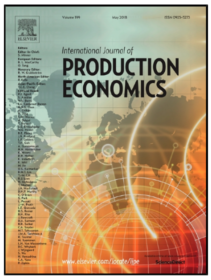
Accepted Manuscript



Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms

Gudbrand Lien, Subal C. Kumbhakar, Habtamu Alem

PII:	S0925-5273(18)30185-3
DOI:	10.1016/j.ijpe.2018.04.023
Reference:	PROECO 7023
To appear in:	*International Journal of Production Economics*
Received Date:	10 January 2018
Revised Date:	23 April 2018
Accepted Date:	26 April 2018

Please cite this article as: Gudbrand Lien, Subal C. Kumbhakar, Habtamu Alem, Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms, *International Journal of Production Economics* (2018), doi: 10.1016/j.ijpe.2018.04.023

Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms

Gudbrand Lien
Norwegian Institute of Bioeconomy Research, Oslo Norway and Inland Norway University College, Lillehammer, Norway, e-mail: Gudbrand.Lien@inn.no

Subal C. Kumbhakar*
Department of Economics, State University of New York, Binghamton, NY, USA, and Stavanger Business School, Stavanger, Norway, Phone +1 607 777 4762, E-mail: kkar@binghamton.edu

Habtamu Alem
Norwegian Institute of Bioeconomy Research, Oslo Norway, E-mail: Habtamu.Alem@nibio.no

Abstract

This paper addresses the endogeneity of inputs and output (which is mostly ignored in the stochastic frontier (SF) literature) in the SF panel data model under the behavioral assumption that firms maximize returns to the outlay. We consider a four component SF panel data model in which the four components are: firms' latent heterogeneity, persistent inefficiency, transient inefficiency and random shocks. Second, we include determinants in transient inefficiency. Finally, to avoid the impact of distributional assumptions in estimating the technology parameters, we apply a multi-step estimation strategy to an unbalanced panel dataset from Norwegian crop-producing farms observed from 1993 to 2014. Distributional assumptions are made in second and third steps to predict both persistent and transient inefficiency, and their marginal effects.

Keywords: Efficiency, endogeneity, returns to the outlay, panel data

JEL Classification No.: D21, D22, Q12

*Corresponding author

Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms

Abstract

This paper addresses the endogeneity of inputs and output (which is mostly ignored in the stochastic frontier (SF) literature) in the SF panel data model under the behavioral assumption that firms maximize returns to the outlay. We consider a four component SF panel data model in which the four components are: firms' latent heterogeneity, persistent inefficiency, transient inefficiency and random shocks. Second, we include determinants in transient inefficiency. Finally, to avoid the impact of distributional assumptions in estimating the technology parameters, we apply a multi-step estimation strategy to an unbalanced panel dataset from Norwegian crop-producing farms observed from 1993 to 2014. Distributional assumptions are made in second and third steps to predict both persistent and transient inefficiency, and their marginal effects.

Keywords: Efficiency, endogeneity, returns to the outlay, panel data

JEL Classification No.: D21, D22, Q12

Introduction

A common problem in macro- and microeconomic models is that one or more of the explanatory variables is jointly determined with the dependent variable. This is especially true in production models where inputs and outputs are jointly (simultaneously) determined. The simultaneity issue causes an endogeneity problem in econometric estimation (the input variables being correlated with the error term in the production function). In such a case the OLS regression gives inconsistent estimators of the technology parameters because consistency of the standard OLS estimator requires that the explanatory variables are either exogenously given or uncorrelated with the error term (e.g., Nerlove, 1965; Griliches and Mairesse, 1995). In practice one or more of the explanatory variables are determined jointly with the dependent variable because we typically use data that reflect the behaviour of producers, who normally seek to maximize or minimize some objective function. Hence, for a given technology, the levels of inputs and outputs are determined by a producer such that the value of the objective function is optimized. Thus, the inputs are neither randomly decided nor exogenously given. And unless one takes this into account in estimation, the results are likely to be inconsistent.

In farm-level productivity related studies of agriculture the endogeneity problem is more the rule rather than the exception. Typically, variables are generated from farm accounting data but the production data used in these studies reflect the optimizing behaviour of the farmers. In other words, the data are generated from the technology along with the behaviour of the farmers. A farmer typically choses the level of some or many of the inputs and outputs with an objective in mind, which makes inputs and output choice variables economically endogenous. The economic endogeneity in all most all the cases leads to econometric endogeneity in which the choice variables are correlated with the composite error term in the production function. Since our goal is to estimate the production technology consistently, we need to deal with econometric endogeneity problem first.

To deal with the endogeneity problem in production function estimation, one can use either the traditional instrumental variable method or the maximum likelihood (ML) method to the equation of interest in addition to unrestricted reduced form equations of endogeneous variables (e.g., Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Shee and Stefanou, 2015; Amsler et al., 2016). Kutlu (2010) and Tran and Tsionas (2013) addressed the endogeneity problem in the Battese and Coelli (1992) formulation by decomposing the error term into two parts, one correlated with the explanatory variables and the other not. Following Levinsohn and Petrin (2003), Shee and Stefanou (2015) introduced a model that has a production shock term which is endogenous to the inputs. Technical inefficiency and noise are assumed to be uncorrelated to the inputs. The energy input in their study is assumed to be endogenous and is used as a proxy for production shock. To address the endogeneity issue in our study we follow Kumbhakar (2011) and Kumbhakar et al. (2013), and assume that producers maximize return to the outlay (RO). We use the first-order conditions of RO maximization to derive an estimating equation that does not suffer from the endogeneity problem. Our approach, based on behavioural assumptions, allows inputs and output to be correlated with inefficiency as well as noise. More specifically, each and every error component is allowed to be correlated with the endogenous inputs and output.[1]

The production function estimation model used in this study is an extension of the recently introduced stochastic frontier (SF) panel data model[2] (Colombi et al. (2014), Kumbhakar et al. (2014) and Tsionas and Kumbhakar (2014)).[3] In these models the error term is split into four components. The model includes one component that captures the latent heterogeneity of farms, one that captures transient (time-varying) inefficiency, one that captures persistent inefficiency and a fourth component that captures random shocks. All the four components are assumed to be random. Tsionas and Kumbhakar (2014) named the four components model the generalized true random-effects model (GTRE), since it captures the persistent inefficiency in addition to the latent heterogeneity of a farm, and transient inefficiency (as in the true random effect model (TRE) by Greene (2005a, b)). Filippini and Greene (2016) showed how such a model can be estimated using a simulated maximum likelihood method. The GTRE was further extended to accommodate determinants of inefficiency in Badunenko and Kumbhakar (2017), Lai and Kumbhakar (2016), and as done in this study.

While there are many SF models that deal with modelling/examining the impact of exogenous variables on the level of inefficiency (e.g., Kumbhakar et al., 1991; Reifschneider and Stevenson, 1991; Caudill and Ford, 1993; Huang and Liu, 1994; Battese and Coelli, 1995; Hadri, 1999; Wang, 2002; Lai and Kumbhakar, 2016; Badunenko and Kumbhakar, 2017), the four-component model described above does not include determinants of either transient or persistent inefficiency, with the exception of Badunenko and Kumbhakar (2017), Lai and Kumbhakar (2016) and in this study. Without determinants (when inefficiency is random and independently and identically distributed) one cannot explain systematic differences in inefficiency within and between farms, especially when the

[1] Färe et al. (2002) and Zofío and Prieto (2006) use maximization of return to the dollar using the distance function formulation. Since their model is estimated using a deterministic linear programming approach, the endogeneity issue did not arise. This is because enodogeneity issue is typical in econometric models, especially when economic behaviors are believed to affect the regressors (quantities of inputs and/or outputs in our case).
[2] Reviews of SF models can be found in, e.g., Kumbhakar and Lovell (2000), Coelli et al. (2005), Greene (2008), Parmeter and Kumbhakar (2014) and Kumbhakar et al. (2015).
[3] The model was proposed in working papers independently by Colombi et al. (2011) and by Kumbhakar et al. (2011).

inefficiency components are assumed to independently and identically distributed random variables. Three groups of variables have generally been studied as inefficiency determinants, namely: characteristics of farm and technology employed; location and environmental variables characterizing the conditions for farming; and human capital variables (Latruffe et al., 2004). Farm size, soil quality, organizational and ownership structure, technology choice (e.g., conventional vs. organic), farm location, physical infrastructure (e.g., levels of modernization), the farmer’s age and education, part-time vs. part-time farming and policy intervention's effect on efficiency levels are examples of aspects investigated in the literature. Some of these or other relevant determinants will have a long-run effect on efficiency (persistent) while other are more likely to be time-variant and short term (transient). Better understanding of the determinants of transient and persistent inefficiency components can be useful. For example, establishing determinants of persistent inefficiency could help decision makers to develop strategies to remove long-term impediments, such as too rigid regulations or other structural rigidities. On the other hand, transient inefficiency can be due to bad luck, management mistakes, etc., that can get corrected. Knowledge about these drivers of transient inefficiency may help in improving the efficiency of individual farms in the short run.

The two main contributions of this article are the following: We address endogeneity of inputs and output using the behavioural assumption that farmers seek to maximise RO; and we present an extension of the above-mentioned four-component SF model that includes determinants of transient inefficiency using an extension of multi-step estimation procedure used in Kumbhakar et al. (2014). We used the model on an unbalanced panel dataset from Norwegian crop-producing farmers observed from 1993 to 2014. Although Badunenko and Kumbhakar (2017) and Lai and Kumbhakar (2016) have shown how to estimate the heteroscedastic four component model in a single step, we use a multi-step procedure so that our estimates of the technology parameters are not contaminated by distributional assumptions which are central to the single-step procedure.

The rest of the article is organised as follows. We first briefly describe some characteristics of Norwegian crop farming, then we outline how we deal with the endogeneity problem, followed by a brief overview of the state-of-the-art efficiency models, especially as applied to panel data. Then, we introduce the four-component SF model with determinants of transient inefficiency and explain how we deal with the endogeneity problem. Further, we briefly describe the data that is used in the model, followed by a presentation of empirical results. Finally, we draw some conclusions.

Norwegian crop farming

Farmers in Norway face unfavourable production environments such as harsh climate, extensive areas of rugged terrain, and short growing seasons. These contribute to the high costs of production. High production costs make it difficult for Norwegian farmers to compete in an open market. As a result, the Norwegian government has sought to offset the disadvantages faced by farmers by large subsidies (OECD, 2009), plus import and other regulations, which in turn have enabled many smaller farms to survive. So, even though there has been some structural development in Norwegian crop farming[4], the above-mentioned features may have influenced both the scale, productivity and efficiency aspects of production.

Even if the government policy with high level subsidies has kept farm incomes at high levels, the value or support payments has been decreasing in real terms, and farm income represents a relatively small and decreasing share of the total income of Norwegian farm-family households. In 2013, only 19% of the average total household income for farmers and their partners came from agriculture, forestry and fishing, down from 35% in 1999 (NIBIO, 2015).

As in many other countries, the objectives of agricultural policy in Norway are diverse and complex. However, important policy goals include encouraging production to contribute towards a high level of national food security and supporting farm incomes to enable farm families to enjoy standards of living in line with the rest of the population. In pursuit of the latter objective, governments have generally sought to encourage efficient farm resource use (NILF, 2014).

An overview of the model

In this section we first introduce the model to deal with the endogeneity of inputs in estimating the production function. Then we explain how that model is adapted to incorporate the latest advances in panel data efficiency models. Finally, we briefly describe different ways to introduce efficiency determinants into these models.

The model under return to the outlay maximization

Instead of addressing the endogeneity problem following Shee and Stefanou (2015) who added inefficiency to the model of Levinsohn and Petrin (2003), we follow a standard neoclassical approach in which economic endogeneity is formally modeled. Note that Shee and Stefanou (2015) assumed the productivity term to be correlated with inputs but inefficiency and the noise terms are assumed to be

[4] In 2014, the average crop farm size was 40.6 ha, up from 27.7 ha in 1997 (NILF, 2014).

independent of the inputs. Our concern in stochastic frontier models is correlation between inputs and inefficiency. To address this concern, we assume that the producers' objective is maximization of return to outlay (RO), defined as total revenue divided by total cost, R/C. In doing so we allow both inputs and output to be freely correlated with inefficiency and the noise terms. Furthermore, we extend the model to introduce persistent inefficiency. With a single output y, $R = py$ and $C = w'x$, where p is output price, w is the input price vector and x is the vector of variable inputs.[5]

The producers are assumed to maximize $py/w'x$ with respect to y and x subject to the production function $y = f(x,t)A$, where $A = e^{v-u}$, v is stochastic noise, u is inefficiency and t is time. This specification of A fits into the model for both cross-sectional and some panel data models. The Lagrangean is $L = py/w'x + \mu\left(y - f(x,t)A\right)$ and the first-order conditions (FOCs) of maximization with respect to y and x_j are:

$$p/C + \mu = 0 \Rightarrow \mu = -p/C;$$

and

$$-\left(R/C^2\right)w_j - \mu A\left(\partial f/\partial x_j\right) = 0 \Rightarrow -(R/C)w_j x_j + R(d\ln y/d\ln x_j) = 0.$$

The last equality upon addition over j (j = 1,…,J) implies $\sum_j d\ln y/d\ln x_j = 1$. This means that the production function f(.) is homogeneous of degree 1 in inputs. Note that linear homogeneity is not an assumption. It follows from the FOCs of maximization of RO.

Assuming that the underlying production is Cobb-Douglas (for simplicity), the linear homogeneity results above can be used to write the production function as

$$\ln\left(\frac{y}{x_1}\right)_{it} = \beta_0 + \sum_{j=2}\beta_j \ln\left(\frac{x_j}{x_1}\right)_{it} + \ln A_{it}$$

which can be rewritten as

$$\ln \tilde{y}_{it} = \beta_0 + \sum\nolimits_{j=2} \beta_j \ln \tilde{x}_{jit} + \ln A_{it} \tag{1}$$

where $\tilde{y} = y/x_1$ and $\tilde{x}_j = x_j/x_1, j = 2,...,J$. Finally, $\ln A$ is the error term and we assume it to have four components (see below).

[5] We can easily include quasi-fixed inputs (that is the inputs that are considered fixed, not a choice variable, in the short run) in the model. For example, different farmers are likely to have different years of farming experience but it is not a choice variable. In such a case the decision variables will be the output and variable inputs with the quasi-fixed inputs added in as additional regressors in the estimating function in the same way as the time trend variable. Everything goes through even if cost is defined as cost of quasi-fixed and variable inputs.

Irrespective of how we specify the error term, the specification of the production function in the form of (1) solves the endogeneity problem. This can be shown using the FOCs above, viz., $-(R/C)w_j x_j + R(\partial \ln y / \partial \ln x_j) = 0 \Rightarrow w_j x_j / w_1 x_1 = (\partial \ln y / \partial \ln x_j)/(\partial \ln y / \partial \ln x_1) = \beta_j / \beta_1$ where the last equality comes from the use of CD production function for which $\partial \ln y / \partial \ln x_j = \beta_j$. From the above relationship, it follows that the input ratios, viz., $\tilde{x}_j \equiv x_j / x_1 = (w_1 / w_j)(\beta_j / \beta_1)$ are independent of ln A. Thus the regressors in (1) are independent of the error terms in the production function, which solves the endogeneity problem. In other words, the use of (1) as the estimating equation instead of the production function solves the endogeneity problem. The solution is that, although x_j are correlated with the error term in the production function ($\ln A$), the input ratios $(\tilde{x}_j)$ are independent of $\ln A$, because all the inputs are affected by A by the same way. This holds for the translog production function as well. Note that we allow every component in $\ln A$ (formally introduced in the next section) to be correlated with all the variable inputs.

Panel data efficiency models

The state-of-the-art stochastic production frontier model for panel data, as introduced simultaneously by Colombi et al. (2014), Kumbhakar et al. (2014) and Tsionas and Kumbhakar (2014), can now be built on equation (1), resulting in:

$$\ln \tilde{y}_{it} = \beta_0 + \sum\nolimits_{j=2} \beta_j \ln \tilde{x}_{it} + \beta_t t + b_i - \eta_i + v_{it} - u_{it} \tag{2}$$

We added the time trend variable as an additional regressor in (2) to accommodate technical change (shift in the production function). The error term ($\ln A_{it}$ in Eq. (1)) in this model is composed of four components. The first component (b_i) captures latent heterogeneity of the firms, the second component (η_i) captures persistent inefficiency, the third component (v_{it}) captures random shocks and the last component (u_{it}) captures transient inefficiency. It was shown in Kumbhakar et al. (2014) that all the four components can be identified with appropriate distributional assumptions on them.

Many interesting SF panel models can be obtained as special cases of the model specified in Eq. (2) by excluding one or more error components. For example, the 'true' random-effects model by Greene (2005a, b) is obtained by dropping the η_i term from (2). Similarly, the short- and long-run inefficiency model by Kumbhakar and Heshmati (1995) and Kumbhakar and Hjalmarsson (1993) can be obtained by dropping the b_i component in (2). Finally, time-invariant inefficiency models of Pitt and Lee (1981), Schmidt and Sickles (1984), Kumbhakar (1987) and Battese and Coelli (1988) are obtained by dropping both the b_i and u_{it} terms in (2).

The four-component SF (the GTRE) model in (2) improves upon the models described earlier in several ways. First, although some of the time-varying inefficiency models presented above can accommodate firm effects (Greene, 2005a, b), these models fail to take into account the possible presence of some unobserved factors that might have persistent effects on a firm's inefficiency. Second, SF models often imply that the inefficiency of a firm at time t is independent of its previous level of inefficiency (distributed as iid). This formulation is too restrictive. It is more sensible to assume that inefficiency is correlated over time and there are some determinants that can explain this dependence. Thus a firm may eliminate part of its inefficiency by removing some of the short-run rigidities, while some other sources of inefficiency might stay with the firm over time. In that case, overall inefficiency $(\eta_i + u_{it})$ in (2) will be correlated over time for each firm. Finally, the SF panel models that include permanent/time-invariant inefficiency effects (Kumbhakar and Heshmati, 1995; Kumbhakar and Hjalmarsson, 1993) confound permanent/time-invariant inefficiency with firm effects (heterogeneity) since they do not take into account the effect of unobserved firm heterogeneity on output.

Estimation of the GTRE model can be done in several ways. One is the single-step full maximum likelihood procedure first proposed in Colombi et al. (2014), and extended in Badunenko and Kumbhakar (2017) and Lai and Kumbhakar (2016) to accommodate heteroscedasticity in some or all the error components. All the parameters of the model are estimated simultaneously in a single-step full maximum likelihood (ML) method. That approach is cumbersome to implement in practice (Filippini and Greene, 2016). Kumbhakar et al. (2014) proposed estimating the four-component model using a multi-step procedure. However, that approach is not as efficient as the one-step ML method. Filippini and Greene (2016) used a simulation-based one-step maximum likelihood estimator that circumvented most of the challenges associated with the classical full information ML procedure used by Colombi et al. (2014). Now we introduce the model by Kumbhakar et al. (2014) in which inefficiency components are not necessarily iid but their variances are functions of determinants.

Models with determinants of inefficiency

In addition to estimating inefficiency for each farm over time, it is of interest to learn about factors that affect inefficiency between and within farms. The GTRE models can be extended to discern the impact of exogenous determinants of the level of inefficiency for a given farm over time. In our model we assume that there is a vector of exogenous variables/determinants z_{it} that influence u_{it} in equation (2) above.

The first comprehensive models to investigate exogenous determinants of inefficiency were introduced by Kumbhakar et al. (1991) and Reifschneider and Stevenson (1991). In this cross-sectional model η_i is distributed as $\eta_i(z_i) \sim N^+(z_i'\delta, \sigma_\eta^2)$ so that the model can be used to parameterizes the pre-truncation mean of the truncated normal distribution for η_i.

Instead of parameterizing the mean of the truncated normal distribution, Caudill and Ford (1993), Caudill et al. (1995) and Hadri (1999) parametrized the variance of the half-normal distribution, i.e., $\eta_i(z_i) \sim N^+(0, \exp(z_0 + z_i'\omega))$. Wang (2002) proposed a further generalization in which both the mean and variance of η_i are functions of z_i variables. The models of Huang and Liu (1994) and Battese and Coelli (1995), in which variances are assumed to be constant, are special cases of the Wang (2002) model.

All the models with determinants of inefficiency mentioned above can be extended to panel data settings, especially in the context of the four component models (see Badunenko and Kumbhakar, 2017; Lai and Kumbhakar, 2016).

A four-component GTRE model with determinants of inefficiency

The extensions introduced here are to allow for determinants of inefficiency in the transient inefficiency component while dealing with the problem of endogeneity of inputs. The four-component model in equation (2) extended to accommodate determinants for transient inefficiency[6] is:

$$\ln \tilde{y}_{it} = \beta_0 + \sum\nolimits_{j=2} \beta_j \ln \tilde{x}_{it} + \beta_t t + b_i - \eta_i + v_{it} - u_{it}(z_{it}) \tag{3}$$

We propose estimating the model in multiple steps. The main advantage of using multiple steps is that the production function parameters, β (estimated in step 1) are not affected by the distributional assumptions on the error components. Furthermore, the estimation is simpler than the single-step procedure. The three-step procedure we use is as follows:

[6] Since there is no natural time-invariant variable in our data that can be used as determinants of persistent inefficiency in our application, we assume persistent inefficiency to be iid. The model we propose here can be extended to include determinants of persistent inefficiency. See Badunenko and Kumbhakar (2017) and Lai and Kumbhakar (2016) for determinants of both persistent and transient inefficiency model that is estimated in a single-step.

Step 1: The firm-effects b_i can be either random or fixed. Here we assume them to be random and iid with mean zero. The random production shocks v_{it} are also assumed to have zero mean and constant variance. The persistent inefficiency, η_i, is assumed to be iid with $E(\eta_i)=a$ which is a constant. This can be easily generalized to have a non-constant mean (see Badunenko and Kumbhakar, 2017; Lai and Kumbhakar, 2016). The mean of the transient inefficiency, $u_{it}(z_{it})\geq 0$ is $E(u_{it}(z_{it}))=g(z_{it})\geq 0$, where the z variables are determinants of transient inefficiency. Note that no distributional assumption is made for any of the error components in step 1. We included the $\beta_t t$ part in $\beta'\ln\tilde{x}_{it}$ and rewrote (3) as[7]:

$$\begin{aligned}\ln\tilde{y}_{it} &= [\beta_0-a-g(z_{it})]+\beta'\ln\tilde{x}_{it}+[b_i-(\eta_i-a)]+[v_{it}-(u_{it}(z_{it})-g(z_{it}))]\\ &= h(z_{it})+\beta'\ln\tilde{x}_{it}+\alpha_i+\varepsilon_{it}\end{aligned} \tag{4}$$

where $h(z_{it})=\beta_0-a-g(z_{it}),\alpha_i=b_i-(\eta_i-a)$ and $\varepsilon_{it}=v_{it}-[u_{it}(z_{it})-g(z_{it})]$. With this reformulation $E(\alpha_i)=0,E(\varepsilon_{it})=0$. The model in (4) is a partial linear model for random effects panel data (Robinson, 1988). We follow Robinson's suggestion to estimate the parametric component $(\beta'\ln\tilde{x}_{it})$ in (4) via a two-step procedure. For this we take the conditional expectation of each side of (4) with respect to z_{it}, leading to:

$$\begin{aligned}E(\tilde{y}_{it}|z_{it}) &= E((h(z_{it})+\beta'\ln\tilde{x}_{it}+\alpha_i+\varepsilon_{it})|z_{it})\\ &= E(h(z_{it})|z_{it})+\beta'E(\ln\tilde{x}_{it}|z_{it})+E(\alpha_i|z_{it})+E(\varepsilon_{it}|z_{it})\\ &= h(z_{it})+\beta'E(\ln\tilde{x}_{it}|z_{it})\end{aligned} \tag{5}$$

since $E(\alpha_i|z_{it})=0$ and $E(\varepsilon_{it}|z_{it})=0$. We then subtract (5) from (4) to obtain:

$$\begin{aligned}\ln\tilde{y}_{it}-E(\ln\tilde{y}_{it}|z_{it}) &= h(z_{it})+\beta'\ln\tilde{x}_{it}+\alpha_i+\varepsilon_{it}-(h(z_{it})+\beta'E(\ln\tilde{x}_{it}|z_{it}))\\ &= \beta'[\ln\tilde{x}_{it}-E(\ln\tilde{x}_{it}|z_{it})]+\alpha_i+\varepsilon_{it}\end{aligned} \tag{6}$$

[7] It is worth noting that one can add any number of quasi-fixed inputs. The difference between the variable and quasi-fixed inputs in our model is that the variable inputs are endogenous (chosen by the firm to maximize the objective function) while quasi-fixed inputs are exogenous (predetermined). This will make the production function homogeneous of degree one in variable inputs. Consequently the estimated translog function will be of the form $\ln\tilde{y}=\psi(\ln\tilde{x},q)+\ln A$, where $\psi(.)$ is a translog function with arguments $\ln\tilde{x}$ and quasi-fixed inputs q (including time trend, experience, and other environmental variables such as location, regions, etc.).

The conditional means $E\left(\ln \tilde{y}_{it} \middle| z_{it}\right)$ and $E\left(\ln \tilde{x}_{it} \middle| z_{it}\right)$ are estimated non-parametrically using the NP package in 'R'.[8] Using these estimates, we rewrite (6) as

$$y_{it}^{*} = \beta' x_{it}^{*} + \alpha_i + \varepsilon_{it} \tag{7}$$

where $y_{it}^{*} = \ln \tilde{y}_{it} - E\left(\ln \tilde{y}_{it} \middle| z_{it}\right)$ and $x_{it}^{*} = \ln \tilde{x}_{it} - E\left(\ln \tilde{x}_{it} \middle| z_{it}\right)$. The model in (7) is a linear random-effects panel data model. Estimation of (7) gives consistent estimates of β irrespective of the distributions of the error components (Hsiao (2014), Baltagi(2008)). It also gives predicted values of α_i (which is consistent when T $\to \infty$) and ε_{it}, which we use in steps 2 and 3.

Step 2: Using the predicted value of α_i from step 1 (and ignoring the difference between the true value and predicted values of α because of the fact the β parameters in step 1 are consistent) we have

$$\alpha_i = b_i - \left(\eta_i - a\right) = a + b_i - \eta_i \tag{8}$$

We then made the distributional assumptions about η_i and b_i. Assuming b_i is iid $N\left(0, \sigma_b^2\right)$ and η_i iid $N^{+}\left(0, \sigma_\eta^2\right)$ (which means that $E\left(\eta_i\right) = \sqrt{2/\pi}\,\sigma_\eta$) we estimate (8) using the standard cross-sectional SF technique and get the predicted values of η_i using the Jondrow et al. (1982) procedure. We also obtain predicted values of persistent technical efficiency, defined as $\exp(-\eta_i)$.

Step 3: In this step we use the predicted value of ε_{it} from step 1 to estimate u_{it}. In doing so we ignore the difference between the predicted and true values of ε_{it} because of the fact the β parameters in step 1 are consistent. Note that

$$\varepsilon_{it} = v_{it} - \left(u_{it}\left(z_{it}\right) - g\left(z_{it}\right)\right) = g\left(z_{it}\right) + v_{it} - u_{it}\left(z_{it}\right) \tag{9}$$

We assume v_{it} to be iid $N\left(0, \sigma_v^2\right)$ and $u_{it}\left(z_{it}\right) \sim N^{+}\left(0, \sigma_u^2\left(z_{it}\right)\right)$, which means $E\left(u_{it}\left(z_{it}\right)\right) = \sqrt{2/\pi}\,\sigma_u\left(z_{it}\right) \equiv g\left(z_{it}\right)$. We then estimate (9) using the pooled SF technique in which the dependent variable is ε_{it} and the regression part is $g(z_{it})$. Note that $g(z_{it})$ is related to the variance of

[8] In this study we used a local-constant kernel regression estimator. However, a local-linear kernel regression estimator could also have been used. Note that the nonparametric estimate for the conditional means are used for the transformation only. Although we used the NP package in R, Stata 15 has also the non-parametric regression package available.

u_{it} and to make it non-negative it is parameterized as $\exp(\delta_0 + \delta' z_{it})$. Since $g(z_{it})$ has no new parameters other than those in the variance of u_{it}, we needed to make sure that the exact relationship between $g(z_{it})$ and the varaince of $u_{it}(z_{it})$ is maintained in estimating the model. In this step we get predicted inefficiency values of $u_{it}(z_{it})$, as well as the marginal effects of the z_{it} variables on transient inefficiency (using either the Wang (2002) or the Kumbhakar and Sun (2013) procedure). Sometimes the interest is in obtaining estimates of transient technical efficiency, defined as $\exp(-u_{it}(z_{it}))$.

The main advantage of the three-step procedure over the single-step method, as mentioned earlier, is that the regression coefficients (β) in the first-step are not affected by distributional assumptions on the error components. This is especially important to those who are primarily interested in estimating the technology parameters consistently, and estimation of inefficiency might be secondary or of no interest. Distributional assumptions are, however, necessary to predict absolute measures of persistent and transient inefficiency. However, the important thing to note is that no matter whether one uses a single or multiple steps, the endogeneity problem is eliminated by using the transformed model in (2) in which the regressors are independent of the error terms in the production function under the behavioural assumption that farms maximize returns to the outlay. Similarly, even if one is not interested in estimating inefficiency, the endogeneity problem is eliminated when the step 1 is used in the transformed model in (2).

The data

The data source is the Norwegian Farm Accountancy Survey. This is an unbalanced set of farm-level panel data, collected by the Norwegian Institute of Bioeconomy Research (NIBIO). It includes farm production and economic data collected annually from about 1000 farms. There is no limit on the number of years a farm may be included in the survey. However, for various reasons, approximately 10% of the survey farms are replaced per year. The farms are classified according to their main category of farming. Only crop farms are included in this study. The data used in the analysis consists of 918 observations from 46 crop farms observed from 1993 to 2014. The average duration of farms in the survey was 20 years.

Crop farms usually produce several types of crops, and by the classification system applied, these farms have few (if any) activities besides crops. The total crop output (Y_1) is aggregated and measured in revenue terms adjusted to 2010 prices using the consumer price index (CPI) in Norwegian Kroner (NOK).

Three input variables are included in the estimated model. Farmland (X_1) is measured in hectares. Labour (X_2) is defined as hours of owner and wage labour. Materials and capital inputs (X_3) are adjusted (using the CPI) to 2010 NOK prices. The time trend variable (t) is used to accommodate shift in the production function (technical change).

In addition to output and input variables we included two variables as determinants of transient inefficiency. These are: i) off-farm activity (Z_1), defined as the ratio of time allocated to off-farm activity to the total time allocated to agriculture; and ii) subsidies (in NOK, adjusted to 2010 price level) (Z_2).[9] In Table 1 the descriptive statistics are listed.

[Table 1 about here]

In Fig. 1 the trends in the crop output and the three inputs are outlined. The crop output as well as land and materials and capital inputs have increased during the period 1993 to 2014, while the labour input has decreased.

[Figure 1 about here]

Application

Empirical models

Although our discussion of the theoretical model and endogeneity was in terms of a CD production, the argument follows through if the underlying production is translog. Since maximization of RO implies that the underlying production function is homogeneous of degree 1 in inputs, the linear homogeneous translog function can be written as

$$\ln \tilde{y}_{it} = \beta_0 + \sum_{j=2}^{J} \beta_j \ln \tilde{x}_{it} + \frac{1}{2} \sum_{j=2}^{J} \sum_{k=2}^{J} \beta_{jk} \ln \tilde{x}_{jit} \ln \tilde{x}_{kit} + \beta_t t + \frac{1}{2} \beta_{tt} t^2 + \sum_{j=2}^{J} \beta_{jt} \ln \tilde{x}_{it} t + \varepsilon_{it} \qquad (10)$$

where ε_{it} is the composite error term. Symmetry restrictions imply that $\beta_{jk} = \beta_{kj}$. Using the FOCs of maximization of RO it can be shown that $\tilde{x}_j$ are independent of the error terms in the production function thereby solving the endogeneity problem (correlation between $\ln \tilde{x}_{jit}$ and ε_{it}).

[9] We also included dummies for region in our preliminary analysis. None of the region dummies were found to have significant impact on the output and consequently were not included in the analysis.

Using the estimated production function parameters from (10), we computed the following elasticities:

$$\frac{\partial \ln \tilde{y}_{it}}{\partial \ln \tilde{x}_{jit}} = \frac{\partial \ln y_{it}}{\partial \ln x_{jit}} = \beta_j + \sum_{k=2}^{J} \beta_{jk} \ln \tilde{x}_{kit} + \beta_{jt} t, \quad j = 2,...,J \tag{11}$$

$$\frac{\partial \ln \tilde{y}_{it}}{\partial t} = \frac{\partial \ln y_{it}}{\partial t} = \beta_t + \beta_{tt} t + \sum_{j=2}^{J} \beta_{jt} \ln \tilde{x}_{jit} \tag{12}$$

The formulae in Eq. (11) is used to compute output elasticities, and the formula in Eq. (12) is used to estimate technical change (TC).

We estimated three models, each with different specification of the composite error term ε_{it}. Land was used as the numeraire input (x_1) in all the models. Model 1 is the True Random-Effects (TRE) model with determinants of transient inefficiency (Greene, 2005a, b). Model 2 is the four-component SF model by Colombi et al. (2014), Kumbhakar et al. (2014) and Tsionas and Kumbhakar (2014). Model 3 is the four-component model with determinants of transient inefficiency. Specifications of these models[10] are given in Table 2.

[Table 2 about here]

Results

For the three models described above, the estimated output elasticities with respect to land, labour and materials and capital inputs all differed from zero at the 1% significance level (Table 3). The elasticity of land is the largest at 0.48-0.57, being about twice the elasticities with respect to materials and capital inputs and more than twice the elasticities with respect to labour. Hence, as expected, the value of total crop production depends strongly on the area of land used. The elasticity estimates in the three models are quite consistent.

Estimates of technical change, at 1.3% -1.9% per year on average, are statistically significant and positive for all models (Table 3). These findings are quite similar to the technical change estimates by Kumbhakar et al. (2014) and somewhat lower than those found by Lien et al. (2010). Fig. 2 shows that all three models consistently indicate steady acceleration of technical progress over the period 1993 to 2014.

[10] Models 1 and 2 were estimated using Limdep, while Model 3 was estimated with the package "np" in R (to get the transformed output and inputs) and Stata.

[Table 3 about here]

[Figure 2 about here]

Technical efficiency

The mean transient technical efficiency values for all three models are quite close, ranging from 0.82 for Model 2 to 0.88 for Model 1 (Table 3). The distributions of the efficiency scores are shown in Fig 3, indicating that the distributions are quite similar for the models. The figure also shows quite different levels of estimates of persistent efficiency between Model 2 and Model 3 (Fig. 3).

[Figure 3 about here]

Efficiency determinants

The results for Models 1 and 3 both show a negative association between transient technical efficiency and off-farm time ratio. As shown in Table 3 and Fig. 4, the off-farm time ratio increased the variance of the transient inefficiency function (the δ coefficient being positive), which means increased (decreased) inefficiency (efficiency). Our finding that off-farm work decrease efficiency (significant for Model 1, not for Model 3) supports earlier results obtained by Karagiannis and Sarris (2005) as well as Kumbhakar and Lien (2010) and Lien et al. (2010), but is at variance with some other earlier findings (e.g., Brümmer, 2001; Goodwin and Mishra, 2004; Kumbhakar et al., 2014). The mechanisms causing our results are unclear. However, one possibility is that off-farm activity broadens the farmer's experience, which then leads to improved farm management. Another possibility is that farmers with off-farm work tend to be less efficient, because they spend less time to manage the farm. In our study, perhaps the first effect is the stronger of these two, even though the results in Model 3 are not statistically significant.

Subsidy payment has a positive and statistically significant effect on technical efficiency both for Model 1 and Model 3 (Table 3). A meta-analysis of the association between subsidies and technical efficiency, by Minviel and Latruffe (2017), found mixed results. The overall effect of subsidies were negative, but 46% of the studies found statistically significant positive or not any significant effect. The findings of a recent study by Latruffe et al. (2017) of diverse western EU countries support the above mentioned meta-analysis, and found positive, zero, or negative effect of subsidies on technical efficiency, depending on analysed country. Piesse and Thirtle (2000), Giannakas et al. (2001) and Karagiannis and Sarris (2005) found a negative effect of subsidies on technical efficiency, while

McCloud and Kumbhakar (2008), Kumbhakar and Lien (2010) and Zhu and Oude Lansink (2010) found a positive effect, which is consistent with our findings.

Figure 4 shows that the positive efficiency effect of subsidy payment decreased with increasing subsidies (less negative marginal effect on inefficiency), suggesting that too much subsidy may reduce the motivation of farmers to work efficiently. However, one should be careful in drawing any general conclusion about this issue because in Norway there are a range of subsidies that may have different effects, some of which might cancel each other out. Minviel and Latruffe (2017) stressed how the subsidy variables are defined (total subsidies received at the farm, or specific subsides such as production or environmental subsidies, etc.) in the studies. To get more informative and useful analysis, one may need analysis of less aggregated subsidy variables.

[Figure 4 about here]

Discussion and conclusions

In this article, we have dealt with the well-known but mostly neglected endogeneity problem. We have also extended the four-component SF model by including determinants of transient inefficiency. This comprehensive model yielded estimates of transient inefficiency and the determinants thereof, while controlling for heterogeneity among firms. Consequently, it provided a more thorough and comprehensive analysis, compared to earlier models applied in this field. Within agricultural economics/productivity analysis the endogeneity problem is common, and knowledge of the drivers of inefficiency is of importance, implying that the model presented would be suitable for future studies in this field.

The model presented here could easily be extended to also include determinants of persistent inefficiency (given access to relevant and reliable data). In that regard, since transient inefficiency and persistent inefficiency measure different things, it is logical to use different Z variables as determinants for transient and persistent inefficiency. The variables explaining persistent inefficiency should naturally be time-invariant (e.g., regional location, education, experience, a period with persistent policy regime, etc.).

The model was estimated via a multi-step approach, related to the approach proposed by Kumbhakar et al. (2014), which proved easy to implement and understand. While one-step estimation approaches are normally preferred (and is presented in Budunenko and Kumbhakar, 2017), the multi-step approach used in this article is more intuitive and easy to implement. Furthermore, the technology parameters estimated in step 1 are robust to distributional assumptions on the error components (including inefficiency).

The approach outlined was applied on unbalanced panel data on Norwegian crop farms observed from 1993 to 2014. The output elasticities and technical change results seem intuitive and quite consistent with the results from the two additional competing models applied (the 'true random effects SF model' and the 'four-component SF model'). The mean transient technical efficiency was found to range from 0.82 to 0.88, implying a quite large share of farmers below the best, and there are potential for improvement. The results for determinants of inefficiency seemed intuitive, in that off-farm work ratio was found to have a negative effect on transient efficiency, while subsidies was found to have a positive effect on transient efficiency.

The model used offers the option to drop either transient inefficiency determinants or (more typically) persistent inefficiency determinants in the analysis, if one or other seems less important to model, or if suitable variables are difficult to obtain. For the future, given access to reliable and relevant output, input and *Z* variables over several years, interesting and useful analyses will be possible based on the extended model introduced in this article.

References

Ackerberg, D.A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. *Econometrica* 83, 2411-2451.

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

Badunenko, O., Kumbhakar, S.C., 2017. Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research* 260, 789–803.

Badi Baltagi (2008), Econometric Analysis of Panel Data, John Wiley & Sons, UK.

Battese, G.E., Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38, 387–399.

Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis* 3(1), 153–169.

Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20, 325–332.

Brümmer, B., 2001. Estimating confidence intervals for technical efficiency: The case of private farms in Slovenia. *European Review of Agricultural Economics* 28, 285–306.

Caudill, S.B., Ford, J.M., 1993. Biases in frontier estimation due to heteroskedasticity. *Economics Letters* 41, 17–20.

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

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., Battese, G.E., 2005. *An Introduction to Efficiency and Productivity Analysis, Second Edition*. Springer, New York.

Colombi, R., Martini, G., Vittadini, G., 2011. A stochastic frontier model with short-run and long-run inefficiency random effects. WP 1101, Department of Economics and Technology Management, University of Bergamo, Italy.

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

Färe, R., Grosskopf, S., Zaman, O., 2002. Hyperbolic efficiency and return to the dollar. *European Journal of Operational Research* 136, 671–679.

Filippini, M., Greene, W.H., 2016. Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis* 45, 187–196.

Giannakas, K., Schoney, R., Tzouvelekas, V., 2001. Technical efficiency, technological change and output growth of wheat farms in Saskatchewan. *Canadian Journal of Agricultural Economics* 49, 135–152.

Greene, W., 2005a. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23, 7–32.

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

Greene, W., 2008. The econometric approach to efficiency analysis. In: H.O. Fried, C.A.K. Lovell and S.S. Shelton (eds.): *The Measurement of Productivity Efficiency and Productivity Growth*, pp. 92-250. Oxford University Press, New York.

Griliches, Z., Mairesse, J., 1995. Production functions: The search for identification. NBER Working Paper 5067.

Goodwin, B.K., Mishra, A.K., 2004. Farming efficiency and the determinants of multiple job holding by farm operators. *American Journal of Agricultural Economics* 86, 722–729.

Hadri, K., 1999. Estimation of a doubly heteroscedastic stochastic frontier cost function. *Journal of Business & Economic Statistics* 17, 359–363.

Huang, C.J., Liu, J.-T., 1994. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis* 5, 171–180.

Cheng Hsiao (2014), Analysis of Panel Data, 3rd edition, Cambridge University Press, New York.

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

Karagiannis, G., Sarris A., 2005. Measuring and explaining scale efficiency with the parametric approach: the case of Greek tobacco growers. *Agricultural Economics* 33, 441–451.

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

Kumbhakar, S.C., 2011. Estimation of production technology when the objective is to maximize return to the outlay. *European Journal of Operational Research 208*, 170–176.

Kumbhakar, S.C., Heshmati, A., 1995. Efficiency measurement in Swedish dairy farms: An application of rotating panel data, 1976-88. *American Journal of Agricultural Economics* 77, 660–674.

Kumbhakar, S.C., Hjalmarsson, L., 1993. Technical efficiency and technical progress in Swedish dairy farms. In: H.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency – Techniques and Applications*, pp. 256–270. Oxford University Press, Oxford, UK.

Kumbhakar, S.C., Lien, G., 2010. Impacts of subsidies on farm productivity and efficiency. In Ball, E., Fanfani, R., Gutierrez, L. (eds), *The Economic Impact of Public Support to Agriculture, An International Perspective*, New York: Springer, 109–124.

Kumbhakar, S.C., Lovell, C.A.K., 2000. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S.C., Sun, K., 2013. Derivation of marginal effects of determinants of technical inefficiency. *Economics Letters* 120, 305–309.

Kumbhakar, S.C., Asche, F., Tveteras, R., 2013. Estimation and decomposition of inefficiency when producers maximize return to the outlay: an application to Norwegian fishing trawlers. *Journal of Productivity Analysis* 40, 307–321.

Kumbhakar, S.C., Ghosh, S., McGuckin, J.T., 1991. A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business & Economic Statistics* 9, 279–286.

Kumbhakar, S.C., Lien, G., Hardaker, J.B., 2011. Technical efficiency in competing panel data models: A study of Norwegian grain farming. International Congress of the European Association of Agricultural Economics, Zurich.

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

Kumbhakar, S.C., Wang, H.-J., Horncastle, A.P., 2015. *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge University Press, New York, NY.

Kutlu, L., 2010. Battese-coelli estimator with endogeneous regressors. *Economics Letters* 109, 79–81.

Lai, H-P., Kumbhakar, S.C., 2016. Panel data stochastic frontier model with determinants of persistent and transient inefficiency, Unpublished paper.

Latruffe, L., Balcombe, K., Davidova, S., Zawalinska, K., 2004. Determinants of technical efficiency of crop and livestock farms in Poland. *Applied Economics* 36, 1255–1263.

Latruffe, L., Bravo-Ureta, B.E., Carpentier, A., Desjeux, Y., Moreira, V.H., 2017. Subsidies and technical efficiency in agriculture: Evidence from European dairy farms. *American Journal of Agricultural Economics* 99, 783–799.

Lien, G., Kumbhakar, S.C., Hardaker, J.B., 2010. Determinants of off-farm work and its effects on farm performance: The case of Norwegian grain farmers. *Agricultural Economics* 41, 577–586.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2), 317–341.

McCloud, N., Kumbhakar, S.C., 2008. Do subsidies drive productivity? A cross-country analysis of Nordic dairy farms. In S. Chib, W. Griffiths, G. Koop, D. Terrell (ed.) *Bayesian Econometrics (Advances in Econometrics, Volume 23)* Emerald Group Publishing Limited, pp. 245–274.

Minviel, J.J., Latruffe, L., 2017. Effect of public subsidies on farm technical efficiency: a meta-analysis of empirical results. Applied Economics 49, 213–226.

Nerlove, M., 1965. *Estimation and Identification of Cobb-Douglas Production Functions*. Rand McNally & Company, Chicago.

NIBIO, 2015. Resultatkontroll for gjennomføringen av landbrukspolitikken. Budsjettnemda for jordbruket. (In Norwegian).

NILF, 2014. *Norwegian Agriculture – Status and Trends 2014*. Norwegian Agricultural Economics Research Institute (NILF), Oslo, Norway.

OECD, 2009. *Agricultural Policies in OECD Countries: Monitoring and Evaluation 2009*. OECD Publishing, Paris. Parmeter, D.F., Kumbhakar, S.C., 2014. *Efficiency Analysis: A Primer on Recent Advances*. Foundations and Trends in Econometrics 7(3–4), 191–385.

Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64, 1263–1297

Parmeter, C.F., Kumbhakar, S.C., 2014. *Efficiency Analysis: A Primer on Recent Advances*. Foundations and Trends (R) in Econometrics, 7(3-4), 191-385.

Piesse, J., Thirtle, C., 2000. A stochastic frontier approach to firm level efficiency, technological change and productivity during the early transition in Hungary. *Journal of Comparative Economics* 28, 473–501.

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

Reifschneider, D., Stevenson, R., 1991. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review* 32, 715–723.

Robinson, P.M., 1988. Root-n consistent semiparametric regression. *Econometrica* 56, 931–954.

Schmidt, P., Sickles, R., 1984. Production frontiers and panel data. *Journal of Business and Economic Statistics* 2, 367–374.

Shee, A., Stefanou, S., 2015. Endogeneity corrected stochastic production frontier and technical efficiency. *American Journal of Agricultural Economics* 97, 939–952.

Tran, K.C., Tsionas, E.G., 2013. GMM estimation of stochastic frontier model with endogeneous regressors. *Economics Letters* 118, 233–236.

Tsionas, E.G., Kumbhakar, S.C., 2014. Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random effects model. *Journal of Applied Econometrics* 29, 110–132.

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

Zhu X., Oude Lansink, A., 2010. Impact of CAP subsidies on technical efficiency of crop farms in Germany, the Netherlands and Sweden. *Journal of Agricultural Economics*, 61, 545–564.

Zofío, J., Prieto, A., 2006. Return to dollar, generalized distance function and the Fisher productivity index. *Spanish Economic Review* 8, 113–138.

Tables and Figures

Table 1. Descriptive statistics of the sample (*n* = 918)

Variable	Label	Mean	Std. dev.	Min.	Max.
Output					
Y_1	Crop output (2010 1000 NOK)	413.1	375.5	19.7	3597.8
Inputs					
X_1	Land (ha)	32.4	17.9	5.5	126.4
X_2	Labour (hours)	1270	907	30	5461
X_3	Materials and capital input (2010 1000 NOK)	254.5	207.4	28.0	2140.4
t	Year	2004	6	1	22
Determinants of transient inefficiency					
Z_1	Off-farm activity	0.57	0.29	0.00	0.98
Z_2	Subsidies (1000 NOK)	91.30	62.22	2.50	357.90

Table 2. Econometric specification of the stochastic production frontier.

	Model 1	Model 2	Model 3
Composed error ε_{it}	$\varepsilon_{it} = b_i + v_{it} - u_{it}(z_{it})$ $b_i \sim iidN(0, \sigma_b^2)$ $v_{it} \sim iidN(0, \sigma_v^2)$ $u_{it}(z_{it}) \sim N^+(0, \sigma_u^2(z_{it}))$ $= N^+(0, \exp(\delta_{u0} + \delta'_{uit} z_{it}))$	$\varepsilon_{it} = b_i + \eta_i + v_{it} - u_{it}$ $b_i \sim iidN(0, \sigma_b^2)$ $\eta_i \sim iidN^+(0, \sigma_\eta^2)$ $v_{it} \sim iidN(0, \sigma_v^2)$ $u_{it} \sim iidN^+(0, \sigma_u^2)$	$\varepsilon_{it} = b_i + \eta_i + v_{it} - u_{it}(z_{it})$ $b_i \sim iidN(0, \sigma_b^2)$ $\eta_i \sim iidN^+(0, \sigma_\eta^2)$ $v_{it} \sim iidN(0, \sigma_v^2)$ $u_{it}(z_{it}) \sim N^+(0, \sigma_u^2(z_{it}))$ $= N^+(0, \exp(\delta_{u0} + \delta'_{uit} z_{it}))$
Main reference(s)	Greene (2005a, b)	Colombi et al. (2014), Kumbhakar et al. (2014), Tsionas and Kumbhakar (2014)	Special case of Badunenko and Kumbhakar (2017), Lai and Kumbhakar (2016)

Table 3. Elasticities, returns to scale, technical change and inefficiency estimates

Label	Model 1 Estimate	SE	Model 2 Estimate	SE	Model 3 Estimate	SE
Elasticity						
Land (calc.)	0.554		0.566		0.475	
Labour	0.173***	0.011	0.179***	0.010	0.270***	0.027
Materials and capital input	0.273***	0.015	0.255***	0.015	0.255***	0.025
Returns to scale	1.000		1.000		1.000	
Technical change	0.017***	0.003	0.019***	0.002	0.013***	0.003
Technical efficiency						
Transient	0.877	0.074	0.824	0.087	0.874	0.073
Persistent			0.887	0.004	0.966	0.002
Determinants for variance in the pre-truncated transient inefficiency function						
Off-farm time ratio	0.387***	0.087			0.185	0.210
Subsidies	-0.009***	0.002			-0.060***	0.019

***, **, * => Significance at 1%, 5%, 10% level

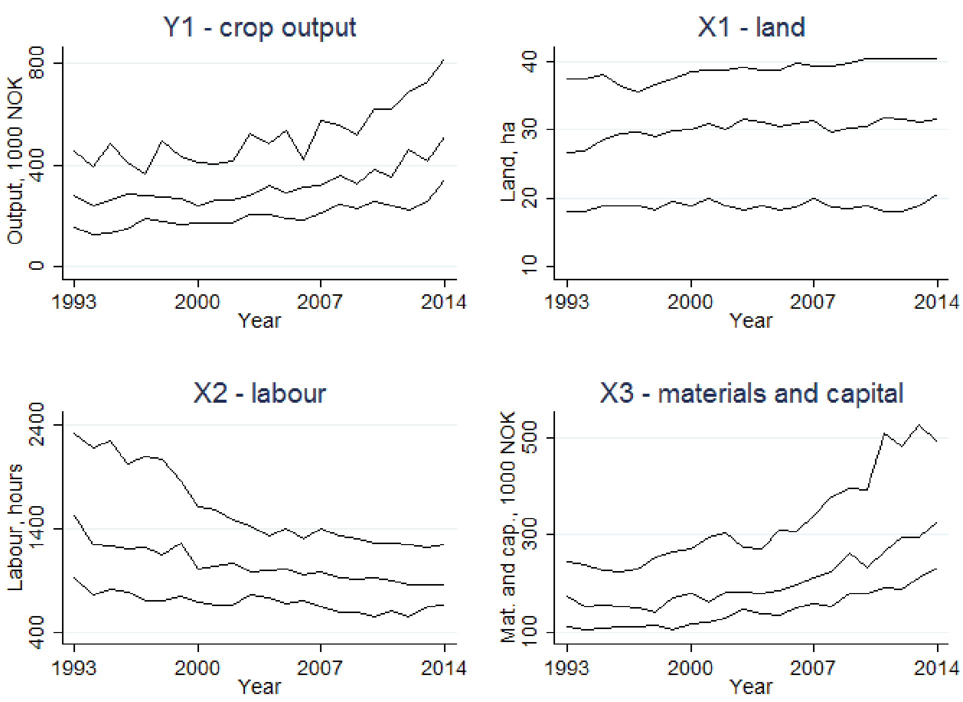


Fig 1. The median, first and third quantile values (middle, bottom and top lines) of output and inputs in the sample over the sample period.

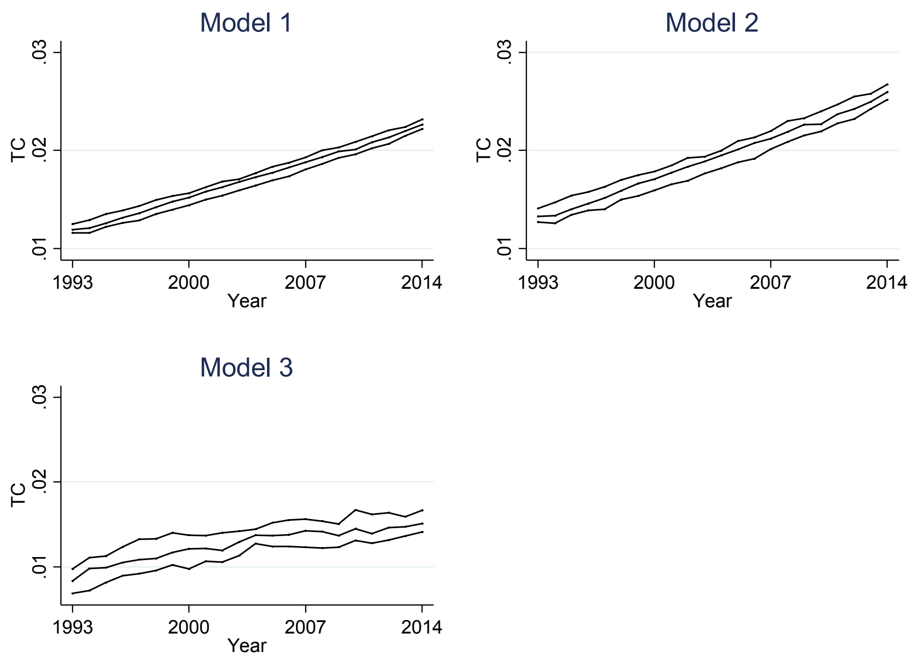


Fig 2. The median, first and third quantile values (middle, bottom and top lines) of the distribution of technical change (TC) for the sample observations in Models 1 to 3

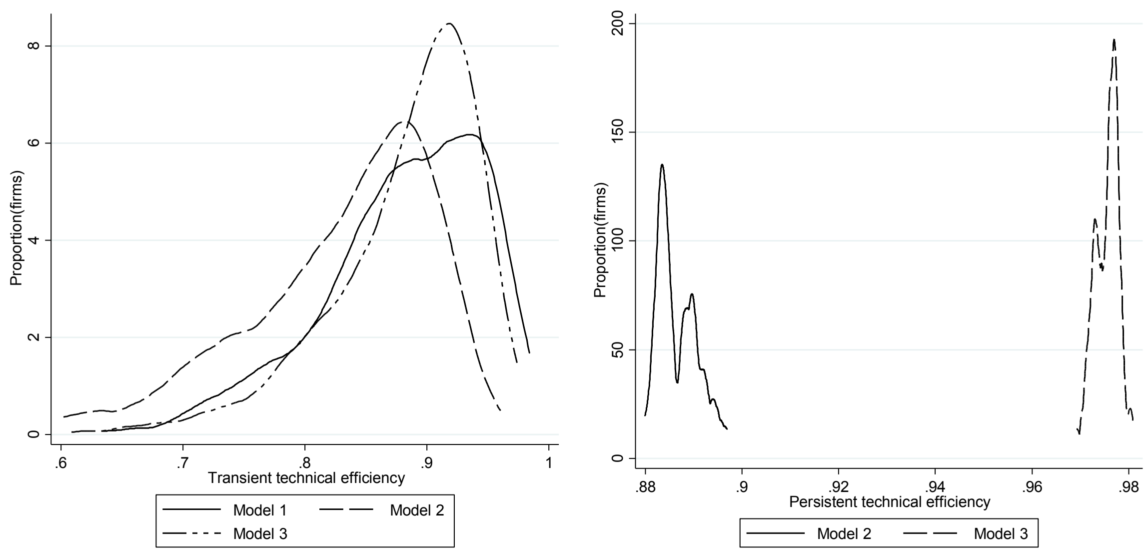


Fig 3. Distributions of transient efficiency (left) and persistent efficiency (right) for Models 1-3

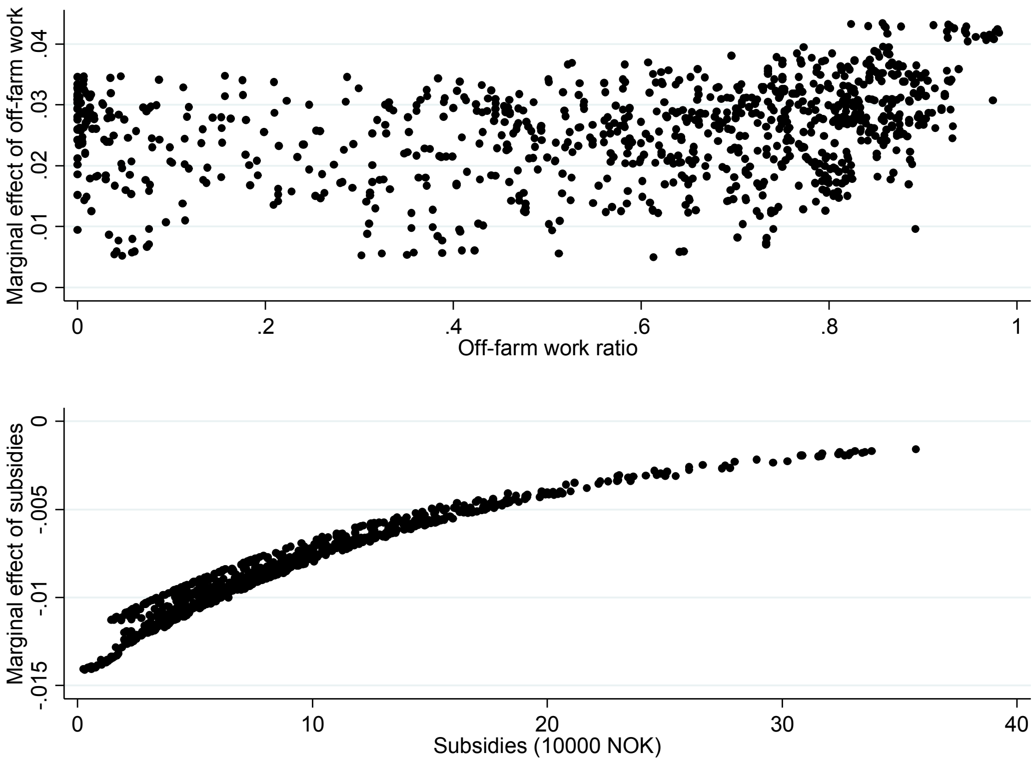


Fig 4. Scatter plot of marginal effects on expected value of inefficiency, based on Model 3. Upper panel shows marginal effects of off-farm work ratio on transient inefficiency. Lower panel shows marginal effects on transient inefficiency of subsidies.