Pitfalls of Normal-Gamma Stochastic Frontier Models

Christian Ritter and Léopold Simar*

Center for Operations Research and Econometrics

and

Institut de Statistique

Université Catholique de Louvain

Belgium

October 19, 1994

Correspondence: Christian Ritter, Institut de Statistique, Université Catholique de Louvain, 34, voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium, tel. 0032-10-47-43-14, fax. 0032-10-47-30-32.

Abstract

Although conceptually pleasing, normal-gamma frontier models lead to difficult estimation problems. It is shown here that unless the sample size reaches several thousands of observations the shape parameter of the gamma density is hard to estimate, and that this carries over to estimates of the stochastic frontier, the individual inefficiencies, and the allocation of the overall variance to the stochastic frontier and to the inefficiencies.

Keywords: Identifiability, least squares, likelihood, profile, simulation.

1 Introduction

Stochastic frontier models with composed errors (Meeusen and van den Broeck, 1977; Aigner, Lovell and Schmidt, 1977; Stevenson, 1980) can provide interesting interpretations to econometric data. Such a stochastic frontier model can be written as

$$y_i = h(\mathbf{x}_i) + \nu_i - z_i,$$

where $h(\mathbf{x}_i)$ denotes the frontier as a function of the explanatory variables $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, and where z_i and ν_i denote the inefficiency of the i-th firm, and the fluctuation of the frontier, respectively. The function $h(\cdot)$, the model of the ideal frontier, is deterministic, the z_i and the ν_i are random. A natural choice for the distribution of the fluctuations, ν_i, is normal with zero mean and variance σ^2. The choice of the distribution of the inefficiencies, z_i, is not as obvious.

Ideally, one would like to use families of distributions covering a large range of shapes and estimate the shape from the data. The gamma density has this flexibility: depending on its shape parameter, it can resemble an exponential, a Pareto, or a normal density, or anything in between. Therefore, stochastic frontier models with normal fluctuations of the frontier and gamma inefficiencies are often used for data with presumed frontier-inefficiency structure (Stevenson, 1980; Beckers and Hammond, 1987; Greene, 1990; van den Broeck, Koop, Osiewalski and Steel, 1993).

Here, we investigate how well one can estimate the four parameters of the simplest normal-gamma stochastic frontier model,

$$y_i = \beta + \nu_i - z_i, \quad \nu_i \overset{iid}{\sim} N(0, \sigma^2), \quad z_i \overset{iid}{\sim} \Gamma(\alpha, \lambda). \tag{1}$$

The four parameters are the frontier (β), the variance of the fluctuations of the frontier (σ^2), the shape parameter (α), and the scale parameter (λ) of the gamma distribution of the inefficiencies. The parameters of particular interest to the investigator are the frontier and the quantity $\rho = \sigma^2/(\sigma^2 + \alpha\lambda^2)$, the part of the total variance explained by the stochastic fluctuations of the frontier.

We show that the problem of estimating the four parameters β, σ^2, α and λ is poorly conditioned for samples of up to several hundreds of observations. We also show that as a consequence estimates of the stochastic frontier, the individual inefficiencies, and the allocation of the overall variance to the stochastic frontier and to the inefficiencies suffer from substantial imprecision, are ambiguous, or can't be calculated at all. This concerns especially maximum likelihood estimation.

In Section 2, we describe several limiting cases of the normal-gamma model. We show that if these limiting models already provide good fits, the full model is hard to identify. In Section 3, we show how likelihood profiles (Ritter and Bates, 1993) can be used to conduct analyses of the likelihood of the problem even if a maximum likelihood estimate cannot be found easily. In Section 4, we study simulated data sets of varying sample sizes under the full model and analyze them using likelihood profiles. We show how the identifiability problems appear in practice and how they can be detected. As an example we show the likelihood profile of a model for data on American electric utilities. We conclude by summarizing and interpreting our results.

2 Limiting Cases of The Model

2.1 General

The normal-gamma stochastic frontier model generalizes the ordinary least squares model (no inefficiencies), the pure gamma model (no fluctuations of the frontier), and the normal-exponential model (fixed shape). The normal-gamma model has two parameters more than the ordinary least squares model and one parameter more than the pure gamma or the normal-exponential model. In this section, we use the simple model (1) to investigate how the normal-gamma model accommodates its sub-models.

Ideally, if the data originating from one of the sub-models are treated using the normal-gamma model the resulting fit should produce similar interpretations as a fit obtained using the correct sub-model. In our context, this applies to the estimate of the frontier and to the allocation of variance to the inefficiencies and to the stochastic frontier. If the interpretations drawn from the true sub-model and the enlarged model can differ, serious ambiguities arise.

2.2 Ordinary Least Squares

The simplest limiting case of (1) is the ordinary least squares model

$$y_i = \beta + \epsilon_i, \tag{2}$$

where the ϵ_i are presumed independent and normally distributed with mean zero and equal variance σ^2. If this is the "truth" behind the data and if the data are interpreted using a normal-gamma model

$$y_i = \beta' + \nu_i - z_i, \quad \nu_i \overset{iid}{\sim} N(0, \sigma'^2), \quad z_i \overset{iid}{\sim} \Gamma(\alpha, \lambda) \tag{3}$$

instead, good fits matching the true distribution of the y_i in at least its first two moments are achieved in very distant regions of the parameter space. This will be explained in the following three examples.

First, for fixed scale parameter λ, the normal-gamma model resembles an ordinary least squares model if the shape parameter α of the gamma distribution tends to zero. The gamma distribution we obtain is a narrow spike at zero, and the density of $\nu_i - z_i$, the convolution of the normal with the gamma density, is close to normal. A similar situation occurs for fixed but arbitrary shape parameter α as the scale parameter λ approaches zero. Again, the limiting gamma density is a spike. These two situations are innocuous, since the estimates for the frontier, the inefficiencies and of the allocation of variance are the same as in the "true" model.

Second, the ordinary least squares case can be approached while keeping the mean of the gamma density, $\mu = \alpha\lambda$, constant. Again, we let λ tend to zero. Two factors play together to assure that the density of $\nu_i - z_i$ swiftly approaches a normal density; the variance $\alpha\lambda^2 = \mu\lambda$ tends to zero and α increases without bound. The first lets the gamma distribution degenerate to a Dirac-delta distribution at μ, the second assures that the gamma density by itself resembles a normal density. Thus, for small values of λ the density of $\beta' + \nu_i - z_i$ is nearly normal with mean $\beta' - \mu$ and variance σ'^2. Matching the first two moments of this density to the true density of y_i we obtain a new frontier $\beta' = \beta + \mu$ with nearly the correct variance allocation of $\sigma'^2 \approx \sigma^2$ for the stochastic frontier and close to zero for the gamma component.

Third, the ordinary least squares model can be approached while keeping the variance $\omega^2 = \alpha\lambda^2$ of the gamma density constant and letting α go to infinity. This forces again the scale parameter λ to zero, yet not as fast as before. The gamma density approaches now a normal density with variance ω^2; its mean increases without bound. Thus, for large values of α, the distribution of $\beta' + \nu_i - z_i$ looks like $N(\beta' - \alpha\lambda, \sigma'^2 + \omega^2)$. Matching the first two moments of this density to the true density of y_i yields a new frontier $\beta' = \beta + \alpha\lambda$ and a modified variance allocation, $\sigma'^2 = \sigma^2 - \omega^2$ for the stochastic frontier and ω^2 for the inefficiencies.

The three cases described above are examples of a large continuum of normal-gamma models in the vicinity of the ordinary least squares model. They imply that if the data originate from an ordinary least squares model and if we use a normal-gamma model to estimate the frontier and the inefficiencies, we can receive a great variety of different interpretations. Adding the gamma extension to the ordinary least squares model makes it therefore hard to recover a genuine least squares situation. Thus, it is very important that an ordinary least squares model is fitted alongside with the normal-gamma model to provide a baseline. If the normal-gamma model yields only little improvement in the explanation of the data, the ordinary least-squares fit should be retained and presented together with the fit of the normal-gamma model. What is a sufficient amount of improvement to clearly offset the normal-gamma model from its least squares competitor?

The normal-gamma model clearly generalizes the ordinary least squares model but it does not contain it as a special case. Therefore, the usual χ^2 approximation with two degrees of freedom to twice the log-likelihood ratio statistic is not strictly valid. However, the likelihood of the normal-gamma model approaches the likelihood of a ordinary least squares model in many ways, including the three possibilities described above.

Let us denote the parameter vector $(\beta', \sigma'^2, \alpha, \lambda)$ by θ, the normal-gamma log-likelihood by $l(\theta \mid \mathbf{y})$, and the ordinary least squares likelihood by $l_{ols}(\beta, \sigma^2 \mid \mathbf{y})$. Suppose that the true model is normal-gamma with parameter vector θ and that a maximum likelihood estimate $\hat{\theta}$ exists. Suppose also that the estimates under the ordinary least squares model are $\hat{\beta}$ and $\hat{\sigma}^2$ and that they correspond to the likelihood $\hat{l}_{ols} = l_{ols}(\hat{\beta}, \hat{\sigma}^2 \mid \mathbf{y})$.

Joint $1-\eta$ likelihood regions with approximate frequency coverage $1-\eta$ for θ can be defined using regions bounded by level contours of $2\left[l(\hat{\theta} \mid \mathbf{x}) - l(\theta \mid \mathbf{x})\right]$ corresponding to $1-\eta$ critical values of a χ^2 distribution with four degrees of freedom. If α and λ correspond to one of the case described above in which the normal-gamma model mimics an ordinary least squares model, the normal-gamma likelihood taken as a function of β' and σ'^2 only, will have a maximum near $\hat{l}_{ols}$. Thus, if $l(\hat{\theta} \mid \mathbf{y})$ does not surpass $\hat{l}_{ols}$ by at least several units, the above defined likelihood regions will be open even for low coverage probabilities and therefore indicate poor precision of the maximum likelihood estimates.

2.3 Closer Competitors

Two closer competitors of the normal-gamma frontier model are the normal-exponential model and the pure gamma model, both having one parameter less than the normal-gamma model. Suppose at first that the "truth" behind our data is a normal exponential model with scale parameter one and that we fit it with a normal-gamma model instead. The normal-exponential model corresponds to a normal-gamma frontier with shape parameter equal to one. For shape parameters $\alpha < (\sigma^2+1)^3$, normal gamma models can be found matching the normal-exponential model in its first three moments. The equations defining the parameters λ', σ'^2, and β' of a normal-gamma model with shape parameter α matching the first three moments of a normal-exponential model with scale parameter $\lambda = 1$, frontier $\beta = 0$, and variance σ^2 of the normal component are: $\lambda' = \alpha^{-1/3}$, $\sigma'^2 = \sigma^2 + 1 - \alpha^{1/3}$, and $\beta' = \alpha^{2/3} - 1$. If the variance of the normal component in the normal-exponential model is small, that is, if the density of $\nu_i - z_i$ resembles an exponential density, the difference between the normal-gamma models and the normal-exponential model is sufficiently pronounced to be detectable even from small data sets. Correspondingly, the variance $\sigma^2 + 1$ of $\nu_i - z_i$ is close to one, and therefore the range of shape parameters α which produce gamma distributions which match the normal-exponential distribution in its three moments does not reach much beyond one.

In practice, the variance of the normal component often constitutes between 1/3 and 1/2 of the total variance. In this case, the convolution of the exponential density with the normal or the gamma density dilutes considerably the difference. Figure 1 testifies to this claim. In the top plot, the normal component holds a share of 1/2 the variance of the normal-exponential model; in the bottom plot the share is 1/3. The plots show empirical distribution functions for data sets with 250000 observations simulated under the normal-exponential model ($\alpha = 1$) and under normal-gamma models with shape parameters $\alpha = 0.33$, 0.5, 2, 3, and 4, for which the remaining parameters have been chosen to match the normal-exponential model in its mean, variance, and skewness as much as possible. That is, for $\alpha > (\sigma^2 + 1)^3$, a pure gamma model was used, i.e., σ'^2 was set to zero. The two figures show how close the models are and suggest the sample sizes required to achieve a significant distinction. For example, a sample of only a few hundred observations is insufficient to reliably distinguish a normal-exponential model with frontier $\beta = 0$ and variance allocation of $1:1$ from a normal-gamma model with frontier $\beta' = 1.51$ and variance allocation

1 : 4.

However, as opposed the ordinary least squares model, the normal-exponential model does occur in the interior of the parameter space of the normal-gamma model and therefore the χ^2 approximation to the log likelihood ratio statistic is directly applicable. It is therefore comparatively easy to assess whether the normal-gamma model provides a significant improvement over the normal-exponential model.

As we mentioned above, the identification problems are serious if they imply ambiguity in the interpretation of the results. Our considerations suggest that the normal-gamma model is quite prone to such ambiguous answers. This is not the case when the normal-exponential is the super-model and the ordinary least squares model is the sub-model, since the normal-exponential model needs to have a vanishing exponential scale parameter to mimic an ordinary least squares model, and thus has to produce equivalent interpretations of the data in terms of the frontier, the inefficiencies, and the allocation of the variance. It is the step from the normal-exponential model to the normal-gamma model which is responsible for the identification problems discussed here.

A third competitor is the plain gamma frontier model (with deterministic frontier). This model is approached by the normal-gamma frontier model when the variance of the stochastic frontier tends to zero. It provides another benchmark.

3 Maximum Likelihood and Likelihood Profiles

The likelihood for our problem can be written as a product of convolutions of normal and gamma densities:

$$p(y_i \mid \lambda, \alpha, \sigma^2) = \int_0^\infty \frac{\exp\left[-\frac{(y_i - \beta + z)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}} \frac{z^{(\alpha-1)} \exp\left(-\frac{z}{\lambda}\right)}{\lambda^\alpha \Gamma(\alpha)} dz.$$

These integrals are improper, since the domain for z extends to ∞. If $\alpha < 1$, they are also improper at $z = 0$. For integer values of α, they can be written in terms of moments of a truncated normal density (Greene, 1990). In general, they have to be evaluated numerically. Unfortunately, the shape of the integrand can vary considerably depending on y_i, making it difficult to define a common grid or interval in z for numerical integration. Care is therefore needed to construct an algorithm for evaluation of the likelihood which performs well for a large enough part of the parameter domain in order to allow maximum likelihood estimation. Ritter and Simar (1994) give suggestions for how this can be done in practice. Here we use the transformation $g(\zeta) = \zeta^3$ in the integrand and a common integration grid from 10^{-7} to 8.

A proper maximum likelihood estimate need not exist. This occurs, for example, if the data originate from a simple least squares model. Then, the parameters will likely diverge into a direction which corresponds to a least squares model, i.e., α and β may tend to infinity, and λ to zero, maintaining $\alpha\lambda^2$ approximately constant. In this case, the log-likelihood will approach a limit which is slightly higher than the maximum of the log-likelihood of under the simple least squares model, since there are two additional parameters. Even if it is hard or impossible to obtain a maximum likelihood estimate under the normal-gamma model, the estimation under the normal-exponential or the ordinary least squares sub-model may not pose great difficulty.

Regardless of whether a maximum likelihood estimate of the normal-gamma model can be found or not, but conditional on the availability of a maximum likelihood estimate for the normal-exponential model ($\alpha = 1$ in the normal-gamma model), we can select a set of shape parameters $\alpha_{-s} < \cdots < \alpha_{-1} < \alpha_0 = 1 < \alpha_1 \cdots < \alpha_r$ and try to compute conditional maximum likelihood estimates by fixing the shape parameter at any of the values in the sequence. This should be possible at least in a neighborhood of $\alpha = 1$ and, hopefully, in an interval reaching from less than .5 to more than 5, thus covering a range reaching from Pareto-type inefficiency distributions, i.e, $\alpha < 1$ over the exponential case $\alpha = 1$ and genuine gamma cases $1 < \alpha < 4$ to cases representing approximate normality $\alpha \geq 4$.

Let us denote the corresponding conditional maximum likelihood estimates by $\widetilde{\theta}_{-s}, \ldots, \widetilde{\theta}_r$ and the achieved log-likelihoods by $\tilde{l}_{-s}, \ldots, \tilde{l}_r$. A graph of $2\tilde{l}_i$ versus α_i, $i = -s, \ldots, r$ shows the profile of twice the log-likelihood over this important range of shapes. If there is a unique overall maximum likelihood estimate whose shape component $\hat{\alpha}$ lies between α_{-s} and α_r, the profile will be unimodal in the neighborhood of $\hat{\alpha}$. Using the χ^2 approximation to the profile statistic $2\left[\tilde{l}(\widehat{\alpha}) - \tilde{l}(\alpha)\right]$ with one degree of freedom we can construct approximate $1 - \eta$ confidence intervals for α by studying the graph of the profile and by selecting the interval of α values above $2\tilde{l}(\hat{\alpha}) - \chi^2_1(1 - \eta)$. If such intervals are unbounded even for small coverage probabilities, the shape parameter α cannot be estimated reliably from the data, and the normal-gamma model cannot be identified sufficiently well to allow good inference about the frontier, the inefficiencies, and the allocation of variance. How this presents itself in practice is shown in the simulations in Section 5. A quick

evaluation of a profile of $2\tilde{l}(\alpha)$ can be done by adding twice the likelihood under the ordinary least squares model and under the pure gamma model to the graph. As we detailed in the previous section, the ordinary least squares model is approached by the normal-gamma model when the shape parameter α becomes very small or very large. Therefore, twice the log likelihood of the ordinary least square case can be added to the graph either at the left or the right end. Twice the log-likelihood under the pure gamma model is most conveniently added at the estimated value of the shape parameter. If the highest observed values of the profile do not sufficiently exceed the corresponding values under the sub-models, the data provide insufficient information to determine a normal-gamma model and to set it apart from the sub-models.
What we just described can be implemented in the following procedure:

- Compute maximum likelihood estimates under the simple least squares model and the pure gamma model. Call the resulting estimates $\hat{l}_{ols}$ and $\hat{l}_{\gamma}$.

- Compute constrained maximum likelihood estimates of the normal-gamma model for fixed values $\alpha_{-s}, \ldots, \alpha_r$ including $\alpha_0 = 1$, the normal-exponential case, and denote them by $\tilde{l}_i$, $i = -s, \ldots, r$.

- Construct a plot of $2\tilde{l}_i$ versus α_i. We find it convenient to use a logarithmic or a square-root scale for α in order to avoid a crowded picture in the region $0 < \alpha < 1$.

- Add $2\hat{l}_{\gamma}$ at the value of α obtained with the pure gamma estimation. Add also $2\hat{l}_{ols}$ to the picture. We suggest doing this either at the left or at the right end of the picture.

- Draw a line d units below the observed maximum of the profile, where d is defined as the upper $1 - \eta$ percentile of a χ^2 distribution with one degree of freedom and where $1-\eta$ is the desired coverage probability.

- Define the confidence interval for α as the interval containing the set of α values which correspond to ordinate values above the line.

- If there is strong correlation between the shape parameter α and the estimates of the frontier β and of the proportion ρ of the combined variance attributed to the stochastic frontier, approximate confidence intervals for those parameters can be constructed as the images of the confidence interval for α.

Remark 1: If the pure gamma or the simple least squares case are contained in the approximate confidence set for α, the normal-gamma model cannot be identified for practical purposes and the reduced models provide more parsimonious explanations of the data.

Remark 2: The likelihood profile for α can be regarded as a pseudo-likelihood, and the latter be used for marginal inference on the shape parameter. However, this method suffers from the fact that the variance of the conditional estimate $\widetilde{\theta}$ given α can vary with α. This implies that the profile likelihood may not give a balanced account of the overall variability of α. If this is of concern, modifications of profile likelihoods can be used adjusting for the changing levels of variation of the conditional maximizer of the likelihood (Barndorff-Nielsen, 1983; Barndorff-Nielsen, 1986; Cox and Reid, 1987; McCullagh and Tibshirani, 1990). In the Bayesian framework, the Laplacian approximation to the marginal posterior density of α provides the desired adjustment (Leonard, 1982; Tierney and Kadane, 1986; Tierney, Kass and Kadane, 1989; Leonard, Hsu and Tsui, 1989; Leonard, Hsu and Ritter, 1994).

Remark 3: Constructing confidence intervals for the frontier and for the proportion of variance contained in the stochastic frontier from the confidence intervals of α and the corresponding conditional maxima is only a crude approach. For a refinement, profiles could also be computed for those parameters. However, if there is high correlation between α and the other parameters, their profiles are close to the α profile and the above approximation is good.

Example: Greene (1990) shows how a normal-gamma stochastic frontier model can be estimated from data on American electrical utilities presented in Christensen and Greene (1976). Besides the estimates of the parameters of the frontier, he obtains an estimate of 2.45 for α with a standard error of 1.01. Unfortunately, as a consequence of numerical problems in the calculation of the likelihood function, this parameter estimate is incorrect (Ritter and Simar, 1994). With improved numerical methods one finds that the likelihood favors shape parameters smaller than one. The α component of the overall maximum likelihood estimator is about 0.2.

Figure 5 shows the profile of twice the log-likelihood. The value for the ordinary least squares estimate was added to the left of the plot. One sees that the profile is very flat and that the value of the ordinary least squares case lies just barely below the line defining an approximate 95% confidence region for α. Following the above arguments we conclude that the shape parameter cannot be estimated from the data and that there is little evidence that the normal-gamma model improves over the simpler sub-models. Table 2 shows the individual inefficiencies estimated using the method by Greene (1990) for firm 1 through 5 as a function of α. We see that the individual inefficiencies depend considerably on α although their ranking remains unchanged except for the extreme case of the pure gamma model. Note that firm 1 is identified as inefficient regardless of α, whereas the other firms are considered nearly efficient for α close to zero and clearly inefficient if α is large.

4 Simulation Study

Here we simulate data sets from a simple normal-gamma model of from (1) and analyze them using likelihood profiles. We always use a gamma distribution with shape parameter $\alpha = 2$ and scale parameter $\lambda = 1$. It corresponds to a mean inefficiency of $\alpha\lambda = 1$ and a variance of the inefficiency distribution of $\alpha\lambda^2 = 2$. Such a gamma distribution is clearly different from either an exponential distribution or a normal distribution. For the stochastic frontier we always use a mean $\beta = 0$, but we vary the variance in order to express a range of different allocations of the combined variance to the stochastic frontier and the inefficiencies.

The simulations are conducted for the seven scenarios shown in Table 1. Here n denotes the number of observations and ρ the share of the total variance attributed to the stochastic fluctuations of the frontier, i.e., $\rho = \sigma^2/(\alpha\lambda^2 + \sigma^2)$. The choice $\rho = 1/3$ represents a normal-gamma stochastic frontier model with an appreciable effect of the stochastic frontier for which the variance inefficiencies is still the dominating part.

4.1 Simulation

For each of the scenarios in Table 1 sixty data sets are simulated. For example, for scenario $(n, \rho) = (400, 1/5)$ sixty data sets with 400 observations each are simulated from a normal-gamma model with $\beta = 0$, $\alpha = 2$, $\lambda = 1$, and $\sigma^2 = 1/4$. For each of data set, the value of the log-likelihood is calculated at the true parameter values, yielding l^*. In addition the maximum of the likelihood under the ordinary least squares model, the pure gamma model and of the models corresponding to $\alpha_1 = 0.5, \alpha_0 = 1, \alpha_1 = 2, \alpha_2 = 3, \alpha_3 = 4$ is computed providing $\hat{l}_{ols}$, $\hat{l}_\gamma$, $\tilde{l}_{-1}$, $\tilde{l}_0$, $\tilde{l}_1$, $\tilde{l}_2$, and $\tilde{l}_3$. To make these results comparable between data sets and scenarios, the likelihood ratio statistics $r_{ols} = 2\left[\hat{l}_{ols} - l^*\right]$, $r_\gamma = 2\left[\hat{l}_\gamma - l^*\right]$, $r_{-1} = 2\left[\tilde{l}_{-1} - l^*\right]$, $r_0 = 2\left[\tilde{l}_0 - l^*\right]$, $r_1 = 2\left[\tilde{l}_1 - l^*\right]$, $r_2 = 2\left[\tilde{l}_2 - l^*\right]$, $r_3 = 2\left[\tilde{l}_3 - l^*\right]$ are computed. As a result, for a given scenario we obtain 60 values of r_{ols}, 60 values of r_γ, etc.. For each scenarios, these values are summarized by taking the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} quantiles of r_{ols}, r_γ, etc. The same is done for the estimated frontier $\tilde{\beta}$ and variance allocation $\tilde{\rho}$.

A fast comparison of the scenarios is given by a plot of the medians of the likelihood ratio statistics. It is shown in Figure 2. More detail can be obtained by plotting all quantiles and by adding the corresponding pictures of the estimated frontier and of the variance allocation. These are shown in Figures 3 and 4 for the cases of 100 observations with $\rho = 1/2$ and 800 observations with $\rho = 1/3$. In each case, the top plot depicts the percentiles of the likelihood ratio statistics, the center plot the percentiles of the estimated frontier, and the bottom plot the quantiles of the estimated share of the variance taken up by the stochastic frontier. The plots of the likelihood-ratio statistic have been enhanced by four lines. The three dashed lines give the 25^{th}, 50^{th}, and 75^{th} percentiles of a χ^2 distribution with three degrees of freedom. The fat line is drawn d units below the median line where d is the 95^{th}percentile of a χ^2distribution with one degree of freedom. These lines serve the following purposes. First, since the true α is 2, the distribution of $2(\tilde{l}_1 - l^*)$ is approximately χ^2 with three degrees of freedom. Therefore, the dashed lines allow a comparison of the theoretical percentiles of this approximation with the percentiles observed in the simulation. Second, the fat line gives a cutoff for confidence intervals for α. If the median of r_1 corresponding to $\alpha = 2$ is above the line while the medians of r_{-1}, r_3 and r_γ are below the line the simulation-estimation procedure will generate estimates of α close to the true $\alpha = 2$ most of the time. If, however, all the medians are above the line, the estimation of

α is a hopeless endeavor.

4.2 Analysis of the Simulation Results

The top plots of Figures 3 and 4 suggest that the χ^2 approximation to twice the log-likelihood ratio statistic at $\alpha = 2$ works rather well. When inspecting Figure 2 for the quality of the estimation we notice that for $\rho \leq 1/3$ the median of r_{ols} lies always much below the fat line. Therefore, at $\rho \leq 1/3$ only 100 observations are sufficient to exclude the ordinary least squares model as a probable explanation. However, if the normal component accounts for at least half of the variance ($\rho \geq 1/2$), one hundred observations might not be enough to exclude the ordinary least squares model. Here, r_{ols} still lies below the fat line but not very much. Note in this context that the ordinary least squares model has two parameters as opposed to four in the normal-gamma model but does not occur in the interior of the parameter space. The fat line is based on a χ^2 approximation to the twice the log-likelihood ratio with one degree of freedom whereas between one and two degrees of freedom might be more appropriate. Therefore, the ordinary least squares model should still be considered a serious competitor to the normal-gamma model.

The normal-exponential model, the pure gamma model and a large range of normal-gamma models provide explanations to the data which are as probable as the "true" model even for several hundreds of observations. Larger samples size clearly improve the estimation of the shape parameter. Improvement is also obtained if the variance component corresponding to the stochastic frontier is weaker as shown in the cases (f) and (g). In the latter circumstances, we are able to exclude values of α below 1. However, the pure gamma model is now a more serious competitor.

Difficulties in the estimation of α may not be serious if the estimation of the frontier and the variance allocation are not affected. Yet, Figures 3 and 4 show that the estimation of these quantities may be strongly related to the estimation of α. In our model, the complete uncertainty in the estimation of α for 100 observations translates into an almost complete uncertainty in the estimation of the frontier and of the variance allocation.

5 Discussion

Our study raises serious doubt on whether one can reliably estimate the shape of a normal-gamma stochastic frontier model unless the sample size is large. This follows from our discussion of limit cases and from the simulation study we conducted. In particular, our simulations lead to the following conclusions.

1. A normal-gamma stochastic frontier model with shape parameter $\alpha = 2$ and less than half of the variance allocated to the normal component can be distinguished from an ordinary least squares model with as few as 100 observations. If the variance of the normal component is larger, and if there are only 100 observations, such a normal-gamma model does not differs sufficiently from an ordinary least squares model to allow a reliable distinction.

2. Even for are several hundreds of observation, a normal-gamma stochastic frontier model with shape parameter $\alpha = 2$ cannot be distinguished from a normal-exponential stochastic frontier model or from other normal-gamma models with shape parameters as low as 0.5 and as high as 4. This large range of competing models corresponds to a large range of interpretations in terms of the frontier, the individual inefficiencies, and the allocation of variance.

3. The shape parameter of the gamma distribution is somewhat better determined if the variance of the normal component (fluctuations of the frontier) is much smaller than the variance of the inefficiencies. Even then, much larger sample sizes than presented here are needed for a reliable estimation.

4. Since, as shown in Figures 3 through 4, the estimates for α, for the stochastic frontier β, and for the allocation of the combined variance to the stochastic frontier and to the inefficiencies can be strongly correlated, the uncertainty in estimating α can carry over directly to the estimates of the stochastic frontier, the variance allocation, and subsequently to estimates of the individual inefficiencies.

The above results were obtained from simulations on several versions of one particular example only. However, the choice of this example was such that the results for most other examples of normal-gamma models will be at least as bad as the ones presented here.

We believe that the same can be said in similar types of analysis, such as estimation by methods of moments in normal-gamma models, estimation of frontier-inefficiency structures using normal-truncated-normal models, or non- and semiparametric methods which allow fluctuations of the frontier and which do not fix the shape of the distribution of the inefficiencies.

In the composed-error framework, fixed-shape models, such as normal-exponential, normal-half-normal, or normal-gamma with predetermined shape, may be undesirable since they are based on a rather arbitrary choice about the distributions of the inefficiencies and therefore provide biased results. However, experience has shown that they are relatively easy to use and that the precision (not the bias) of their results is under control. Comparisons of groups of firms based on inefficiencies estimated using the same fixed-shape model may often be unaffected by the bias. Thus, the estimation of inefficiencies using fixed-shape composed error models may define a valid measurement process for the purpose of comparisons.

The free-shape normal-gamma model promises reduced bias, since the shape is estimated jointly with the other parameters from the data. However, as our analysis shows, the estimation of the shape parameter requires large sample sizes, that is several thousand observations. If one attempts to glean the shape parameter from small sample sizes one is left with an imprecise estimate of the shape parameter and an almost complete loss of precision regarding the quantities of interest, i.e., the frontier, the allocation of variance, and the individual inefficiencies. Such results would be completely worthless for the purpose of comparisons, and particularly so, if the shape parameter is estimated within each group of firms. We believe that in general algorithms aimed at estimating frontiers and inefficiencies using free-shape composed error models from small to medium sized samples do not constitute valid measurement processes.

6 Notes

* This research was partly funded by the European Community under grant ERBCHBGCT920167, by CORE, and by the Ministery of Education and Research of the Communauté Française de Belgique under grant 93/98-164. The authors would like to thank Knox Lovell, Mark Steel, Gary Koop and Philippe Van den Eeckaut for their comments, suggestions, and encouragements.

References

Aigner, D., Lovell, C. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* **6**: 21–37.

Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika* **70**: 343–365.

Barndorff-Nielsen, O. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio, *Biometrika* **73**: 307–322.

Beckers, D. and Hammond, C. (1987). A tractable likelihood function for the normal-gamma stochastic frontier model, *Econometric Letters* **24**: 33–38.

Christensen, L. and Greene, R. (1976). Economics of scale in US electric power generation, *Journal of Political Economy* **84**: 653–667.

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society Series B* **49**: 1–18.

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

Leonard, T. (1982). Comments on "A simple predictive density function", *Journal of the American Statistical Association* **77**: 657–658.

Leonard, T., Hsu, J. and Ritter, C. (1994). The Laplacian-t approximation in Bayesian inference, *Statistica Sinica.* in press.

Leonard, T., Hsu, J. S. J. and Tsui, K. W. (1989). Bayesian marginal inference, *Journal of the American Statistical Association* **84**: 1051–1058.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods, *Journal of the Royal Statistical Society Series B* **52(2)**: 325–344.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* **8**: 435–444.

Ritter, C. and Bates, D. (1993). Profile methods, *Technical Report 93–31*, Institut de Statistique, Universié Catholique de Louvain, B-1348, Louvain-la-Neuve, Belgium.

Ritter, C. and Simar, L. (1994). Another look at the American Electrical Utility data, *Technical report*, Institut de Statistique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.

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

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and densities, *Journal of the American Statistical Association* **81**: 82–86.

Tierney, L., Kass, R. and Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions, *Biometrika* **76**: 425–433.

van den Broeck, J., Koop, G., Osiewalski, J. and Steel, M. F. J. (1993). Stochastic frontier models: a Bayesian perspective, *Journal of Econometrics*.

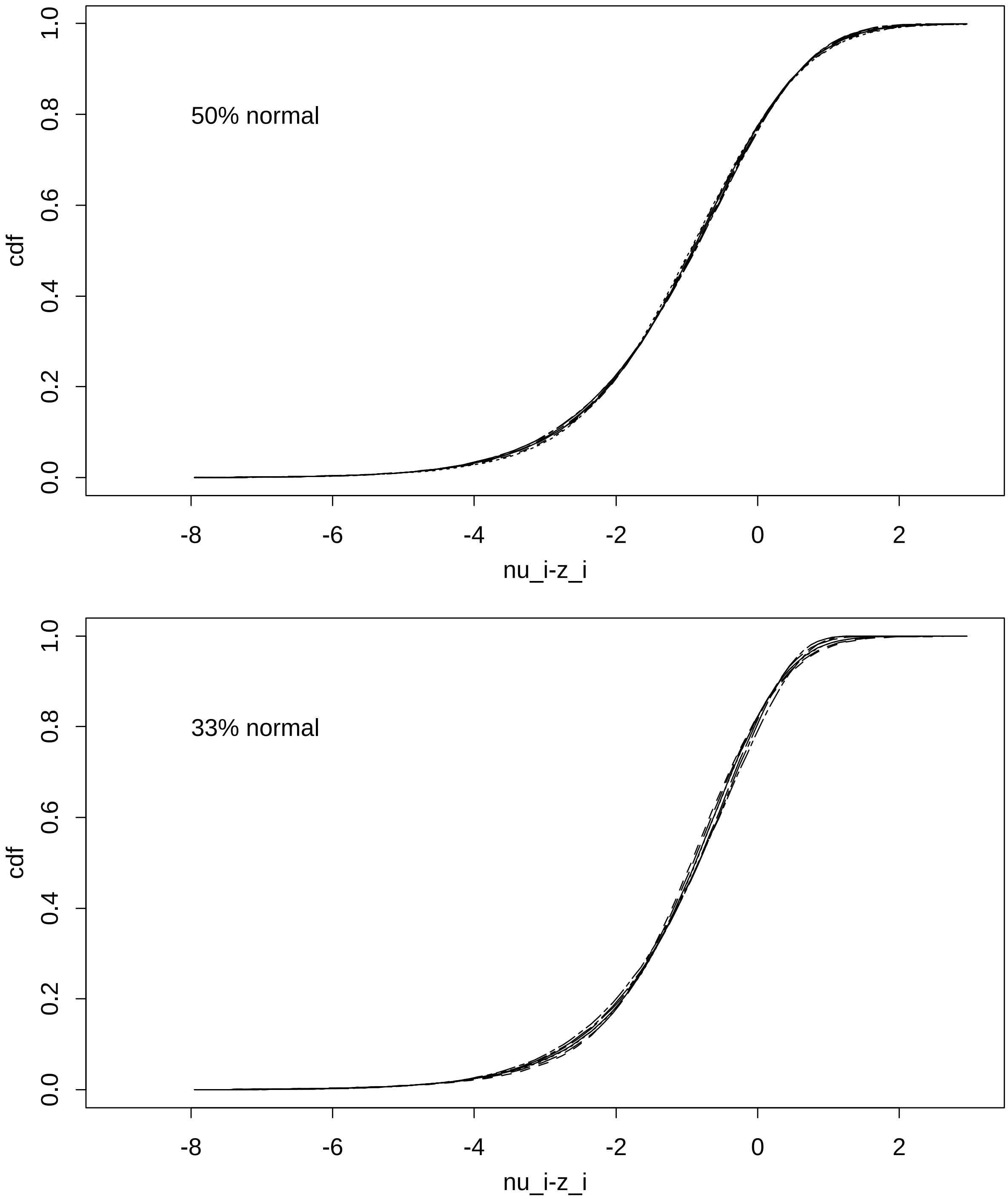

Figure 1: Empirical cumulative distribution functions for samples of 250000 observations for normal-gamma models matched to a normal-exponential model. Top plot: normal component makes 50% of the variance in the normal-exponential model; bottom plot: normal component accounts for 33% of the variance of the normal-exponential model.

	ρ			
n	1/9	1/5	1/3	1/2
100			x	x
200			x	
400	x	x	x	
800			x	

Table 1: Simulation scenarios

α	firm 1	firm 2	firm 3	firm 4	firm 5
0.1	64.9	99.7	99.2	98.4	99.5
0.2	65.8	99.4	98.5	97.2	99.1
0.5	66.8	98.5	96.6	94.6	97.9
1.0	67.1	97.3	94.2	91.7	96.3
2.0	66.2	95.1	90.2	87.6	93.6
3.0	65.0	93.2	87.2	84.6	91.2
4.0	63.7	91.4	84.5	82.0	89.1
62.2	78.3	59.2	70.4	71.5	70.0

Table 2: Estimated individual inefficiencies in percent for different values of α. The last line corresponds to an approximate pure gamma fit.

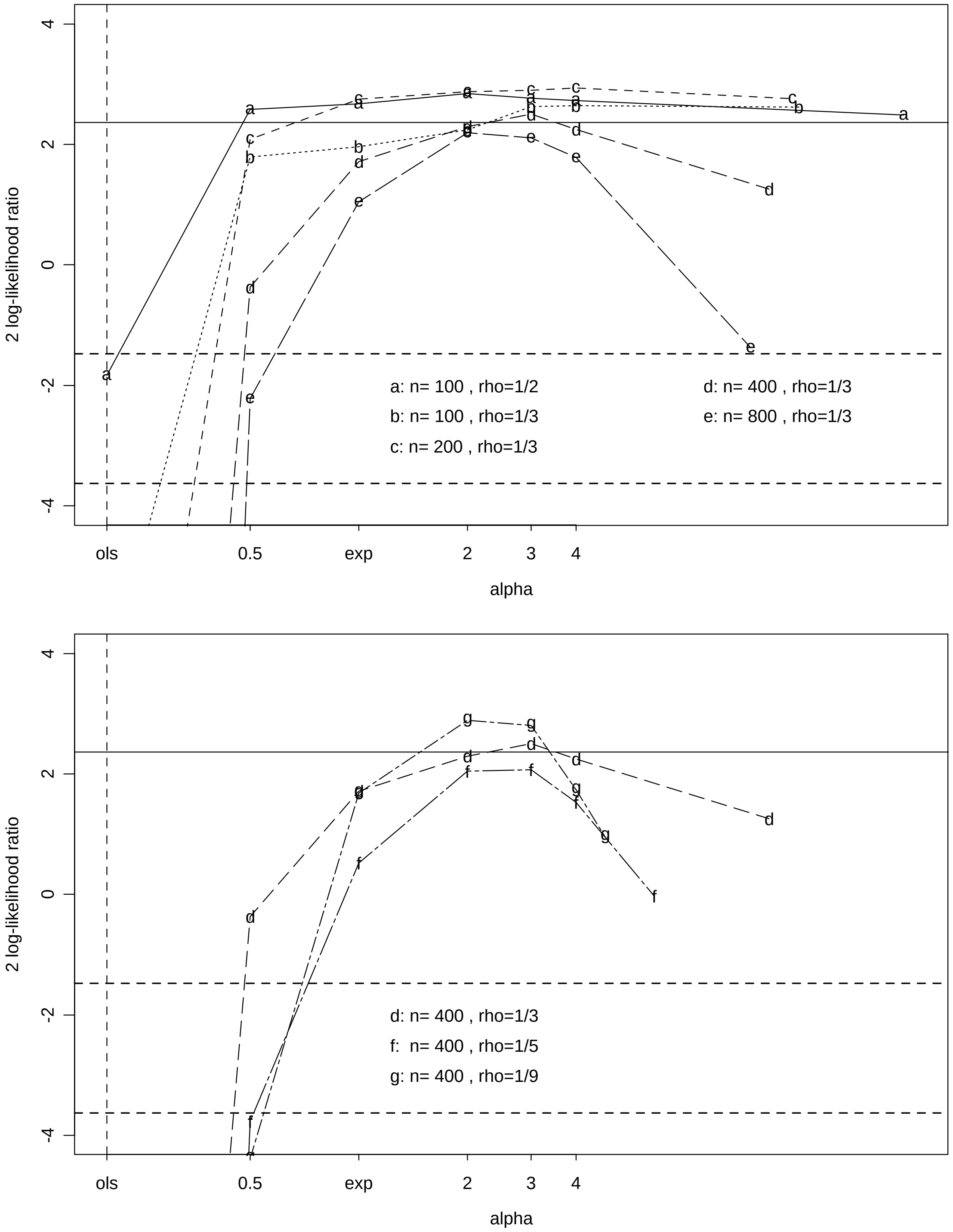

Figure 2: Medians of r_{ols}, r_γ, r_{-1}, ..., r_3 versus α for scenarios (a) through (f). Top plot: effect of sample size; bottom plot: effect of variance allocation. The case (d) is shown in both plots and serves as a reference. The ordinary least squares case is added to the left of the picture.

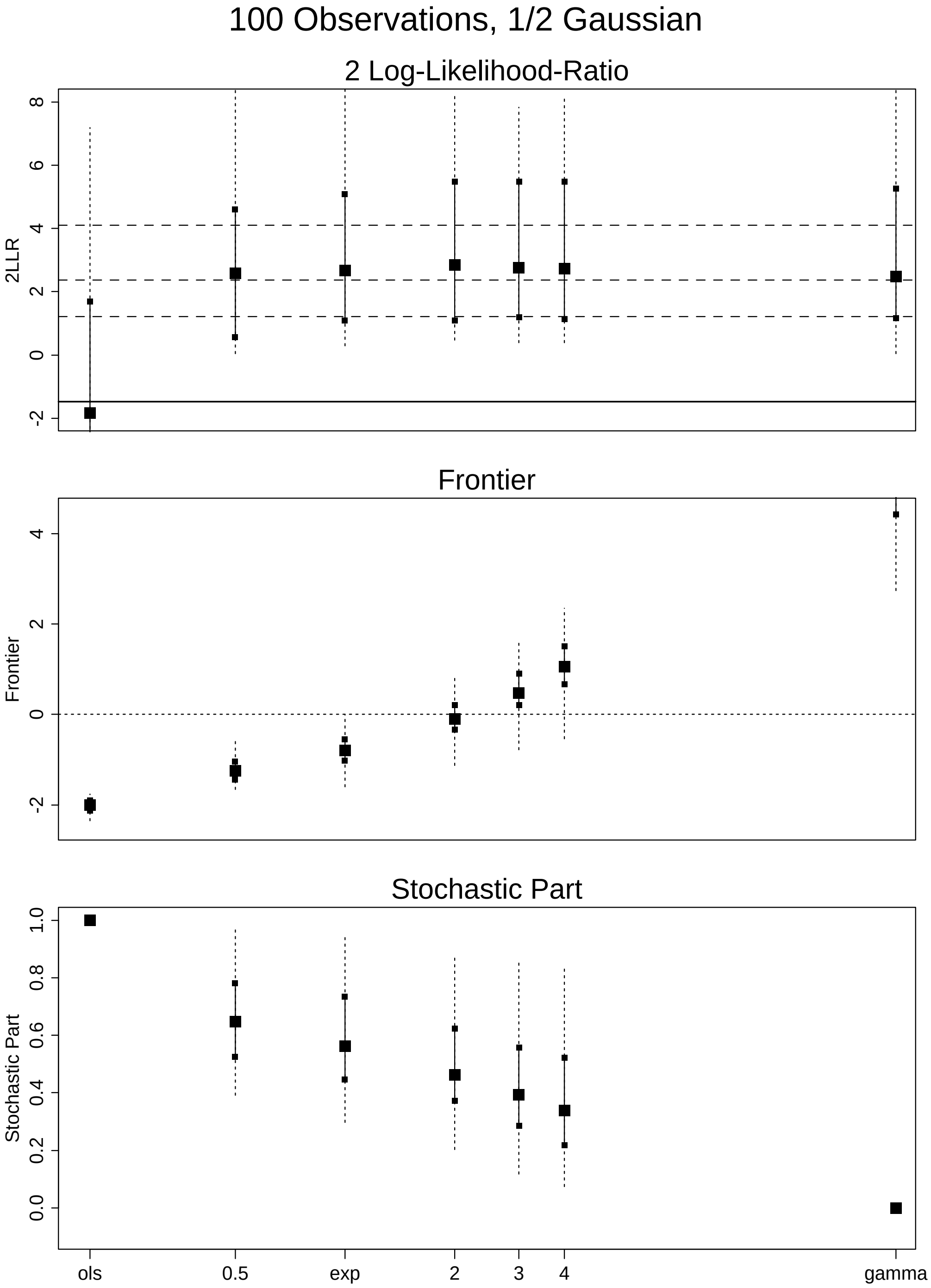

Figure 3: 100 observations, 60 simulations, $\rho = 1/2$. Percentiles 5, 25, 50, 75, 95 of twice the log likelihood ratio (top), the estimated frontier (middle), and the proportion of the variance corresponding to the stochastic frontier.

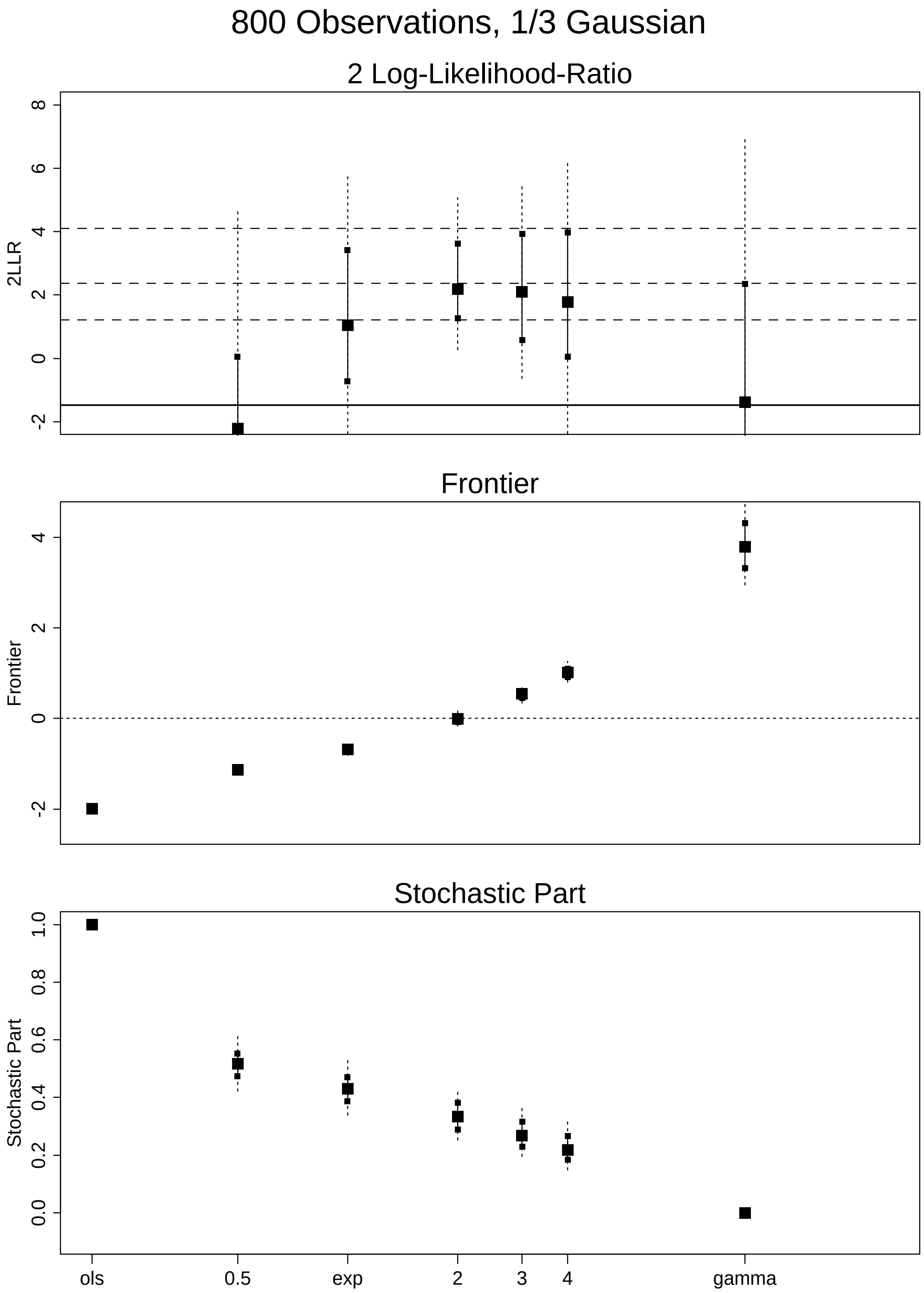

Figure 4: 800 observations, 60 simulations, $\rho = 1/3$. Percentiles 5, 25, 50, 75, 95 of twice the log likelihood ratio (top), the estimated frontier (middle), and the proportion of the variance corresponding to the stochastic frontier.

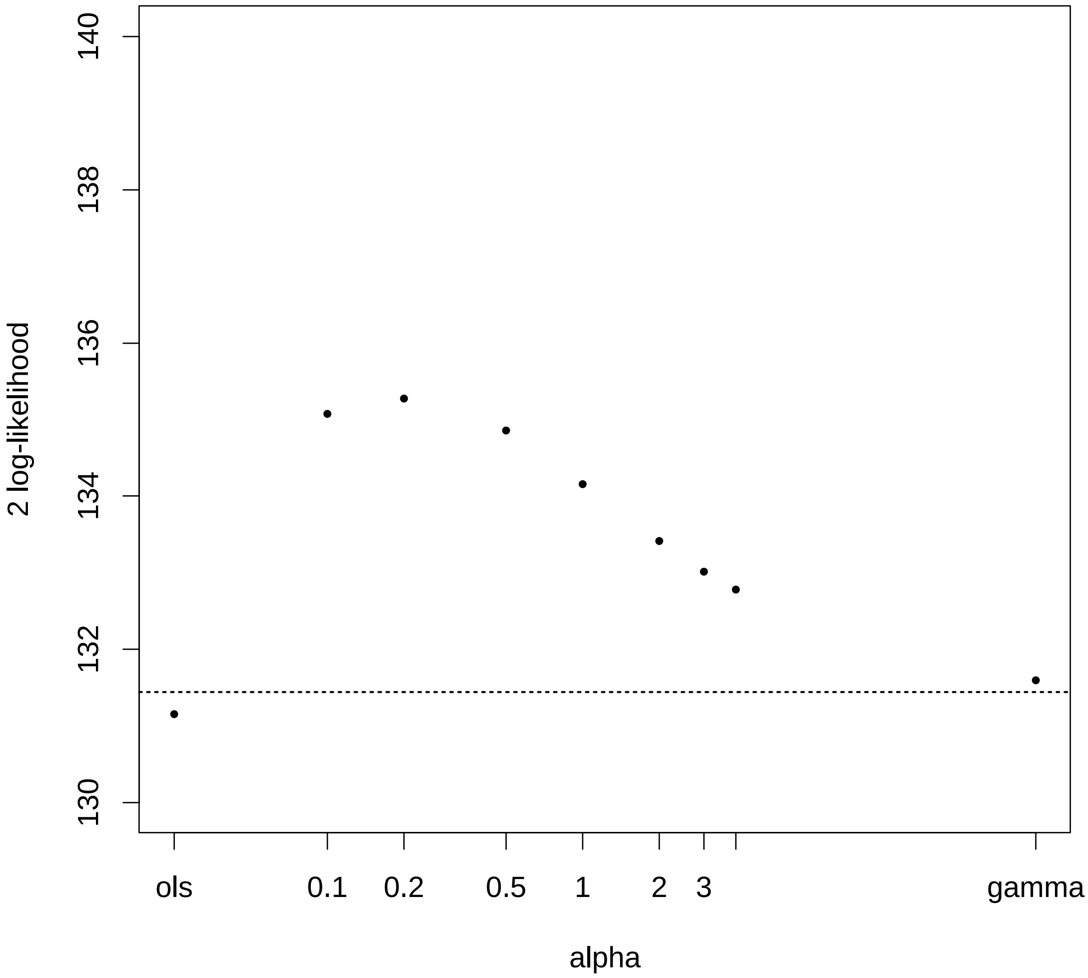

Figure 5: Likelihood profile for American electric utility data. The line is drawn $\chi_1^2(0.95) = 3.84$ units below the observed maximum.