CrossMark

Are all firms inefficient?

Seunghwa Rho · Peter Schmidt

Abstract In the usual stochastic frontier model, all firms
are inefficient, because inefficiency is non-negative and the
probability that inefficiency is exactly zero equals zero. We
modify this model by adding a parameter p which equals
the probability that a firm is fully efficient. We can estimate
this model by MLE and obtain estimates of the fraction of
firms that are fully efficient and of the distribution of
inefficiency for the inefficient firms. This model has also
been considered by Kumbhakar et al. (J Econ 172:66–76,
2013). We extend their paper in several ways. We discuss
some identification issues that arise if all firms are ineffi-
cient or no firms are inefficient. We show that results like
those of Waldman (J Econ 18:275–279, 1982) hold for this
model, that is, that the likelihood has a stationary point at
parameters that indicate no inefficiency and that this point
is a local maximum if the OLS residuals are positively
skewed. Finally, we consider problems involved in testing
the hypothesis that $p = 0$. We also provide some simula-
tions and an empirical example.

Keywords Stochastic frontier model · Finite
mixture model · Latent class model · Technical
inefficiency

JEL Classification C10 · C46 · C52

S. Rho
Department of Economics, Louisiana State University, Baton
Rouge, LA, USA
e-mail: srho@lsu.edu

P. Schmidt (✉)
Department of Economics, Michigan State University, East
Lansing, MI, USA
e-mail: schmidtp@msu.edu

1 Introduction

In the basic stochastic frontier model of Aigner et al.
(1977) and Meeusen and van den Broeck (1977), all firms
are inefficient to some degree. The one-sided error that
represents technical inefficiency has a distribution (for
example, half normal) for which zero is in the support, so
that zero is a possible value, but it is still the case that the
probability is zero that a draw from a half normal exactly
equals zero. This may be restrictive empirically, since it is
plausible, or at least possible, that an industry may contain
a set of firms that are fully efficient.

In this paper we allow the possibility that some firms are
fully efficient. We introduce a parameter p which repre-
sents the probability that a firm is fully efficient. So the
case of $p = 0$ corresponds to the usual stochastic frontier
model and the case of $p = 1$ corresponds to the case of full
efficiency (no one-sided error), while if $0 < p < 1$ a frac-
tion p of the firms are fully efficient and a fraction
$1 - p$ are inefficient. This may be important because if
some of the firms actually are fully efficient, the usual
stochastic frontier model is misspecified and can be
expected to yield biased estimates of the technology and of
firms' inefficiency levels.

This model is a special form of the latent class model
considered by Caudill (2003), Orea and Kumbhakar
(2004), Greene (2005) and others. It has the special fea-
ture that the frontier itself does not vary across the two
classes of firms; only the existence or non-existence of
inefficiency differs. Our model has previously been con-
sidered by Kumbhakar, Parmeter, and Tsionas (2013),
hereafter KPT. See also Grassetti (2011). Our results were
derived without knowledge of the KPT paper, but in this
paper we will naturally focus on our results which are not
in their paper.

The plan of the paper is as follows. In Sect. 2 we will present the model and give a brief summary of the basic results that are also in the KPT paper. These include the likelihood to be maximized, the form of the posterior probabilities of full efficiency for each firm, and the expression for the estimated inefficiencies for each firm. In Sect. 3 we provide some new results. We discuss identification issues. We give the generalization of the results of Waldman (1982), which establish that there is a stationary point of the likelihood at a point of full efficiency and that this point is a local maximum of the likelihood if the OLS residuals are positively skewed. We propose using logit or probit models to allow additional explanatory variables to affect the probability of a firm being fully efficient. We also discuss the problem of testing the hypothesis that $p = 0$. In Sect. 4 we present some simulations, and in Sect. 5 we give an empirical example. Finally, Sect. 6 gives our conclusions.

Since this paper is in a volume in honor of Lennart Hjalmarsson, it is appropriate to comment on Lennart's contributions to efficiency and productivity analysis, and to ask whether there is any link between his work and this paper. Lennart was of course an extremely productive scholar and one of very broad interests. His most remarkable paper is arguably Førsund and Hjalmarsson (1974), "On the Measurement of Productive Efficiency". This paper laid out in careful terms what it is that we seek to measure, several years before the first non-stone-age tools were developed to measure it. It is harder to find links between his work and this paper, because this is an SFA paper and he was at heart a DEA person. He did some influential work with Almas Heshmati and Subal Kumbhakar comparing the results from different models, including SFA and DEA, but when it came to actual applications he generally picked DEA. His applications were amazing diverse, in terms of industry if not geographical location: Swedish banking, Swedish dairy plants, Swedish electrical distribution, Swedish social insurance offices, Swedish dairy farms, Swedish cement plants, and the Swedish pork industry, to give a partial list. He did occasionally peek at data from the rest of the world, for example banks of other Nordic countries or Columbian cement plants, but this was rare. The best specific link between his work and this paper is that in some of his applications he took interest in the percentage of firms that were 100 % efficient, which of course occurs naturally in DEA, or in the percentage of output produced by firms that were 100 % efficient. For example, in Berg et al. (1993), he commented on the fact that the percentage of loans produced by banks that are 100 % efficient was 44 % in Finland, 52 % in Norway and 72 % in Sweden. So we would like to think that he would have regarded the topic of this paper as interesting.

2 The model and basic results

We begin with the standard stochastic frontier model of the form:

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i, \quad \varepsilon_i = v_i - u_i, \quad u_i \geq 0. \tag{1}$$

Here $i = 1, \ldots, n$ indexes firms. We have in mind a production frontier so that y is typically log output and $\mathbf{x}$ is a vector of functions of inputs. The v_i are iid $N(0, \sigma_v^2)$, the u_i are iid $N^+(0, \sigma_u^2)$ (i.e., half-normal), and $\mathbf{x}$, v, and u are mutually independent (so $\mathbf{x}$ can be treated as fixed). We will refer to this model as the basic stochastic frontier (or basic SF) model.

We now define some standard notation. Let ϕ be the standard normal density, and Φ be the standard normal cdf. Let f_v and f_u represent the densities of v and u:

$$
\begin{aligned}
f_v(v) &= \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{v^2}{2\sigma_v^2}\right) = \frac{1}{\sigma_v}\phi\left(\frac{v}{\sigma_v}\right), \\
f_u(u) &= \frac{2}{\sqrt{2\pi}\sigma_u}\exp\left(-\frac{u^2}{2\sigma_u^2}\right) = \frac{2}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right), \quad u \geq 0.
\end{aligned}
\tag{2}
$$

Also define $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$. This implies that $\sigma_v^2 = \sigma^2/(1 + \lambda^2)$ and $\sigma_u^2 = \sigma^2\lambda^2/(1 + \lambda^2)$. Finally, we let f_ε represent the density of $\varepsilon = v - u$:

$$f_\varepsilon(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right]. \tag{3}$$

Now we define the model of this paper. Suppose there is an *unobservable* variable z_i such that

$$z_i = 1(u_i = 0) = \begin{cases} 1 & \text{if } u_i = 0 \\ 0 & \text{if } u_i > 0. \end{cases}$$

Define $p = P(z_i = 1) = P(u_i = 0)$. We assume that $u_i | z_i = 0$ is distributed as $N^+(0, \sigma_u^2)$, that is, half normal. Thus

$$u_i = \begin{cases} 0 & \text{with probability } p \\ N^+(0, \sigma_u^2) & \text{with probability } 1 - p. \end{cases}$$

This model contains the parameters β, σ_u^2, σ_v^2, and p or β, λ, σ^2, and p.

We will follow the terminology of KPT and call this model the "zero-inefficiency stochastic frontier" (ZISF) model. The name refers to the fact that, in this model, the event $u_i = 0$ can occur with non-zero frequency. Note that

$$f(\varepsilon|z = 1) = f_v(\varepsilon),$$
$$f(\varepsilon|z = 0) = f_\varepsilon(\varepsilon),$$

(where f_v and f_ε are defined in (2) and (3) above) and so the marginal (unconditional) density of ε is

$$f_p(\varepsilon) = pf_v(\varepsilon) + (1 - p)f_\varepsilon(\varepsilon). \tag{4}$$

Using this density, we can form the (log) likelihood for the model:

$$\ln L\big(\beta, \sigma_u^2, \sigma_v^2, p\big) = \sum_{i=1}^{n} \ln f_p\big(y_i - \mathbf{x}_i'\beta\big). \qquad (5)$$

We will estimate the model by MLE; that is, by maximizing $\ln L$ with respect to β, σ_u^2, σ_v^2, and p. Or, alternatively, the model may be parameterized in terms of β, λ, σ^2, and p, with maximization over that set of parameters.

When we have estimated the model, we can obtain $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\beta}$, an estimate of $\varepsilon_i = y_i - \mathbf{x}_i'\beta$. Using Bayes rule, we can now update the probability that a particular firm is fully efficient, because ε_i is informative about that possibility. That is, we can calculate

$$P(z_i = 1|\varepsilon_i) = \frac{P(z_i=1)\, f(\varepsilon_i|z_i=1)}{f_p(\varepsilon_i)} = \frac{pf_v(\varepsilon_i)}{f_p(\varepsilon_i)}$$
$$= \frac{pf_v(\varepsilon_i)}{pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)}. \qquad (6)$$

We will call this the "posterior" probability that firm i is fully efficient. It is evaluated at $\hat{p}$, $\hat{\varepsilon}_i$ and also $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$, which enter into the densities of f_v and f_ε. We put quotes around "posterior" because it is not truly the posterior probability of $z_i = 1$ in a Bayesian sense. (A true Bayesian posterior would give $P(z_i = 1|y_i, x_i)$ and would have started with a prior distribution for the parameters β, σ_u^2, σ_v^2, and p.)

We now wish to estimate (predict) u_i for each firm. Following the logic of Jondrow, Lovell, Materov and Schmidt (1982) (hereafter JLMS), we define $\hat{u}_i = E(u_i|\varepsilon_i)$. Now

$$\begin{aligned} E(u_i|\varepsilon_i) &= E_{z|\varepsilon} E(u_i|\varepsilon_i, z_i) \\ &= P(z_i=1|\varepsilon_i)E(u_i|\varepsilon_i, z_i=1) \\ &\quad + P(z_i=0|\varepsilon_i)E(u_i|\varepsilon_i, z_i=0) \\ &= P(z_i=0|\varepsilon_i)E(u_i|\varepsilon_i, z_i=0) \end{aligned} \qquad (7)$$

since $u_i \equiv 0$ when $z_i = 1$. But $E(u_i|\varepsilon_i, z_i = 0)$ is the usual expression from JLMS, and $P(z_i = 0|\varepsilon_i) = 1 - P(z_i = 1|\varepsilon_i)$ which can be evaluated using Eq. (6) above. Therefore,

$$\begin{aligned} \hat{u}_i &= E(u_i|\varepsilon_i) \\ &= \frac{(1-p)f_\varepsilon(\varepsilon_i)}{pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)} \times \sigma_* \left[\frac{\phi(a_i)}{1-\Phi(a_i)} - a_i \right], \end{aligned} \qquad (8)$$

where $a_i = \varepsilon_i\lambda/\sigma$ and $\sigma_* = \sigma_u\sigma_v/\sigma = \lambda\sigma/(1+\lambda^2)$.

A slight extension of this result, which is not in KPT, is to follow Battese and Coelli (1988) and define technical efficiency as $TE = \exp(-u)$. Correspondingly technical inefficiency would be $1 - TE = 1 - \exp(-u)$, which is only approximately equal to u (for small u). They provide the expression for $E(TE|\varepsilon)$. In our model the expression is a little more complicated. The equivalent of (7) is that

$$\begin{aligned} E(e^{-u_i}|\varepsilon_i) &= P(z_i=1|\varepsilon_i)E(e^{-u_i}|\varepsilon_i, z_i=1) \\ &\quad + P(z_i=0|\varepsilon_i)E(e^{-u_i}|\varepsilon_i, z_i=0). \end{aligned}$$

But now $E(e^{-u_i}|\varepsilon_i, z_i = 1) = 1$ and so both terms are non-zero. This leads to the expression

$$\begin{aligned} \widehat{TE}_i &= E(e^{-u_i}|\varepsilon_i) \\ &= \frac{(1-p)f_\varepsilon(\varepsilon_i)}{pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)} \times \frac{\Phi\left(\frac{\mu_i^*}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_i^*}{\sigma_*}\right)} \exp\left(\frac{\sigma_*^2}{2} - \mu_i^*\right) \\ &\quad + \frac{pf_v(\varepsilon_i)}{pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)}, \end{aligned} \qquad (9)$$

where $\mu_i^* = -\varepsilon_i\sigma_u^2/\sigma^2$, $\sigma_* = \sigma_u\sigma_v/\sigma$ (as above), and correspondingly $\mu_i^*/\sigma_* = -a_i$ where $a_i = \varepsilon_i\lambda/\sigma$ (as above).

Note that the expression for $\hat{u}_i$ is just a simple scaling of the JLMS expression. However, for $\widehat{TE}_i$, this is not the case. $\widehat{TE}_i$ is a scaling of the Battese–Coelli estimate, *plus* a non-zero additive term reflecting $P(z_i=1|\varepsilon_i)E(e^{-u_i}|\varepsilon_i, z_i = 1) = P(z_i=1|\varepsilon_i)$. As in Jondrow et al. (1982), the expression in either (8) or (9) would need to be evaluated at the estimated values of the parameters (p, σ_u^2, and σ_v^2) and at $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\beta}$.

In this paper we have maximized the likelihood by a direct optimization (numerical search) with respect to all of the parameters. An alternative would be the EM algorithm, which is applicable when the model can be viewed as one with missing data. In our case the missing data are the z_i that indicate whether or not the firm is fully efficient. If we knew the z_i, the resulting likelihood (the "complete data likelihood") would be simple. The EM algorithm alternates between two steps. The first (E) step is to replace the z_i by their expectations given the observed data and the tentative parameter values. These expectations are just the posterior probabilities given in (6) above. Then, taking these values as given, the second (M) step is to maximize the likelihood with respect to the remaining parameters. Also the value of p is updated as the average of the posterior probabilities. This procedure is continued until convergence. Greene (2012, pp. 1104–1106) gives computational details for the EM algorithm for a latent class model which is very similar to our model. A conventional wisdom [e.g. Greene (2012, pp. 1104–1106)] is that the EM algorithm can be very slow (many iterations until convergence) but that it is numerically very stable. In particular it is guaranteed that each iteration raises the likelihood value. We did not encounter any serious computational issues in estimating the model, so in our view the choice of method is mostly a matter of personal taste and/or software availability.

3 Extensions of the basic model

We now investigate some extensions of the basic results of the previous section. Most of the results in this section are not in KPT.

3.1 Identification issues

Some of the parameters are not identified under certain circumstances. When $p = 1$, so that all firms are fully efficient, σ_u^2 is not identified. Conversely, when $\sigma_u^2 = 0$, p is not identified. In fact, the likelihood value is exactly the same when (i) $\sigma_u^2 = 0$, $p =$ anything as when (ii) $p = 1$, $\sigma_u^2 =$ anything. More generally, we might suppose that σ_u^2 and p will be estimated imprecisely when a data set contains little inefficiency, since it will be hard to determine whether there is little inefficiency because σ_u^2 is small or because p is close to one.

This issue of identification is relevant to the problem of testing the null hypothesis that $p = 1$ against the alternative that $p < 1$. This is a test of the null hypothesis that all firms are efficient against the alternative that some fraction (possibly all) of them are inefficient, and that is an economically interesting hypothesis. KPT suggest a likelihood ratio test of this hypothesis. As they note, the null distribution of their statistic is affected by the fact that the null hypothesis is on the boundary of the parameter space. They refer to Chen and Liang (2010, Case 2, p. 608) to justify an asymptotic distribution of $1/2\chi_0^2 + 1/2\chi_1^2$ for the likelihood ratio statistic. However, it is not clear that this result applies, given that one of the parameters (σ_u^2) is not identified under the null that $p = 1$. Specifically, the argument of Chen and Liang (2010) depends on the existence and asymptotic normality of the estimator $\hat{\eta}(\gamma_0)$ (see p. 606, line 4) where γ_0 corresponds to p_0 ($= 1$), and where η corresponds to the other parameters of our model, including σ_u^2.

A more relevant reference, which KPT note but do not pursue, is Andrews (2001). This paper explicitly allows the case in which the parameter vector under the null may lie on the boundary of the maintained hypothesis *and* there may be a nuisance parameter that appears under the alternative hypothesis, but not under the null. See his Theorem 4, p. 707, for the relevant asymptotic distribution result, which unfortunately is considerably more complicated than the simple result (50–50 mixture of chi-squareds) of Chen and Liang (2010).

3.2 A stationary point for the likelihood

For the basic stochastic frontier model, let the parameter vector be $\theta = (\beta', \lambda, \sigma^2)'$. Then Waldman (1982)

established the following results. First, the log likelihood always has a stationary point at $\theta^* = (\hat{\beta}', 0, \hat{\sigma}^2)'$, where $\hat{\beta} = $ OLS and $\hat{\sigma}^2 = ($OLSsum of squared residuals$)/n$. Note that these parameter values correspond to $\hat{\sigma}_u^2 = 0$, that is, to full efficiency of each firm. Second, the Hessian matrix is singular at this point. It is negative semi-definite with one zero eigenvalue. Third, these parameter values are a local maximizer of the log likelihood if the OLS residuals are positively skewed. This is the so-called "wrong skew problem".

The log likelihood for the ZISF model has a stationary point very similar to that for the basic stochastic frontier model. This stationary point is also a local maximum of the log likelihood if the least squares residuals are positively skewed.

Theorem 1 *Let* $\theta = (\beta', \lambda, \sigma^2, p)'$ *and let* $\theta^{**} = (\hat{\beta}', 0, \hat{\sigma}^2, \hat{p})'$, *where* $\hat{\beta} = $ OLS, $\hat{\sigma}^2 = ($OLS sum of squared residuals$)/n$, *and where* $\hat{p}$ *is any value in [0,1]. Then*

1. θ^{**} *is a stationary point of the log likelihood.*
2. *The Hessian matrix is singular at this point. It is negative semi-definite with two zero eigenvalues.*
3. θ^{**} *with* $\hat{p} \in [0, 1)$ *is a local maximizer of the log likelihood function if and only if* $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$, *where* $\hat{\varepsilon}_i = y_i - \mathbf{x}_i' \hat{\beta}$ *is the OLS residual.*
4. θ^{**} *with* $\hat{p} = 1$ *is a local maximizer of the log likelihood function if* $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$.

Proof See Appendix.   □

As is typically done for the basic stochastic frontier model, we will presume that θ^{**} is the global maximizer of the log likelihood when the residuals have positive ("wrong") skew. Note that at θ^{**}, we have $\hat{\lambda} = 0$ or equivalently $\hat{\sigma}_u^2 = 0$, and p is not identified when $\sigma_u^2 = 0$. We get the same likelihood value for any value of p. In our simulations (in Sect. 4) we will set $\hat{p} = 1$ in the case of wrong skew, since $\hat{p} = 1$ is another way of reflecting full efficiency. However, for a given data set, the value of $\hat{p}$ does not matter when $\theta = \theta^{**}$.

Since $\text{plim}\left((1/n)\sum \hat{\varepsilon}_i^3\right) = E(\varepsilon_i - E(\varepsilon_i))^3 = \sigma_u^3 \sqrt{2/\pi}$ $(1 - p)(-4p^2 + (8 - 3\pi)p + \pi - 4)/\pi \leq 0$ for any $p \in [0, 1]$, as the number of observations increases, the probability of a positive third moment of the OLS residuals goes to zero asymptotically. In a finite sample, the probability of a positive third moment increases when λ is small and/or p is near 0 or 1. See Table 1 below. The entries in Table 1 are based on simulations with 100,000 replications, with $\sigma_u = 1$, $\lambda = \sigma_u/\sigma_v$, $\lambda \in \{0.5, 1, 2\}$, and $p \in \{0, 0.1, \ldots, 0.9\}$, for sample sizes 50, 100, 200, and 400.

Table 1 Frequency of a positive third moment of the OLS residuals

	$n = 50$			$n = 100$			$n = 200$			$n = 400$		
	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
$p = 0$	0.476	0.363	0.114	0.463	0.300	0.038	0.447	0.224	0.006	0.421	0.139	0.000
$p = 0.1$	0.475	0.352	0.102	0.460	0.286	0.031	0.443	0.210	0.003	0.416	0.123	0.000
$p = 0.2$	0.472	0.338	0.080	0.456	0.266	0.019	0.438	0.185	0.001	0.407	0.101	0.000
$p = 0.3$	0.469	0.322	0.058	0.451	0.245	0.011	0.431	0.161	0.000	0.398	0.079	0.000
$p = 0.4$	0.466	0.308	0.042	0.447	0.226	0.006	0.424	0.141	0.000	0.391	0.062	0.000
$p = 0.5$	0.465	0.300	0.034	0.444	0.215	0.004	0.421	0.129	0.000	0.386	0.053	0.000
$p = 0.6$	0.466	0.299	0.033	0.445	0.215	0.004	0.420	0.128	0.000	0.387	0.052	0.000
$p = 0.7$	0.468	0.311	0.043	0.449	0.229	0.006	0.427	0.143	0.000	0.394	0.063	0.000
$p = 0.8$	0.474	0.342	0.076	0.458	0.268	0.017	0.439	0.185	0.001	0.414	0.098	0.000
$p = 0.9$	0.485	0.399	0.178	0.474	0.348	0.079	0.463	0.280	0.019	0.446	0.200	0.001

3.3 Models for the distribution of u_i

The ZISF model can be extended by allowing the distribution of u_i to depend on some observable variables w_i. For example, in our empirical analysis of Sect. 5, the w_i will include variables like the age and education of the farmer and the size of his household. These variables can be assumed to affect either $P(z_i = 1)$ or $f(u_i|z_i = 0)$ or both.

First consider the case in which we assume that w_i affects the distribution of u_i for the inefficient firms. A general assumption would be that the distribution of u_i conditional on w_i and on $z_i = 0$ is $N^+(\mu_i, \sigma_i^2)$ where μ_i and/or σ_i^2 depend on w_i. For example, in Sect. 5 we will assume the RSCFG model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995), under the specific assumptions that $\mu_i = 0$ and $\sigma_i^2 = \exp(w_i'\gamma)$. Another possible model is the KGMHLBC model of Kumbhakar et al. (1991), Huang and Liu (1994) and Battese and Coelli (1995), with $\sigma_i^2 = \sigma_u^2$ constant and with $\mu_i = w_i'\psi$ or $\mu_i = c \exp(w_i'\psi)$. Wang (2002) proposes parameterizing both μ_i and σ_i^2. See also Alvarez et al. (2006).

A second case is the one in which we assume that w_i affects $P(z_i = 1)$. For example, we could assume a logit model:

$$P(z_i = 1|w_i) = \frac{\exp(w_i'\delta)}{1 + \exp(w_i'\delta)}.$$

KPT, p. 68, make the same suggestion. A probit model would be another obvious possibility.

Finally, we can consider a more general model in which both $P(z_i = 1|w_i)$ and $f(u_i|z_i = 0, w_i)$ depend on w_i, as above. We will estimate such a model in our empirical section.

3.4 Testing the hypothesis that $p = 0$

In this section, we discuss the problem of testing the null hypothesis $H_0: p = 0$ against the alternative $H_A: p > 0$. The null hypothesis is that all firms are inefficient, so the basic stochastic frontier model applies. The alternative is that some firms are fully efficient and so the ZISF model is needed.

It is a standard result that, under certain regularity conditions, notably that the parameter value specified by the null hypothesis is an interior point of the parameter space, the likelihood (LR), Lagrange multiplier (LM), and Wald tests all have the same asymptotic χ^2 distribution. However, in our case p cannot be negative, and therefore the null hypothesis that $p = 0$ lies on the boundary of the parameter space. This is therefore a non-standard problem. Unlike the case of testing the hypothesis that $p = 1$, however, there is no problem with the identification of the other parameters (nuisance parameters) β, σ_u^2, and σ_v^2, or β, λ, and σ^2. We need to restrict $\sigma_u^2 > 0$ and $\sigma_v^2 > 0$ so that the nuisance parameters are in the interior of the parameter space, and also because p would not be identified if $\sigma_u^2 = 0$. However, with these modest restrictions, this is only a mildly non-standard problem, which has been discussed by Rogers (1986), Self and Liang (1987), and Gouriéroux and Monfort (1995, chapter 21), for example.

We consider five test statistics: the likelihood ratio (LR), Wald, Lagrange multiplier (LM), modified Lagrange multiplier (modified LM), and Kuhn–Tucker (KT) tests. All of these except the LM test will have asymptotic distributions that are different from the usual (χ_1^2) distribution.

We will assume that the likelihood function $L_n(\theta)$ satisfies the usual conditions,

$$\frac{1}{\sqrt{n}} \frac{\partial l_n(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_0),$$

$$\frac{1}{n} \frac{\partial^2 L_n(\theta_0)}{\partial \theta \partial \theta} \xrightarrow{p} \mathcal{H}_0 = -\mathcal{I}_0,$$

where $\theta = (\beta', \sigma_u, \sigma_v, p)'$, and the parameters other than p are away from the boundary of their parameter spaces. Define the restricted estimator $(\tilde{\theta})$ and the unrestricted estimator $(\hat{\theta})$:

$$\tilde{\theta} = \operatorname*{argmax}_{\sigma_u \ge 0, \sigma_v \ge 0, p=0} \ln L_n(\theta),$$

$$\hat{\theta} = \operatorname*{argmax}_{\sigma_u \ge 0, \sigma_v \ge 0, 0 \le p \le 1} \ln L_n(\theta).$$

We also define $l_i = \ln f(\varepsilon_i), \hat{s}_i = \partial l_i(\hat{\theta})/\partial \theta, \tilde{s}_i = \partial l_i(\tilde{\theta})/\partial \theta, \hat{h}_i = \partial^2 l_i(\hat{\theta})/\partial \theta \partial \theta'$, and $\tilde{h}_i = \partial^2 l_i(\tilde{\theta})/\partial \theta \partial \theta'$.

3.4.1 LR test

The LR statistic when testing $H_0: p = 0$ is $\xi^{LR} = 2(\ln L_n(\hat{\theta}) - \ln L_n(\tilde{\theta}))$. Under standard regularity conditions, the asymptotic distribution of ξ^{LR} is a mixture of χ_0^2 and χ_1^2, with mixing weights 1/2, where χ_0^2 is defined as the point mass distribution at zero. That is $\xi^{LR} \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$. This follows, for example, from Chen and Liang (2010), as cited by KPT.

3.4.2 Wald test

The Wald statistic for $H_0: p = 0$ is $\xi^W = \hat{p}^2/se(\hat{p})^2$. Note that $se(\hat{p})^2$ can be computed using the outer product of the score form of the variance matrix of $\hat{\theta}$, $[(\sum_{i=1}^{n} \hat{s}_i \hat{s}_i')^{-1}]$, the Hessian form, $[(\sum_{i=1}^{n} -\hat{h}_i)^{-1}]$, or the Robust form, $[(\sum_{i=1}^{n} -\hat{h}_i)^{-1}(\sum_{i=1}^{n} \hat{s}_i \hat{s}_i')(\sum_{i=1}^{n} -\hat{h}_i)^{-1}]$. As with the LR statistic, $\xi^W \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$. Note that the non-standard nature of this result means that the "significance" of an estimated $\hat{p}$ from the ZISF model cannot be assessed using standard results.

3.4.3 LM test

The LM statistic for $H_0: p = 0$ is $\xi^{LM} = (\sum_{i=1}^{n} \tilde{s}_i)' \tilde{M}^{-1}(\sum_{i=1}^{n} \tilde{s}_i)$. $\tilde{M}$ can be either $[(\sum_{i=1}^{n} \tilde{s}_i \tilde{s}_i')]$ or $[(\sum_{i=1}^{n} -\tilde{h}_i)]$, in either case evaluated at $\tilde{\theta}$. Unlike the other statistics considered here, the LM statistic has the usual χ_1^2 distribution. It ignores the one-sided nature of the

alternative, because it rejects for a large (in absolute value) positive or *negative* value of $\tilde{s}_i$. As pointed out by Rogers (1986), this may result in a loss in power relative to tests that take the one-sided nature of the alternative into account.

3.4.4 Modified LM test

The LM statistic has the usual χ_1^2 distribution because it does not take account of the one-sided nature of the alternative. By taking account of the one-sided nature of the alternative, the LM test might have better power. The Modified LM statistic proposed by Rogers (1986) is motivated by this point. The modified LM statistic is:

$$\xi^{\text{modified LM}} = \begin{cases} \xi^{\text{LM}}, & \text{if } \sum_{i=1}^{n} \tilde{s}_i > 0 \\ 0, & \text{otherwise.} \end{cases}$$

In the modified LM statistic, a positive score is taken as evidence against the null and in favor of the alternative $p > 0$, whereas a negative score is not. So a negative score is simply set to zero. The asymptotic distribution of $\xi^{\text{modified LM}}$ is $1/2\chi_0^2 + 1/2\chi_1^2$.

3.4.5 KT test

Another variant of the score test statistic that takes account of the one-sided nature of the alternative is the KT (Kuhn–Tucker) statistic proposed by Gouriéroux et al. (1982). The KT statistic for $H_0: p = 0$ is:

$$\xi^{KT} = \left(\sum_{i=1}^{n} \tilde{s}_i - \sum_{i=1}^{n} \hat{s}_i\right)' \tilde{M}^{-1} \left(\sum_{i=1}^{n} \tilde{s}_i - \sum_{i=1}^{n} \hat{s}_i\right).$$

Here, $\tilde{s}_i = {}_i(\tilde{\theta})/\partial \theta$ and $\hat{s}_i = {}_i(\hat{\theta})/\partial \theta$, as given just before the beginning of Sect. 3.4.1. Also $\tilde{M}$ can be either $\sum_{i=1}^{n} \tilde{s}_i \tilde{s}_i'$ (OPG form) or $\sum_{i=1}^{n}(-\tilde{h}_i)$ (Hessian form). When $\hat{p} = 0$, which occurs with probability one half under the null, $\sum_{i=1}^{n} \tilde{s}_i = \sum_{i=1}^{n} \hat{s}_i$, and the statistic will equal zero. Otherwise, when $\hat{p} > 0, \sum_{i=1}^{n} \hat{s}_i = 0$ and the test statistic has the usual χ_1^2 distribution. Therefore, $\xi^{KT} \xrightarrow{d} 1/2\chi_0^2 + 1/2\chi_1^2$.

3.4.6 The wrong skew problem, revisited

When the OLS residuals are positively skewed ($\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$), we have $\hat{\sigma}_u^2 = 0$ (or equivalently, $\hat{\lambda} = 0$) and $\hat{p}$ is not well defined. Also the information matrix, whether evaluated at $\hat{\theta}$ or $\tilde{\theta}$, is singular. Specifically,

$$\sum_{i=1}^{n} \hat{s}_i \hat{s}_i' = \begin{pmatrix} \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i x_i' & -(1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^2}} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i & \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 x_i & 0 \\ -(1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^2}} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i' & \frac{2}{\pi}(1-\hat{p})^2 \frac{n}{\hat{\sigma}_v^2} & -(1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^5}} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 & 0 \\ \frac{1}{\hat{\sigma}_v^5} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 x_i' & -(1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^5}} \sum_{i=1}^{n} \hat{\varepsilon}_i^3 & \frac{1}{\hat{\sigma}_v^6} \sum_{i=1}^{n} \hat{\varepsilon}_i^4 - \frac{n}{\hat{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \tilde{s}_i \tilde{s}_i' = \begin{pmatrix} \frac{1}{\tilde{\sigma}_v^2} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i x_i' & -\sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^2}} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i & \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 x_i & 0 \\ -\sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^2}} \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 x_i' & \frac{2}{\pi}\frac{n}{\tilde{\sigma}_v^2} & -\sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^5}} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 & 0 \\ \frac{1}{\tilde{\sigma}_v^5} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 x_i' & -\sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^5}} \sum_{i=1}^{n} \tilde{\varepsilon}_i^3 & \frac{1}{\tilde{\sigma}_v^6} \sum_{i=1}^{n} \tilde{\varepsilon}_i^4 - \frac{n}{\tilde{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \hat{h}_i = \begin{pmatrix} -\sum_{i=1}^{n} \frac{x_i x_i'}{\hat{\sigma}_v^2} & (1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^2}} \sum_{i=1}^{n} x_i & 0 & 0 \\ (1-\hat{p})\sqrt{\frac{2}{\pi}\frac{1}{\hat{\sigma}_v^2}} \sum_{i=1}^{n} x_i' & -(1-\hat{p})^2 \frac{2n}{\pi \hat{\sigma}_v^2} & 0 & 0 \\ 0 & 0 & \frac{-2n}{\hat{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sum_{i=1}^{n} \tilde{h}_i = \begin{pmatrix} -\sum_{i=1}^{n} \frac{x_i x_i'}{\tilde{\sigma}_v^2} & \sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^2}} \sum_{i=1}^{n} x_i & 0 & 0 \\ \sqrt{\frac{2}{\pi}\frac{1}{\tilde{\sigma}_v^2}} \sum_{i=1}^{n} x_i' & -\frac{2}{\pi}\frac{n}{\tilde{\sigma}_v^2} & 0 & 0 \\ 0 & 0 & \frac{-2n}{\tilde{\sigma}_v^2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

All the matrices above are singular for any $\hat{p} \in [0, 1]$. Therefore, when the third moment of the OLS residuals is positive, only the LR statistic can be defined, and equals zero. It remains to decide if we should reject the null hypothesis or not when the OLS residuals have wrong skew. Clearly, the LR test will not reject the null hypothesis, since the statistic equals zero under wrong skew. But for the other tests, the statistic is undefined and it is not clear what to conclude. If we consider the wrong skew cases as indicating that all firms are efficient, then it would be reasonable to reject the null hypothesis. However, as a practical matter, whether we reject the null hypothesis or not does not affect anything, because the estimated model whether $p = 0$ or not collapses to the same model. It might be reasonable to simply say that p is not identified with incorrectly skewed OLS residuals. For a given data set, both the null and the alternative hypothesis would lead to same results.

Assuming that $\sigma_u^2 > 0$, the wrong skew problem occurs with a probability that goes to zero asymptotically. However, as shown in Table 1, it can occur with non-trivial probability in finite samples. Also, the discussion above may be relevant even when the data do not have the wrong

skew problem. The log likelihood has a stationary point at θ^{**} regardless of the skew of the residuals. In the wrong skew case, the likelihood is perfectly flat in the p direction with $\hat{\beta} = $ OLS, $\hat{\lambda} = 0$, and $\hat{\sigma}^2 = 1/nSSE$. In the correct skew case, this is not true, but when λ is small, we expect that the partial of log likelihood with respect to p (evaluated at the MLE of the other parameters) would often be small in the vicinity of $p = 0$, so that the LM test and its variants might have low power. We will investigate this issue in the simulations of the next section.

3.5 Panel data

A complete treatment of this model with panel data is beyond the scope of this paper but we will make a few general comments. Now we have data $y_{it}, x_{it}, z_{it}, i = 1, \ldots, n, t = 1, \ldots, T$. We will think in terms of n being large and T being small (for example, 43 rice farms each observed for eight years).

If T is small, it is feasible to let any of the parameters of the model be different for different values of t. For example, we might let p be different for different time

Table 2 Basic SF model versus ZISF model, all replications: $n = 200$

	Basic SF model			ZISF model			Basic SF model			ZISF model			Basic SF model			ZISF model		
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
	$\lambda=1(\sigma_u=1,\sigma_v=1),p=0$						$\lambda=2(\sigma_u=1,\sigma_v=0.5),p=0$						$\lambda=5(\sigma_u=1,\sigma_v=0.2),p=0$					
β_0	0.87	−0.13	0.20	0.60	−0.40	0.29	0.98	−0.02	0.02	0.71	−0.29	0.14	1.00	0.00	0.00	0.85	−0.15	0.04
σ_u	0.84	−0.16	0.30	0.91	−0.09	0.40	0.98	−0.02	0.03	0.97	−0.03	0.05	1.00	0.00	0.01	0.97	−0.03	0.01
σ_v	0.99	−0.01	0.02	1.02	0.02	0.02	0.50	0.00	0.01	0.54	0.04	0.01	0.19	−0.01	0.00	0.23	0.03	0.00
p	—			0.53	0.53	0.43	—			0.33	0.33	0.19	—			0.16	0.16	0.05
λ	0.93	−0.07	0.44	0.95	−0.05	0.45	2.08	0.08	0.48	1.89	−0.11	0.41	5.61	0.61	7.04	4.79	−0.21	6.21
σ	1.39	−0.02	0.04	1.47	0.06	0.12	1.11	−0.01	0.01	1.12	0.00	0.03	1.02	0.00	0.00	1.00	−0.02	0.01
$\log L$	−313.00			−312.85			−229.93			−229.65			−176.43			−176.06		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.67	−0.13	0.50	0.40	−0.40	0.60	0.78	−0.02	0.17	0.51	−0.29	0.30	0.80	0.00	0.04	0.65	−0.15	0.08
$\hat{p}(z_i=1|\hat{\varepsilon}_i)$				0.53						0.33						0.16		
	$\lambda=1(\sigma_u=1,\sigma_v=1),p=0.25$						$\lambda=2(\sigma_u=1,\sigma_v=0.5),p=0.25$						$\lambda=5(\sigma_u=1,\sigma_v=0.2),p=0.25$					
β_0	1.16	0.16	0.19	0.85	−0.15	0.15	1.26	0.26	0.08	0.98	−0.02	0.05	1.23	0.23	0.05	1.03	0.03	0.01
σ_u	0.95	−0.05	0.26	1.00	0.00	0.33	1.08	0.08	0.03	1.04	0.04	0.03	1.05	0.05	0.01	1.02	0.02	0.00
σ_v	0.97	−0.03	0.02	1.01	0.01	0.02	0.45	−0.05	0.01	0.50	0.00	0.01	0.14	−0.06	0.01	0.19	−0.01	0.00
p	—			0.49	0.24	0.19	—			0.30	0.05	0.06	—			0.23	−0.02	0.02
λ	1.07	0.07	0.45	1.06	0.06	0.43	2.53	0.53	0.88	2.21	0.21	0.55	8.32	3.32	21.81	6.06	1.06	10.77
σ	1.44	0.03	0.04	1.51	0.09	0.11	1.18	0.06	0.01	1.16	0.05	0.02	1.05	0.03	0.01	1.04	0.02	0.00
$\log L$	−314.71			−314.55			−232.55			−232.22			−174.14			−172.76		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.76	0.16	0.51	0.45	−0.15	0.47	0.86	0.26	0.23	0.58	−0.02	0.20	0.82	0.23	0.09	0.63	0.03	0.04
$\hat{p}(z_i=1|\hat{\varepsilon}_i)$				0.49						0.30						0.23		
	$\lambda=1(\sigma_u=1,\sigma_v=1),p=0.5$						$\lambda=2(\sigma_u=1,\sigma_v=0.5),p=0.5$						$\lambda=5(\sigma_u=1,\sigma_v=0.2),p=0.5$					
β_0	1.40	0.40	0.30	1.04	0.04	0.12	1.45	0.45	0.21	1.07	0.07	0.05	1.30	0.30	0.09	1.00	0.00	0.00
σ_u	1.00	0.00	0.22	1.07	0.07	0.31	1.07	0.07	0.02	1.06	0.06	0.02	0.93	−0.07	0.01	1.01	0.01	0.01
σ_v	0.93	−0.07	0.02	0.98	−0.02	0.02	0.39	−0.11	0.02	0.47	−0.03	0.01	0.11	−0.09	0.01	0.20	0.00	0.00
p	—			0.51	0.01	0.12	—			0.44	−0.06	0.06	—			0.50	0.00	0.01
λ	1.16	0.16	0.45	1.15	0.15	0.43	2.83	0.83	1.29	2.34	0.34	0.57	8.96	3.96	36.70	5.17	0.17	0.61
σ	1.44	0.03	0.04	1.52	0.11	0.12	1.15	0.03	0.01	1.16	0.04	0.02	0.94	−0.08	0.01	1.03	0.01	0.01
$\log L$	−311.00			−310.77			−221.64			−220.71			−148.06			−138.15		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.80	0.40	0.59	0.44	0.05	0.51	0.85	0.45	0.35	0.47	0.08	0.18	0.70	0.30	0.13	0.40	0.00	0.03
$\hat{p}(z_i=1|\hat{\varepsilon}_i)$				0.51						0.44						0.50		
	$\lambda=1(\sigma_u=1,\sigma_v=1),p=0.75$						$\lambda=2(\sigma_u=1,\sigma_v=0.5),p=0.75$						$\lambda=5(\sigma_u=1,\sigma_v=0.2),p=0.75$					
β_0	1.51	0.51	0.40	1.16	0.16	0.13	1.52	0.52	0.28	1.05	0.05	0.03	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.89	−0.11	0.23	1.01	0.01	0.34	0.91	−0.09	0.02	1.03	0.03	0.04	0.71	−0.29	0.09	1.00	0.00	0.02
σ_v	0.90	−0.10	0.03	0.95	−0.05	0.02	0.37	−0.13	0.02	0.48	−0.02	0.00	0.11	−0.09	0.01	0.20	0.00	0.00

Table 2 continued

	Basic SF model			ZISF model			Basic SF model			ZISF model			Basic SF model			ZISF model		
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
p		–		0.57	−0.18	0.16		–		0.67	−0.08	0.05		–		0.75	0.00	0.00
λ	1.08	0.08	0.43	1.12	0.12	0.44	2.54	0.54	0.72	2.18	0.18	0.28	6.96	1.96	19.52	5.05	0.06	0.43
σ	1.35	−0.07	0.04	1.47	0.06	0.13	0.99	−0.13	0.03	1.14	0.02	0.03	0.72	−0.30	0.09	1.02	0.00	0.02
log L	−300.69			−300.40			−196.18			−193.77			−99.12			−74.13		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.71	0.51	0.62	0.36	0.16	0.33	0.72	0.52	0.39	0.25	0.05	0.11	0.52	0.32	0.13	0.20	0.00	0.02
$\hat{p}(z_i=1|\hat{\varepsilon}_i)$				0.57						0.67						0.75		

periods. This creates more parameters to estimate but no conceptual issues of estimation.

Let the errors be v_{it} and u_{it} and let $\varepsilon_{it} = v_{it} - u_{it}$, the obvious generalization of (1) above. Then the density of ε_{it} is still as given in (4), and we can form the "likelihood" in (5), except that the sum would now be over $t = 1, \ldots, T$ as well as over i. This would commonly be called a *quasi-likelihood*. It is in fact the likelihood if the v_{it} and u_{it} are independent and identically distributed (iid) over t as well as i, in which case ε_{it} is also iid over t as well as i. However, if ε_{it} is not independent over t, then the true likelihood would depend on the joint distribution of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, and (5) is just an approximation that does not reflect the dependence over t of the ε_{it}. It is a standard econometric result that maximization of the quasi-likelihood yields a consistent estimate, although it is not efficient unless we really do have independence over t, and the conventionally-calculated standard errors of the estimates are wrong. Correct standard errors comes from the so-called "sandwich form" for robust standard errors. See e.g., Hayashi (2000, section 8.7, p. 544).

It is hard to specify a joint distribution for $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ because there is no natural joint distribution when the marginal distributions are non-normal. This problem is discussed, for the simpler case of the standard SFA model, by Amsler et al. (2014). They specify a joint distribution using a copula. This leads to conceptual and computational issues, for which the reader is referred to their paper. In the present context the quasi-MLE approach is probably the best we can reasonably hope for.

4 Simulations

We conducted simulations in order to investigate the finite sample performance of the ZISF model, and to compare it to the performance of the basic stochastic frontier model. We are interested both in parameter estimation and in the performance of tests of the hypothesis $p = 0$.

We consider a very simple data generating process: $y_i = \beta + \varepsilon_i$, where as in Sect. 2 above, $\varepsilon_i = v_i - u_i$ and u_i is half-normal with probability $1 - p$ and $u_i = 0$ with probability p. We pick $n = 200$ and 500, $\beta = 1$, and $\sigma_u = 1$. We consider $p = 0$, 0.25, 0.5, and 0.75, and $\lambda = 1$, 2, and 5 (i.e., $\sigma_v = 1$, 0.5, and 0.2). Our simulations are based on 1,000 replications. Because the MLE's were sensitive to the starting values used, we used several sets of starting values and chose the results with the highest maximized likelihood value.

Our experimental design was similar to that in KPT. They included a non-constant regressor, but in our experiments that made little difference. A more substantial

Table 3 Basic SF model versus ZISF model, correct skew replications: $n = 200$

Block λ = 1 ($\sigma_u = 1$, $\sigma_v = 1$)

	Basic SF model mean	bias	mse	ZISF model mean	bias	mse
$p = 0$						
β_0	1.06	0.06	0.07	0.71	−0.29	0.20
σ_u	1.08	0.08	0.10	1.17	0.17	0.22
σ_v	0.94	−0.06	0.02	0.98	−0.02	0.02
p		—		0.40	0.40	0.26
λ	1.20	0.20	0.28	1.23	0.23	0.30
σ	1.46	0.05	0.03	1.56	0.15	0.13
log L	−313.16			−312.97		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.86	0.06	0.36	0.51	−0.29	0.49
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.40		
$p = 0.25$						
β_0	1.29	0.29	0.16	0.93	−0.07	0.12
σ_u	1.13	0.13	0.12	1.19	0.19	0.20
σ_v	0.94	−0.06	0.02	0.98	−0.02	0.02
p		—		0.39	0.14	0.12
λ	1.27	0.27	0.35	1.26	0.26	0.32
σ	1.50	0.08	0.03	1.57	0.16	0.11
log L	−315.09			−314.90		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.90	0.30	0.47	0.54	−0.06	0.43
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.39		
$p = 0.5$						
β_0	1.51	0.51	0.32	1.10	0.10	0.12
σ_u	1.14	0.14	0.11	1.22	0.22	0.21
σ_v	0.91	−0.09	0.02	0.96	−0.04	0.02
p		—		0.44	−0.06	0.11
λ	1.33	0.33	0.38	1.31	0.31	0.35
σ	1.48	0.07	0.03	1.58	0.16	0.12
log L	−311.17			−310.91		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.91	0.51	0.61	0.51	0.11	0.40
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.44		
$p = 0.75$						
β_0	1.64	0.64	0.46	1.22	0.22	0.15
σ_u	1.05	0.05	0.09	1.19	0.19	0.23
σ_v	0.87	−0.13	0.03	0.93	−0.07	0.02

Block λ = 2 ($\sigma_u = 1$, $\sigma_v = 0.5$)

	Basic SF model mean	bias	mse	ZISF model mean	bias	mse
$p = 0$						
β_0	0.99	−0.01	0.02	0.72	−0.28	0.14
σ_u	0.98	−0.02	0.03	0.97	−0.03	0.04
σ_v	0.49	−0.01	0.01	0.54	0.04	0.01
p		—		0.32	0.32	0.18
λ	2.10	0.10	0.46	1.90	−0.10	0.39
σ	1.11	−0.01	0.01	1.12	0.00	0.03
log L	−229.94			−229.66		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.78	−0.01	0.16	0.52	−0.28	0.29
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.32		
$p = 0.25$						
β_0	1.26	0.26	0.08	0.98	−0.02	0.05
σ_u	1.08	0.08	0.03	1.04	0.04	0.03
σ_v	0.45	−0.05	0.01	0.50	0.00	0.01
p		—		0.30	0.05	0.06
λ	2.53	0.53	0.88	2.21	0.21	0.55
σ	1.18	0.06	0.01	1.16	0.05	0.02
log L	−232.57			−232.24		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.86	0.26	0.23	0.58	−0.02	0.20
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.30		
$p = 0.5$						
β_0	1.45	0.45	0.21	1.07	0.07	0.05
σ_u	1.07	0.07	0.02	1.06	0.06	0.02
σ_v	0.39	−0.11	0.02	0.47	−0.03	0.01
p		—		0.44	−0.06	0.06
λ	2.83	0.83	1.29	2.34	0.34	0.57
σ	1.15	0.03	0.01	1.16	0.04	0.02
log L	−221.64			−220.71		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.85	0.45	0.35	0.47	0.08	0.18
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.44		
$p = 0.75$						
β_0	1.52	0.52	0.28	1.05	0.05	0.03
σ_u	0.91	−0.09	0.02	1.03	0.03	0.04
σ_v	0.37	−0.13	0.02	0.48	−0.02	0.00

Block λ = 5 ($\sigma_u = 1$, $\sigma_v = 0.2$)

	Basic SF model mean	bias	mse	ZISF model mean	bias	mse
$p = 0$						
β_0	1.00	0.00	0.00	0.85	−0.15	0.04
σ_u	1.00	0.00	0.01	0.97	−0.03	0.01
σ_v	0.19	−0.01	0.00	0.23	0.03	0.00
p		—		0.16	0.16	0.05
λ	5.61	0.61	7.04	4.79	−0.21	6.21
σ	1.02	0.00	0.00	1.00	−0.02	0.01
log L	−176.43			−176.06		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.80	0.00	0.04	0.65	−0.15	0.08
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.16		
$p = 0.25$						
β_0	1.23	0.23	0.05	1.03	0.03	0.01
σ_u	1.04	0.04	0.01	1.02	0.02	0.00
σ_v	0.14	−0.06	0.01	0.19	−0.01	0.00
p		—		0.23	−0.02	0.02
λ	8.32	3.32	21.81	6.06	1.06	10.77
σ	1.05	0.03	0.01	1.04	0.02	0.00
log L	−174.14			−172.76		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.82	0.23	0.09	0.63	0.03	0.04
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.23		
$p = 0.5$						
β_0	1.30	0.30	0.09	1.00	0.00	0.00
σ_u	0.93	−0.07	0.01	1.01	0.01	0.01
σ_v	0.11	−0.09	0.01	0.20	0.00	0.00
p		—		0.50	0.00	0.01
λ	8.96	3.96	36.70	5.17	0.17	0.61
σ	0.94	−0.08	0.01	1.03	0.01	0.01
log L	−148.06			−138.15		
$\hat{E}[u_i|\hat{\varepsilon}_i]$	0.70	0.30	0.13	0.40	0.00	0.03
$\hat{p}(z_i = 1|\hat{\varepsilon}_i)$				0.50		
$p = 0.75$						
β_0	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.71	−0.29	0.09	1.00	0.00	0.02
σ_v	0.11	−0.09	0.01	0.20	0.00	0.00

Table 3 continued

	Basic SF model			ZISF model			Basic SF model			ZISF model			Basic SF model			ZISF model		
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
p		–		0.49	–0.26	0.18		–		0.67	–0.08	0.05		–		0.75	0.00	0.00
λ	1.27	0.27	0.33	1.31	0.31	0.35	2.54	0.54	0.72	2.19	0.19	0.28	6.96	1.96	19.52	5.05	0.06	0.43
σ	1.40	–0.02	0.03	1.54	0.13	0.13	0.99	–0.13	0.03	1.14	0.02	0.03	0.72	–0.30	0.09	1.02	0.00	0.02
$\log L$	–301.06			–300.72			–196.20			–193.79			–99.12			–74.13		
$\hat{E}[u_i\|\hat{\varepsilon}_i]$	0.84	0.64	0.69	0.42	0.22	0.34	0.72	0.52	0.39	0.25	0.05	0.11	0.52	0.32	0.13	0.20	0.00	0.02
$\hat{p}(z_i=1\|\hat{\varepsilon}_i)$				0.49						0.67						0.75		

difference is that we used $n = 200$ and 500 whereas they used $n = 500$ and 1,000.

There were some technical problems related to the facts that σ_u^2 is not identified when $p = 1$, and p is not identified when $\sigma_u^2 = 0$. We define $\hat{p} = 1$ when $\hat{\sigma}_u = 0$ and $\hat{\sigma}_u = 0$ when $\hat{p} = 1$. This would imply that when the OLS residuals have incorrect skew, the MLE would be θ^{**} with $\hat{p} = 1$. It was very seldom the case that $\hat{\sigma}_u^2 = 0$ or $\hat{p} = 1$ other than in the wrong skew cases.

4.1 Parameter estimation

Table 2 contains the mean, bias, and MSE of the various parameter estimates, for the basic stochastic frontier model and for the ZISF model, for the case that $n = 200$. We also present the mean, bias, and MSE of the technical inefficiency estimates, and the mean of the "posterior" probabilities of full efficiency.

Unsurprisingly, the basic stochastic frontier model performs poorly except when $p = 0$ (in which case it is correctly specified). This is true for all three values of λ. We over-estimate technical inefficiency, because we act as if all firms are inefficient, whereas in fact they are not. This bias is naturally bigger when p is bigger.

For the ZISF model, the results depend strongly on the value of λ. When $\lambda = 1$, the results are not very good. Note in particular the mean values of $\hat{p}$, which are 0.53, 0.49, 0.51, and 0.57 for $p = 0$, 0.25, 0.50, and 0.75, respectively. It is disturbing that the mean estimate of p does not appear to depend on the true value of p.

These problems are less severe for larger values of λ. The mean value of $\hat{p}$ when $p = 0$ is 0.33 for $\lambda = 2$ and 0.16 for $\lambda = 5$. The estimates are considerably better for the other values of p. So basically the model performs reasonably well when λ is large enough and p is not too close to zero.

Table 3 is similar to Table 2 except that it reports the results only for the cases of correct skew (i.e., wrong skew cases are not included). This makes almost no difference for $\lambda = 2$ or 5, because there are very few wrong skew cases when $\lambda = 2$ or 5. For $\lambda = 1$, it matters more. However, the conclusions given above really do not change.

Table 4 contains the same information as Table 2, except that now we have $n = 500$ rather than $n = 200$. The results are better for $n = 500$ than for $n = 200$, but a larger sample size does not really solve the problems that the ZISF model has in estimating p when $p = 0$ and/or $\lambda = 1$. For example, when $p = 0$, the mean $\hat{p}$ for $\lambda = 1, 2, 5$ is $0.53, 0.33, 0.16$ when $n = 200$ and $0.50, 0.30, 0.12$ when $n = 500$. Reading the table in the other direction, when $\lambda = 1$, the mean $\hat{p}$ for $p = 0, 0.25, 0.5, 0.75$ is $0.53, 0.49, 0.51, 0.57$ when $n = 200$ and $0.50, 0.46, 0.48, 0.58$

Table 4 Basic SF model versus ZISF model, all replications: $n = 500$

Panel $p = 0$

| | Basic SF model ($\lambda=1$, $\sigma_u=1$, $\sigma_v=1$) | | | ZISF model | | | Basic SF model ($\lambda=2$, $\sigma_u=1$, $\sigma_v=0.5$) | | | ZISF model | | | Basic SF model ($\lambda=5$, $\sigma_u=1$, $\sigma_v=0.2$) | | | ZISF model | | |
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	0.90	−0.10	0.11	0.59	−0.41	0.26	0.99	−0.01	0.01	0.74	−0.26	0.11	1.00	0.00	0.00	0.88	−0.12	0.03
σ_u	0.88	−0.12	0.17	0.94	−0.06	0.24	0.99	−0.01	0.01	0.95	−0.05	0.01	1.00	0.00	0.00	0.97	−0.03	0.00
σ_v	1.01	0.01	0.01	1.04	0.04	0.01	0.50	0.00	0.00	0.54	0.04	0.01	0.20	0.00	0.00	0.22	0.02	0.00
p		—		0.50	0.50	0.38		—		0.30	0.30	0.15		—		0.12	0.12	0.03
λ	0.91	−0.09	0.21	0.93	−0.07	0.23	2.02	0.02	0.15	1.79	−0.21	0.19	5.18	0.18	0.99	4.55	−0.45	1.35
σ	1.39	−0.02	0.02	1.46	0.05	0.08	1.11	0.00	0.00	1.10	−0.02	0.01	1.02	0.00	0.00	0.99	−0.03	0.00
$\log L$	−785.23			−785.03			−577.71			−577.38			−442.32			−441.94		
$\hat{E}[u_i\vert\hat{\varepsilon}_i]$	0.70	−0.10	0.40	0.39	−0.41	0.55	0.79	−0.01	0.15	0.54	−0.26	0.26	0.80	0.00	0.03	0.68	−0.12	0.06
$\hat{p}(z_i=1\vert\hat{\varepsilon}_i)$				0.50						0.30						0.12		

Panel $p = 0.25$

| | Basic SF model ($\lambda=1$, $\sigma_u=1$, $\sigma_v=1$) | | | ZISF model | | | Basic SF model ($\lambda=2$, $\sigma_u=1$, $\sigma_v=0.5$) | | | ZISF model | | | Basic SF model ($\lambda=5$, $\sigma_u=1$, $\sigma_v=0.2$) | | | ZISF model | | |
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.19	0.19	0.13	0.85	−0.15	0.12	1.26	0.26	0.07	1.00	0.00	0.04	1.22	0.22	0.05	1.01	0.01	0.01
σ_u	0.99	−0.01	0.13	1.03	0.03	0.19	1.08	0.08	0.01	1.02	0.02	0.01	1.04	0.04	0.00	1.01	0.01	0.00
σ_v	0.99	−0.01	0.01	1.03	0.03	0.01	0.46	−0.04	0.00	0.50	0.00	0.00	0.14	−0.06	0.00	0.19	−0.01	0.00
p		—		0.46	0.21	0.16		—		0.27	0.02	0.04		—		0.24	−0.01	0.01
λ	1.04	0.04	0.19	1.03	0.03	0.19	2.39	0.39	0.32	2.08	0.08	0.18	7.55	2.55	8.46	5.36	0.36	1.87
σ	1.44	0.03	0.02	1.50	0.09	0.07	1.18	0.06	0.01	1.15	0.03	0.00	1.05	0.03	0.00	1.03	0.01	0.00
$\log L$	−790.14			−789.94			−584.97			−584.58			−436.65			−433.87		
$\hat{E}[u_i\vert\hat{\varepsilon}_i]$	0.79	0.19	0.42	0.45	−0.15	0.42	0.86	0.26	0.22	0.60	0.00	0.18	0.82	0.22	0.08	0.61	0.01	0.04
$\hat{p}(z_i=1\vert\hat{\varepsilon}_i)$				0.46						0.27						0.24		

Panel $p = 0.5$

| | Basic SF model ($\lambda=1$, $\sigma_u=1$, $\sigma_v=1$) | | | ZISF model | | | Basic SF model ($\lambda=2$, $\sigma_u=1$, $\sigma_v=0.5$) | | | ZISF model | | | Basic SF model ($\lambda=5$, $\sigma_u=1$, $\sigma_v=0.2$) | | | ZISF model | | |
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.45	0.45	0.26	1.05	0.05	0.09	1.45	0.45	0.21	1.05	0.05	0.03	1.30	0.30	0.09	1.00	0.00	0.00
σ_u	1.06	0.06	0.10	1.10	0.10	0.14	1.07	0.07	0.01	1.02	0.02	0.01	0.93	−0.07	0.01	1.00	0.00	0.00
σ_v	0.94	−0.06	0.01	0.99	−0.01	0.01	0.40	−0.10	0.01	0.49	−0.01	0.00	0.12	−0.08	0.01	0.20	0.00	0.00
p		—		0.48	−0.02	0.10		—		0.46	−0.04	0.04		—		0.50	0.00	0.00
λ	1.17	0.17	0.19	1.13	0.13	0.16	2.72	0.72	0.68	2.14	0.14	0.15	8.01	3.01	10.75	5.03	0.03	0.17
σ	1.45	0.04	0.02	1.51	0.10	0.06	1.15	0.03	0.00	1.14	0.02	0.01	0.93	−0.09	0.01	1.02	0.01	0.00
$\log L$	−779.58			−779.28			−556.16			−554.57			−371.51			−347.55		
$\hat{E}[u_i\vert\hat{\varepsilon}_i]$	0.85	0.45	0.53	0.45	0.05	0.36	0.85	0.45	0.34	0.45	0.05	0.15	0.70	0.30	0.12	0.40	0.00	0.02
$\hat{p}(z_i=1\vert\hat{\varepsilon}_i)$				0.48						0.46						0.50		

Panel $p = 0.75$

| | Basic SF model ($\lambda=1$, $\sigma_u=1$, $\sigma_v=1$) | | | ZISF model | | | Basic SF model ($\lambda=2$, $\sigma_u=1$, $\sigma_v=0.5$) | | | ZISF model | | | Basic SF model ($\lambda=5$, $\sigma_u=1$, $\sigma_v=0.2$) | | | ZISF model | | |
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
β_0	1.56	0.56	0.38	1.14	0.14	0.09	1.52	0.52	0.27	1.02	0.02	0.01	1.32	0.32	0.10	1.00	0.00	0.00
σ_u	0.95	−0.05	0.11	1.05	0.05	0.19	0.91	−0.09	0.01	1.00	0.00	0.02	0.71	−0.29	0.09	1.00	0.00	0.01
σ_v	0.91	−0.09	0.02	0.97	−0.03	0.01	0.38	−0.12	0.02	0.49	−0.01	0.00	0.11	−0.09	0.01	0.20	0.00	0.00

Table 4 continued

	Basic SF model			ZISF model			Basic SF model			ZISF model			Basic SF model			ZISF model		
	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse	mean	bias	mse
p	–			0.58	−0.17	0.14	–			0.72	−0.03	0.02	–			0.75	0.00	0.00
λ	1.08	0.08	0.19	1.10	0.10	0.21	2.43	0.43	0.32	2.04	0.04	0.07	6.36	1.36	2.77	5.00	0.00	0.16
σ	1.35	−0.06	0.02	1.47	0.06	0.08	0.99	−0.13	0.02	1.12	0.00	0.01	0.72	−0.30	0.09	1.02	0.00	0.01
log L	−754.00			−753.53			−492.83			−487.45			−250.31			−188.10		
$\hat{E}[u_i\|\hat{\varepsilon}_i]$	0.76	0.56	0.58	0.34	0.14	0.27	0.72	0.52	0.39	0.21	0.02	0.09	0.52	0.32	0.13	0.20	0.00	0.01
$\hat{p}(z_i=1\|\hat{\varepsilon}_i)$	0.76			0.58						0.72						0.75		

when $n = 500$. So again there are problems in estimating p when $p = 0$ or when λ is small.

It is perhaps not surprising that we encounter problems when we estimate the ZISF model when the true value of p is zero. Essentially, we are estimating a latent-class model with more classes than there really are. It is true that the class with zero probability contains no new parameters. If it did, they would not be identified and the results would presumably be much worse.

These results do not always agree with the summary of the results in KPT. KPT concentrate on the technical inefficiency estimates, and the only results they show explicitly for the parameter estimates (their Figure 3) are for $n = 1,000$, and $\lambda = 5$ and $p = 0.25$. We did successfully replicate their results, but $n = 1,000$ and $\lambda = 5$ is a very favorable parameter configuration. In their Sect. 3.1, they say the following about the case when the true p equals zero: "The ML estimator from the ZISF model is found to perform quite well. … Estimates of p were close to zero." It is not clear what parameter configuration this refers to, but in our simulations this is not true except when $\lambda = 5$. For smaller values of λ, the ZISF estimates of p when the true $p = 0$ are not very close to zero.

4.2 Testing the hypothesis $p = 0$

We now turn to the results of our simulations that are designed to investigate the size and power properties of the tests of the hypothesis $p = 0$, as discussed in Sect. 3.4 above. This hypothesis is economically interesting, and it is also practically important to know whether $p = 0$, since our model does not appear to perform well in that case. We would like to be able to recognize cases when $p = 0$ and just use the basic SF model in these cases.

The data generating process and parameter values for these simulations are as discussed above (in the beginning of Sect. 4). Specifically, the simulations are for $n = 200$ and $n = 500$.

We begin with the likelihood ratio (LR) test, which is the test that we believed ex ante would be most reliable. The results for $n = 200$ are given in Table 5. For each value of λ and p, we give the mean of the statistic (over the full set of 1,000 replications), the number of rejections and the frequency of rejection. The rejection rates in the rows corresponding to $p = 0$ are the size of the test, whereas the rejection rates in the rows corresponding to the positive values of p represent power.

Look first at the set of results for all replications. The size of the test is reasonable. It is undersized for $\lambda = 1$ and approximately correctly sized for $\lambda = 2$ and 5. However,

Table 5 Likelihood ratio test, $n = 200$

	All			Correct skew			Incorrect skew		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	0.29	21 (0.02)	1,000	0.38	21 (0.03)	776	0.00	0 (0.00)	224
$p = 0.25$	0.32	20 (0.02)	1,000	0.37	20 (0.02)	842	0.00	0 (0.00)	158
$p = 0.5$	0.46	40 (0.04)	1,000	0.53	40 (0.05)	878	0.00	0 (0.00)	122
$p = 0.75$	0.57	53 (0.05)	1,000	0.67	53 (0.06)	850	0.00	0 (0.00)	150
$\lambda = 2$									
$p = 0$	0.56	42 (0.04)	1,000	0.56	42 (0.04)	994	0.00	0 (0.00)	6
$p = 0.25$	0.66	63 (0.06)	1,000	0.66	63 (0.06)	999	0.00	0 (0.00)	1
$p = 0.5$	1.87	244 (0.24)	1,000	1.87	244 (0.24)	1,000		–	0
$p = 0.75$	4.81	596 (0.60)	1,000	4.81	596 (0.60)	999	0.00	0 (0.00)	1
$\lambda = 5$									
$p = 0$	0.73	60 (0.06)	999[+]	0.73	61 (0.06)	999[+]		–	0
$p = 0.25$	2.76	393 (0.39)	996[+]	2.76	395 (0.40)	996[+]		–	0
$p = 0.5$	19.82	988 (0.99)	997[+]	19.82	988 (0.99)	997[+]		–	0
$p = 0.75$	49.98	997 (1.00)	997[+]	49.98	997 (1.00)	997[+]		–	0

[+] Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined

Table 6 Likelihood ratio test, $n = 500$

	All			Correct skew			Incorrect skew		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	0.39	30 (0.03)	1,000	0.45	30 (0.03)	879	0.00	0 (0.00)	121
$p = 0.25$	0.41	28 (0.03)	1,000	0.45	28 (0.03)	921	0.00	0 (0.00)	79
$p = 0.5$	0.60	48 (0.05)	1,000	0.62	48 (0.05)	964	0.00	0 (0.00)	36
$p = 0.75$	0.95	102 (0.10)	1,000	1.01	102 (0.11)	939	0.00	0 (0.00)	61
$\lambda = 2$									
$p = 0$	0.66	56 (0.06)	1,000	0.66	56 (0.06)	1,000		–	0
$p = 0.25$	0.77	63 (0.06)	1,000	0.77	63 (0.06)	1,000		–	0
$p = 0.5$	3.19	461 (0.46)	1,000	3.19	461 (0.46)	1,000		–	0
$p = 0.75$	10.76	911 (0.91)	1,000	10.76	911 (0.91)	1,000		–	0
$\lambda = 5$									
$p = 0$	0.75	70 (0.07)	1,000	0.75	70 (0.07)	1,000		–	0
$p = 0.25$	5.55	689 (0.69)	1,000	5.55	689 (0.69)	1,000		–	0
$p = 0.5$	47.94	1,000 (1.00)	1,000	47.94	1,000 (1.00)	1,000		–	0
$p = 0.75$	124.42	1,000 (1.00)	1,000	124.42	1,000 (1.00)	1,000		–	0

the power is disappointing, except when λ is large. There is essentially no power, even against the alternative $p = 0.75$, when $\lambda = 1$. When $\lambda = 2$, power is 0.60 against $p = 0.75$, but only 0.24 against $p = 0.50$ and 0.06 against $p = 0.25$. Power is more reasonable when $\lambda = 5$.

Table 6 gives the same results for $n = 500$. Increasing n has little effect on the size of the test, but it improves the

power. Power is still low when $\lambda = 1$ or when $\lambda = 2$ and p is not large.

In either case ($n = 200$ or 500), looking separately at the correct-skew cases does not change our conclusions.

In Tables 7 and 8, we give results for the Wald test, for $n = 200$ and 500, respectively. Since the Wald test is undefined in wrong-skew cases, we show the results only

Table 7 Wald test, $n = 200$

	OPG			Hessian			Robust		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	5.92	128 (0.17)	773[*]	57.97	189 (0.24)	776	143.75	546 (0.70)	776
$p = 0.25$	4.42	147 (0.18)	838[*]	36.90	215 (0.26)	842	104.61	592 (0.70)	842
$p = 0.5$	6.98	179 (0.21)	873[*]	40.61	270 (0.31)	878	135.57	629 (0.72)	878
$p = 0.75$	9.74	247 (0.29)	849[*]	63.53	334 (0.39)	850	146.11	607 (0.71)	850
$\lambda = 2$									
$p = 0$	6.18	215 (0.22)	994	21.23	290 (0.29)	994	45.46	620 (0.62)	994
$p = 0.25$	4.26	264 (0.26)	999	10.21	320 (0.32)	997[◇]	19.28	618 (0.62)	999
$p = 0.5$	10.91	580 (0.58)	1,000	14.37	639 (0.64)	1,000	19.46	735 (0.73)	1,000
$p = 0.75$	45.72	856 (0.86)	999	61.79	883 (0.88)	998[◇]	80.71	906 (0.91)	999
$\lambda = 5$									
$p = 0$	3.25	266 (0.27)	999[+]	3.58	315 (0.32)	996[◇+]	5.03	490 (0.49)	999[+]
$p = 0.25$	8.63	696 (0.70)	998[+]	9.16	725 (0.73)	997[◇ +]	9.90	727 (0.73)	998[+]
$p = 0.5$	59.73	997 (1.00)	998[+]	60.75	997 (1.00)	998[+]	61.12	996 (1.00)	998[+]
$p = 0.75$	247.62	1,000 (1.00)	1,000	254.99	1,000 (1.00)	1,000	257.77	1,000 (1.00)	1,000

[*] Some iterations are dropped due to a singular OPG variance matrix

[◇] Some of the iterations where MLE is at the boundary ($\hat{p} = 0$) are dropped due to not negative definite Hessian

[+] Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined

Table 8 Wald test, $n = 500$

	OPG			Hessian			Robust		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	12.05	201 (0.23)	878[*]	112.28	250 (0.29)	877[◇]	286.90	620 (0.71)	878[°]
$p = 0.25$	10.34	203 (0.22)	921	94.59	264 (0.29)	921	215.21	639 (0.69)	921
$p = 0.5$	12.69	275 (0.29)	963[*]	47.99	347 (0.36)	964	121.96	661 (0.69)	964
$p = 0.75$	24.74	368 (0.39)	938[*]	120.86	447 (0.48)	937[◇]	258.85	678 (0.72)	939
$\lambda = 2$									
$p = 0$	5.32	262 (0.26)	1,000	26.94	310 (0.31)	1,000	47.13	618 (0.62)	1,000
$p = 0.25$	3.94	306 (0.31)	1,000	0.47	373 (0.37)	999[◇]	8.46	642 (0.64)	1,000
$p = 0.5$	17.10	800 (0.80)	1,000	19.10	831 (0.83)	1,000	22.49	837 (0.84)	1,000
$p = 0.75$	93.38	988 (0.99)	1,000	105.45	987 (0.99)	1,000	121.46	983 (0.98)	1,000
$\lambda = 5$									
$p = 0$	3.02	282 (0.28)	1,000	3.24	311 (0.31)	1,000	4.79	496 (0.50)	1,000
$p = 0.25$	17.87	890 (0.89)	1,000	18.79	894 (0.89)	1,000	20.02	893 (0.89)	1,000
$p = 0.5$	142.55	1,000 (1.00)	1,000	143.99	1,000 (1.00)	1,000	144.73	1,000 (1.00)	1,000
$p = 0.75$	609.81	1,000 (1.00)	1,000	618.15	1,000 (1.00)	1,000	621.49	1,000 (1.00)	1,000

[*] Some iterations are dropped due to a singular OPG variance matrix

[◇] Some of the iterations where MLE is at the boundary ($\hat{p} = 0$ or $\hat{p} = 1$) are dropped due to not negative definite Hessian

[°] One iteration dropped due to $\hat{p} = 1$

for the correct-skew cases. We consider separately the OPG, Hessian, and Robust forms of the test, as defined in Sect. 3.4.2 above. Regardless of which form of the test is used, the test is considerably over-sized. This is true for both sample sizes. The problem is worst for the Robust form and least serious for the OPG form, but there are serious size distortions in all three cases. Based on these results, the Wald test is not recommended.

In Tables 9 and 10, we give the results for the score-based tests (LM, modified LM, and KT). Once again the tests are undefined for wrong-skew cases so we report results only for the correct-skew cases. The (two-sided) LM test is the best of the three. It shows moderate size distortions and no power when $\lambda = 1$, but only modest size distortions when $\lambda = 2$ or 5. The modified LM test has bigger size distortions and less power when $\lambda = 2$ or 5. The KT test has the largest size distortions and is therefore not recommended.

Our results are easy to summarize. The likelihood ratio test is the best of the five tests we have considered, at least for these parameter values. It is the only one of the tests that does not over-reject the true null that $p = 0$. However,

Table 9 Score-based tests, $n = 200$

	LM			Modified LM			KT		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	1.83	99 (0.13)	776	1.47	122 (0.16)	776	1.82	146 (0.19)	776
$p = 0.25$	1.83	101 (0.12)	840[*]	1.48	135 (0.16)	840[*]	1.83	160 (0.19)	840[*]
$p = 0.5$	1.74	112 (0.13)	878	1.24	116 (0.13)	878	1.74	161 (0.18)	878
$p = 0.75$	1.74	100 (0.12)	849[*]	1.07	88 (0.10)	849[*]	1.74	152 (0.18)	849[*]
$\lambda = 2$									
$p = 0$	1.30	73 (0.07)	994	0.94	105 (0.11)	994	1.29	135 (0.14)	994
$p = 0.25$	1.20	75 (0.08)	999	0.77	96 (0.10)	999	1.19	130 (0.13)	999
$p = 0.5$	1.69	140 (0.14)	1,000	0.18	16 (0.02)	1,000	1.68	222 (0.22)	1,000
$p = 0.75$	3.82	422 (0.42)	999	0.04	3 (0.00)	999	3.82	531 (0.53)	999
$\lambda = 5$									
$p = 0$	1.16	58 (0.06)	999[+]	0.71	79 (0.08)	999[+]	1.11	113 (0.11)	999[+]
$p = 0.25$	1.78	139 (0.14)	996[+]	0.12	7 (0.01)	996[+]	1.73	216 (0.22)	996[+]
$p = 0.5$	13.10	961 (0.96)	997[+]	0.00	0 (0.00)	997[+]	13.10	978 (0.98)	997[+]
$p = 0.75$	29.46	996 (1.00)	997[+]	0.00	0 (0.00)	997[+]	29.46	996 (1.00)	997[+]

[*] Some iterations are dropped due to a singular OPG variance matrix

[+] Some iterations dropped due to $\hat{\sigma}_v$ being too small such that $\hat{\lambda}$ is not well defined

Table 10 Score-based tests, $n = 500$

	LM			Modified LM			KT		
	Mean	Rejection	Total	Mean	Rejection	Total	Mean	Rejection	Total
$\lambda = 1$									
$p = 0$	1.39	84 (0.10)	878[*]	0.99	104 (0.12)	878[*]	1.39	130 (0.15)	878[*]
$p = 0.25$	1.36	81 (0.09)	921	0.98	100 (0.11)	921	1.36	127 (0.14)	921
$p = 0.5$	1.26	73 (0.08)	964	0.70	79 (0.08)	964	1.26	125 (0.13)	964
$p = 0.75$	1.37	86 (0.09)	939	0.45	50 (0.05)	939	1.37	151 (0.16)	939
$\lambda = 2$									
$p = 0$	1.21	76 (0.08)	1,000	0.79	89 (0.09)	1,000	1.20	127 (0.13)	1,000
$p = 0.25$	1.07	48 (0.05)	1,000	0.62	68 (0.07)	1,000	1.06	107 (0.11)	1,000
$p = 0.5$	2.59	249 (0.25)	1,000	0.03	2 (0.00)	1,000	2.57	370 (0.37)	1,000
$p = 0.75$	8.14	795 (0.80)	1,000	0.00	0 (0.00)	1,000	8.14	887 (0.89)	1,000
$\lambda = 5$									
$p = 0$	1.06	56 (0.06)	1,000	0.62	69 (0.07)	1,000	0.97	109 (0.11)	1,000
$p = 0.25$	2.97	280 (0.28)	1,000	0.03	1 (0.00)	1,000	2.92	415 (0.41)	1,000
$p = 0.5$	30.94	1,000 (1.00)	1,000	0.00	0 (0.00)	1,000	30.94	1,000 (1.00)	1,000
$p = 0.75$	69.77	1,000 (1.00)	1,000	0.00	0 (0.00)	1,000	69.77	1,000 (1.00)	1,000

[*] Some iterations are dropped due to a singular OPG variance matrix

it does not have much power. That is, we will have trouble rejecting the hypothesis that the basic SF model is correctly specified, even if the ZISF model is needed and p is not close to zero. The exception to this pessimistic conclusion is the case when both p and λ are large, in which case the power of the test is satisfactory.

$$
\begin{aligned}
u_i &\sim N^+(0, \sigma_i^2), \\
\sigma_i^2 &= \exp\left(\gamma_0 + age_i\gamma_1 + edyrs_i\gamma_2 + hhsize_i\gamma_3 \right. \\
&\quad \left. + nadult_i\gamma_4 + banrat_i\gamma_5\right),
\end{aligned}
\tag{11}
$$

$$
P(z_i = 1|w_i) = \frac{\exp\left(\delta_0 + age_i\delta_1 + edyrs_i\delta_2 + hhsize_i\delta_3 + nadult_i\delta_4 + banrat_i\delta_5\right)}{1 + \exp\left(\delta_0 + age_i\delta_1 + edyrs_i\delta_2 + hhsize_i\delta_3 + nadult_i\delta_4 + banrat_i\delta_5\right)}
\tag{12}
$$

5 Empirical example

We apply the models defined in Sects. 2 and 3 to the Philippine rice data used in the empirical examples of Coelli et al. (2005, chapters 8–9). The Philippine data are composed of 43 farmers over 8 years and Coelli et al. (2005) estimate the basic stochastic frontier model with a trans-log production function, ignoring the panel nature of the observations. Their output variable is tonnes of freshly threshed rice, and the input variables are planted area (in hectares), labor, and fertilizer used (in kilograms). These variables are scaled to have unit means so the first-order coefficients of the trans-log function can be interpreted as elasticities of output with respect to inputs evaluated at the variable means. We follow the basic setup of Coelli et al. (2005) but estimate the extended models where some farms are allowed to be efficient, and the probability of farm i being efficient and/or the distribution of u_i depend on farm characteristics. Data on age of household head, education of household head, household size, number of adults in the household, and the percentage of area classified as bantog (upland) fields are used as farm characteristics that influence the probability of a farm begin fully efficient and/ or the distribution of the inefficiency. See Coelli et al. (2005, Appendix 2) for a detailed description of the data.

5.1 Model

We consider models based on the following specification:

$$
\begin{aligned}
\ln y_i &= \beta_0 + \theta t + \beta_1 \ln area_i + \beta_2 \ln labor_i + \beta_3 \ln npk_i \\
&\quad + \beta_{11}\frac{1}{2}(\ln area_i)^2 + \beta_{12}\ln area_i \ln labor_i \\
&\quad + \beta_{13}\ln area_i \ln npk_i + \beta_{22}\frac{1}{2}(\ln labor_i)^2 \\
&\quad + \beta_{23}\ln labor_i \ln npk_i + \beta_{33}\frac{1}{2}(\ln npk_i)^2 + v_i - u_i,
\end{aligned}
\tag{10}
$$

where $area_i$ is the size of planted area in hectares, $labor_i$ is a measure of labor, npk_i is fertilizer in kilograms, age_i is the age of household head, $edyrs_i$ is the years of education of the household head, $hhsize_i$ is the household size, $nadult_i$ is the number of adults in the household, and $banrat_i$ is the percentage of area classified as bantog (upland) fields.

We assume a trans-log production function with time trend as in (10). We estimate the following models:

a. the basic stochastic frontier model, in which σ_i^2 is constant ($\equiv \sigma_u^2$) and $P(z_i = 1|w_i) = 0$;
b. the ZISF model in which σ_u^2 is constant and $P(z_i = 1|w_i)$ is constant ($\equiv p$) but not necessarily equal to zero;
c. the "heteroskedasticity" model in which $p = 0$ but σ_i^2 is as given in (11);
d. the "logit" model in which σ_u^2 is constant but $P(z_i = 1|w_i)$ is as given in (12);
e. the "logit + heteroskedasticity" model in which σ_i^2 is as given in (11) and $P(z_i = 1|w_i)$ is as given in (12).

5.2 The estimates

The MLEs and their OPG standard errors are reported in Table 11.

Consider first the results for the basic stochastic frontier model (first column of results in the table). The inputs are productive and there are roughly constant returns to scale. Average technical efficiency is about 70 %. The estimated value of λ is 2.75, and both that value and the sample size ($n = 344$) are big enough to feel confident about proceeding to the ZISF model and its extensions.

The next block of column of results is for the ZISF model. Here we have $\hat{p} = 0.58$, so a substantial fraction of the observations (farm-time period combinations) are characterized by full efficiency. The technology (effect of inputs on output) is not changed much from the basic SF model, but the intercept is lower and the level of technical

Table 11 Model comparison

Variable	Basic SF			ZISF			Heteroskedasticity			Logit			Logit+Heteroskedasticity		
	Coef	SE	t stat	Coef	SE	t stat	Coef	SE	t stat	Coef	SE	t stat	Coef	SE	t-stat
cons (β_0)	0.27	0.04	6.68	0.08	0.05	1.67	0.25	0.04	5.50	0.07	0.04	1.80	0.05	0.04	1.38
time period (θ)	0.02	0.01	2.27	0.01	0.01	2.19	0.02	0.01	2.60	0.02	0.01	2.57	0.02	0.01	2.60
area (β_1)	0.53	0.08	6.38	0.52	0.08	6.58	0.57	0.09	6.43	0.59	0.09	6.85	0.56	0.08	7.04
labor (β_2)	0.23	0.09	2.71	0.25	0.08	3.07	0.21	0.09	2.46	0.22	0.08	2.65	0.24	0.08	3.07
fertilizer (β_3)	0.20	0.05	3.95	0.21	0.05	4.54	0.19	0.05	3.83	0.18	0.05	3.82	0.18	0.04	3.96
β_{11}	-0.48	0.21	-2.27	-0.45	0.20	-2.26	-0.35	0.25	-1.39	-0.29	0.25	-1.13	-0.31	0.26	-1.19
β_{12}	0.61	0.17	3.60	0.55	0.17	3.21	0.58	0.20	2.87	0.53	0.21	2.60	0.50	0.21	2.35
β_{13}	0.06	0.15	0.43	0.08	0.13	0.62	0.01	0.15	0.07	0.00	0.14	0.02	0.02	0.14	0.16
β_{22}	-0.56	0.24	-2.33	-0.50	0.25	-2.04	-0.60	0.27	-2.21	-0.56	0.27	-2.08	-0.47	0.27	-1.73
β_{23}	-0.14	0.13	-1.04	-0.12	0.12	-1.02	-0.11	0.15	-0.78	-0.11	0.13	-0.80	-0.12	0.13	-0.91
β_{33}	-0.01	0.09	-0.08	-0.03	0.08	-0.45	0.01	0.09	0.15	0.00	0.09	-0.03	-0.01	0.09	-0.13
σ_u	0.44	0.03	13.86	0.44	0.04	10.87				0.42	0.04	10.43			
σ_{ui}							0.42						0.53		
cons (γ_0)							-3.05	0.93	-3.27				2.61	3.11	0.84
age (γ_1)							0.03	0.01	2.06				-0.07	0.04	-1.60
edyrs (γ_2)							0.11	0.03	3.36				-0.03	0.12	-0.28
hhsize (γ_3)							0.08	0.09	0.99				-0.28	0.22	-1.26
nadult (γ_4)							-0.10	0.11	-0.98				-0.02	0.23	-0.08
banrat (γ_5)							-1.30	0.43	-2.98				1.32	1.20	1.10
σ_v	0.16	0.02	8.23	0.20	0.02	12.22	0.17	0.02	8.34	0.20	0.01	14.15	0.21	0.01	15.84
λ	2.75			2.18			2.43			2.06			2.49		
p				0.58	0.11	5.42									
p_i				0.58						0.55			0.57		
$\widehat{\mathrm{pr}}[z_i = 1|\widehat{\varepsilon}]$				0.58						0.55			0.57		
cons (δ_0)										3.88	2.29	1.69	8.80	3.42	2.57
age (δ_1)										-0.10	0.03	-2.88	-0.16	0.05	-3.36
edyrs (δ_2)										-0.14	0.15	-0.93	-0.27	0.21	-1.30
hhsize (δ_3)										-0.44	0.23	-1.93	-0.81	0.32	-2.56
nadult (δ_4)										0.35	0.27	1.30	0.44	0.35	1.25
banrat (δ_5)										4.22	1.31	3.23	5.46	1.55	3.53
$\widehat{E}[u_i|\widehat{\varepsilon}_i]$	0.35			0.15			0.33			0.15			0.14		
$\exp\left(-\widehat{E}[u_i|\widehat{\varepsilon}_i]\right)$	0.70			0.86			0.72			0.86			0.87		
$\widehat{E}[\exp(-u_i)|\widehat{\varepsilon}_i]$	0.73			0.89			0.74			0.88			0.89		

Table 11 continued

Variable	Basic SF			ZISF			Heteroskedasticity			Logit			Logit+Heteroskedasticity		
	Coef	SE	t stat	Coef	SE	t stat	Coef	SE	t stat	Coef	SE	t-stat	Coef	SE	t-stat
ln L	−74.41			−71.88			−65.89			−59.90			−54.33		
Num. of parameter	13			14			18			19			24		
AIC	174.82			171.75			167.78			157.79			156.67		
BIC	224.75			225.52			236.91			230.76			248.84		
HQIC	194.70			193.17			195.31			186.86			193.38		

efficiency is higher (between 85 and 90 %). Based on our simulations, this is a predictable consequence of finding that a substantial number of observations are fully efficient.

The next block of columns of results is for the heteroskedasticity model in which all farms are inefficient but the level of inefficiency depends on farm characteristics. A number of farm characteristics (age of the farmer, education of the farmer and percentage of bantog fields) have significant effects on the level of inefficiency. In this parameterization, a positive coefficient indicates that an increase in the corresponding variable makes a farm more *in*efficient. The model implies that farms where the farmer is older and more educated, and where the percentage of bantog fields is lower, tend to be more inefficient (less efficient). Or, saying the same thing the other way around, farms are more efficient on average if the farmer is younger and less educated and the percentage of bantog fields is higher. The effect of education is perhaps surprising. Because this model does not allow any farms to be fully efficient, we once again have a low level of average technical efficiency, about 72 %, which is similar to that for the basic SF model.

Next we consider the logit model in which the distribution of inefficiency is the same for all firms that are not fully efficient, but the probability of being fully efficient depends on farm characteristics according to a logit model. Now age of the farmer and percentage of bantog fields have significant effects on the probability of full efficiency, and the coefficient of household size is almost significant at the 5 % level (t statistic $= -1.93$). The results indicate that farms with younger farmers, smaller household size, and a larger proportion of bantog fields are more likely to be fully efficient. The results for age of the farmer and percentage of bantog fields are similar in nature to those for the heteroskedasticity model. The average level of inefficiency is once again higher, about 86 %, which is very similar to the result for the ZISF model with constant p.

Finally, the last set of results are for the logit + heteroskedasticity model in which farm characteristics influence both the probability of being fully efficient and the distribution of inefficiency for those farms that are not fully efficient. Now none of the farm characteristics considered have significant effects on the distribution of inefficiency for the inefficient farms, but three of them (age of the farmer, household size and proportion of bantog fields) do have significant effects on the probability of being fully efficient. The coefficients for these three variables have the same signs as in the logit model without heteroskedasticity. It is interesting that we can estimate a model this complicated and still get significant results. Also, we note that, because this model allows the probability of full efficiency, we are back to a high average level of technical inefficiency, between 85 and 90 %.

5.3 Model comparison and selection

We will now test the restrictions that distinguish the various models we have estimated. Based on the results of our simulations, we will use the likelihood ratio (LR) test. We immediately encounter some difficulties because, to use the LR test (or the other tests we considered in Sect. 3.4), the hypotheses should be nested, whereas not all of our models are nested. There are two possible nested hierarchies of models: (a) basic SF$\subset$ ZISF$\subset$logit$\subset$logit-heteroskedasticity, and (b) basic SF$\subset$ heteroskedasticity.

We begin with hierarchy (a). When we test the hypothesis that $p = 0$ in the ZISF model, we obtain LR = 5.07, which exceeds the 5 % critical value of 2.71 for the distribution $(1/2\chi_1^2 + 1/2\chi_1^2)$. So we reject the basic SF model in favor of the ZISF model. Next we test the ZISF model against the logit model. This is a standard test of the hypothesis that $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$ in the logit model. The LR statistic of 23.96 exceeds the 5 % critical value for the χ_5^2 distribution (11.07), so we reject the ZISF model in favor of the logit model. Finally, we test the logit model against the logit-heteroskedasticity model. This is a standard test of the hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$ in the logit-heteroskedasticity model. The LR test statistic of 11.12 very marginally exceeds the 5% critical value, so we reject the logit model in favor of the logit-heteroskedasticity model, but not overwhelmingly. Note that the logit model is rejected even though, in the logit-heteroskedasticity model, none of the individual γ_j in the heteroskedasticity portion of the model is individually significant.

Now consider hierarchy (b). We test the basic SF model against the heteroskedasticity model. This is a standard test of the hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$ in the heteroskedasticity model. The LR statistic of 17.04 exceeds the 5 % critical value, so we reject the basic SF model in favor of the heteroskedasticity model.

We cannot test the heteroskedasticity model against the logit-heteroskedasticity model, at least not by standard methods, since the restriction that would convert the logit-heteroskedasticity model into the heteroskedasticity model is $\delta_0 = -\infty$ and under this null the other δ_j are unidentified. Still, the difference in log-likelihoods, which is 11.56, would appear to argue in favor of the logit-heteroskedasticity model.

In order to compare the models in a slightly different way, and to amplify on the comment at the end of the preceding paragraph, we will also consider some standard model selection criteria. We consider AIC = $-2LF + 2d$ (Akaike 1974), BIC = $-2LF + d\ln n$ (Schwarz 1978) and HQIC = $-2LF + 2d\ln(\ln n)$ (Hannan and Quinn 1979), where d is the number of estimated parameters, n is the number of observations, and LF is the log-likelihood value. Smaller values of these criteria indicate a "better" model. We note that all three criteria favor the logit model over the heteroskedasticity model, two of the three favor the logit-heteroskedasticity model over the heteroskedasticity model, and two of the three favor the logit model over the logit-heteroskedasticity model.

Based on the results of our hypothesis tests and the comparison of the model selection procedures, we conclude that a case could be made for either the logit model or the logit-heteroskedasticity model as the preferred model. As we saw above, the substantive conclusions from these two models were basically the same.

6 Concluding remarks

In this paper we considered a generalization of the usual stochastic frontier model. In this new "ZISF" model, there is a probability p that a firm is fully efficient. This model was proposed by Kumbhakar et al. (2013), who showed how to estimate the model by MLE, how to update the probability of a firm being fully efficient on the basis of the data, and how to estimate the inefficiency level of a specific firm.

We extend their analysis in a number of ways. We show that a result similar to that of Waldman (1982) holds in the ZISF model, namely, that there is always a stationary point of the likelihood at parameter values that indicate no inefficiency, and that this point is a local maximum if the OLS residuals are positively skewed. We show how to test the hypothesis that $p = 0$. We also provide a more comprehensive set of simulations than Kumbhakar et al. (2013) did.

Let $\lambda = \sigma_u/\sigma_v$, a standard measure in the stochastic frontier literature of the relative size of technical inefficiency and statistical noise. The main practical implication of our simulations is that the ZISF model works well when neither λ nor p is small. However, we have trouble estimating p reliably, or testing whether it equals zero, when λ is small. And if the true p equals zero, we have trouble estimating it reliably unless λ is larger than is empirically plausible (e.g., $\lambda = 5$). Larger sample size obviously helps, but the above conclusions do not depend strongly on sample size in our simulations. Situations where the ZISF model may be useful therefore have the characteristics that (1) it is reasonable to suppose that some firms are fully efficient, and (2) the inefficiency levels of the inefficient firms are not small relative to statistical noise. Such situations do not seem implausible, and it is an empirical question as to how common they are.

Appendix

We will use the following notation. Let $f_v(\varepsilon_i) = \sqrt{\frac{1+\lambda^2}{\sigma^2}}\phi\left(\varepsilon_i\sqrt{\frac{1+\lambda^2}{\sigma^2}}\right)$, $f_\varepsilon(\varepsilon_i) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)\left(1-\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right)$, $f_p(\varepsilon_i) = pf_v(\varepsilon_i) + (1-p)f_\varepsilon(\varepsilon_i)$, $\ln L = \sum\ln f_p(\varepsilon_i)$, $m_i = \frac{\phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)}{1-\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)}$, $\theta = (\beta', \lambda, \sigma^2, p)'$, $\beta = k \times 1$ vector, $\theta^{**} = \left(\hat{\beta}', \hat{\lambda}, \hat{\sigma}^2, \hat{p}\right)'$, where $\hat{\beta} = $ OLS, $\hat{\lambda} = 0$, $\hat{\sigma}^2 = \frac{1}{n}\sum\hat{\varepsilon}_i^2$, $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$, $\hat{p} \in [0, 1]$. $\vee$ indicates maximum.

Result 1 θ^{**} *is a stationary point of the log likelihood function.*

Proof The first derivative of $\ln L$ is:

$$S(\theta) = \begin{pmatrix} \frac{\partial\ln L}{\partial\beta} \\ \frac{\partial\ln L}{\partial\lambda} \\ \frac{\partial\ln L}{\partial\sigma^2} \\ \frac{\partial\ln L}{\partial p} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(\frac{1+\lambda^2}{\sigma^2}\varepsilon_i\mathbf{x}_i\right)+(1-p)f_\varepsilon(\varepsilon_i)\left(\frac{\varepsilon_i\mathbf{x}_i}{\sigma^2}+\frac{m_i\mathbf{x}_i\lambda}{\sigma}\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(\frac{\lambda}{1+\lambda^2}-\frac{\lambda}{\sigma^2}\varepsilon_i^2\right)-(1-p)f_\varepsilon(\varepsilon_i)\left(\frac{1}{\sigma}m_i\varepsilon_i\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{pf_v(\varepsilon_i)\left(-\frac{1}{2\sigma^2}+\frac{1+\lambda^2}{2\sigma^4}\varepsilon_i^2\right)+(1-p)f_\varepsilon(\varepsilon_i)\left(-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\varepsilon_i^2+\frac{\lambda}{2\sigma^3}m_i\varepsilon_i\right)}{f_p(\varepsilon_i)} \\ \sum_{i=1}^{n} \frac{f_v(\varepsilon_i)-f_\varepsilon(\varepsilon_i)}{f_p(\varepsilon_i)} \end{pmatrix}.$$

When $\lambda = 0$,

$$S(\theta)|_{\lambda=0} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}\varepsilon_i\mathbf{x}_i \\ -(1-p)\sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\sum_{i=1}^{n}\varepsilon_i \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\varepsilon_i^2 \\ 0 \end{pmatrix}.$$

It is straightforward that $S(\theta^{**}) = 0$, since, with $\varepsilon_i = \hat{\varepsilon}_i$, $\sum_{i=1}^{n}\hat{\varepsilon}_i = 0$ and $\sum_{i=1}^{n}\hat{\varepsilon}_i\mathbf{x}_i = 0$. Therefore, θ^{**} is a stationary point. $\square$

Result 2 *Evaluated at the stationary point, θ^{**} the Hessian of the log likelihood is negative semi-definite with two zero eigenvalues.*

Proof The Hessian evaluated at the stationary point θ^{**} is

$H(\theta^{**})$

$$= \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i & 0 & 0 \\ (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i' & -(1-\hat{p})^2\frac{2n}{\pi} & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

When $\hat{p} = 1$,

$$H(\theta^{**}) = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Because $-\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'$ is a negative definite matrix, $H(\theta^{**})$ is a negative semi-definite matrix with two zero eigenvalues.

Now suppose that $\hat{p} \neq 1$. Note that the first row of $H(\theta^{**})$, $\left(-\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i', (1-\hat{p})\sqrt{\frac{2n}{\pi}}\frac{1}{\hat{\sigma}}, 0, 0\right)$, is linearly dependent with the $(k+1)$th row of $H(\theta^{**})$. Multiplying the first row by $(1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}$ and adding to the $(k+1)$th row results in a row vector of zeros. Hence,

$$H(\theta^{**}) \sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\sim$ stands for an elementary row operation. Again, the first column and the $(k+1)$th column of the transferred matrix are linearly dependent. Similarly, multiplying the first column by $(1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}$ and adding to the $(k+1)$th column results in a column vector of zeros. In other words,

$$H(\theta^{**}) \sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & (1-\hat{p})\sqrt{\frac{2}{\pi}}\frac{1}{\hat{\sigma}}\sum_{i=1}^{n}\mathbf{x}_i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{n}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Elementary operations preserve the rank of a matrix. Hence, the rank of $H(\theta^{**})$ is $k + 1$, i.e., $H(\theta^{**})$ has two zero eigenvalues.

Now we will show that $H(\theta^{**})$ is negative semi-definite. Let $\alpha = (\alpha_1', \alpha_2, \alpha_3, \alpha_4)'$ be an arbitrary non-zero $(k + 3) \times 1$ vector, where α_1 is a $k \times 1$ vector, and $\alpha_2, \alpha_3, \alpha_4$ are scalars. Then,

$$\alpha' H(\theta^{**})\alpha = -\left(\frac{1}{\hat{\sigma}}\frac{1}{\sqrt{n}}\alpha_1' \sum_{i=1}^{n} \mathbf{x}_i - \alpha_2(1-p)\sqrt{\frac{2}{\pi}}\sqrt{n}\right)^2$$

$$-\frac{1}{\hat{\sigma}^2}\alpha_1'\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i' - \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \sum_{i=1}^{n} \mathbf{x}_i'\right)\alpha_1 - \frac{n}{2\hat{\sigma}^4}\alpha_3^2$$

$$\leq 0,$$

because $\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i' - \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \sum_{i=1}^{n} \mathbf{x}_i'$

$$= \sum_{i=1}^{n}\left(\mathbf{x}_i - \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j\right)\left(\mathbf{x}_i - \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j\right)'$$

is positive semi-definite.

Therefore $H(\theta^{**})$ is negative semi-definite. $\square$

Result 3 θ^{**} with $\hat{p} \in [0, 1)$ is a local maximizer of the log likelihood function if and only if $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$.

Proof From Result 2, we know that the Hessian evaluated at θ^{**} is negative semi-definite. Therefore, if the log likelihood decreases in the direction of the two eigenvectors associated with zero eigenvalues, θ^{**} is a local maximizer of the log likelihood. The two eigenvectors that are associated with the two zero eigenvalue are

$$\begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma} \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Let

$$\Delta\theta = \mu\begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma} \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \phi\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} (1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu \\ 0 \\ \mu \\ 0 \\ \phi \end{pmatrix}.$$

Because $\lambda \geq 0$, $\mu > 0$. $\Delta\theta$ has only three non-zero arguments. Thus, relevant parameters would be β_0, λ, and p. By Taylor's expansion,

$$L(\theta^* + \Delta\theta) - L(\theta^*) = \frac{1}{6}\left[L_{\beta_0\beta_0\beta_0}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^3\right.$$

$$+3L_{\beta_0\beta_0\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^2\mu$$

$$+3L_{\beta_0\lambda\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\mu^2$$

$$+3L_{\beta_0\beta_0p}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)^2\phi$$

$$+3L_{\beta_0pp}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\phi^2 + L_{\lambda\lambda\lambda}\mu^3$$

$$+ 3L_{\lambda\lambda p}\mu^2\phi + 3L_{\lambda pp}\mu\phi^2 + L_{ppp}\phi^3$$

$$\left.+6L_{\beta_0p\lambda}\left((1-\hat{p})\sqrt{\frac{2}{\pi}}\hat{\sigma}\mu\right)\mu\phi\right]$$

$$+ o\left((\mu \vee \phi)^4\right)$$

$$= (1-\hat{p})\frac{1}{6\pi}\sqrt{\frac{2}{\pi\hat{\sigma}^3}}(-4\hat{p}^2$$

$$+\hat{p}(8-3\pi)+\pi-4)\sum_{i=1}^{n}\hat{\varepsilon}_i^3\mu^3$$

$$+ o\left((\mu \vee \phi)^4\right).$$

The 1st order term is zero because θ^{**} is a stationary point (Result 1). The 2nd order term is zero by the definition of the eigenvector. Note that $(-4\hat{p}^2 + \hat{p}(8-3\pi) + \pi - 4)$ has its maximum, $\pi - 4 < 0$, when $\hat{p} = 0$. Since $\mu > 0$, $L(\theta^* + \Delta\theta) - L(\theta^*) < 0$ if and only if $\sum \hat{\varepsilon}_i^3 > 0$. Therefore, θ^{**} with $\hat{p} \in [0, 1)$ is a local maximizer if and only if $\sum \hat{\varepsilon}_i^3 > 0$. When $\hat{p} = 0$, the expression goes back to the one in Waldman (1982). $\square$

Result 4 θ^{**} with $\hat{p} = 1$ is a local maximizer of the likelihood function if $\sum_{i=1}^{n} \hat{\varepsilon}_i^3 > 0$.

Proof The two eigenvectors associated with the zero eigenvalues are

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Let

$$\Delta\theta = \mu\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \phi\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \\ 0 \\ \phi \end{pmatrix}.$$

Because $\lambda \geq 0$ and $p \leq 1$, $\mu > 0$ and $\phi < 0$. $\Delta\theta$ has only two non-zero arguments. Thus, the relevant parameters would be λ and p. By Taylor's expansion,

$$
\begin{aligned}
L(\theta^* + \Delta\theta) - L(\theta^*) = & \frac{1}{24}\big[L_{\lambda\lambda\lambda\lambda}\mu^4 + 4L_{\lambda\lambda\lambda p}\mu^3\phi \\
& + 6L_{\lambda\lambda pp}\mu^2\phi^2 + 4L_{\lambda ppp}\mu\phi^3 + L_{pppp}\phi^4\big] \\
& + o\left((\mu \vee \phi)^5\right) \\
= & -\frac{1}{4}\mu^4 + \frac{1}{3\hat{\sigma}^3}\sqrt{\frac{2}{\pi}}\sum_{i=1}^{n}\hat{\varepsilon}_i^3\mu^3\phi - \frac{n}{\pi}\mu^2\phi^2 \\
& + o\left((\mu \vee \phi)^5\right)
\end{aligned}
$$

The 1st order term is zero because θ^{**} is a stationary point (Result 1). The 2nd order term is zero by the definition of the eigenvector. The third order term is zero because $L(\theta^* + \Delta\theta) - L(\theta^*)$ in Result 3 is zero when $\hat{p} = 1$. Since $\phi < 0$ and $\mu > 0$, $\frac{1}{3\hat{\sigma}^3}\sqrt{\frac{2}{\pi}}\sum_{i=1}^{n}\hat{\varepsilon}_i^3\mu^3\phi < 0$ when $\sum \hat{\varepsilon}_i^3 > 0$. Therefore, if $\sum \hat{\varepsilon}_i^3 > 0$, $L(\theta^* + \Delta\theta) - L(\theta^*) < 0$ and θ^{**} with $\hat{p} = 1$ is a local maximizer. $\square$

References

Aigner D, Lovell CK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econ 6:21–37

Akaike H (1974) A new look at the statistical model identication. IEEE Trans Autom Control 19:716–723

Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. J Prod Anal 25:201–212

Amsler C, Prokhorov A, Schmidt P (2014) Using copulas to model time dependence in stochastic frontier models. Econ Rev (forthcoming)

Andrews DWK (2001) Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69:683–734

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. J Econ 38:387–399

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir Econ 20:325–332

Berg SA, Førsund FR, Hjalmarsson L, Suominen M (1993) Banking efficiency in the Nordic countries. J Banking Financ 17:371–388

Caudill SB (2003) Estimating a mixture of stochastic frontier regression models via the em algorithm: a multiproduct cost function application. Empir Econ 28:581–598

Caudill SB, Ford JM (1993) Biases in frontier estimation due to heteroskedasticity. Econ Lett 41:17–20

Caudill SB, Ford JM, Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. J Bus Econ Stat 13:105–111

Chen Y, Liang K-Y (2010) On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. Biometrika 97:603–620

Coelli TJ, Rao DP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis. Springer, New York

Førsund FR, Hjalmarsson L (1974) On the measurement of productive efficiency. Swed J Econ 76:141–154

Gouriéroux C, Holly A, Monfort A (1982) Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50:63–80

Gouriéroux C, Monfort A (1995) Statistics and econometric models. In: Statistics and econometric models: testing, confidence regions, model selection, and asymptotic theory, vol 2. Cambridge University Press, New York

Grassetti L (2011) A novel mixture based stochastic frontier model with application to hospital efficiency. unpublished manuscript, University of Udine

Greene W (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econ 126:269–303

Greene WH (2012) Econometric analysis. Prentice Hall, New York

Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. J R Stat Soc Ser B (Methodological) 41:190–195

Hayashi F (2000) Econometrics. Princeton University Press, Princeton

Huang CJ, Liu J-T (1994) Estimation of a non-neutral stochastic frontier production function. J Prod Anal 5:171–180

Jondrow J, Lovell CK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econ 19:233–238

Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production Frontier approach for estimating determinants of inefficiency in U.S. dairy farms. J Bus Econ Stat 9:279–286

Kumbhakar SC, Parmeter CF, Tsionas EG (2013) A zero inefficiency stochastic Frontier model. J Econ 172:66–76

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb–Douglas production function with composed error. Int Econ Rev 18:435–444

Orea L, Kumbhakar SC (2004) Efficiency measurement using a latent class stochastic frontier model. Empir Econ 29:169–183

Reifschneider D, Stevenson R (1991) Systematic departures from the Frontier: a framework for the analysis of firm inefficiency. Int Econ Rev 32:715–723

Rogers AJ (1986) Modified lagrange multiplier tests for problems with one-sided alternatives. J Econ 31:341–361

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Waldman DM (1982) A stationary point for the stochastic frontier likelihood. J Econ 18:275–279

Wang H-J (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. J Prod Anal 18:241–253