A Bayesian Latent Variable Mixture Model for Filtering Firm Profit Rates

Gregor Semieniuk and Ellis Scharfenaker

Schwartz Center for Economic Policy Analysis
Department of Economics
The New School for Social Research
6 East 16th Street, New York, NY 10003
www.economicpolicyresearch.org
* This research benefited from discussions with Duncan Foley, Sid Dalal as well as from suggestions by Mishael Milakovic, Amos Golan, Anwar Shaikh, Sid Chib, Fernando Rugitsky, participants in Sid Dalal's Columbia course "Bayesian Statistics", at the ASSRU seminar in Trento, at the New School Economics Dissertation Workshop and at the meeting of the International Society for Bayesian Analysis at Duke University. We are grateful to Columbia University Library for access to the COMPUSTAT dataset. Gregor gratefully acknowledges financial support from INET.

† Both Department of Economics, New School for Social Research, New York City, USA. Email addresses Email address: schae854@newschool.edu (Ellis) and semig068@newschool.edu (Gregor).
Abstract: By using Bayesian Markov chain Monte Carlo methods we select the proper subset of competitive firms and find striking evidence for Laplace shaped firm profit rate distributions. Our approach enables us to extract more information from data than previous research. We filter US firm-level data into signal and noise distributions by Gibbs-sampling from a latent variable mixture distribution, extracting a sharply peaked, negatively skewed Laplace-type profit rate distribution. A Bayesian change point analysis yields the subset of large firms with symmetric and stationary Laplace distributed profit rates, adding to the evidence for statistical equilibrium at the economy-wide and sectoral levels.
Keywords: Firm competition, Laplace distribution, Gibbs sampler, Profit rate, Statistical equilibrium

JEL codes: C15, D20, E10, L11

1 Firm Competition and Bayesian Methods

We propose Bayesian computational methods, in particular Markov chain Monte Carlo simulations for selecting the subset of firms for analysis of firm profitability under competition. This enables us to extract more information from firm level data than the exogenously selected subset yielded in previous research by Alfarano et al. (2012) on profit rates and e.g. Stanley et al. (1996); Bottazzi et al. (2001); Malevergne et al. (2013) on firm growth rates. We obtain striking results of Laplace-shaped profit rate distribution by applying our selection method to US firms in the COMPUSTAT dataset from 1950-2012. This confirms the findings by Alfarano and Milaković (2008) and Alfarano et al. (2012) for a much larger economy-wide dataset. Moreover, we extend the sectoral results of Bottazzi and Secchi (2003a) for growth rates to profit rates.

The formation of a general rate of profit, around which profit rates gravitate, takes place through competition and capital mobility. This point was stressed by classical political economists beginning with Adam Smith, who theorized that through competition capital would instinctively move from one area to another in its endless search for higher rates of profit. From this instinctive movement, a tendency of the equalization of profit rates across all industries would emerge. Excluded from
this process of equalization would be industries that are shielded from competition by legislative measures and monopolies or public enterprises that strive for other objectives than profitability.

Recently, Alfarano et al. (2012) in this journal and also Alfarano and Milaković (2008), examine firm-level data and find that private firms in all but the financial services participate in the competitive processes as described by Adam Smith, which they call “classical competition.” Interested in determining the shape of the distribution of profit rates, Alfarano et al. (2012) fit a generalized error distribution also known as the Subbotin distribution to a sample of firm-level data. They find that throughout their sample the Subbotin’s shape parameter is close to one for a subset of long-lived firms, implying profit rates are distributed as a Laplace distribution.

The Laplace distribution is already a common result in research on firm growth rates. Stanley et al. (1996); Amaral et al. (1997); Bottazzi et al. (2001); Bottazzi and Secchi (2003a,b) find Laplace distributed, or nearly Laplace distributed (Buldyrev et al., 2007) firm growth rates in particular industries. These studies are part of a larger literature that examines distributions of various firm characteristics (Farmer and Lux, 2008; Lux, 2009). A preview of our results in Figure 1 shows that for our data we find strong evidence for Laplace distributed profit rates by sector as well as for the total economy.

One problem in many such analyses is selecting the proper subset of the data. Some papers examine specific sectors, e.g. the pharmaceutical sector in Bottazzi et al. (2001) or use preselected datasets such as the Fortune 2000 (Alfarano and Milaković, 2008). For more general datasets, subsets are commonly selected using an exogenously determined lower size bound or minimum age. For example, in their
Figure 1: Profit rate histograms on semi-log plot with maximum likelihood fit in the class of Laplace distributions. Data is pooled over the years 1961-2012, with filtering by Bayesian mixture and change point models having endogenously removed noise and entry and exit dynamics, as described in sections 3 and 4 of this paper. Results look similar for plots of single years.

A seminal paper on the statistical theory of firm growth rates, Simon and Bonini (1958) disregard firms with a capital stock below a lower bound. Subsequently, Hymer and Pashigian (1962) investigate only the thousand largest U.S. firms, and Amaral et al. (1997) and Malevergne et al. (2013) subset their data using various lower bounds on firm revenue. These exogenous cut-offs appear to be necessary to rid the dataset of noise as well as of new or dying firms, which distort the distribution of firms engaged in the competitive process. In their analysis of *profit rates*, Alfarano et al. (2012)...
consider only firms that are alive at least 27 years or for the entire period spanned by the dataset. While their approach is a good first approximation, they necessarily discard valuable information.

In this paper we argue that instead of a more or less arbitrary cut-off, a Bayesian approach enables us to extract more information from the data. As a natural starting point our prior is based off of the previous research which takes the distribution of profit rates as Laplace. This we take as one of two components of a mixture distribution into which all firms are mapped, regardless of whether they partake in competition or are disruptive noise. In order to decide for each firm whether it is in the Laplace distribution, which we designate as the signal distribution, or confounding noise, we will give a latent variable to each firm observation that is one if the firm belongs to the signal distribution and zero if it does not. We then compute each latent variable’s posterior probability of taking on the values one or zero, and thus the firm’s probability of being part of the competitive process. If the probability of its being in the signal distribution is not sufficiently high, we discard it. For our dataset, after discarding the noise, we retain nearly 97% of all observations regardless of age or size. Unlike previous studies, this allows us to analyze the profit rate distribution throughout the entire competitive process preserving significantly more information. As a second step, we apply a Bayesian change point analysis to filter out small firms not subject to entry-exit dynamics, again using the Laplace distribution as our prior, to demonstrate an alternative way of obtaining a subset of stationary Laplace distributions.¹ Both steps are made possible by using Markov chain Monte Carlo (MCMC) methods.

¹After this more restrictive selection we discard 45% of our original data compared to Alfarano et al. (2012) who discard 79% of their data.
The Gibbs sampler, a widely-used MCMC method, is an extremely useful tool in this type of economic analysis. It has been introduced in this journal by Waggoner and Zha (2003) in the context of structural vector auto regressions, but to the best of our knowledge has never been applied to the problem of firm distributions. In general, MCMC methods are computational procedures for estimating complicated, multi-parameter densities in statistical problems making them particularly useful for Bayesian analysis. They were first used by Metropolis et al. (1953), but the increased computing capacity has led to a proliferation of these techniques since the 1990s. MCMC methods reproduce densities by sampling a large number of values from the full conditional marginal densities of a joint, high-dimensional, stationary distribution whose analytical form may be difficult to find. Since the parameter value sampled at \( t \) depends on the one at \( t - 1 \), the resulting sequence of samples is a Markov chain.

In section two, we will introduce our Bayesian approach to the profit rate filtering problem and the Gibbs sampler for a mixture distribution using latent variables. Section three will show the results that accrue from applying it to the COMPUSTAT database. Section four will use additional Gibbs sampling to analyze the existence of a Laplace distribution in a subset of firms and relate the result to statistical equilibrium arguments in the literature. Section five will summarize the results and point to further research opportunities.
2 A mixture model for profit rates

2.1 A Bayesian approach

Using prior knowledge and a model for the likelihood, we would like to infer each firm’s posterior probability of being in the signal distribution. From previous research, our prior for the shape of the likelihood of firm profit rates is a two component mixture distribution of a Laplace signal and an arbitrary noise distribution. We specify the noise as a Gaussian distribution with a large variance.\(^2\) The density function for the Laplace/Gaussian mixture is

\[
f(y|q, \mu_1, \sigma_1, \mu_2, \sigma_2) = q \cdot \mathcal{L}(y|\mu_1, \sigma_1) + (1 - q) \cdot \mathcal{N}(y|\mu_2, \sigma_2)
\]

\(^{2}\text{Sensitivity tests using uniform and Laplace noise distributions made no qualitative difference for the sorting process.}\)

\(\mathcal{L}\) denotes a Laplace distribution, \(\mathcal{N}\) denotes a Gaussian distribution with \(\mu_1\) and \(\mu_2\) being location parameters and \(\sigma_1\) and \(\sigma_2\) being scale parameters; the parameter \(q\) weighs the two distributions. The mixture distribution can be rewritten with latent variables. For any observation \(i\) of \(y\), and \(i = 1, \ldots, n\), write (1) as

\[
f(y_i|z, \mu_1, \sigma_1, \mu_2, \sigma_2) = [\mathcal{L}(y_i|\mu_1, \sigma_1)]^{z_i}[\mathcal{N}(y_i|\mu_2, \sigma_2)]^{(1-z_i)}
\]

where \(z\) is Bernoulli distributed, \(z_i \sim \text{Bern}(q)\).

We assume that the priors of the parameters are distributed independently of each other, and that all parameters have non-informative prior distributions with those of \(\sigma_1, \sigma_2\) restricted to the interval \((0, \infty)\) and that of \(q\) to \((0, 1)\). We use uniform priors.
for $\mu_1$ and $\sigma_1$ because the Laplace distribution does not belong to the exponential family of distribution and thus has no computationally convenient conjugate priors. Having little preconception about the prior distribution of the noise, we use Jeffrey’s prior for $\mu_2$ and $\sigma_2$. Finally, we use a uniform prior for $q$.

Since (2) has a parameter, $z_i$, for each firm observation, the $n$ latent variables, $z_1, ... z_n$, allow us to distinguish in which part of the distribution individual firms lie. This requires that we know the posterior distribution of each $z_i$. Since the joint density for (2) has $n + 5$ dimensions, it is difficult to derive analytically. We will use the Gibbs sampler to sample from the full conditional posteriors of each parameter density. If 95% of all sample values from the Gibbs sampler of some $z_i$ are equal to one, the ‘null hypothesis’ that a firm is noise is refuted and the observation is kept in the signal distribution. This way, we avoid arbitrary cut-offs but let the mixture distribution determine endogenously whether a firm belongs to the competitive signal distribution or not. It remains to derive the full conditional posterior distributions for the Gibbs sampler.

### 2.2 The Gibbs sampler

The Gibbs sampler is one of a set of Markov chain Monte Carlo (MCMC) methods that, for collections of random variables with an analytically intractable joint density, finds the marginal densities without using the joint density. Instead, only the full conditional posterior distributions of all parameters are required.\(^3\) Using (2) and our

\(^3\)See Gelfand and Smith (1990) and Casella and George (1992) for the concept. MCMC methods were originally use by Metropolis et al. (1953) and developed by Hastings (1970). The Gibbs sampler in particular was named and used by Geman and Geman (1984). Its first application to an econometric problem was to the best of our knowledge in Koop (1992). Bayesian econometrics goes back much further, at least to the now classical textbook of Zellner (1971). Bayesian Monte Carlo
priors we obtain all full conditional posterior densities. They are:

\[ q \sim \text{Beta}\left(\sum_i z_i + 1, \ n - \sum_i z_i + 1\right) \] (3)

\[ z_i \sim \text{Bern}\left(\frac{q \mathcal{L}(y_i|\mu_1, \sigma_1)}{q \mathcal{L}(y_i|\mu_1, \sigma_1) + (1 - q) \mathcal{N}(y_i|\mu_2, \sigma_2)}\right), \text{ for } i \text{ in } 1, \ldots, n \] (4)

\[ \mu_1 \sim c \prod_{i=1}^{n} \mathcal{L}(\mu_1|y_i, \sigma_1)^{z_i} \] (5)

\[ \sigma_1 \sim \mathcal{IG}(1 + \sum_i z_i, \sum_i z_i|y_i - \mu_1|) \] (6)

\[ \mu_2 \sim \mathcal{N}\left(\frac{\sum_i (1 - z_i)x_i}{n - \sum_i z_i}, \ \frac{\sigma_2}{n - \sum_i z_i}\right) \] (7)

\[ \sigma_2 \sim \mathcal{IG}\left(\frac{n - \sum_i z_i}{2}, \ \frac{1}{2} \sum_i (1 - z_i)(x_i - \mu_2)^2\right) \] (8)

The detailed derivations of these density and the Gibbs sampler pseudo code are explained in Appendix B.

### 3 Empirical Results

#### 3.1 The data

Our data is a set of firm profit rates from COMPUSTAT, Standard and Poor’s (2013), spanning the years 1950-2012. We calculate the profit rate by dividing the difference of net sales and operating costs, which equals operating income before depreciation,
by total assets.\textsuperscript{4}

\[
\frac{\text{Net Sales} - \text{Operating Cost}}{\text{Total Assets}} = \frac{\text{Operating Income before Depreciation}}{\text{Total Assets}} = r
\]

Our selection of variables is the same as that of Alfarano et al. (2012). Government and financial services, real estate and insurance have been excluded because the former does not partake in competition and the latter adheres to different accounting conventions for revenue calculation that makes this part of the industries incomparable. The summary statistics of the remaining 290,931 observations or on average 4618 observations per year are in Table 1.

Figure 2 reveals enormous outliers in the box plot of the distribution for each year. While obscured in the plot, the summary statistics relate that the interquartile range is “well-behaved” and that the median is at the historically credible level of 10%. A mixture distribution with latent variables removes these outliers by identifying them with appropriate indicators. As the profit rate has the unit 1/time, it is insensitive to changes in prices except if the price index in the numerator develops substantially differently from that in the denominator. Therefore we sometimes display data pooled across years, even while we do all of our analysis and sampling year by year to avoid muddling the results through cyclical influences such as business cycles.

\textsuperscript{4}Details for our data are in Appendix A.
Table 1: Summary statistics, raw profit rates

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>-9860</td>
<td>0.02</td>
<td>0.10</td>
<td>6.67</td>
<td>0.17</td>
<td>153,600.00</td>
</tr>
</tbody>
</table>

Figure 2: The uncensored year by year box plots of the raw profit rates.

3.2 Results from the Gibbs Sampler

We run the Gibbs sampler separately for each year to avoid artifacts from pooling heterogeneous years. For each year, the Gibbs sampler yields 500 samples of the latent variable Bernoulli posterior distribution of each firm, $z_{i,t}$, where $t$ is the time subscript. Discarding any observation with $z_{i,t} = 1$ less than 95% of the iterations, we use the remaining sample for our subsequent analysis.

Figure 3 shows box plots of the posterior distributions of $q$, $\mu_1$, and $\sigma_1$ for every year after discarding the first fifty “burn-in” samples to allow the Markov chain

\[\text{We compute all sampling algorithms with the software } R \ (R \text{ Core Team, 2013).}\]
to converge to the stationary distribution. All distributions are sharply peaked, confirming convergence of the distribution.6

The first thing that stands out in both Figure 4 and Figure 3 is a structural break in the profit rates and the Gibbs-sampled parameter around 1960. If the results are divided into sectors - an analysis we cannot pursue further here - the most obvious departure from later years' distributions is in the transport sector. This might be due to competition-stifling legislation. Moreover, the number of annual data points before 1960 is less than a thousand, but rises rapidly to more than 5000 in 1974, reaching peaks of over 10,000 in the 1990s. Therefore, section 4 of this paper will focus on the period after 1960.

Secondly, after 1960 the Gibbs parameters display trends. $q$, the parameter that indicates the share of firms in the signal distribution, hovers at almost one and gradually drops to around 95%. This implies that less than 5% of the data are noise according to our algorithm. $\mu$, the location parameter of the Laplace signal distribution, fluctuates around a falling trend, its median falling from 17% in 1965 to a low of 7% in 2001, preliminary evidence for a falling rate of profit. The $\sigma$ series, showing the scale parameter of the Laplace signal, follows a rising trend, that peaks

---

6See Appendix C for trace plots of the sampled signal and mixture parameters.
Table 2: Summary statistics, raw (r) and signal (r*) profit rates

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>-9860</td>
<td>0.02</td>
<td>0.10</td>
<td>6.67</td>
<td>0.17</td>
<td>153,600.00</td>
</tr>
<tr>
<td>r*</td>
<td>-1.65300</td>
<td>0.04265</td>
<td>0.11630</td>
<td>0.07506</td>
<td>0.17810</td>
<td>2.42200</td>
</tr>
</tbody>
</table>


3.3 Analyzing the Signal Distribution

As Table 2 shows, the very dispersed distribution has been reduced to a well-behaved one without having made any arbitrary cut-offs as regards firm size or length of survival. We believe this captures the entire gamut of competition. The comparative summary statistics, moreover, show that the maximum and minimum values have been substantially shrunk towards the median, while the interquartile range has remained virtually unaffected. This reflects the setup of the mixture distribution, that keeps the signal, and discards highly dispersed noise, which is illustrated in Figure 4.

The semi-log tent-plot of the pooled histogram of the signal distribution in Figure 5 shows an extremely peaked mode and the tent shape characteristic of the Laplace distribution.\(^7\) The strong clustering around the mean profit rate, which we refer to as general rate of profit, shows minimal fluctuation over time (see below Figure 7 for single year plots). The second striking feature in Figure 5 is a left skew

\(^7\)It is important to realize that this result is independent of the choice of the Laplace distribution as a signal. It is the data that has this tent shape, it is not the sampling distribution that “forms” this shape, since the error part of the mixture takes out none of the data around the mode. This shape confirms the results of the previous literature and our prior conjecture of Laplace distributed data was well-chosen.
in the distribution. This skew is likely due to entry and exit dynamics, where un-
competitive firms realize very low profit rates and either die or are absorbed by more
competitive firms, or new firms realize initially little income.

Thus, unlike previous authors, the latent variable Gibbs sampler procedure yields
a distribution without having to make additional assumptions about minimum size or age. In particular we keep more than 96% of our data\textsuperscript{8} despite our restrictive selection condition of more than 95% of $z$ samples being one. In the next section, we use Bayesian methods to explore removing the skew in the distribution in order to locate statistical regularities as in Alfarano et al. (2012).

4 Change Point Analysis

4.1 Filtering for Larger Firms

Figure 5 shows that a tent shape is clearly perceptible in a semi-log plot of the signal distribution; however, some asymmetry remains. Alfarano et al. (2012) conjecture that this skew is due to entry and exit dynamics of young firms, and that the competitive process of long lived, more established firms is characterized by a symmetric Laplace distribution. We extend this conjecture to firm size and continuing with our Bayesian approach we investigate this hypothesis for the present, noiseless data. Using a simple Bayesian change point analysis on the first moment of the distribution, conditional on capital stock instead of age, we determine which subset of our data is distributed symmetrically about the mode. The change point in the mean thus indicates the change from asymmetric to symmetric distribution.

The use of the first moment in our change point analysis is motivated by the idea that smaller firms are subject to entry and exit dynamics resulting in their average profit rates being significantly different from the general rate of profit, while larger

\textsuperscript{8}The exact percentage varies slightly depending on which pseudorandom number seed is set for the simulation.
firms’ profit rates will tend to be clustered around it. We partition firms by their size and use the conditional mean profit rate to determine from which size onwards firm profit rates converge to the general rate of profit.

This is achieved by computing the posterior probability of a change point between each two adjacent partition blocks. We partition the firm capital stock space into vigintiles (twenty evenly sized quantiles) and after estimating the posterior probability of a change point in the mean between each quantile for each year separately, we discard all firms in those vigintiles below high posterior probabilities of a change point. As an example, Figure 6 illustrates that for 1995 firms below the eleventh vigintile are discarded. By using vigintiles of the capital stock for each year instead of a dollar value, we avoid the problem of adjusting for inflation. We set up the change point analysis with MCMC methods as in Barry and Hartigan (1993). Using the Erdman and Emerson (2007) package that implement Barry and Hartigan’s change point algorithm in R we run a Gibbs sampler for 500 iterations excluding the first 50 burn-in samples.

This analysis provides a lower bound for firm size and selects a subset of large, well-established firms. The distribution of this part of the competitive process is strikingly symmetric and Laplace-shaped in every year, as Figure 7 shows. Moreover, the symmetry is present in every industry, and all distributions are peaked at almost the same value, as was already shown in Figure 1 in the introduction.

This further process of filtration is not very restrictive, as on average, the change point analysis only discards 45% of the initial, raw dataset. That is, the subset of the competitive process characterized by the Laplace distribution, which filters out the entry and exit dynamics of small firms, still comprises 55% of the original, entire
Figure 6: Example of the output of the change point analysis for the mean profit rate conditional on capital stock for 1995. The upper panel shows how the conditional mean profit rate converges to a steady value and the lower panel shows significant posterior probabilities of a change point until the 10th vigintile.

4.2 Statistical Equilibrium of Large Firms

While the change point analysis is only a rough one and could be refined to tease out more precisely which firms belong to the symmetric distribution, it highlights the striking symmetry and stationary of this subset’s distribution. Comparing its regular features to the strongly varying statistics of the entire competitive process in Figure 8, emphasizes this important regularity.

By way of comparison, where we lose only 45% of the data, Alfarano et al. (2012) discard 79% of their observations with their method of selecting firms that are alive the entire span of their dataset, 27 years, a number which Mishael Milaković kindly shared with us. Had we instead filtered out firms that are alive less than 27 years, we would have kept only 35.5% of the original sample. Moreover, had we focused on firms surviving the entire period, 63 years, we would have kept only 2.2% of the dataset.
The stationarity of the distribution can be interpreted as a *statistical equilibrium*, and its non-Gaussian shape entails that there are other constraints than the mean and variance acting on the shape of the distribution. Other authors have explained such phenomena in economic data based on maximum entropy reasoning, e.g. Farjoun and Machover (1983); Stutzer (1983); Foley (1994); Dragulescu and Yakovenko (2000); Milaković (2003); Alfarano and Milaković (2008); Alfarano et al. (2012); Chakrabarti et al. (2013). It is attractive to explain the empirical result by analogy to statistical physics: The distribution of firm profit rates achieves its maximum entropy distribution, that is we see the distribution of profit rates in or
Figure 8: The year by year box plots of the profit rates of large firms remaining after the change point analysis (left panel) and for comparison the box plots of the entire competitive process, which is reproduced from Figure 4 (right panel). The regularity of the distribution of large firms emerges after 1960.

very near to the macro state (that determines the shape of the distribution) that can be achieved by the largest number of combinations of micro states subject to constraints.\textsuperscript{10} The shape of the macro state is determined by constraints acting on the ensemble of firms. The Laplace distribution is the maximum entropy distribution given a constraint of a constant mean of the absolute values of profit rates.\textsuperscript{11}

\textsuperscript{10}An alternative, information-theoretic foundation for this reasoning on which some of the cited literature in fact draws is in Jaynes (2003).

\textsuperscript{11}Mathematically the Laplace distribution is the density that maximizes the entropy $H$ of the probability distribution $p(y)$ subject to the normalizing and first moment of absolute values constraints, that is

$$
\max_p H[p(y)]
$$

subject to

$$
\begin{align*}
\int_{-\infty}^{\infty} p(y) \, dy &= 1 \\
\int_{-\infty}^{\infty} |y| p(y) \, dy &= c
\end{align*}
$$

where the Lagrange multiplier of the moment of absolute values is equal to the entropy maximizing Laplace scale parameter.
A stationary Laplace distribution over more than five decades invites an entropic viewpoint and conveys important information about the strong competitive forces active in the economy on both an intra- and inter-industry level. The challenge is to motivate the constraint more thoroughly with economic theory. Numerically, the Laplace constraint implies that for any profit rate that moves away from the mean, another has to move towards it. Classical competition focusing on how opposing forces of competition and individual firms’ innovative efforts result in a distribution, appears a good starting point for more theoretical work.

5 Conclusions

This paper has contributed striking evidence for Laplace distributed profit rates. We have shown that by using a Bayesian mixture model with latent variables and MCMC methods it is possible to filter out confounding noise and extract a signal distribution of firm profit rates. This endogenous selection process results in an effective filter with little information lost, allowing us to investigate the distributional form of profit rates in the U.S. economy that captures the competitive process. We find that the Laplace distribution with a negative skew characterizes the general economy-wide rate of profit. Using a Bayesian change point analysis we find that large firms are characterized by a stationary symmetric Laplace distribution. We also find this to be the case at the sectoral level, suggesting similar inter- and intra- industry dynamics. Our analysis both confirms Alfarano et al. (2012), who similarly find that for a sample of long-lived firms the Laplace distribution is consistent with their data, and extends the finding of sectoral Laplace distributed growth rates in Bottazzi and
Secchi (2003a) to profit rates.

The Bayesian mixture and change point models have the advantage of being very versatile and are by no means restricted to firm profit rate analysis. In particular, mixtures can be applied to any distribution of objects to be partitioned, where it is necessary to know for each observation into which subset of the partition it belongs, while Bayesian change point analysis can be applied to any regime change problem where the threshold value is not immediately obvious. Further methodological research can be directed towards a mixture distribution with more than two components, where a third one might capture the small firms affected by entry and exit dynamics.

On the theoretical level, interpreting the large firm distribution as a statistical equilibrium in terms of maximum entropy appears attractive, but needs further theoretical investigation. In addition, the empirical transition dynamics that give rise to the distribution need consideration, which would also offer insight into the inter-industry dynamics, and whether the profit rate gravitation around a general rate of profit is observed.

References


Standard, Poor’s, 2013. COMPUSTAT.


6 Appendix A: Data Sources

Data is gathered from the COMPUSTAT Annual Northern American Fundamentals database. We extract yearly observations of the variables \( AT = \) Total Assets, \( REVT = \) total revenue, \( XOPR = \) operating cost, \( SIC = \) Standard Industry Code, \( FYR = \) year, \( CONM = \) company name, from 1950 through 2012. At the time of download, data for 2012 was incomplete. Subtracting \( XOPR \) from \( REVT \) and dividing by \( AT \) gives a rate of income per capital stock, and thus a profit rate.

\[
\frac{REVT - XOPR}{AT} = \frac{OIBDP}{AT} = r
\]

Our raw data set consists of 430,209 observations. Subtracting completely missing values, government and finance, insurance and real estate - government because it is not engaged in a search for profit maximization but pursues other objectives and finance et al. because the accounting methods are different for them and part of income is not recorded in \( REVT \), leading to profit rates almost zero - we impute at random the remaining missing values. Our completed case dataset contains 290,931 observations or roughly 4,600 observations per year. The actual yearly number is much smaller in earlier years and above 10,000 in the 1990s.

7 Appendix B: Mixture Gibbs Sampler Derivation

This appendix derives the Gibbs sampler for a latent variable augmented Laplace-Gaussian distribution. Consider
\[ f(y|q, \mu_1, \sigma_1, \mu_2, \sigma_2) = q \cdot \mathcal{L}(y|\mu_1, \sigma_1) + (1-q) \cdot \mathcal{N}(y|\mu_2, \sigma_2) \quad \text{(B.1)} \]

\( \mathcal{L} \) denotes a Laplace distribution and \( \mathcal{N} \) is a Normal or Gaussian distribution and assume that the parameters are all independent of each other a priori, and we also assume that all parameters have uniform prior distributions. Introducing latent variables, for every observation \( i \) of \( y \), and \( i = 1, ..., n \), this can be written as

\[ f(y_i|z, \mu_1, \sigma_1, \mu_2, \sigma_2) = [\mathcal{L}(y|\mu_1, \sigma_1)]^{z_i} [\mathcal{N}(y|\mu_2, \sigma_2)]^{(1-z_i)} \quad \text{(B.2)} \]

where \( z_i \sim \text{Bern}(q) \).

We derive full conditional posteriors for each variable.

### 7.1 Full conditional posterior densities

We will derive the conditional posterior densities using a few rules of conditional and joint probabilities. They are

\[ p(A|B)p(B) = p(A, B) \Leftrightarrow p(A|B) = \frac{p(A, B)}{p(B)} \quad \text{(B.3)} \]

\[ \int p(A, B) dA = p(B) \Rightarrow p(A|B) = \frac{p(A, B)}{\int p(A, B) dA} \quad \text{(B.4)} \]
We want to find the density of a parameter conditional on priors and likelihood. Hence, our generic \( p(A, B) \) is the joint hyperprior-prior-likelihood density. The integral integrates out the parameter whose conditional density we are interested in. Everywhere except for the Laplace, nice properties of the distributions will enable us to avoid the brute force integral. However, the Laplace distribution is not of the exponential family (Kotz et al., 2001) and will require use of explicit integration.

7.2 The joint density, Laplace/Normal

For uniform priors for Laplace distribution and the mixture parameter \( q \) and Jeffrey’s prior for the Gaussian distribution, our joint density can be written as follows. Let \( Z = \{z_1, \ldots, z_n\} \) and \( Y = \{y_1, \ldots, y_n\} \).

\[
p(q, \mu_1, \sigma_1, \mu_2, \sigma_2, Z, Y) = p(q)p(\mu_1)p(\sigma_1)p(\mu_2)p(\sigma_2) \prod_{i=1}^{n} p(z_i|q) \prod_{i=1}^{n} p(y_i|z_i) \tag{B.5}
\]

\[
\propto \frac{1}{\sigma} \prod_{i=1}^{n} q^{z_i}(1-q)^{1-z_i} \prod_{i=1}^{n} [\mathcal{L}(y|\mu_1, \sigma_1)]^{z_i} [\mathcal{N}(y|\mu_2, \sigma_2)]^{(1-z_i)} \tag{B.6}
\]

In what follows we will omit priors, if they are uniform.
7.3 The full conditional posterior densities, Laplace/Normal

7.3.1 z posterior

The posterior for each $z_i$ is Bernoulli because it can only take on two values. Let $Z_{-i} = \{z_1, ..., z_{i-1}, z_{i+1}, ..., z_n\}$.

\[
p(z_i = 1 | q, \mu_1, \sigma_1, \mu_2, \sigma_2, Z_{-i}, Y) \sim \text{Bern}\left(\frac{q \mathcal{L}(y_i | \mu_1, \sigma_1)}{q \mathcal{L}(y_i | \mu_1, \sigma_1) + (1-q)\mathcal{N}(y_i | \mu_2, \sigma_2)}\right)
\]

Thus $z_i \sim \text{Bern}\left(\frac{q \mathcal{L}(y_i | \mu_1, \sigma_1)}{q \mathcal{L}(y_i | \mu_1, \sigma_1) + (1-q)\mathcal{N}(y_i | \mu_2, \sigma_2)}\right)$.

7.3.2 q posterior

\[
p(q | \mu_1, \sigma_1, \mu_2, \sigma_2, Z, Y) \propto \prod_{i=1}^{n} q^{z_i}(1-q)^{1-z_i}
\]

This is the kernel of a beta distribution, hence

$q \sim \text{Beta}(\sum_i z_i + 1, n - \sum_i z_i + 1)$.
7.3.3 $\mu_1$ posterior

Since the Laplace is a non-exponential family distribution, one cannot directly derive a closed-form conditional posterior distribution.$^{12}$

The conditional posterior density for $\mu_1$ is

\[
p(\mu_1|q, \sigma_1, \mu_2, \sigma_2, Z, Y)
= p(\mu_1|\sigma_1, Z, Y) \quad (B.12)
= \frac{\prod_{i=1}^{n} \mathcal{L}(y|\mu_1, \sigma_1)^{z_i}}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} \mathcal{L}(y|\mu_1, \sigma_1)^{z_i} d\mu_1} \quad (B.13)
\]

\[
= \frac{exp \left[ \sum_i z_i \left( \frac{-|y_i - \mu_1|}{\sigma_1} \right) \right]}{\int_{-\infty}^{\infty} exp \left[ \sum_i z_i \left( \frac{-|y_i - \mu_1|}{\sigma_1} \right) \right] d\mu_1} \quad (B.14)
\]

\[
= c exp \left[ \sum_i z_i \left( \frac{-|y_i - \mu_1|}{\sigma_1} \right) \right] \quad (B.15)
\]

Since $|y_i - \mu_1| = |\mu_1 - y_i|$, we have a product of Laplace distributions and the posterior density is:

\[
p(\mu_1|\sigma_1, Z, Y) \sim c \prod_{i=1}^{n} \mathcal{L}(\mu_1|y_i, \sigma_1)^{z_i} \quad (B.17)
\]

$^{12}$Kozumi and Kobayashi (2011) use a mixture representation of the Laplace distribution from Kotz et al. (2001) to find a closed form solution for the posterior $\mathcal{L}(0, \sigma)$. However, our location parameter is unequal zero and therefore we have to resort to numerical integration at every iteration of the Gibbs sampler.
7.3.4 $\sigma_1$ posterior

\[
p(\sigma_1 | \mu_1, Z, Y) \propto \prod_{i=1}^{n} \mathcal{L}(y_i | \mu_1, \sigma_1)^{z_i}
\]

\[
\propto \prod_{i=1}^{n} \sigma^{-z_i} \exp \left[ z_i \left( -\frac{\sqrt{2} | y_i - \mu_1 |}{\sigma_1} \right) \right]
\]

\[
= \sigma^{-\sum z_i} \exp \left[ \sum_{i} z_i \left( -\frac{\sqrt{2} | y_i - \mu_1 |}{\sigma_1} \right) \right]
\]

This is an inverse gamma kernel. Therefore

\[
p(\sigma_1 | \mu_1, Z, Y) \sim IG(1 + \sum_{i} z_i, \sum_{i} z_i | y_i - \mu_1 |)
\]  

(B.21)

7.3.5 $\mu_2$ posterior

The derivation of Jefrey’s prior is given implicitly in many textbooks e.g. (Gelman et al., 2003, p. 46), when the Gaussian conjugate prior is derived. The result is a normal distribution for $\mu_2$ and an inverse gamma distribution for $\sigma_2$.

\[
p(\mu_2 | q, \mu_1, \sigma_1, \sigma_2, Z, Y) = \mathcal{N}(y_i | \mu_2, \sigma_2)^{(1-z_i)}
\]

\[
\propto \exp \left[ -\frac{1}{2} \left( \frac{1}{\frac{1}{\sigma_2^2} \left( n - \sum_{i} z_i \right)^2} \left( \frac{\sum_{i} (1 - z_i) x_i}{n - \sum_{i} z_i} - \mu_2 \right)^2 \right) \right]
\]

Hence \( p(\mu_2 | q, \mu_1, \sigma_1, \sigma_2, Z, Y) \sim \mathcal{N} \left( \frac{\sum_{i} (1 - z_i) x_i}{n - \sum_{i} z_i}, \frac{\sigma_2^2}{n - \sum_{i} z_i} \right) \)
7.3.6 \( \sigma_2 \) posterior

\[
p(\sigma_2 | q, \mu_1, \sigma_1, \mu_2, Z, Y) = \frac{1}{\sigma_2^n} \prod_{i=1}^{n} \left[ \mathcal{N}(y_i | \mu_2, \sigma_2) \right]^{(1-z_i)} \tag{B.24}
\]

\[
\propto \sigma_2^{-2\left(\frac{n-\sum_i z_i}{2} + 1\right)} \exp \left[ \frac{1}{2} \sum_i (1-z_i)(x_i - \mu_2)^2 \right] \tag{B.25}
\]

Hence \( p(\sigma_2 | q, \mu_1, \sigma_1, \mu_2, Z, Y) \sim IG \left( \frac{n-\sum_i z_i}{2}, \frac{1}{2} \sum_i (1-z_i)(x_i - \mu_2)^2 \right) \).

This completes the derivation of the Gibbs sampler. It generates a value for each \( z_i \) from its full conditional posterior. After a Monte Carlo number of runs, the ratio of \( z_i \) can be taken as the probability that the underlying firm observation belongs to the Laplace (the signal) distribution. The pseudocode is presented below.\(^{13}\)

7.4 The Gibbs sampler pseudocode, Laplace/Normal

Select initial values for \( q_j, \mu_{1,j}, \sigma_{1,j}, \mu_{2,j}, \sigma_{2,j}, Z_j \), \( j = 0 \)

for (j in 0 through J) :

\(^{13}\)This can be easily transformed into e.g. a Laplace/Laplace Gibbs sampler, by repeating steps 3 and 4 for the noise distribution instead of carrying out 5 and 6, and replacing \( \mathcal{N} \) in step 2 by \( \mathcal{L} \).
1. \( q_{j+1} \sim Beta(\sum_i z_{i,j} + 1, n - \sum_i z_{i,j} + 1) \)

2. \( z_{1,j+1} \sim Bern(\frac{qL(y_i|\mu_1, \sigma_1)}{qL(y_i|\mu_1, \sigma_1) + (1-q)N(y_i|\mu_2, \sigma_2)}) \)

3. \( \mu_{1,j+1} \sim c_1 \prod_{i=1}^n L(\mu_{1,j+1}|y_i, \sigma_{1,j})^{z_{i,j+1}} \)

4. \( \sigma_{1,j+1} \sim IG(1 + \sum_i z_{i,j+1}, \sum_i z_{i,j+1}|y_i - \mu_{1,j+1}|) \)

5. \( \mu_{2,j+1} \sim N\left(\frac{\sum_i (1 - z_{i,j+1})x_i}{n - \sum_i z_{i,j+1}}, \frac{\sigma_{2,j}}{n - \sum_i z_{i,j+1}}\right) \)

6. \( \sigma_{2,j+1} \sim IG\left(\frac{n - \sum_i z_{i,j+1}}{2}, \frac{1}{2} \sum_i (1 - z_{i,j+1})(x_i - \mu_{2,j+1})^2\right) \)

### 8 Appendix C: Convergence of Gibbs Sampler

The convergence of the marginal distributions for each parameter in the signal distribution from the mixture model is shown in the trace plot in Figure C.1 for selected years. Trace plots show sampled values against the simulation index and give a good visual diagnostic for the convergence of the Markov chain to its stationary distribution. It is clear that after 10 iterations the chain converges to a stationary distribution, fluctuating narrowly around the mode of the posterior distribution. While the mode for each parameter is non-stationary over time, the distribution clearly is once the initial burn-in of 50 iterations is discarded.
Figure C.1: Trace plots for the Gibbs sampled $q$, $\mu_1$, $\sigma_1$ for years 1950, 1960,... 2010

- Graph showing trace plots for $q$ with iterations and years.
- Graph showing trace plots for $\mu_1$ with iterations and years.
- Graph showing trace plots for $\sigma_1$ with iterations and years.