



On hypothesis testing in latent class and finite mixture stochastic frontier models, with application to a contaminated normal-half normal model

Alexander D. Stead ¹ · Phill Wheat ¹ · William H. Greene ²

Accepted: 13 March 2023
© The Author(s) 2023

Abstract

Latent class and finite mixture stochastic frontier models have been proposed as a means of allowing either for technological heterogeneity or more flexible distributions of noise and inefficiency. As in the wider literature on latent class and finite mixture models, we are interested in class enumeration, particularly testing against homogeneity. We apply a modified likelihood ratio test for homogeneity in a stochastic frontier setting, based on established results for non-Gaussian latent class and finite mixture models, and provide evidence from Monte Carlo experiments which suggest the applicability of results regarding a modified likelihood ratio test to the stochastic frontier setting. We demonstrate an application to testing a model with a contaminated normal noise term against a model with a normally distributed noise term, finding that the former is preferred, with significant implications for efficiency prediction.

JEL Codes C12 · C46 · D24

Keywords Stochastic frontier analysis · Latent classes · Finite mixtures · Modified likelihood ratio

1 Introduction

In recent years, a number of latent class and finite mixture (LCFM) stochastic frontier (SF) models have been proposed. Orea and Kumbhakar (2004), Greene (2005), and Kutlu et al. (2020) propose latent class SF models in which there are J classes of firms with differing technologies, and class membership is unobserved. This is one way of allowing for technological heterogeneity between firms. The zero inefficiency stochastic frontier (ZISF) model proposed by Kumbhakar et al. (2013) and Rho and Schmidt (2015), in which an unknown proportion of firms are fully efficient, is also an LCFM model. Our illustrative example

for the present study is a SF model in which the noise distribution is contaminated normal—i.e. a scale mixture of normal distributions.

The use of LCFM models raises the issue class enumeration, i.e. determining the correct number of classes or mixture components to include in the model. A natural way to do this would be to use the likelihood ratio (LR) statistic to test more complex models against simpler null models with fewer classes or mixture components. However the well-known Wilks (1938) result that the asymptotic distribution of the LR statistic is χ_k^2 under the null hypothesis, where k is the number of restrictions, does not apply in this context because of identification issues and boundary issues under the null hypothesis. For this reason, a number of alternative approaches to class enumeration—see Nyland et al. (2007) for discussion and comparison—are common, such as the use of information criteria, as in Orea and Kumbhakar (2004), Greene (2005), and Kutlu et al. (2020) in the SF literature. Many information criteria penalise model complexity, though some others such as the complete likelihood classification (CLC) and the integrated classification likelihood (ICL) criterion (McLachlan and Peel

✉ Alexander D. Stead
a.d.stead@leeds.ac.uk

¹ Institute for Transport Studies, University of Leeds, Leeds LS2 9LJ, UK

² Department of Economics, Stern School of Business, New York University, New York, NY 10012, USA

2000) instead penalise poorly separated clusters—see Fonseca and Cardoso (2007) for a discussion.

We prefer hypothesis tests over information criteria for two reasons. First, although Kutlu et al. (2020) show that ICs can perform well in identifying the number of classes in an SF setting, Monte Carlo evidence suggests that likelihood-based tests outperform ICs in terms of selecting the number of classes or mixture components (Nylund et al. 2007). Second, we wish to conduct tests where the null hypothesis is interesting in economic or statistical terms. For example, in the LCFM SF model proposed of Orea and Kumbhakar (2004) and Greene (2005), testing against the nested standard SF model can be interpreted as testing the null hypothesis of technological homogeneity.

In this paper, we discuss the problem of hypothesis testing in LCFM models in the specific context of SF analysis. Although the problem in general can be very complex, we utilise results from the literature on likelihood-based tests for LCFM models for a class of models, including SF models, to derive results on the asymptotic distribution of a modified likelihood ratio (MLR) statistic under the null hypothesis. It is the non-Gaussian error distribution in the SF setting that allows us to derive these results, results that may not hold in a Gaussian setting. These results have broad applicability to LCFM SF models in general. We explore the applicability of these results, providing some evidence from Monte Carlo simulations.

The remainder of this paper is structured as follows: The section “Latent class and finite mixture stochastic frontier models” reviews the use of LCFM specifications in SF modelling, and issues that arise in hypothesis testing. The section “Hypothesis testing in latent class and finite mixture stochastic frontier models” discusses hypothesis testing in LCFM models generally, focussing on results regarding the asymptotic distribution of the MLR test statistic under the null hypothesis that are relevant to the SF context. The section “Applications” presents some Monte Carlo evidence on the sampling distribution of the MLR statistic and compares this to the asymptotic results found in the literature. The section “Empirical application” provides a brief empirical application of a particular LCFM SF model. The section “Summary and conclusions” concludes.

2 Latent class and finite mixture stochastic frontier models

In this section, we describe a LCFM SF model of the form estimated by Orea and Kumbhakar (2004) and Greene (2005). In subsequent sections, our discussion of hypothesis testing will focus on a restricted versions of the model in which only one parameter varies across classes or mixture

components; however it is useful at this point to describe the general model.

The basic cross-sectional stochastic frontier model, as proposed by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) is

$$y_i = x_i' \beta + \varepsilon_i, \quad \varepsilon_i = v_i - su_i, \\ i = 1, \dots, N, \quad (1)$$

where y_i is the dependent variable, x_i is a vector of independent variables including a constant term, and β is a vector of coefficients. The error term, ε_i , is composed of a symmetric component $v_i \sim N(0, \sigma_v^2)$ capturing noise, and an inefficiency term following some one-sided distribution, e.g. $u_i = |u_i^*|$, $u_i^* \sim N(0, \sigma_u^2)$. For a production frontier, $s = 1$, while for a cost frontier $s = -1$.

SF models are typically estimated via maximum likelihood (ML) by making some assumptions about the distribution of the error components. In the case of more general models, the likelihood function may lack a closed-form expression, or may be hard to operationalise. For example in the generalised true random effects (GTRE) model, the composed error follows a closed skew normal distribution (Colombi et al., 2014), the density function of which is challenging to evaluate. An alternative approach to incorporating heterogeneity into SF models is to use an LCFM framework.

Amending (1) to allow for latent class technological heterogeneity gives

$$y_{ij} = x_{ij}' \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} = v_{ij} - su_{ij}, \\ i = 1, \dots, N, \quad j = 1, \dots, J, \quad (2)$$

where J is the number of classes or mixture components. Note that the β_j vector varies between classes. We can distinguish between latent class models, in which the functional form is identical between classes, and finite mixture models more generally, which may involve mixtures of fundamentally different models. As noted by Greene (2005), this model is a discrete analogue¹ to the random parameters SF model, which assumes that the frontier parameters follow a continuous multivariate distribution. Here, however, derivation of the likelihood function is straightforward. We can express the model using

¹ However note that, while the unconditional mixing distributions (or probabilities of class membership) are discrete, the conditional (on ε_i) distributions (probabilities) are not, so that we obtain observation-specific conditional parameter estimates, as in the random parameters case. Alternatively, following Greene (2005), we may simply choose the class with the highest conditional probability.

Iverson bracket notation, so that

$$y_i = \sum_{j=1}^J [z_i = j] x'_i \beta_j + \varepsilon_i, \quad \varepsilon_i = v_i - su_i, \quad (3)$$

$$i = 1, \dots, N, j = 1, \dots, J,$$

where z_i is an iid random variable drawn from a categorical (a.k.a. generalised Bernoulli) distribution with probability mass function

$$f_z(z_i) = \sum_{j=1}^J [z_i = j] p_j, \quad \sum_{j=1}^J p_j = 1, \quad (4)$$

such that p_j can therefore be as a mixing proportion or unconditional probability of belonging to class j from the point of view of the outside observer. We can generalise the model further by also allowing the scaling of the error components v_i and u_i to vary across classes or mixture components, e.g.

$$v_i = \sum_{j=1}^J [z_i = j] \frac{v_j^*}{\sigma_{vj}}, \quad v_j^* \sim N(0, 1) \forall j, \quad (5)$$

$$u_i = \left| \sum_{j=1}^J [z_i = j] \frac{u_j^*}{\sigma_{uj}} \right|, \quad u_j^* \sim N(0, 1) \forall j. \quad (6)$$

It is straightforward to derive the following joint density:

$$f_{v,u,z}(y_i - x'_i \beta_j - su_i, u_i, z_i) = \begin{cases} 2 \sum_{j=1}^J \frac{[z_i = j] p_j}{\sigma_{vj} \sigma_{uj}} \phi\left(\frac{y_i - x'_i \beta_j - su_i}{\sigma_{vj}}\right) \phi\left(\frac{u_i}{\sigma_{uj}}\right), & u_i > 0 \\ 0, & u_i \leq 0, \end{cases} \quad (7)$$

and the marginal density of ε_i is found by integrating over all possible values of u_i and summing over all possible values of z_i , which gives

$$f_\varepsilon(\varepsilon_i) = \sum_{j=1}^J \frac{2p_j}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}} \phi\left(\frac{y_i - x'_i \beta_j}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right) \Phi\left[-\frac{s(y_i - x'_i \beta_j) \sigma_{uj} / \sigma_{vj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right], \quad (8)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Efficiency prediction is based on the conditional distribution of $u_i | \varepsilon_i$

$$f_{u_i | \varepsilon_i}(u_i | \varepsilon_i) = \begin{cases} \frac{2}{f_\varepsilon(\varepsilon_i)} \sum_{j=1}^J \frac{p_j}{\sigma_{vj} \sigma_{uj}} \phi\left(\frac{y_i - x'_i \beta_j - su_i}{\sigma_{vj}}\right) \phi\left(\frac{u_i}{\sigma_{uj}}\right), & u_i > 0 \\ 0, & u_i \leq 0, \end{cases} \quad (9)$$

with $E(u_i | \varepsilon_i)$ (Jondrow et al. 1982) or $E[\exp(-u_i) | \varepsilon_i]$ (Battese and Coelli 1988) commonly used in the literature, and given in this setting by

$$E(u_i | \varepsilon_i) = \frac{1}{f_\varepsilon(\varepsilon_i)} \sum_{j=1}^J \frac{p_j \sigma_{vj} \sigma_{uj}}{\sigma_{vj}^2 + \sigma_{uj}^2} \sqrt{\frac{2}{\pi}} \phi\left(\frac{\varepsilon_i}{\sigma_{vj}}\right) \left[1 - \frac{s \varepsilon_i \sigma_{uj} / \sigma_{vj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}} \frac{\Phi\left(-\frac{s \varepsilon_i \sigma_{uj} / \sigma_{vj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right)}{\phi\left(\frac{s \varepsilon_i \sigma_{uj} / \sigma_{vj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right)} \right], \quad (10)$$

$$E(e^{-u_i} | \varepsilon_i) = \frac{1}{f_\varepsilon(\varepsilon_i)} \sum_{j=1}^J \frac{2p_j}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}} \phi\left(\frac{\varepsilon_i}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right) \frac{\Phi\left(-\frac{s \varepsilon_i \sigma_{uj} / \sigma_{vj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}} - \frac{\sigma_{vj} \sigma_{uj}}{\sqrt{\sigma_{vj}^2 + \sigma_{uj}^2}}\right)}{\exp\left(\frac{s \varepsilon_i \sigma_{uj}^2}{\sigma_{vj}^2 + \sigma_{uj}^2} - \frac{1}{2} \frac{\sigma_{vj}^2 \sigma_{uj}^2}{\sigma_{vj}^2 + \sigma_{uj}^2}\right)}. \quad (11)$$

Note that for any distributional assumptions under which where these expressions have closed forms in the $J = 1$ case (i.e. under homogeneity), they will also have closed-form expressions for arbitrary J , since given the sum rule of integration, the integrals involved are simply weighted sums of the integrals in the $J = 1$ case.

2.1 Identification and boundary problems

Turning to the issue of class enumeration, a natural approach would be to test down from more general models with more classes for mixture components to more restricted models using the LR statistic, defined as:

$$LR = -2(\ln L_0 - \ln L_A), \quad (12)$$

where $\ln L_0$ is the estimated log-likelihood under the null hypothesis, and $\ln L_A$ is the estimated log-likelihood under the alternative hypothesis. However, as alluded to previously, the standard Wilks (1938) result, that this statistic follows a χ_k^2 distribution under the null hypothesis, with k being the number of restrictions, does not apply in this case, since the standard regularity conditions are violated. Specifically, the model is not identified under the null hypothesis, and the requirement that the true parameter values must be interior to the parameter space is not met. For simplicity of exposition, we will focus on the case where we are testing from a LCFM SF model as in (3) where $J = 2$ down to the standard SF model shown in (1). The LCFM SF model has the parameter vectors $\theta'_1 = (\beta'_1 \quad \sigma_{v1} \quad \sigma_{u1})$, $\theta'_2 = (\beta'_2 \quad \sigma_{v2} \quad \sigma_{u2})$, and p . There are then three different null hypotheses under which we can recover the standard SF model:

1. $H_0: \theta_1 = \theta_2$.
2. $H_0: p = 0$.
3. $H_0: p = 1$.

In each case, there is an identification problem. Under (1), p is not identified. Under (2), θ_1 is not identified, and likewise

under (3), θ_2 is not identified. In the last two cases, an additional problem is that p is at a boundary of the parameter space. Finally, there is an additional identification issue with the model in general, in that the classes or mixture components are interchangeable—that is, for any given θ_1 and θ_2 , these vectors would be swapped and $1-p$ substituted for p , yet the model would remain fundamentally the same. We will now discuss specific applications of LCFM modelling in the SF literature, in which this problem arises.

Previously proposed LCFM SF models can be obtained as special cases of the general specification described in (3). The ZISF model of Kumbhakar et al. (2013) and Rho and Schmidt (2015) is obtained under the restrictions that $J=2$, $\beta_1=\beta_2$, $\sigma_{v1}=\sigma_{v2}$, $\sigma_{u1}=0$. A further example, and the motivation for the present study, is obtained when we impose the restrictions that $\beta_j=\beta$, $\sigma_{ij}=\sigma_u\forall j$, allowing only the noise variance to vary between classes or mixture components. On way of conceptualising this is as a standard SF model in which the probability density function for the noise term is given by

$$f_v(v_i) = \sum_{j=1}^J \frac{p_j}{\sigma_{vj}} \phi\left(\frac{v_i}{\sigma_{vj}}\right), \quad p_j \in [0, 1] \forall j, \quad \sum_{j=1}^J p_j = 1. \quad (13)$$

In the $J=2$ case, this distribution is known as the scale-contaminated normal distribution, so called because a proportion of observations are drawn from a ‘contaminating’ distribution with higher variance. This distribution has heavier tails than the normal distribution—although at the extremes, the tails become roughly Gaussian—and has therefore been used as an alternative to the normal distribution allowing for thicker tails—see Gleason (1993). Indeed, this was the original motivation for proposing this model, as an alternative to our Student’s t -half normal SF model (Wheat et al. 2019).

In each of the examples given, the null hypothesis of homogeneity (i.e. that $J=1$) is of particular interest. In the Orea and Kumbhakar (2004) model, this is effectively a test for technological homogeneity. In the contaminated normal SF model, it implies the noise distribution is normal rather than a scale mixture of normals, and may be interpreted as a test for a heavier tailed noise distribution. In the ZISF model, the interpretation depends upon the null model we are testing against: if we test against a model with the restriction $\text{Var}(u_i)=0$, the null hypothesis is that all firms are fully efficient, whereas if we test against a null model in which $\text{Var}(u_i)>0$, the null hypothesis is that none of the firms are fully efficient.

2.2 Hypothesis testing in latent class and finite mixture stochastic frontier models

In the section ‘‘Identification and boundary problems’’, we explained that the two fundamental problems of testing in LCFM models are

1. Unidentified parameters under the null hypotheses
2. Parameters of interest being at a boundary of the parameter space under the null hypothesis

The latter problem is in fact a familiar one in the SF literature. It arises in the standard model given in (1) when testing the null hypothesis of no efficiency, i.e. $\text{Var}(u_i)=0$, and in the Wheat et al. (2019) model when testing against the standard model, in which the shape parameter in the noise distribution approaches infinity. In both cases, the relevant result is that the LR statistic follows a 50:50 mixture of a χ_0^2 and χ_1^2 distributions, denoted $\chi_{0,1}^2$, under the null hypothesis. We take this result from Lee (1993) and Case 5 in Self and Liang (1987). The discussion in Kumbhakar et al. (2013) and Rho and Schmidt (2015) leaves us in some doubt as to whether the result may be applied to the problem of testing the null hypothesis that *all* firms belong to the fully efficient class in the ZISF model since, as Rho and Schmidt (2015) point out, there is also an unidentified parameter under the null hypothesis.

In fact, the same result arises in the literature on testing for homogeneity in non-Gaussian LCFM models where a parameter of interest differs between classes or mixture components, following the addition of parameter constraints and a penalty function which have the effect of removing identification problems from the model. The following discussion indicates that the result seems to have broad applicability to the problem of testing against homogeneity in LCFM SF models, and suggests an alternative approach to model specification and testing in this setting.

2.3 Hypothesis testing in latent class and finite mixture models

Given the wide application of LCFM modelling approaches, much attention has been paid to the problem of class enumeration, typically in terms of either deriving the asymptotic distribution for the LR statistic under alternative regularity conditions, or by proposing some alternative approach to estimation and testing.

Naturally, much of this literature focuses on the specific case of Gaussian mixture models. This is despite the fact that Gaussian mixture models have particularly unattractive properties that complicate the analysis considerably²: Hartigan (1985) showed that the likelihood function is unbounded. In addition, there is a loss of strong identifiability as defined by Chen (1995), since $\frac{\partial^2}{(\partial\mu)^2} \phi\left(\frac{y-\mu}{\sigma}\right) = 2 \frac{\partial^2}{\partial\sigma^2} \phi\left(\frac{y-\mu}{\sigma}\right)$.

² As Chen and Li (2009) point out, this is counterintuitive given the well-known attractive properties of the normal distribution in most other contexts.

Fortunately, these latter issues with Gaussian mixture models do not apply to the mixture SF model except as $\sigma_u \rightarrow 0$. We therefore do not dwell on the various results concerning Gaussian mixtures, except to note that the LRT when testing for homogeneity is generally complex, involving the squares of the suprema of Gaussian processes—see e.g. Chen and Chen (2001a, 2001b, 2003)—with the precise result apparently depending on which parameters are assumed to be known and unknown, and which parameters do and do not vary by class or mixture component. Nevertheless, the literature on Gaussian mixtures has yielded useful techniques for regularising the problem which carry over to the general case.

As discussed in the section “Identification and boundary problems”, a well-known issue arises in LCFM models, including the LCFM SF model as formulated in the section “Latent class and finite mixture stochastic frontier models”, due to the fact that the parameter vectors θ_1 and θ_2 and the mixing proportions or unconditional probabilities p and $1-p$ are interchangeable. This interchangeability is not a fundamental problem—clearly, it does not affect the value of the log-likelihood and therefore the LR—but it is easily dealt with, and it is convenient to do so. A common solution is to impose parameter constraints such that this identification problem is resolved. For example, Chen and Chen (2001a) note that—where only one parameter ϑ_j varies between two classes or mixture components—we can impose the restriction $p \geq 0.5$ or the restriction $\vartheta_j \geq \vartheta_2$ or similar, and adopt the former restriction. Analogous restrictions are imposed by Chen et al. (2001), Zhu and Zhang (2004). In a more general setting where $J > 2$, a similar labelling restriction is used by O’Donnell and Griffiths (2006) in the context of estimating state-contingent production frontiers in which only the frontier intercepts vary. In their case, the restriction is $\vartheta_j \leq \vartheta_{j+1} \forall j$, where ϑ_j is the intercept for the j th mixture component, so that the classes or mixture components are in descending order of productivity. In cases where more than one parameter may vary between classes or mixture components, we may again impose a restriction such as $\vartheta_j \leq \vartheta_{j+1} \forall j$, where ϑ_j is any arbitrarily chosen member of the set of parameters that vary across classes or mixture components. However, since such the ordering of classes or mixture components under such a restriction may lack an appealing interpretation, it may be more intuitive to place a labelling restriction such as $p_j \geq p_{j+1} \forall j$, so that we order by mixing proportion (or unconditional probability of class membership).

Simple parameter restrictions are therefore used to remove the identification problem associated with the symmetry of classes or mixture components, but the more fundamental identification issue—that there are unidentified parameters under each of the three different null hypotheses

described in the section “Identification and boundary problems”—remains, and requires a different approach.

One such approach is the use of modified likelihood ratio (MLR) test proposed by Chen and Kalbfleisch (1996), Chen (1998), and Chen et al. (2001), under which a modified likelihood function is used, such that

$$ML = \ln L + c \ln P, \quad (14)$$

where c is a positive constant and $\ln P$ is some penalty term which forces the estimates of certain parameters away from the boundary of the parameter space and ensures that the model is identified under the null hypothesis. The use of a penalty term in estimating LCFM models was discussed by Leroux (1992), who suggested penalising the number of parameters. Chen and Kalbfleisch (1996) suggested adding a penalty term to a minimum distance estimator. Rather than penalising the number of parameters, however, the authors proposed penalising small values of the mixing parameters, and showed that their estimator was consistent for the mixing distribution and the number of components. The inclusion of such a penalty term for the purpose of performing a MLR test was first proposed by Chen (1998), who consider comparing a two-component mixture of binomial distributions—differing in terms of their success probability parameters—against a single binomial distribution, and derive a simple asymptotic χ^2_1 distribution for the MLR statistic under the null hypothesis in this case.

The penalty term proposed by Chen (1998) was $c \ln(1-p)$, where p is the mixing proportion or probability corresponding to the first component. Note that this forces p away from 1, since

$$\lim_{p \rightarrow 1} c \ln(1-p) = -\infty,$$

but does not force p away from 0. Hence, the identification issue is not entirely solved, since under the null hypothesis the true density can still be represented by two different cases: $p = 0$, $\theta_2 = \theta_0$, or $\theta_1 = \theta_2 = \theta_0$. Chen et al. (2001) proposed a penalty term $c \ln[4p(1-p)]$, which solves this problem by forcing p away from both 0 and 1—since

$$\lim_{p \rightarrow 0} c \ln[4p(1-p)] = \lim_{p \rightarrow 1} c \ln[4p(1-p)] = -\infty,$$

This penalty function can be generalised to $c \ln(J^J \prod_{j=1}^J p_j)$ for a J class model³. The authors show that, under a two-component mixture of one-parameter densities

³ Chen et al. (2001) state that the penalty function becomes $\sum_{j=1}^J \ln(2p_j)$, but this seems to be an error, since for $J > 2$, this would result in a negative MLR when the null model is recovered.

satisfying a general set of regularity conditions—including the strong identifiability condition proposed by Chen (1995)⁴—the asymptotic distribution of the MLR statistic under the null hypothesis of homogeneity is a 50:50 mixture of χ_0^2 and χ_1^2 distributions—the former being a degenerate distribution at zero—which is denoted $\chi_{0.1}^2$. The intuition behind the preceding result is that, when the penalty term $c \ln [4p(1-p)]$ is included, under the null hypothesis the likelihood is maximised where $p = 0.5$ and $\theta_1 = \theta_2 = \theta_0$. A boundary problem remains, since the identification restriction $p \leq 0.5$ means that this parameter is at a boundary of the parameter space under the null.

The Chen et al. (2001) result is of immediate interest in our context, since it applies to a very general class of LCFM models. However, it is unclear whether or not this result holds if we generalise the model to include one or more ‘structural’ parameters—that is, parameters that are unknown and common across components—as in the SF context. This is a concern, since in the context of Gaussian mixtures, the inclusion of structural parameters has been shown to lead to different asymptotic distributions for the LR and MLR statistics when testing for homogeneity (Chen and Chen 2003; Qin and Smith 2004; Chen and Kalbfleisch 2005). A related issue is that the inclusion of covariates in the model, which may affect the asymptotic distributions of test statistics under the null hypothesis in the Gaussian mixture case (Kasahara and Shimotsu 2015), is not considered.

However, the results of Zhu and Zhang (2004) are encouraging in these regards. The authors consider testing for homogeneity in the context of non-Gaussian LCFM models, allowing for the presence of both covariates and structural parameters, though under stronger assumptions than those of Chen (1995). The mixture density, in general terms, is given by

$$g(y, \vartheta, \theta_1, \theta_2, p) = pf(y, x; \vartheta, \theta_1) + (1 - p)f(y, x; \vartheta, \theta_2), \quad (15)$$

where y is the dependent variable, and x is the vector of covariates, ϑ is a vector of parameters common to each component—which may include coefficients on covariates—and θ_1 and θ_2 are vectors of component-specific parameters. Zhu and Zhang (2004) derive the asymptotic distribution of the MLR statistic—using the penalty term proposed by Chen et al. (2001)—under the null hypothesis of homogeneity. This is complicated in the general case, but when θ_1 and θ_2 are scalars, is the $\chi_{0.1}^2$ distribution. As the authors note, this is identical to the result derived by Chen et al. (2001) in the absence of covariates and structural

parameters. Zhu and Zhang (2004) present the results of a simulation study comparing rejection rates of the MLRT and LRT statistics under the null hypothesis, and conclude that they prefer the MLRT.

So far we have restricted the discussion, for simplicity’s sake, to testing down from a $J = 2$ component model to a $J = 1$ component model. In principle however, we are also interested in the general problem of testing down from any J component model to any simpler model. We note two particularly relevant results.

First, Chen et al. (2001) considered the extension of their results to the case where we have a J component model and we wish to test for homogeneity. The authors find that the same result applies, i.e. the MLR statistic is asymptotically distributed $\chi_{0.1}^2$. This is somewhat counterintuitive, given the increased difference in dimensionality. The authors note that the power of the test is lower the greater is J , since the penalty becomes larger the greater the number of components. Second, Chen et al. (2003) extends the Chen et al. (2001) approach to consider testing down to a two-component model from a J component model, and testing via find that the asymptotic distribution of the MLR statistic under the null hypothesis is a mixture of χ_0^2 , χ_1^2 , and χ_2^2 distributions. Whether or not these results hold in our context is unclear, since Zhu and Zhang (2004) only considered testing for homogeneity in the context of a two-component model.

In summary, while the problem of testing down from models with more to fewer classes or mixture components is in general a complex one, the literature suggests that the problem may be simplified in the case of non-Gaussian LCFM models. Furthermore, proposed MLR tests in which the log-likelihoods are modified to include a term penalising extreme values of the class probabilities or mixing proportions appear to regularise the problem by ensuring that the model parameters are identified under the null hypothesis.

2.4 Choice of value of the tuning parameter c

One issue that arises when using the MLRT approach is how to choose the value of the tuning parameter, c . The higher the value chosen, the more heavily extreme values of the probabilities or mixing proportions are penalised. Chen et al. (2001) state that their simulation exercises suggested that the results were not sensitive to the value of c chosen, but suggest that $c = \ln m$, where there is some restriction on the parameter of interest such that its magnitude cannot exceed m as one possible choice, based on the rate of divergence of the LRT that applies for certain kernel functions. In our Monte Carlo experiments, we try two different values, $c = 1$ and $c = 5$ in order to determine whether or not this choice has a material effect on the results.

⁴ Note therefore that this result cannot be applied in the case of Gaussian mixture models.

2.5 Model estimation

Given the above discussion and the use of the MRLT, the model is estimated by maximising the modified likelihood function whilst imposing the parameter constraints necessary for identification of the model. Discussing the estimation of LCFM models, Zhu and Zhang (2004) suggest either EM or a simple Newton-Raphson algorithm. We find that the latter performs well.

From parts (a) and (c) of Theorem 2 of Zhu and Zhang (2004), which concern convergence rates of estimates when maximising the modified likelihood ratio, $\hat{\theta}$ has the usual $N^{-1/2}$ convergence rate, and $\hat{\theta}_1$ and $\hat{\theta}_2$ have a convergence rate of $N^{-1/4}$, while $\hat{p} = O_p(1)$ and a convergence rate of $N^{-1/4}$ in can be reached in estimating the mixing distribution using the parameter estimates. Convergence rates are therefore comparable to those under standard maximum likelihood estimation, though—intuitively—lower for the parameter of interest.

3 Applications

In order to ascertain the applicability of the aforementioned results from the literature on testing for homogeneity for non-Gaussian latent class and finite mixture models using the MLR test, we carry out a series of Monte Carlo experiments allowing us to estimate the sampling distribution of the MLR statistic and compare it to the $\chi^2_{0.1}$ distribution. If the sampling distribution corresponds closely to the $\chi^2_{0.1}$ distribution, this suggests that the $\chi^2_{0.1}$ result is applicable in the LCFM SF setting.

In keeping with the focus of the preceding discussion, we restrict ourselves to problems in which one parameter varies between two classes or mixture components. The specification we use is a contaminated normal stochastic frontier model, in which the scaling of the noise distribution is allowed to vary between classes or mixture components. We consider two different specifications—the first with $J=2$, the second with $J=3$. Note that, in the latter case, we are being more speculative—although Zhu and Zhang (2004) derived the $\chi^2_{0.1}$ result for the $J=2$ case including covariates, and Chen et al. (2001) derived the result for the case where there is arbitrary J , to our knowledge there has been no study of the case where we have both arbitrary J and structural parameters and covariates. However, we conjecture that, having removed the identification problems from the model, Case 5 in Self and Liang (1987) may be applied.

We assume that the true data generating process—under the null hypothesis—is

$$y_i = 1 + x_i + v_i + u_i \tag{16}$$

where x_i is an observed covariate drawn from a standard normal distribution, $v \sim N(0,1)$ is a symmetric noise term, and $u_i = |w_i|$, $w_i \sim N(0, \sigma_u^2)$ is the inefficiency term. Unlike the more general LCFM SF models introduced by Orea and Kumbhakar (2004) and Greene (2005), which allow all parameters to vary, we restrict all parameters except the variance of the noise distribution to be the same across classes or mixture components—the focus of this model is on allowing for greater flexibility of the noise distribution, rather than technological heterogeneity. We then estimate a standard normal-half normal SF model and penalised two-component and three-component contaminated SF models, in order to obtain the log-likelihood under the null hypothesis and the modified log-likelihoods under the alternative hypotheses, respectively. To be specific, the contaminated normal SF model is specified as follows:

$$y_i = x'_i\beta + \varepsilon_i, \varepsilon_i = v_i - su_i, v_i = \sum_{j=1}^J [z_i = j]v_{ij}, \tag{17}$$

$$i = 1, \dots, N, j = 1, \dots, J,$$

where

$$u_i \sim |N(0, \sigma_u^2)|, \tag{18}$$

$$v_{ij} \sim N(0, \sigma_{vj}^2). \tag{19}$$

The model is estimated by maximising the log modified likelihood function

$$\ln ML_J = \ln L + c \ln \left(J^J \prod_{j=1}^J p_j \right), \tag{20}$$

with the constraint, imposed in order to remove the identification problem associated with the symmetry of classes or mixture components, that

$$\sigma_{vj} \leq \sigma_{vj+1}, \forall j = 1, \dots, J - 1. \tag{21}$$

Note that the penalty term in (20) effectively penalises more complex models with more classes or mixture components, and encourages sparsity. In this respect, the statistic behaves similarly to many information criteria, but contrasts with others such as the CLC or ICL which instead penalise poorly separated clusters.

We repeat this experiment 5000 times for $\sigma_u = 0.5$, $\sigma_u = 1$, and $\sigma_u = 5$ to check that our findings are not dependent upon particular specifications or choices of parameter values. Following the previous discussion of the appropriate choice of c , it may be expected that our results will be somewhat sensitive to the chosen value of this parameter. Given that the model nests a Gaussian mixture as $\sigma_u \rightarrow 0$, and that the $\chi^2_{0.1}$ result cannot be applied to the case of Gaussian mixture models, it

may also be that the sampling distribution of the MLR statistic also depends on the signal-to-noise ratio, σ_u/σ_v .

Having obtained estimated MLR statistics for each replication, we then compare the sampling distributions to the theoretical quantiles of the $\chi^2_{0:1}$ distribution by means of quantile-quantile (Q-Q) plots, i.e. plots of the quantiles of the sampling distributions and the corresponding quantiles of the $\chi^2_{0:1}$ distribution against those of the χ^2_1 distribution. The quantiles of the χ^2_1 distribution are on the horizontal axis. If the MLR followed a χ^2_1 distribution, we would therefore expect the quantiles of the sampling distribution to lie on the 45 degree line, shown by a black line on each plot. A red line then shows the quantiles of the $\chi^2_{0:1}$ distribution as a function of the corresponding quantiles of the χ^2_1 —there is a simple closed-form relationship between the two. The quantile function for the $\chi^2_{0:1}$ distribution, $F_{0:1}^{-1}$, may be expressed in terms of the χ^2_1 quantile function, F_1^{-1} :

$$F_{0:1}^{-1}(1-p) = \begin{cases} 0, & p > 0.5 \\ F_1^{-1}(1-2p), & p \leq 0.5 \end{cases} = z_{0:1}, \quad (22)$$

where

$$\begin{aligned} F_1^{-1}(1-p) &= z_1, \\ p &= 1 - F_1(z_1). \end{aligned} \quad (23)$$

Plugging this into (22), we have an expression for the quantiles of the $\chi^2_{0:1}$ distribution in terms of the corresponding quantiles of the χ^2_1 distribution for the same p :

$$z_{0:1} = \begin{cases} 0, & 1 - F_1(z_1) > 0.5 \\ F_1^{-1}[2F_1(z_1) - 1], & 1 - F_1(z_1) \leq 0.5 \end{cases}. \quad (24)$$

Given that we believe the MLR statistic is asymptotically distributed $\chi^2_{0:1}$, we therefore expect the quantiles of the sampling distribution of the MLR statistic, plotted against the theoretical quantiles of the χ^2_1 distribution, to closely follow $z_{0:1}$. That is, we expect a close correspondence between the estimated quantiles of the sampling distribution and the red line representing the quantiles of the $\chi^2_{0:1}$ distribution.

3.1 Quantile-quantile plots—contaminated normal SF model

The figures below show the Q-Q plots comparing the sampling distribution of the MLR statistic for various DGPs for the cases where $J=2$ and $J=3$. Bearing in mind that the 95th percentile of the $\chi^2_{0:1}$ distribution is 2.706, we can see that the correspondence between the quantiles of the distributions is very close across the range of interest. This is less true of the extreme tails of the distribution, where the quantiles of the sampling distribution tend to be greater than

those of the $\chi^2_{0:1}$. The difference between the two seems to be greater the higher the value of σ_u . It appears, however, that this is sensitive to the chosen value of the tuning parameter c , since when the higher value of 5 is used, the correspondence at the right tail of the distribution is closer.

In addition to the Q-Q plots in Fig. 1, Table 1 shows the median absolute deviation of the empirical quantiles of the MLR statistic from the theoretical quantiles of the $\chi^2_{0:1}$ and χ^2_1 distributions under each DGP. The median absolute deviations from the $\chi^2_{0:1}$ quantiles are in each case zero, compared to around 0.45 from the χ^2_1 distribution.

It is of particular interest that the results do not appear to substantially change when the alternative model has $J=3$, which lends support to our conjecture that the $\chi^2_{0:1}$ result for the asymptotic distribution of the MLRT under the null hypothesis may be applied rather broadly to testing against homogeneity in LCFM SF models including J classes or mixture components, structural parameters, and covariates.

4 Empirical application

In this section, we apply the contaminated normal-half normal SF model outlined in the section “Applications” to the Christensen and Greene (1976) dataset on US electricity generation costs, consisting of cost, output, and input price data for electricity generating firms in 1970. Following the frontier specification in Greene (1990), the model we estimate is

$$\begin{aligned} \ln\left(\frac{c_i}{e_i}\right) &= \beta_0 + \beta_1 \ln q_i + \beta_2 \ln^2 q_i + \beta_3 \ln\left(\frac{w_i}{e_i}\right) + \beta_4 \ln\left(\frac{r_i}{e_i}\right) + v_i + u_i, \\ v_i &= \sum_{j=1}^J [z_i = j] v_{ij}, \quad v_{ij} \sim N(0, \sigma_{vj}^2), \quad u_i \sim |N(0, \sigma_u^2)|, \end{aligned} \quad (25)$$

where c is total cost, q is millions of kilowatt-hours generated (normalised by the sample mean), and w , r , and e are labour, capital, and energy prices, respectively. The subscripts i and j respectively denote the firm number and class or mixture component. We estimate the model for $J=1$ and $J=2$, respectively. Parameter estimates are shown in Table 2.

Estimated frontier parameters are, as expected, similar in the $J=1$ and $J=2$ cases. Of greater interest here are the estimated scale parameters σ_{v1} , σ_{v2} , and σ_u , and the probability or mixing proportion ρ_1 . The $J=2$ model yields distinct estimates of σ_{v1} and σ_{v2} , between which the estimate of σ_{v1} from the $J=1$ model is intermediate. Under a latent class interpretation, the $J=2$ model has a low-noise class and a high-noise class, and the estimated unconditional probability of an observation belonging to the former is estimated to be 0.798. Under a mixture interpretation, the noise term in the $J=2$ model is a 0.798:0.202 mixture of $N(0, 0.067^2)$ and $N(0, 0.174^2)$ distributions. The estimate of

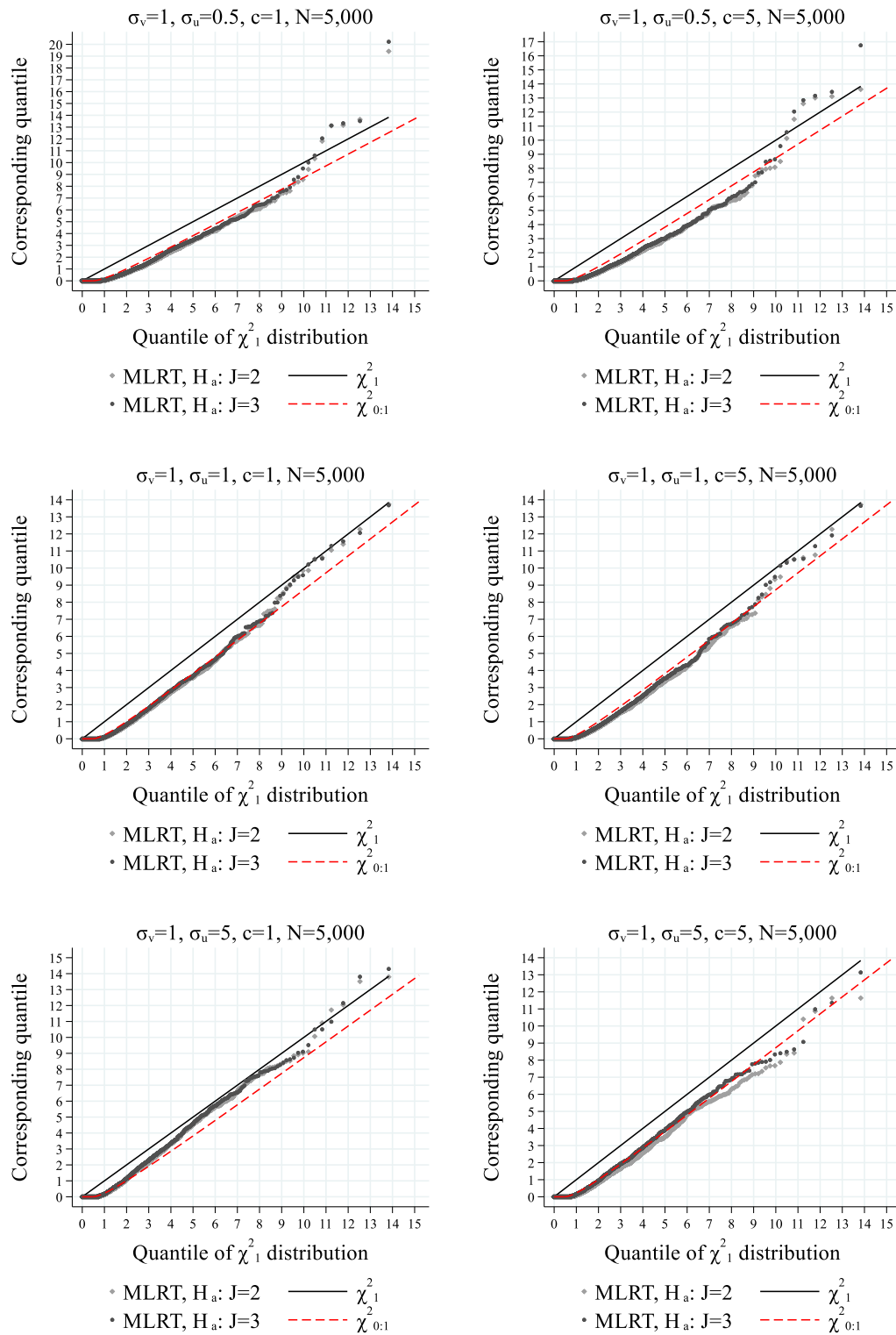


Fig. 1 Quantile-quantile plots

Table 1 Median absolute deviations of empirical quantiles from theoretical quantiles

		$c = 1$		$c = 5$	
		$J = 2$	$J = 3$	$J = 2$	$J = 3$
$\chi_{0:1}^2$	$\sigma_u = 0.5$	0.000	0.000	0.000	0.000
	$\sigma_u = 1$	0.000	0.000	0.000	0.000
	$\sigma_u = 5$	0.000	0.000	0.000	0.000
χ_1^2	$\sigma_u = 0.5$	0.455	0.454	0.454	0.455
	$\sigma_u = 1$	0.450	0.450	0.453	0.453
	$\sigma_u = 5$	0.427	0.412	0.455	0.455

Table 2 Model estimates

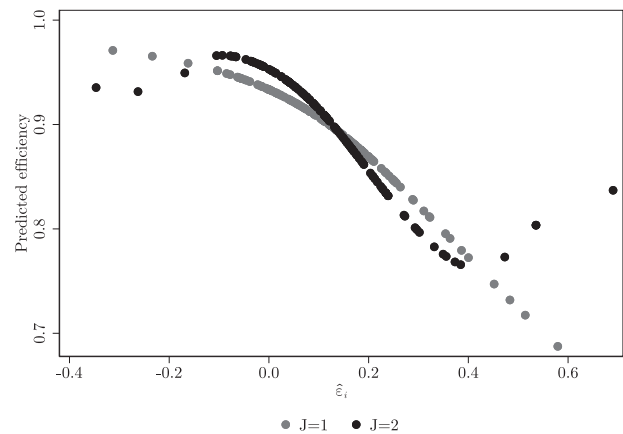
	$J = 1$	$J = 2$
$\beta_1 (\ln q_i)$	0.966*** (0.013)	0.960*** (0.011)
$\beta_2 (\ln^2 q_i)$	0.030*** (0.003)	0.027*** (0.002)
$\beta_3 (\ln w_i)$	0.261*** (0.066)	0.326*** (0.056)
$\beta_4 (\ln w_i)$	0.055 (0.062)	0.027 (0.051)
β_0 (constant)	3.735*** (0.035)	3.748*** (0.038)
σ_{v1}	0.109 (0.023)	0.067 (0.023)
σ_{v2}	–	0.174 (0.067)
σ_u	0.149 (0.049)	0.131 (0.051)
ρ_1	1.000 –	0.798 (0.110)
$\ln ML$	66.865	70.809

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

σ_u from the $J = 2$ model is slightly lower than the estimate from the $J = 1$ model.

Predicted efficiency scores—following the formula in (11)—differ considerably between the two models. As shown in Fig. 2, the efficiency predictions from the $J = 2$ model have not only a smaller range than the predictions from the $J = 1$ model, but also differing behaviour for extreme values of $\hat{\varepsilon}_i$; predicted efficiency from the $J = 1$ model decreases monotonically as $\hat{\varepsilon}_i$ increases due to the log-concavity of the normal distribution (Ondrich and Ruggiero, 2001), while the efficiency predictions from the $J = 2$ model are non-monotonic owing to the log-convex regions of the contaminated noise density—similar behaviour has been noted in the case of SF models with Student's t noise (Wheat et al., 2019), and models with heavy-tailed noise distributions generally appear to handle outlying observations differently⁵, which could have important consequences for regulatory benchmarking in terms of the robustness of efficiency estimates and the

⁵ Note, however, that the further we venture into the tails of the scale-contaminated normal distribution, the more Gaussian (and log-concave) they become, so 'heavy-shouldered' may be a more accurate description than 'heavy-tailed'.

**Fig. 2** Scatter plot of efficiency predictions vs. estimated residuals

incentives firms face with respect to the benchmarking process.

For the purposes of efficiency prediction, the choice between the two models is therefore clearly an important one. The MLR statistic is 7.887 and, given the $\chi_{0:1}^2$ distribution of the statistic under the null hypothesis, the associated p-value is 0.0025. We therefore strongly reject the null hypothesis of homogeneity, and favour the $J = 2$ model. This suggests a sub-Gaussian noise distribution, which as discussed has a substantial impact on efficiency predictions—cf. Wheat et al. (2019) who, using a dataset on English local authority road maintenance costs, similarly find that a Student's t distribution is preferred to a normal distribution for the noise component.

5 Summary and conclusions

In this paper, we have discussed hypothesis testing on class enumeration, and particularly on testing against homogeneity, in the context of latent class and finite mixture stochastic frontier models. Given identification problems and boundary issues under the null hypothesis, the standard Wilks (1938) result that the likelihood ratio statistic follows an asymptotic χ^2 distribution under the null hypothesis cannot be applied. However, the literature on testing against homogeneity in non-Gaussian latent class and finite mixture models offers some alternative results and ways of regularising the problem.

Introducing restrictions on parameter values across classes or mixture components can help resolve some of the issues with parameter identification, as can adding a penalty term to the log-likelihood function in order to force the estimated class probabilities or mixing proportions away from 0 and 1, so that homogeneity can only be recovered under a single combination of parameter values. However the issue then remains that the parameter of interest lies at a

boundary of the parameter space under the null hypothesis. This is a familiar problem in the stochastic frontier literature, and in cases in which one parameter varies between classes or mixture components, findings from Chen et al. (2001) and Zhu and Zhang (2004) suggest a similarly familiar result that the asymptotic distribution of the modified likelihood ratio under the null hypothesis is a 50:50 mixture of χ_0^2 and χ_1^2 distributions, denoted $\chi_{0:1}^2$, under the null hypothesis. This result matches that for the similar problem of testing the null hypothesis of no inefficiency in the standard stochastic frontier model.

We present evidence from Monte Carlo experiments on the sampling distribution of the modified log-likelihood ratio which suggests that this result does indeed seem to be applicable to the stochastic frontier setting, with the quantiles of the sampling distribution corresponding closely to those of the $\chi_{0:1}^2$ distribution within the range of interest. These experiments consider two different latent class or finite mixture stochastic frontier specifications in which a single parameter differs between classes or mixture components, and a range of combinations of parameter values in both cases. Overall it therefore appears that the existing literature on testing for homogeneity in non-Gaussian latent class and finite mixture models yields a set of techniques and results applicable in the stochastic frontier context, enabling us to test important hypotheses regarding technological heterogeneity and the distributions of the error terms.

We also provide an empirical application in which we test down from a stochastic cost frontier model in which the noise term follows a scale-contaminated normal distribution—i.e. a scale mixture of normal distributions—to one in which the noise term is assumed to be normally distributed. The choice between the mixture model and the single class case is shown to have a significant impact on efficiency predictions. Applying the modified likelihood ratio test, we strongly reject the null hypothesis of homogeneity.

In this paper, we restricted the scope of our discussion and Monte Carlo experiments to the problem of testing for homogeneity from a model with two classes or mixture components. While the results discussed here do seem applicable to testing for homogeneity from any J class model, this still leaves the more general question of class enumeration open, and this would be a natural line of enquiry for future research.

Acknowledgements The authors acknowledge funding from the CQC Efficiency Network (see <http://www.nhtnetwork.org/cqc-efficiency-network/home/>).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6:21–37
- Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J Econom* 38:387–399
- Chen H, Chen J (2001a) Large sample distribution of the likelihood ratio test for normal mixtures. *Statistics Probab Lett* 52:125–133
- Chen H, Chen J (2001b) The likelihood ratio test for homogeneity in finite mixture models. *Can. J. Statistics* 29:201–215
- Chen H, Chen J (2003) Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sin* 13:351–365
- Chen H, Chen J, Kalbfleisch JD (2001) A modified likelihood ratio test for homogeneity in finite mixture models. *J Roy Statistical Soc: Ser B (Statistical Methodol)* 63:19–29
- Chen H, Chen J, Kalbfleisch JD (2003) Testing for a finite mixture model with two components. *J Roy Statistical Soc: Ser B (Statistical Methodol)* 66:95–115
- Chen J (1995) Optimal rate of convergence for finite mixture models. *Ann Statistics* 23:221–233
- Chen J (1998) Penalized likelihood-ratio test for finite mixture models with multinomial observations. *Can J Statistics* 26:583–599
- Chen J, Kalbfleisch JD (1996) Penalized minimum-distance estimates in finite mixture models. *Can J Statistics* 24:167–175
- Chen J, Kalbfleisch JD (2005) Modified likelihood ratio test in finite mixture models with a structural parameter. *J Statistical Planning Inference* 129:93–107
- Chen J, Li P (2009) Hypothesis tests for normal mixture models: the EM approach. *Ann Statistics* 37:2523–2542
- Christensen LR, Greene WH (1976) Economies of scale in U.S. electric power generation. *J Political Econom* 84:655–676
- Colombi R, Kumbhakar SC, Martini G, Vittadini G (2014) Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *J Prod Anal* 42:123–136
- Fonseca JRS, Cardoso MGMS (2007) Mixture-model cluster analysis using information theoretical criteria. *Intell Data Anal* 11:155–173
- Gleason JR (1993) Understanding elongation: the scale contaminated normal family. *J Am Statistical Assoc* 88:327–337
- Greene WH (1990) A gamma-distributed stochastic frontier model. *J Econom* 46:141–163
- Greene WH (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J Econom* 126:269–303
- Hartigan, JA 1985. A failure of likelihood asymptotics for normal mixtures. In: LeCam, LM & OLSHEN, RA (eds) *Proceedings of*

- the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer. Wadsworth, Belmont, CA.
- Jondrow J, Knox Lovell CA, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econom* 19:233–238
- Kasahara H, Shimotsu K (2015) Testing the number of components in normal mixture regression models. *J Am Statistical Assoc* 110:1632–1645
- Kumbhakar SC, Parmeter CF, Tsionas EG (2013) A zero inefficiency stochastic frontier model. *J Econom* 172:66–76
- Kutlu L, Tran KC, Tsionas MG (2020) Unknown latent structure and inefficiency in panel stochastic frontier models. *J Prod Anal* 54:75–86
- Lee L-F (1993) Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. *Econom Theory* 9:413–430
- Leroux BG (1992) Consistent estimation of a mixing distribution. *Ann Statistics* 20:1350–1360
- McLachlan GJ, Peel D (2000) *Finite mixture models*. Wiley, New York.
- Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18:435–444
- Nylund KL, Asparouhov T, Muthén BO (2007) Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Eq Model: Multidisciplinary J* 14:535–569
- O’donnell CJ, Griffiths WE (2006) Estimating state-contingent production frontiers. *Am J Agric Econ* 88:249–266
- Ondrich J, Ruggiero J (2001) Efficiency measurement in the stochastic frontier model. *Eur J Oper Res* 129:434–442
- Orea L, Kumbhakar SC (2004) Efficiency measurement using a latent class stochastic frontier model. *Empirical Econ* 29:169–183
- Qin YS, Smith B (2004) Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sin* 14:1165–1177
- Rho S, Schmidt P (2015) Are all firms inefficient? *J Prod Anal* 43:327–349
- 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
- Wheat P, Stead AD, Greene WH (2019) Robust stochastic frontier analysis: a Student’s t-half normal model with application to highway maintenance costs in England. *J Prod Anal* 51:21–38
- Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Statistics* 9:60–62
- Zhu H-T, Zhang H (2004) Hypothesis testing in mixture regression models. *J Roy Statistical Soc: Ser B (Statistical Methodol)* 66:3–16