULL, control=NULL)

  # Results
  Nobs[k] <<- (Results$nobs)
  Wmatrice[k] <<- (Results$Wmat)
  minXvar[k] <<- (Results$minX)
  maxXvar[k] <<- (Results$maxX)
  dataGP[k] <<- (Results$dgp)
  rho_true[k] <<- (Results$fake_rho)
  lambda_true[k] <<- (Results$fake_lambda)
  splag_true[k] <<- (Results$fake_splag)

```

```

beta_true[k,] <- (Results$bcoef)
sdv_true[k] <- (Results$fake_sdv)
sdu_true[k] <- (Results$fake_sdu)
alpha_true[k] <- (Results$fake_alpha)
sigma_true[k] <- (Results$fake_sigma)
delta_true[k] <- (Results$fake_delta)
Loglik_true[k] <- (Results$fake_Loglik)
U_true[k,] <- (Results$fake_U)
mTE_true[k] <- (Results$fake_mTE)

dtribution[k] <- (Results$approach)
estimation[k] <- (Results$estim)

beta_est[k,] <- (Results$beta_hat)
alpha_est[k] <- (Results$alpha_hat)
sigma_est[k] <- (Results$sigma_hat)
splag_est[k] <- (Results$splag_hat)
delta_est[k] <- (Results$delta_hat)
U_est[k,] <- (Results$U_hat)
mTE_est[k] <- (Results$mTE_hat)
Loglik_est[k] <- (Results$Loglik_hat)

end.time <- Sys.time()
time.taken <- difftime (end.time, start.time, units = "secs")
timespent[k] <- time.taken
print (sprintf ("Simulation %s of %s. Processing took %f seconds.",
                k, repetitions, time.taken ) )
}

simul_results <- list ( Nobs = Nobs, Wmatrice = Wmatrice,
                      minXvar = minXvar, maxXvar = maxXvar,
                      dataGP = dataGP, rho_true = rho_true,
                      lambda_true = lambda_true, splag_true = splag_true,
                      beta_true = beta_true, sdv_true = sdv_true,
                      sigma_true = sigma_true,
                      sdu_true = sdu_true, alpha_true=alpha_true,
                      delta_true = delta_true, Loglik_true = Loglik_true,
                      U_true = U_true, mTE_true = mTE_true,
                      Loglik_est = Loglik_est,
                      ditribution = ditribution,
                      estimation = estimation, beta_est = beta_est,
                      alpha_est = alpha_est, sigma_est = sigma_est,
                      splag_est = splag_est, delta_est = delta_est,
                      U_est = U_est, mTE_est = mTE_est,
                      Loglik_est = Loglik_est, repetitions = repetitions,
                      timespent = timespent )

return ( simul_results )
#saveRDS (simul_results, "simul_results.rds" )
}

```



```

restable_fun <- function (replications=NULL ) {

  # True values
  NobsF <- replications$Nobs[1]
  WmatriceF <- replications$Wmatrice[1]
  dataGPF <- dataGP[1]
  beta_trueF <- replications$beta_true[1,]
  rho_trueF <- replications$rho_true[1]
  lambda_trueF <- replications$lambda_true[1]
  sigma_trueF <- replications$sigma_true[1]
  alpha_trueF <- replications$alpha_true[1]
  delta_trueF <- replications$delta_true[1]
  ditributionF <- replications$ditribution[1]
  estimationF <- replications$estimation[1]
  mTE_true_F <- colMeans ( replications$mTE_true)
  Loglik_trueF <- colMeans (replications$Loglik_true)

  # Means of parameters
  beta_mean <<- colMeans ( replications$beta_est )
  alpha_mean <<- colMeans ( replications$alpha_est )
  sigma_mean <<- colMeans ( replications$sigma_est )
  splag_mean <<- colMeans ( replications$splag_est )
  delta_mean <<- colMeans ( replications$delta_est )
  mTE_est_mean <<- colMeans ( replications$mTE_est )
  Loglik_est_mean <<- colMeans (replications$Loglik_est)

  # standard deviations of parameters
  beta_sd <- GCMC:::colSds(replications$beta_est )
  alpha_sd <<- sd ( replications$alpha_est )
  sigma_sd <<- sd ( replications$sigma_est )
  splag_sd <<- sd ( replications$splag_est )
  delta_sd <<- sd ( replications$delta_est )
  mTE_est_sd <<- sd ( replications$mTE_est )
  Loglik_est_sd <<- sd (replications$Loglik_est)

  beta_dif <<- replications$beta_est - replications$beta_true
  alpha_dif <<- replications$alpha_est - replications$alpha_true
  sigma_dif <<- replications$sigma_est - replications$sigma_true
  splag_dif <<- replications$splag_est - replications$splag_true
  delta_dif <<- replications$delta_est - replications$delta_true
  U_dif <<- replications$U_est - replications$U_true
  mTE_dif <<- replications$mTE_est - replications$mTE_true
  Loglik_dif <<- replications$Loglik_est - replications$Loglik_true

  # Biases
  biais <- function (nrepliq, pardif ) {
    ( 1 / nrepliq ) * sum ( pardif ) }

  # Root Mean square error mse
  rmse <- function (nrepliq, pardif ) {
    sqrt ( ( 1 / nrepliq ) * sum ( ( pardif )^2 )) }
}

```

```

beta_biais <- c(biais ( replications$repetitions, beta_dif[,1]),
               biais ( replications$repetitions, beta_dif[,2]))
alpha_biais <- biais ( replications$repetitions, alpha_dif)
sigma_biais <- biais ( replications$repetitions, sigma_dif)
splug_biais <- biais ( replications$repetitions, splag_dif)
delta_biais <- biais ( replications$repetitions, delta_dif)
Loglik_biais <- biais ( replications$repetitions, Loglik_dif)
mTE_biais <- biais ( replications$repetitions, mTE_dif)

beta_rmse <- c(rmse ( replications$repetitions, beta_dif[,1]),
              rmse ( replications$repetitions, beta_dif[,2]))
alpha_rmse <- rmse ( replications$repetitions, alpha_dif)
sigma_rmse <- rmse ( replications$repetitions, sigma_dif)
splug_rmse <- rmse ( replications$repetitions, splag_dif)
delta_rmse <- rmse ( replications$repetitions, delta_dif)
Loglik_rmse <- rmse ( replications$repetitions, Loglik_dif)
mTE_rmse <- rmse ( replications$repetitions, mTE_dif)

Infos <- c(ditributionF, estimationF,
          dataGPF, WmatriceF, NobsF)
names(Infos) <- c("ditribution", "estimation",
                 "dataGP", "Wmatrice", "Nobs")
pars_true <- c ( beta_trueF, sigma_trueF, alpha_trueF,
                delta_trueF, lambda_trueF,
                mTE_true_F, Loglik_trueF)
pars_est_mean <- c ( beta_mean, sigma_mean, alpha_mean,
                   delta_mean, splag_mean,
                   mTE_est_mean, Loglik_est_mean)
pars_est_sd <- c ( beta_sd, sigma_sd, alpha_sd,
                  delta_sd, splag_sd,
                  mTE_est_sd, Loglik_est_sd)
pars_biais <- c ( beta_biais, sigma_biais, alpha_biais,
                 delta_biais, splag_biais,
                 mTE_biais, Loglik_biais)
pars_rmse <- c ( beta_rmse, sigma_rmse, alpha_rmse,
                delta_rmse, splag_rmse,
                mTE_rmse, Loglik_rmse)

result_table <- matrix ( c ( pars_true, pars_est_mean, pars_est_sd,
                             pars_biais, pars_rmse), ncol=5)
result_table <- round (result_table, 3)
colnames (result_table) <- c('True', 'Estimates M', 'Estimates SD',
                             "Bias", "RMSE")
rownames(result_table) <- c("beta_1", "beta_2", "sigma", "alpha",
                             "delta", "splag",
                             "TE", "Loglik")

all_result <- list (replications = replications,
                   result_table = result_table,
                   Infos = Infos)

return(all_result)

```

```

}

appliesim <- function(param, simul_spsfa, restable_fun) {

  repetitions <- param$repetitions
  seed <- param$seed
  nobs <- param$nobs
  Wmatr <- param$Wmatr
  swdim <- param$swdim
  swneig <- param$swneig
  swp <- param$swp
  geolat <- param$geolat
  geolon <- param$geolon
  minX <- param$minX
  maxX <- param$maxX
  dgp <- param$dgp
  bcoef <- param$bcoef
  fake_rho <- param$fake_rho
  fake_lambda <- param$fake_lambda
  fake_sdv <- param$fake_sdv
  fake_sdu <- param$fake_sdu
  fake_delta <- param$fake_delta
  approach <- param$approach
  estim <- param$estim
  efficiency <- param$efficiency
  method <- param$method
  control <- param$control

  rep_result <- simul_spsfa ( repetitions = repetitions, seed = seed,
                             nobs = nobs, Wmatr = Wmatr,
                             swdim = swdim, swneig = swneig, swp =swp,
                             geolat = geolat, geolon = geolon,
                             minX = minX, maxX = maxX, dgp = dgp,
                             bcoef = bcoef,
                             fake_rho = fake_rho, fake_lambda = fake_lambda,
                             fake_sdv = fake_sdv, fake_sdu = fake_sdu,
                             fake_delta = fake_delta,
                             approach = approach, estim = estim,
                             efficiency = efficiency,
                             method = method, control=control)

  finalres <- restable_fun (replications=rep_result)
  return(finalres)
}

# Running the simulations

# Parameters setting
param_base <- list( repetitions = 10, seed = 2000,
                    swdim = 1, swneig = 6, swp =0.1,
                    geolat = 20, geolon = 20,
                    minX = 5, maxX = 14,
                    bcoef = c(1, 1),
                    fake_sdv = 0.3, fake_sdu = 0.3,

```

```

        efficiency = NULL,
        method = "BFGS", control=NULL)

# Parameters for nobs=250, Wmatr = "geo", dgp = "sem", ...
param_250_geo_02_sem_08 <- c(param_base,
                             list(nobs = 250, Wmatr = "geo", dgp = "sem",
                                   fake_rho = 0.2, fake_lambda = 0.2,
                                   fake_delta = -0.8,
                                   approach = "skewn", estim = "skewn_sem",
                                   name="resim_250_geo_02_sem_08"))

# Simulations for repetitions=10, nobs=250, Wmatr = "geo", dgp = "sem", ....
resim_250_geo_05_sem_08 <- applysim (param = param_250_geo_02_sem_08,
                                     simul_spsfa = simul_spsfa,
                                     restable_fun = restable_fun)

## [1] "Simulation 1 of 10. Processing took 2.392616 seconds."
## [1] "Simulation 2 of 10. Processing took 3.653912 seconds."
## [1] "Simulation 3 of 10. Processing took 1.799989 seconds."
## [1] "Simulation 4 of 10. Processing took 2.280699 seconds."
## [1] "Simulation 5 of 10. Processing took 1.758994 seconds."
## [1] "Simulation 6 of 10. Processing took 1.799990 seconds."
## [1] "Simulation 7 of 10. Processing took 2.116800 seconds."
## [1] "Simulation 8 of 10. Processing took 1.772988 seconds."
## [1] "Simulation 9 of 10. Processing took 1.804969 seconds."
## [1] "Simulation 10 of 10. Processing took 1.976870 seconds."

```