ECONOMICS DEPARTMENT

WORKING PAPER

2009

Department of Economics
Tufts University
Medford, MA 02155
(617) 627-3560
http://ase.tufts.edu/econ

Gibrat's Law for (All) Cities: A Rejoinder

September 30, 2009

Yannis M. Ioannides

yannis.ioannides@tufts.edu

Tufts University

Department of Economics

Medford, Massachusetts 02155, USA

and

Spyros Skouras

skouras@aueb.gr

Athens University of Economics and Business

Department of International and European Economic Studies

76 Patision St. 104 34, Athens, Greece

Gibrat's Law for (All) Cities: A Rejoinder

September 2009

Abstract

We establish that the debate between Eeckhout (2004; 2009) and Levy (2009) has still not resolved the key issue of whether the distribution of large US urban places in 2000 is consistent with a lognormal for the entire size range. We resolve this by introducing a new distribution function which switches between a lognormal and a power distribution and estimating it with the data used by Eeckhout and Levy (2009). We find that there is a sudden transition from lognormality to power behavior as city populations increase above 100,000. Gibrat's law holds for most cities but a power law holds for most of the population.

JEL codes: D30, D51, J61, R12, C24.

Keywords: Gibrat's law, Zipf's law, upper tail, mixture of distributions, switching regressions, urban evolution, urban hierarchy.

1 Introduction

Levy (2009), Eeckhout (2004; 2009) and many other researchers in the area of city size distributions agree that knowledge of the probability law describing the upper tail of the distribution is important because that is where most of the population lives. For example, in the data on US places used by both Eeckhout and Levy, 15% of all places have population above 10,000 people, but they accommodate 80% of the population. More dramatically, 1% of all places are larger than 100,000 people but accommodate 63% of the population.

Levy (2009) presents a mostly graphical analysis to counter Eeckhout's (2004) claim that populations of places are lognormally distributed throughout the range of observed size distributions. Levy's evidence suggests that the tail is in fact power law distributed as was widely accepted before Eeckhout's work. But Eeckhout (2009) in his reply points to several drawbacks in Levy's critique and in his own initial analysis concluding with a reaffirmation of his original claim that "the tail of the distribution is indeed lognormal" (*ibid.*, p.1676). The purpose of this rejoinder is twofold: first, to argue that Eeckhout's analysis in his Reply (which is based on a Lilliefors test) is also problematic; and second and most importantly to resolve conclusively the issue of the behavior of the tail and its relation to the body of the distribution.

The main obstacle to the second goal is that the analysis of the 'tail' of any empirical distribution requires an assumption about where the body ends and the tail begins which is *ad hoc* and can seriously affect the conclusions. As Eeckhout, *ibid.*, p.1682, puts it "[w]ith all the data available, and given that one nonetheless does not want to use all data, the question arises what the *appropriate truncation point* is. The choice of the truncation point becomes *endogenous* and can be chosen subjectively to favor one hypothesis over the other." This is the problem that led both Eeckhout and Levy to methods with serious drawbacks which were ultimately inadequate for the intended analysis. We sidestep fully this problem by modeling the *entire* empirical distribution of city sizes with a distribution that can accommodate both lognormal and power behavior over different ranges. We therefore allow the data to determine where, if anywhere, a transition occurs from one behavior to the other as well as how fast the transition is. Our approach builds on our proposed transition distribution function which allows the entire analysis to be conducted with familiar and unambiguous maximum likelihood methods.

In the remainder of this rejoinder we first show that Eeckhout's (2009) Lilliefors test is too weak as a method for testing the hypothesis of lognormality in the tail. We then present and estimate, using the same data as Eeckhout and Levy, a new distribution function which switches between two regimes [e.g. Maddala, 1983] one being lognormal and one power. We find that there is a sudden transition from lognormality to power behavior (not far from Zipf's law) around cities of population 100,000. In the next subsection we discuss the implications of our finding for the theoretical modeling of urban systems broadly as well as for Eeckhout's (2004) model in particular and we close with a final section collecting the main conclusions of our work. This rejoinder may be summed up as reporting conclusive evidence for Gibrat's law for *most cities* but approximately Zipf's law for *most of the population.*

2 The Lilliefors test has low power

Eeckhout (2009) interprets non-rejection of lognormality by a Lilliefors test as evidence that the size distribution of cities is best modeled as lognormal everywhere, including the tail which is of particular interest. But the reason lognormality cannot be rejected by a Lilliefors test is that this test has very little power to detect deviations from a hypothesized distribution when these deviations occur in the tail. This is immediately apparent from the confidence intervals in Eeckhout (2009), Figure 2, in the tail. These confidence intervals are consistent with almost any imaginable distribution for cities larger than $\exp(12) \simeq 160,000$, including power laws with a very broad range of exponent, e.g. 0.1 to 10. In contrast, several previous estimates of distributions of cities in this range have delivered fairly narrow confidence intervals[1] suggesting that the Lilliefors test has little power for distinguishing

[1] See Gabaix and Ioannides (2004), p. 2345, where they discuss Krugman's estimate of 1.005, with a standard error of 0.010, obtained for the top 135 US metropolitan areas in 1990, and *ibid.,* 2351–2354, for a

between tail behaviors. Indeed, it would be very puzzling if city sizes were actually lognormal and this was the narrowest possible confidence interval. If this were the case, we would expect significant statistical fluctuations of the behavior of distribution tails across different samples usually viewed as independent, such as samples from distant time periods and different countries. However, they are at least somewhat clustered around power laws with exponents close 1 (though often larger).

More fundamentally, a Lilliefors approach to testing the adequacy of the lognormal specification is not quite appropriate. These are omnibus tests which are designed to detect an arbitrary deviation from a hypothesized distribution rather than a specific deviation of interest such as the behavior of the tail. Since we are interested in the consistency of the lognormal with the distribution of the size of large places we can consider this much more directly by estimating confidence intervals for the tail by drawing repeated samples from the lognormal equal in size to the observed sample (25358 observations) and examining the distribution function across simulations.

Indeed, we drew 1000 such samples, calculated the simulated distribution function for each sample and at each population value determined the $x\%$ highest and lowest frequencies observed across distribution functions. These $x\%$ confidence intervals are reported in Figures 1a and 1b below for $x = 5, 1$ and 0.1. Formally these are confidence intervals for the null that the empirical distribution is drawn from a lognormal when we consider the empirical distribution at a single *particular* population size. As such they can also be used to detect at which population values, if any, this null hypothesis can be rejected and therefore where the fit is good and where it is not. It is clear from the Figures that for any population size above 100,000 the empirical distribution function behaves differently to what we expect to occur even with 0.1% probability under a lognormal.

This suggests strongly that Eeckhout's (*ibid*, p.1681) claim that "[g]iven that the tail of a lognormal is indistinguishable from the Pareto in certain circumstances, the researcher

discussion of estimates obtained with similar precision by many others.

who is interested in the tail properties of a size distribution can choose which one to use" is unjustified — the tail is indistinguishable only when a misleadingly weak test is used.

However, this analysis does not solve the more fundamental problem of how to model the distribution of city sizes so as to accommodate the observed behavior across its entire range. This is what we turn to in the next section.

3 Letting the data determine where a power law fits better than the lognormal

Eeckhout, *ibid*, correctly argues that no one has fully formalized the notion that city size distributions are power law distributed in the upper tail while at the same time having another distribution elsewhere. Eeckhout also states that arbitrarily setting the truncation point for what constitutes a tail can lead to biased inference. We eschew subjective decisions about splitting the sample arbitrarily by proposing a distribution that is a mixture of a lognormal distribution and a power law, with the mixing distribution being estimated jointly with the other two distributions. Consequently, the resulting estimate combines a lognormal body and a power law upper tail, while limiting values of certain parameters imply either law with virtual certainty. The distribution satisfies the usual regularity conditions required for maximum likelihood estimation and inference. Its estimation lets the data determine whether there is a place size above which power behavior is statistically preferable to lognormal behavior.

The proposed density function is:

$$l_X\left(x;\mu,\sigma,m,s,\beta,\zeta\right)=\frac{\alpha\left(x\right)\beta g_X\left(x\right)+\left(1-\alpha\left(x\right)\right)f_X\left(x\right)}{\int_0^\infty\left[\alpha\left(x\right)\beta g_X\left(x\right)+\left(1-\alpha\left(x\right)\right)f_X\left(x\right)\right]dx},x\geq 1 \tag{1}$$

where f_X is a lognormal density function:

$$f_X\left(x;\mu,\sigma\right)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\ln x-\mu\right)^2}{2\sigma^2}\right),x>0; \tag{2}$$

g_X is a power law density function with lower cut-off specified at 1:

$$g_X(x;\zeta) = \frac{\zeta}{x^{1+\zeta}}, x \geq 1, \zeta > 0; \tag{3}$$

and $\alpha(x)$ is a normal cumulative distribution function:

$$\alpha(x;m,s) = 0.5 + 0.5 \operatorname{erf}\left(\frac{x-m}{s\sqrt{2}}\right). \tag{4}$$

Note that the restriction $x \geq 1$ in the range of the proposed density and the power law density is innocuous for our purposes since smaller place sizes do not make sense. The denominator in the definition of the density function (1) ensures that l_X integrates to one while l_X is continuous and infinitely differentiable with respect to its parameters $(\mu, \sigma, m, s, \beta, \zeta)$. When m tends to $+\infty$, l_X converges to the lognormal and when it tends to $-\infty$, l_X converges to a power law.

Using Eeckhout's data, we can estimate the parameters of $l_X(x;\mu,\sigma,m,s,\beta,\zeta)$ by maximum likelihood. The maximum likelihood estimator is defined as:

$$\max_{\mu,\sigma,m,s,\beta,\zeta} : \sum_{i=1}^{N} \log\left[\alpha(x_i)\beta g_X(x_i) + (1-\alpha(x_i)) f_X(x_i)\right] \\ - N \log \int_0^\infty \left[\alpha(x)\beta g_X(x) + (1-\alpha(x)) f_X(x)\right] dx.$$

The estimation yields the following:

$$\underset{(0.012)}{\widehat{\mu} = 7.261}, \quad \underset{(0.009)}{\widehat{\sigma} = 1.739}, \quad \underset{(4057)}{\widehat{m} = 100070}, \quad \underset{(1629)}{\widehat{s} = 2580}, \quad \underset{(717)}{\widehat{\beta} = 67616}, \quad \underset{(0.006)}{\widehat{\zeta} = 1.368}$$

The reported standard errors are based on 400 bootstrap samples drawn from the original sample with replacement.[2] The log likelihood is $-234,743$.

[2]As a robustness check we divided our bootstrap sample into four subsamples of size 100 and found that the standard errors in each subsample were very close to that of the entire sample.

The parameters $\widehat{m}$ and $\widehat{s}$ specify where, if anywhere, a switch occurs to a power law from a lognormal. Parameter $\widehat{\zeta}$ is the slope of countercumulative distribution function in log-log space, the exponent of the power law. The $\widehat{\beta}$ parameter ensures continuity of the density in the tail where the transition from the lognormal to the power law density occurs.

Based on these observations we see that the transition to a power law, which is close to but statistically different from Zipf's law ($\zeta = 1$), occurs around a population of 100,000. The transition happens quickly since at these parameters $\alpha(x)$ will be zero for all practical purposes at populations below 90,000 and one above 110,000 suggesting lognormality at places smaller than 90,000 and power law behavior above 110,000. The remarkably good quality of the fit of this distribution is visible in Figure 2a where we see not only that it is much better than the lognormal, but also that it is very good in absolute terms. Additionally the quality of the fit is not an artifact of the large number of estimated parameters: note the very small standard errors of the estimated parameters which are comparable to the standard errors of estimated parameters of the pure lognormal distribution used by Eeckhout. In other words, the estimates obtained are very accurate and the proposed distribution is not too flexible for the purposes of the analysis of interest.

It is worth emphasizing that the deviation from Zipf's law we report is relatively small. This is evident in Figure 2b below where we plot the fit of our distribution constrained to satisfy $\zeta = 1$ (this is required for Zipf's law to hold exactly). Many researchers, and most forcefully Simon (1968), have emphasized that given the extremely sharp predictions of regularities like Zipf's law they should be tested leniently, with evidence of an approximate power law with a parameter not too far from 1 being supportive even if the law can be rejected in the strict statistical sense — after all any null hypothesis can be rejected with enough data and this is a very sharp hypothesis. Relatedly, departures from Zipf's law of the order reported here may easily disappear if one uses an alternative definition of a city. It is worth noting that the tail behavior of the estimated distribution is dominated entirely by Zipf's law which has *no free* parameters and our approach is useful in that it determines where the Zipf behavior begins to be relevant. At the same time, the estimate of ζ obtained here is virtually the same as 1.37, the one reported by Ioannides, Overman, Rossi-Hansberg and Schmidheiny (2008), p. 213, for US metro areas with population exceeding 100,000 in 1990.

4 Implications for the economic theory of urban structure

While several researchers in the literature, including Eeckhout (2004) and Gabaix (1999), have proposed simple economic models for the city size distribution that can predict log-normal or power law distributions of city sizes respectively, we know of no model that can predict a transition from one to the other as observed in the data. Our empirical estimates mean that Eeckhout's (2004) model might work for 99% of cities smaller than 100,000 but this accounts for only 37% of the population; on the other hand, Gabaix's model might work for 63% of the population but only 1% of all cities. To the extent that we seek an explanation of the city size distribution that accounts simultaneously for most cities and most of the population, the literature remains entirely silent.

With the benefit of hindsight this failure of extant models to explain the entire city size distribution is not surprising. These models do not allow for structural differences between larger and smaller agglomerations and so they cannot accommodate changing qualitative behavior of the size distribution across different ranges. But the system of cities literature has argued theoretically and empirically (Black and Henderson (2003); Duranton (2007); Henderson (1974; 1997); Rossi-Hansberg and Wright (2007)) that cities specialize and therefore their growth may be subject to different forces once their characteristics become fixed. Indeed, it is natural that size might be correlated with economic characteristics and Henderson (1997) argues that it is medium size cities that specialize. Additionally, casual empiricism suggests that larger cities host a diverse set of industries, and this is bolstered by the urban hierarchy literature as well. This urban hierarchy principle on the other hand asserts that industries found in a city of a given size will also be found in cities of larger size (Hsu (2009);

Mori and Smith (2009)). Much of the theory of cities has implicitly assumed agglomerations of significant size and simply does not make sense for agglomerations of size, 1, 2 or even one thousand. These considerations also provide a strong theoretical rationale for allowing the data to determine whether different probabilistic laws rule in different parts of the city size distribution. Our result is also economically significant since under lognormality we would have far fewer large cities than we actually do.

Naturally, one would expect that populations of places for different time periods or different countries would yield estimates of different points of transition between lognormality and power law behavior. This transition point may be very meaningful if it is interpreted as the point where a 'town' becomes a 'city' — a problem that e.g. Black and Henderson (2003), p. 347 deal with by imposing an *ad hoc* size threshold that can be applied consistently over time.[3] Beyond potentially meaningful thresholds for cities, estimated transition points are likely to be useful when working, as is usually the case, without data for small agglomerations since the shape of their distribution can be imputed if lognormality is accepted as a plausible distribution for these unobserved agglomerations' sizes.

5 Conclusions

Our results confirm rigorously and in a statistically robust manner the claim made by Levy (2009), *op. cit.,* that "the bottom and middle ranges of the empirical distribution of places fits the lognormal, and the top range ... a power law distribution." Eeckhout's (2004) daring step to use the US data for all places in order to estimate the US city size distribution must be credited for leading to the remarkable finding that the shape of the US city size distribution changes abruptly across its range. Most previous research has used data for the upper tail of the distribution, with the cut-off having been arbitrarily set around a population of 50,000 — dictated roughly by US data availability for metropolitan areas over the twentieth century

[3]In particular they assume the transition occurs at the ratio of the minimum to the mean MSA population as estimated in 1990. Multiplying this with the mean MSA for other years they obtain an operational threshold for where a town becomes a city.

— or 100,000 for international data. In either case, based on such data sets, research could not have detected lognormality of urban populations.

The lognormal-power mixture model we introduce may be useful in modeling many variables beyond population such as wealth, income, internet traffic, ecology, and other instances mentioned by Levy (2009) and others where there have been suggestions of similar behavior. Our analysis not only suggests there is a transition from lognormality to a power law, but also that this happens very rapidly around a population level of 100,000 for the US data on places in 2000, and that the estimated power law in the tail is close to that associated with Zipf's law. As discussed, this heterogeneity of the behavior of the size distribution is not surprising from a theoretical perspective, in that size is likely to play a role in determining cities' economic properties, growth, degree of specialization etc. Yet, this is something the city size distribution literature has not yet dealt with: Eeckhout's theoretical model is refuted for most of the population, whereas theoretical models predicting power laws for all cities (as in Gabaix, 1999) get it right for most of the population but wrong for most cities — Gibrat's law holds for most cities but a power law holds for most people. The estimated power law exponent is not exactly 1, but it is not far and Zipf's law provides a good fit. To return to Krugman's (1996) notable adjective, the power law of city sizes and its exponent remain *spooky.*

6 References

Black, Duncan, and J. Vernon Henderson. 2003. "Urban Evolution in the USA." *Journal of Economic Geography,* 3: 343–372.

Duranton, Gilles. 2007. "Urban Evolutions: the Still, the Fast, and the Slow." *American Economic Review,* 97(1): 197—221.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review*, 94(5): 1429–1451.

Eeckhout, Jan. 2009. "Gibrat's Law for (All) Cities: A Reply." *American Economic Review*, 99(4)): 1676–1683.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–67.

Henderson, J. Vernon 1974. "The Size and Types of Cities." *American Economic Review*, 64(4): 640–656.

Henderson, J. Vernon 1997. "Medium Size Cities." *Regional Science and Urban Economics*, 27: 583-612.

Hsu, Wen-Tai. 2009. "Central Place Theory and City Size Distribution." working paper, Department of Economics, Chinese University of Hong Kong. August.

Ioannides, Yannis M., Henry G. Overman, Esteban Rossi-Hansberg, and Kurt Schmidheiny. 2008. "The Effect of Information and Communication Technologies on Urban Structure," *Economic Policy*, April: 203–242.

Krugman, Paul, 1996, "Confronting the Myth of Urban Hierarchy." *Journal of the Japanese and the International Economies*, 10: 399-418.

Levy, Moshe. 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review*, 99(4): 1672–75.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Mori, Tomoya, and Tony E. Smith. 2009. "On the New Empirical Regularities among Industrial Location Behavior, Industrial Diversities and Population Sizes of Cities in Japan." *Brookings Wharton Papers on Urban Affairs,* 175–216.

Simon, Herbert. 1968. "On Judging the Plausibility of Theories." In *Logic, Methodology and Philosophy of Sciences, Vol III,* ed. B. van Rootsellar and J. F. Staal, 439–459. Amsterdam: North-Holland.

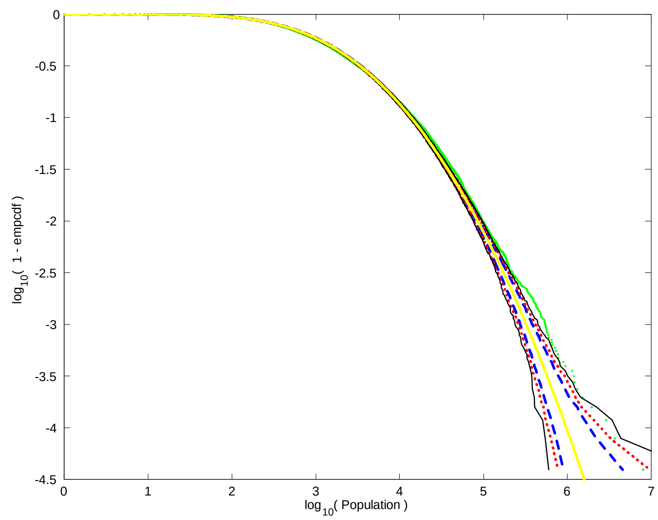

Figure 1a: The countercumulative of the empirical distribution function (circles) is plotted against the countercumulative of Eeckhout's fitted lognormal (central solid line) with confidence intervals for each population level. The confidence intervals are at the 5% (dashed), 1% (dotted) and 0.1% (solid) levels. The countercumulative of the empirical distribution function can also be interpreted as each city's size rank from large to small.

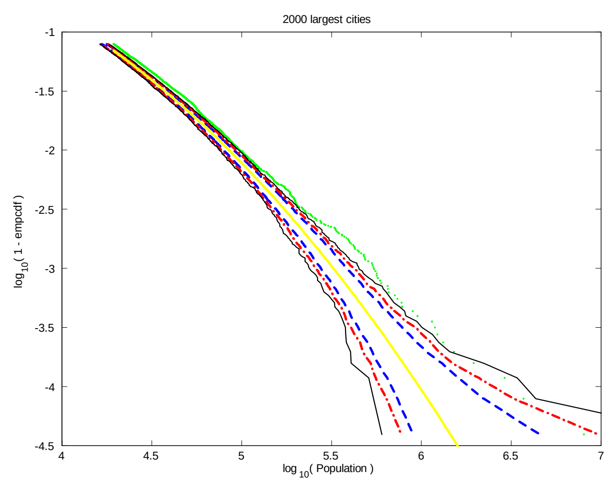

Figure 1b: A magnification into the tail of the countercumulative distributions depicted in Figure 1a. The lognormal provides a poor fit for the 2000 largest cities depicted here (population larger than roughly 20,000).

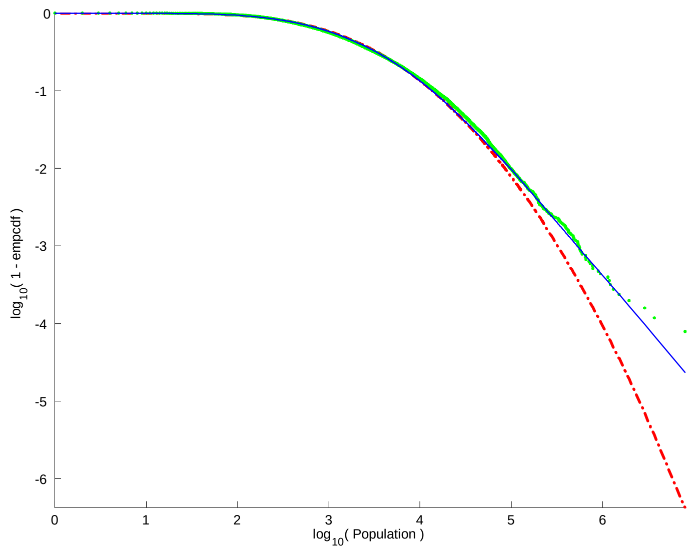

Figure 2a: The countercumulative of the empirical distribution function (circles) is plotted against the countercumulative of Eeckhout's fitted lognormal (dash-dot line) and the estimated proposed distribution (solid). Clearly the proposed distribution with power law tail provides a much better fit. The lognormal misses the empirical distribution by *two orders of magnitude* in the tail.

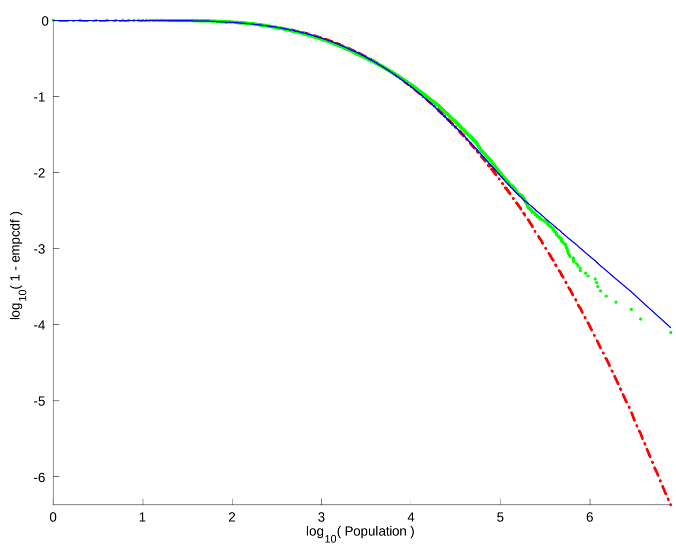

Figure 2b: As in Figure 2a but with the power law exponent constrained to fit Zipf's law exactly. The fit of Zipf's law in the tail is reasonable and certainly better than that of the estimated lognormal.