# A Generalized Gibrat's Law*

By Juan Carlos Córdoba†

*Department of Economics, Rice University, U.S.*

February 2006

Many economic and non-economic variables such as income, wealth, firm size, or city size often distribute Pareto in the upper tail. It is well established that Gibrat's law can explain this phenomenon, but Gibrat's law often does not hold. This note characterizes a wider class of processes, one that includes Gibrat's law as a special case, that can explain Pareto distributions. Of particular importance is a parsimonious generalization of Gibrat's law that allows size to affect the variance of the growth process but not its mean. This note also shows that under plausible conditions Zipf's law is equivalent to Gibrat's law.

*Keywords:* Pareto Distribution, Gibrat's Law, Zipf's Law, City Size Distribution

*JEL Classification:* R11, R12, O41, J10

*This paper is based on the first chapter of my dissertation at the University of Rochester. I thank Per Krusell for his insightful advice, Peter Hartley for the detailed comments, Alan Stockman, Mark Bils, Michael Cranston, Tony Smith, Esteban Rossi-Hansberg, two anonymous referees, and participants in seminars at the universities of Rochester, Pennsylvania, Carnegie Mellon, Pittsburgh, Western Ontario, Cornell, Northwestern, Tulane, and the London School of Economics. I acknowledge financial support provided by the Central Bank of Colombia. All remaining errors are my own.

†Email: jcordoba@rice.edu.

# 1. Motivation

The cross-sectional distribution of various economic and non-economic variables such as wealth, income, firm size, city size, or publications per author in Econometrica often conforms well to a Pareto distribution, or power law, particularly in the upper tail. As an example, Figure 1 illustrates the well-known case of the city size distribution (see for example Gabaix 1999 for details). The upper tail of the distribution, here illustrated by the 100 largest US cities for different periods, conforms well to a linear shape in logs, or to a power shape in levels. Champernowne (1953) and Simon (1955) showed that Pareto distributions arise naturally when the time series behavior of the variable in question satisfies what is known as Gibrat's law: that the current position of the variable does not influence its expected rate of growth nor the variance of its growth rate.

For concreteness, consider a discrete Markov process $\{X_t\}$ taking values in the set $X = \left\{x_i \equiv (1+g)^i\right\}_{i=1}^{N}$, where $g > 0$ and $N \leq \infty$ is a positive integer. Suppose that the conditional distribution of $X_{t+1}$ given $X_t$ is described by the following Markov chain:

$$(1) \qquad \pi = \begin{bmatrix} a_1+b_1 & c_1 & 0 & 0 & .. & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & .. & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & .. & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & 0 & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & 0 & 0 & 0 & 0 & a_N & b_N+c_N \end{bmatrix},$$

where the $\pi_{ij} = \Pr\left\{X_{t+1} = x_j | X_t = x_i\right\}$, $a_i + b_i + c_i = 1$ and $[a_i, b_i, c_i] \gg 0$. Notice that $\{X_t\}$ has a lower barrier, and an upper barrier if $N < \infty$. Denoting $\mu_i$ the conditional expected growth rate of $X_t$ and $\sigma_i^2$ its conditional variance[3], $\mu_i$ and $\sigma_i^2$ satisfy:

$$\mu_i = gc_i - \frac{g}{1+g} a_i \text{ for } i = 2, ..., N-1. \tag{2}$$

$$\sigma_i^2 = a_i \left( \frac{g}{1+g} - \mu_i \right)^2 + b_i \mu_i^2 + c_i (g - \mu_i)^2 \text{ for } i = 2, ..., N-1. \tag{3}$$

Champernowne (1953) showed that if $a_i = a$, $b_i = b$, $c_i = c$, $a > c$, and $N = \infty$, then the unconditional-stationary distribution of $\pi$ has the Pareto form $\Pr\{X_t \geq x\} = Mx^{-\delta}$. This characterization implies that, off boundaries, $\{X_t\}$ satisfies Gibrat's law as $\mu_i$ and $\sigma_i^2$ are independent of $i$.

For diverse economic and non-economic problems Gibrat's law provides an unsatisfactory characterization of the underlying dynamics generating power laws. For example, evidence suggests that small firms face higher proportional risk that large firms (Evans 1987), that richer individuals take more proportional risk that poorer individuals (Carroll 2000), and that large diversified cities experience lower growth volatility than smaller cities (Fujita et. al. 1999). Yet the distributions of firm sizes, incomes, wealth, and city sizes are often well described by Pareto distributions specially in the upper tail (see for example Axtell 2001, Dragulesco & Yakovenko 2001, Champernowne 1953, or Zipf 1949).

This note makes three contributions. First, it characterizes Markov processes of the class illustrated by $\pi$ in equation (1), that is with a quasi-diagonal structure, supporting Pareto distributions. Importantly, this class includes diffusions which are Markov processess with continous state-space and continous time. We show that if $a_i = a\mu^i$ and $c_i = c\mu^i$, where $0 < \mu \leq 1$, then the unconditional distribution associated to $\pi$ is Pareto. This finding generalizes Sims' and Chapernowne's results in that Gibrat's law is included as the special case in which $\mu = 1$. Second, we show

[3] Thus, $\mu_i \equiv E\left[\frac{X_{t+1}-X_t}{X_t} | X_t = x_i\right]$ and $\sigma_i^2 \equiv E\left[\left(\frac{X_{t+1}-X_t}{X_t} - \mu_i\right)^2 | X_t = x_i\right]$.

that if a weak version of Gibrat's law holds, one that requires only mean growth to be independent of size, then the Markov process just described is the only process supporting Pareto distributions. We motivate the importance of this weak restriction on the fact that many variables in economics grow over time and that balanced growth conditions are often required on theoretical or empirical grounds. Our generalized Gibrat's law provides a general process that is scale invariant. Finally, we establish that along balanced growth paths Zipf law can only result from Gibrat's law.

## 2. Main Results

Let $p_t = [p_{1t}, p_{2t}, ..., p_{Nt}]$ be the unconditional distribution of $X_t$ at time $t$ and let $\pi$, defined by (1), be its conditional distribution. Thus, $p_{t+1} = p_t \cdot \pi$. A stationary distribution $p$ satisfies $p = p \cdot \pi$. Given that $[a_i, b_i, c_i] \gg 0$, $p$ exists, it is unique, and $p = \lim_{t\to\infty} p_t$ if $N$ is finite. For $N = \infty$, the additional restriction $a_i > c_i$ guarantees existence, uniqueness, and convergence. By the ergodic theorem, $p$ can also be interpreted as the cross-sectional distribution of population across states of a close system without entry or exit. Notice that $p$ satisfies

$$p_i = a_{i+1}p_{i+1} + b_i p_i + c_{i-1}p_{i-1} \text{ for } i = 2, ..., N-1, \text{ and} \tag{4}$$

$$p_1 = a_2 p_2 + (a_1 + b_1)p_1 \text{ and } p_N = (b_N + c_N)\, p_N + c_{N-1}p_{N-1}. \tag{5}$$

The following result was shown by Champernowne (1953).

**Proposition 1.** Suppose $a_i = a$, $b_i = b$, and $c_i = c$. Then the only solution to the system (4)-(5) is $p_i = M\rho^i = Mx_i^{-\delta}$, where $M$ is a constant, $\rho = \frac{c}{a}$, and $\delta = \frac{\ln(a/c)}{\ln(1+g)}$. Moreover, $\Pr\{X_t \geq x_i\} = \frac{1}{1-(c/a)} Mx_i^{-\delta}$ if $N = \infty$.

A proof of Proposition 1 is given below as part of Proposition 2. Proposition 1 formalizes the idea that Gibrat's law induces Pareto distributions. Notice, however, than Gibrat's law does not hold at the boundaries. Simon (1955) showed that Gibrat's law can induce other stationary distributions depending on boundary conditions.

To uncover additional Markov chains supporting Pareto distributions consider the system of equations (4)-(5) restricted so that $p_i$ adopts the Pareto form $p_i = M\rho^i$, $0 < \rho < 1$. In that case, equations (4) and (5) can be written as:

$$a_{i+1}\rho^{i+1} - (a_i + c_i)\,\rho^i + c_{i-1}\rho^{i-1} = 0 \text{ for } i = 2, ..., N-1, \tag{6}$$

$$c_1 = \rho a_2 \text{ and } a_N = c_{N-1}/\rho. \tag{7}$$

For given $\rho$, (6) and (7) is a system of $N$ equations in $2N$ unknowns and therefore it has potentially multiple solutions. The following is one of them.

**Lemma 1.** For given $\rho$, a solution to the system (6)-(7) is:

$$a_i = a\theta^i,\ c_i = c\theta^i,\ \text{for } i = 1, ..., N, \tag{8}$$

where $\theta \equiv \frac{1}{\rho}\frac{c}{a}$, $a$ and $c$ are strictly positive parameters satisfying $\theta \leq 1$ and $a + c < 1$.

**Proof** To see that equation (6) is satisfied, substitute (8) into the left hand side of (6) to obtain $a\,(\theta\rho)^{i+1} - (a+c)\,(\theta\rho)^i + c\,(\theta\rho)^{i-1} = (\theta\rho)^{i-1}\left[a\,(\theta\rho)^2 - (a+c)\,\theta\rho + c\right]$ $= (\theta\rho)^{i-1}\left[a\,(c/a)^2 - (a+c)\,c/a + c\right] = 0$ for $i = 2, ..., N-1$. Moveover, equations (8) produces $c_1 = c\theta = \rho a\theta\theta = \rho a_2$ and $a_N = a\theta^N = \frac{a\theta}{c}c\theta^{N-1} = \frac{1}{\rho}c_{N-1}$ so that (7) is also satisfied.

The restrictions on $a$ and $c$ in Lemma 1 guarantee that $0 < a_i + c_i < 1$ for $i = 1, ..., N$. Lemma 1 takes $p$ as given to derive a particular Markov chain $\pi$. The following Proposition takes $\pi$ to derive $p$.

**Proposition 2.** Let the Markov chain $\pi$ defined by (1) satisfy $a_i = a\theta^i$, $c_i = c\theta^i$ for $i = 1, .., N$, where $0 < c + a < 1$, $0 < \theta \leq 1$, and $\frac{c}{a\theta} < 1$. Then $p_i = M\rho^i = Mx_i^{-\delta}$

where $M$ is a constant, $\rho \equiv \frac{c}{a\theta}$ and $\delta \equiv \frac{\ln(1/\rho)}{\ln(1+g)} > 0$ is the unique absolute convergent solution to (4) and (5). Moreover, $\Pr\{X_t \geq x_i\} = \frac{M}{1-\rho} x_i^{-\delta}$ if $N = \infty$.

**Proof** By Lemma 1, $p_i = M\rho^i$ satisfy equations (4)-(7) for $\rho \equiv \frac{1}{\theta}\frac{c}{a}$. By the definition of $x_i$, $i = \ln\, x_i / \ln(1+g)$ so that $p_i = M x_i^{\ln\rho/\ln(1+g)} = M x_i^{-\delta}$. Uniqueness and absolute convergence follow from a standard result in stochastic processes (See Cox and Miller 1965, page 108-110). Finally, $\Pr\{X_t \geq x_i\} = \sum_{j=i}^{\infty} p_j = p_i \sum_{j=0}^{\infty} (\rho)^j = \frac{1}{1-\rho} M x_i^{-\delta}$.

Notice that Proposition 1 is obtained as the special case of Proposition 2 in which $\theta = 1$. The Markov chain described by Proposition 2 is just one of many chains that can explain Pareto distributions. However, suppose $\pi$ is further restricted to satisfy a weak version of Gibrat's law: that the expected growth rate of $X_t$ be independent of the position of $X_t$. Some examples in which this is a natural restriction are provided below. In that case, one can show that the only Markov chain $\pi$ supporting Pareto distributions is of the form described by Proposition 2. The following Theorem normalizes the common expected growth rate to zero, but this requirement is relaxed in the Appendix[4].

**Theorem 3.** Suppose the stationary unconditional distribution of $\pi$ has the Pareto form $p_i = M\rho^i = M x_i^{-\delta}$ where $\rho < 1$ and $\delta \equiv \frac{\ln(1/\rho)}{\ln(1+g)} \leq 1$. Suppose, moreover, that $\mu_i = 0$ for $i = 2, 3, ...$ and that $N = \infty$. Then $\pi$ must be of the form

$$a_i = a\theta^i, \; c_i = c\theta^i, \; \text{for } i = 2, 3, .. \tag{9}$$

[4] One can interpret $X_t$ as a detrended variable. The appendix shows that such normalization is withouth loss of generality if $\{X_t\}$ is a diffusion. Diffusions have the advantage of being defined on a continous support and in continous time in constrast to the discrete time discrete support assumed in this Section. The Appendix derives results analogous to Proposition 2 and Theorem 3 for the case of diffusions.

where $\theta \equiv \frac{1}{\rho}\frac{1}{1+g} \leq 1$, $0 < a+c < 1$ and $a = (1+g)c$.. Moreover, $\sigma_i^2 = A\theta^i$ where $A \equiv a\left(\frac{g}{1+g}\right)^2 + cg^2$.

**Proof** From equation (2), $\mu_i = 0$ if and only if $a_i = (1+g)\,c_i$. Substituting this result into equation (6) produces

$$(1+g)\,z_{i+1} - (2+g)\,z_i + z_{i-1} = 0 \text{ for } i = 2, 3, ..$$

where $z_i \equiv c_i \rho^i$. The associated characteristic equation has two roots: 1 and $\frac{1}{1+g}$. Therefore, the solution for $z_i$ and $c_i$ are $z_i = k_1\left(\frac{1}{1+g}\right)^i + k_2$, and $c_i = z_i \rho^{-i} = k_1\left(\frac{1}{\rho}\frac{1}{1+g}\right)^i + k_2\rho^{-i}$. Unless $k_2 = 0$, $c_i > 1$ for large $i$. Thus, $c_i = k_1\left(\frac{1}{\rho}\frac{1}{1+g}\right)^i$. Similarly for $a_i$. Finally, $\sigma_i^2 = A\theta^i$ is obtained by substituing (9) into (3) and using $\mu_i = 0$.

The restriction $\delta \leq 1$ in the theorem is required to guarantee that $c_i + a_i \leq 1$ for all $i$. Theorem 3 provides an even more parsimonous process than Proposition 2 that explains Pareto distributions. This process is relevant in cases for which it is reasonable to expect that a variable's position does not influence its expected rate of growth. For example, models of economic growth typically impose such restriction in the form of "balanced growth" conditions. A stochastic definition of a balanced growth is precisely that the expected growth rate of a variable is constant through time and therefore independent of the size of the variable.

## 3. An Application: The City Size Distribution

A case in which a balanced growth condition is expected to be satisfied is the case of city size illustrated in Figure 1. The common observation is that the tail of the city size distribution has, at least as a first approximation, a linear shape in logs or a power shape in levels. A second observation is that cities in the upper tail of the distribution are growing over time. For example, the minimum size required to enter the set of largest cities was $m_{1840}$ in 1840 but $m_{1990} > m_{1840}$ in 1990. This is mostly due to the fact that urban population grows over time. A third observation is that the fraction of total urban population living in the 100 largest cities had neither increased nor decreased systematically during the last 100 years[5]. We now show that these observations imply $\mu_i = \mu$.

In order to use simple calculus, it is convenient to assume that time and space are continuous rather than discrete. This requires redefining $X_t$ as a diffusion, and $\pi$ is now completely characterized by its drift, $\mu(x)$, and its diffusion, $\sigma^2(x)$, which are the instantaneous versions of $\mu_i$ and $\sigma_i^2$. The observations mentioned above are then summarized by the statement that $\Pr(X(t) \geq x) \equiv a\left(\frac{m(t)}{x}\right)^{\delta}$ where $m(t) = e^{\gamma t}$ is the minimum city size in the upper tail of the distribution, $\gamma > 0$ is the growth rate of the urban population, and $a$ is the fraction of urban population in the upper tail. The probability density in the tail of the distribution is thus given by $p(x,t) = \delta a m(t)^{\delta} x^{-\delta-1}$. Note that $\frac{\partial p(x,t)}{\partial t} = \delta p(x,t)\frac{m'(t)}{m(t)}$.

The average growth rate of cities in the upper tail of the distribution, $g(t)$, is given by (using a law of large numbers):

$$g(t) = \frac{1}{a}\int_{m(t)}^{\infty} \mu(x)p(x,t)dx.$$

Moreover, if the fraction of urban population living in the largest cities is to remain constant over time then $g(t)$ must equal $\gamma$ for all $t$. Thus, $g'(t) = 0$ for all $t$, or by

[5] According to Gibson (1998), the fraction of total population living in the 100 largest cities was 21% both in 1890 and in 1990.

Leibniz's rule,

$$\begin{aligned}
0 &= -\frac{1}{a}\mu(m(t))p(m(t),t)m'(t) + \frac{1}{a}\int_{m(t)}^{\infty}\mu(x)\frac{\partial p(x,t)}{\partial t}dx \\
&= -\frac{1}{a}\mu(m(t))p(m(t),t)m'(t) + \delta\frac{m'(t)}{m(t)}\frac{1}{a}\int_{m(t)}^{\infty}\mu(x)p(x,t)dx \\
&= \left(-\frac{1}{a}\mu(m(t))p(m(t),t) + \frac{\delta}{m(t)}\gamma\right)m'(t) \\
&= \left(-\mu(m(t)) + \gamma\right)\delta\frac{m'(t)}{m(t)} \text{ for all } t.
\end{aligned}$$

Therefore $\mu(m(t)) = \gamma$ for all $t$. Moreover, since for all $x$ in the upper tail there exits $s$ such that $m(s) = x$ and vice versa, then this last result is equivalent to $\mu(x) = \gamma$ for all $x$ in the upper tail. Thus, cities in the upper tail must share the same expected growth rate regardless of their size under the conditions described above.

Theorem 3 applied to the case of the city size distribution has an additional powerful implication. In several cases, in particular for the US, the observed city size distribution in the upper tail conforms well to a Pareto distribution with exponent $\delta$ close to 1, a case known as Zipf's law. According to Theorem 3 such exponent can arise only if the underlying Markov process satisfies $\theta = 1$. Thus, Zipf's law requires that Gibrat's law be satisfied. This result is different Gabaix's (1999) result which states the opposite: that Gibrat's law can produce Zipf's law (Gabaix 1999). We provide a necessary condition while Gabaix (1999) provides a sufficient condition. Córdoba (2004) used these results to evaluate urban models.

## References


[1] Axtell, R. L. (2001), “Zipf Distribution of U.S. Firm Sizes,” Science, Vol. 293. no. 5536, pp. 1818 - 1820.

[2] Carroll, C. (2000), “Portfolios of the Rich,” NBER Working Paper 7826.

[3] Champernowne, D. (1953), “A Model of Income Distribution,” *Economic Journal*, LXIII: 318-351.

[4] Córdoba, Juan-Carlos (2001), "Cities and Growth." Ph.D Dissertation, University of Rochester, July.

[5] Córdoba, Juan-Carlos (2004), "On the Distribution of City Sizes." Mimeo, Rice University.

[6] Cox, D. R. and H. D. Miller (1965), *The Theory of Stochastic Processes,* Methuen & Co. LTD.

[7] Dragulescu, A. and Victor M. Yakovenko (2001), "Exponetial and Power-law probability distributions of wealth and income in the United Kingdom and the United State,." *Physica A* 299: 213-221.

[8] Evans, D. (1987), “The relationship between firm growth, size, and age: estimates for 100 manufacturing industries”, The Journal of Industrial Economics 35(4): 567-581.

[9] Eaton, J. and Zvi Eckstein (1997), “Cities and growth: Theory and evidence from France and Japan”, *Regional Science and Urban Economics 27:* 443-474.

[10] Fujita, Masahisa, Paul Krugman and Anthony J. Venables (1999), *The Spatial Economics: Cities, Regions, and International Trade*, (Cambridge, MA: The MIT Press).

[11] Fujita, Masahisa and J.M. Thisse (2000), "The Formation of Economic Agglomerations:Old Problems and New Perspectives," in *Economics of Cities: Theoretical Perspectives,* Huriot, Jean-Marie and Jacques-François Thisse (Editors), (Cambridge, UK: Cambridge University Press).

[12] Gabaix, Xavier (1999), "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics 114(3)*: 739-766.

[13] Gibrat, R.(1931), *Le inégalités économiques* (Paris, France: Librairie du Recueil Sirey).

[14] Gibson, Campbell (1998), "Population of the 100 largest cities and other urban places in the United States: 1790 to 1990." Population Division Working Paper No. 27, U.S. Bureau of the Census.

[15] Huriot, Jean-Marie and Jacques-François Thisse (Editors) (2000), *Economics of Cities: Theoretical Perspectives* (Cambridge, UK: Cambridge University Press).

[16] Simon, H. (1955), "On a Class of Skew Distribution Functions," *Biometrica*, XXLII: 425-440.

[17] Zipf, G.(1949), *Human Behavior and the Principle of Least Effort*, (Cambridge, MA: Addision-Wesley).

# Appendix: Diffusions

We now prove some additional results for the case in which $X_t$ is a diffusion. Let $X_t$ be a random variable and let $P(x,t)$ denote the probability distribution of $X$ at time $t$, and $p(x,t)$ its corresponding density. The following three assumptions were discussed above.

*Assumption 1: $X_t$ is an i.i.d diffusion process with stationary transition defined by a drift, $\mu(x)$, and diffusion, $\sigma^2(x)$.*

*Assumption 2:* Suppose that for $x \geq m_t$, $m_t \equiv me^{\gamma t}$ and $t \geq 0$ :

$$p(x,t) = \delta a m_t^{\delta} x^{-(1+\delta)}. \tag{A.1}$$

where $a \equiv \Pr\{X_t \geq m_t\}$.

*Assumption 3:*

$$E_t[x\,(\mu(x) - \gamma)\,|X_t \geq m_t] = 0 \text{ for all } t \geq 0. \tag{A.2}$$

Let $p(x_0, x; t)$ be the probability density function of $x_t$, given that at an earlier time, $t_0$, $x = x_0$. Off boundaries, the motion of the conditional probability distribution $p(x_0, x; t)$ is described by the Forward Kolmogorov Equation - *FKE:*

$$\frac{\partial}{\partial t} p(x_0; x, t) = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x_0; x, t)\right] - \frac{\partial}{\partial x}\left[x\mu(x)p(x_0; x, t)\right].$$

Moreover, a limit distribution satisfies

$$\frac{\partial}{\partial t} p(x,t) = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x,t)\right] - \frac{\partial}{\partial x}\left[x\mu(x)p(x,t)\right], \tag{A.3}$$

The following result is easily verified.

**Proposition A.I** Let Assumptions 1 and 2 hold. Then $\mu(x) = \gamma$ and $\sigma^2(x) = Ax^{\delta-1}$ satisfies A.3.

**Definition:** A diffusion process, described by $\mu(x)$and $\sigma^2(x)$, supports a Pareto Equilibrium if it satisfies equations A.1, A.2, and A.3.

### 3.1. Stationary Case: $\gamma = 0$

Substituting A.1 into A.3 and dropping time subscripts produces

$$\frac{1}{2}\frac{\partial}{\partial x^2}\left[x^{1-\delta}\sigma^2(x)\right] - \frac{\partial}{\partial x}\left[\mu(x)x^{-\delta}\right] = 0. \tag{A.4}$$

Integrating and solving for $\mu(x)$, gives

$$\mu(x) = \frac{1}{2}\left[x\frac{\partial}{\partial x}\sigma^2(x) + (1-\delta)\sigma^2(x) + Ax^\delta\right], \tag{A.5}$$

where $A$ is a constant of integration. This equation characterizes the drift, $\mu(x)$, as a function of $\sigma^2(x)$. Multiplying A.5 by $x$, taking expected value with respect to $P(\cdot)$ conditional on $X_t \geq m$ and using condition A.2 one obtains:

$$E\left[x^2\frac{\partial}{\partial x}\sigma^2(x)\right] + (1-\delta)E\left[x\sigma^2(x)\right] + AE\left[x^{1+\delta}\right] = 0, \tag{A.6}$$

where $E$ is the conditional expectation. The first term of the last expression can be re-expressed as:

$$\begin{aligned} E[x^2\frac{\partial}{\partial x}\sigma^2(x)] &= \delta a m^\delta \int_m^\infty x^{1-\delta}\frac{\partial}{\partial x}\sigma^2(x)dx \\ &= \delta a m^\delta\left[x^{1-\delta}\sigma^2(x)\right]_m^\infty - \delta\left(1-\delta\right)am^\delta\int_m^\infty x^{-\delta}\sigma^2(x)dx \\ &= \delta a m^\delta\left[x^{1-\delta}\sigma^2(x)\right]_m^\infty - \left(1-\delta\right)E\left[x\sigma^2(x)\right]. \end{aligned}$$

Plugging this result into A.6 produces

$$\delta a m^\delta\left[x^{1-\delta}\sigma^2(x)\right]_m^\infty + AE\left[x^{1+\delta}\right] = 0.$$

Now, $Ex^{1+\delta} = \int_m^\infty x^{1+\delta}\delta a m^\delta x^{-\delta-1}dx = \delta a m^\delta \int_m^\infty dx = \delta a m^\delta\left[x\right]_m^\infty$. Thus, we can write the previous equation as $\delta m^\delta\left[x^{1-\delta}\sigma^2(x) + Ax\right]_m^\infty = 0$, or

$$\sigma^2(m) = -Am^\delta + m^{\delta-1}\lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x) + A\right]. \tag{A.7}$$

The following is proposition provide a general characterization of the type of diffusions supporting Pareto Distributions.

**Proposition A.II** A diffusion process with drift $\mu(x)$ and diffusion $\sigma^2(x)$ supports a Pareto equilibrium if and only if $\sigma^2(x)$ is a positive differentiable function satisfying A.7, and $\mu(x)$ satisfies A.5.

**Proof.** For sufficiency, notice that equations A.1, A.2, and A.3 are satisfied if A.5 and A.7 are satified. Necessity has already been established since A.5 and A.7 were obtained from A.1, A.2, and A.3. ■

**Example.** Let $r(x)$ be a positive constinously differentiable function satisfying $r(m) = 0$, and $\lim_{x\to\infty} x^{1-\delta} r(x) = 0$. Then, a diffusion process with drift $\mu(x) = \frac{1}{2}\left[xr'(x) + (1-\delta)r(x)\right]$ and variance $\sigma^2(x) = \beta x^{\delta-1} + r(x)$ supports a Pareto distribution.

## 3.2. Non-Stationary Case: $\gamma > 0$

**Theorem A.III** Let $X_t$ be a random variable satisfying assumptions 1, 2, and 3. If $\gamma > 0$ then, $\mu(x) = \gamma$ and $\sigma^2(x) = Ax^{\delta-1} + Bx^{\delta}$ where $A$ and $B$ are non-negative constants.

**Proof** In our case $p(x,t) = \delta a m_t^{\delta} x^{-\delta-1}$. Then $\frac{\partial}{\partial t} p(x,t) = \gamma\delta p(x,t)$. The KFE reads

$$\gamma\delta^2 a \left(me^{\gamma t}\right)^{\delta} x^{-\delta-1} = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x,t)\right] - \frac{\partial}{\partial x}\left[x\mu(x)p(x,t)\right],$$

and integrating once (with respect to $x$) produces

$$-\gamma x p(x,t) + \frac{1}{2}A(t) = \frac{1}{2}\frac{\partial}{\partial x}\left[x^2\sigma^2(x)p(x,t)\right] - \mu(x)xp(x,t), \text{ or}$$

$$[\mu(x) - \gamma]xp(x,t) = \frac{1}{2}\frac{\partial}{\partial x}\left[x^2\sigma^2(x)p(x,t)\right] - \frac{1}{2}A(t). \tag{A.8}$$

Moreover, integrating in the interval $[m_t, \infty)$ we have

$$\int_{m_t}^{\infty} [\mu(x) - \gamma]xp(x,t)dx = \frac{1}{2}\left[x^2\sigma^2(x)p(x,t) - A(t)x\right]_{m_t}^{\infty}. \tag{A.9}$$

Now, according to A.2 the left hand side of the previous equation must be zero for all $t$. Below we show that there are only two possible cases: either $A(t) = 0$ for all $t$ or $A(t) \neq 0$ for all $t$. Consider first the case $A(t) = 0$ for all $t$. Then the following equality must hold for all $t$: $\frac{1}{2}\left[x^2\sigma^2(x)p(x,t)\right]_{m_t}^{\infty} = 0$ or

$$\sigma^2(m_t)m_t = m_t^{\delta} \lim_{v\to\infty} v^{1-\delta}\sigma^2(v) \text{ for all } t.$$

Define $\beta = \lim_{v\to\infty} v^{1-\delta}\sigma^2(v)$. Then, $\sigma^2(m_t) = \beta m_t^{\delta-1}$ for all $t \geq 0$. We can replace the condition "for all t" by the expression "for all $m_t$", but then it is the same as "for all $x$" since $m_t$ grows continuously and unboundedly overtime. Thus, we conclude that

$$\sigma^2(x) = \beta x^{\delta-1} \text{ for all } x. \tag{A.10}$$

Rreplacing this expression into A.8 given that $A(t)$ is zero, one confirms that $\mu(x) = \gamma$ for all $x$. Now consider the case $A(s) \neq 0$ for some $s \geq 0$ in A.8. In that case, A.9 implies $\left[x^{1-\delta}\sigma^2(x)\delta m_t^{\delta} - A(t)x\right]_{m_t}^{\infty} = 0$ for all $t$ or

$$m_t\left[\sigma^2(m_t)\delta - A(t)\right] = \lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x)\delta m_t^{\delta} - A(t)\right] \text{ for all } t. \tag{A.11}$$

This condition requires that $\lim_{x\to\infty} x^{-\delta}\sigma^2(x)\delta m_t^{\delta} - A(t) = 0$ for all $t$, or

$$A(t) = \delta h m_t^{\delta} \text{ for all } t, \tag{A.12}$$

where $h :=\lim_{x\to\infty} x^{-\delta}\sigma^2(x)$. Substituting A.12 into A.1, we obtain

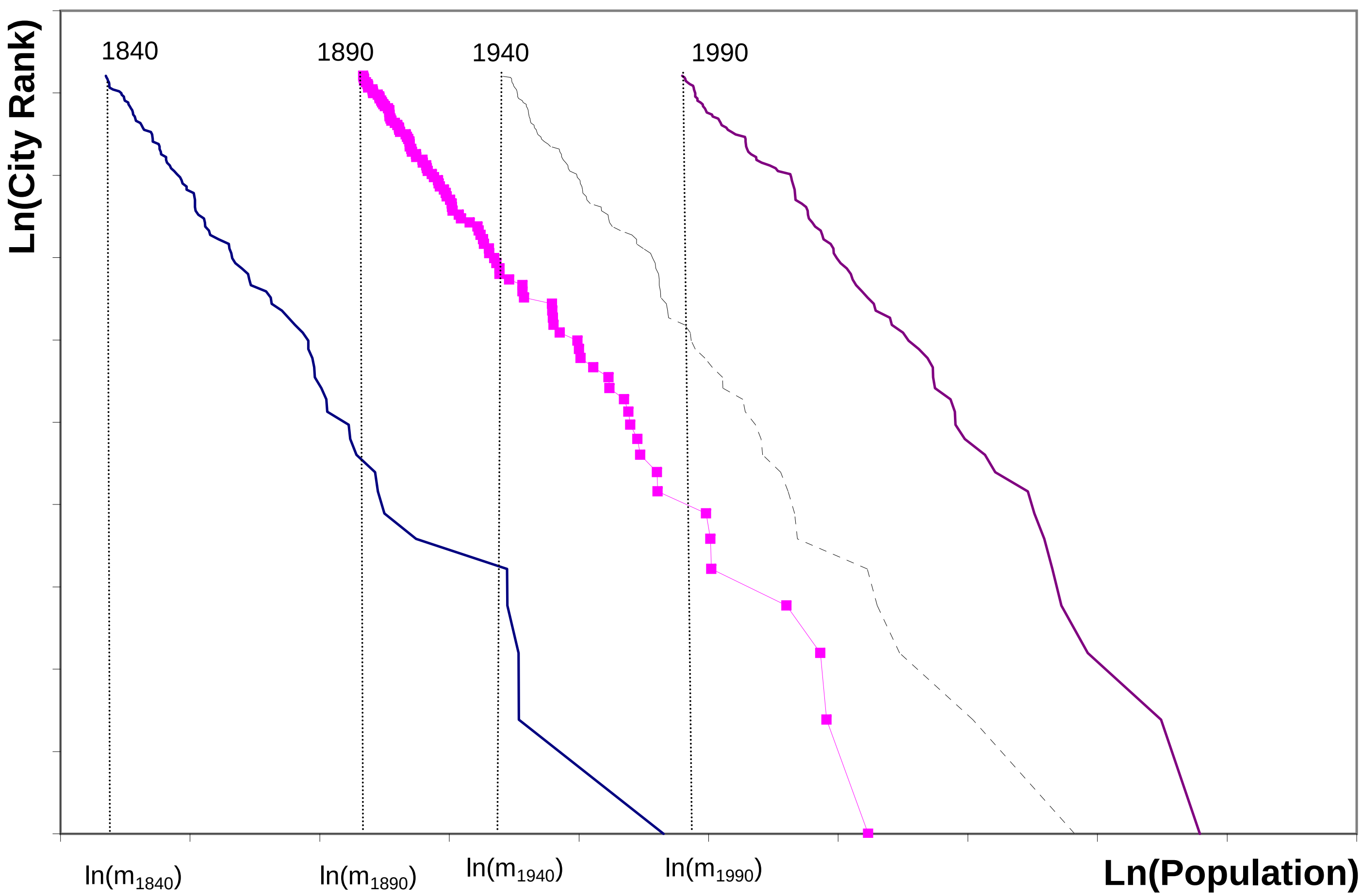
$$m_t\left[\sigma^2(m_t) - h m_t^{\delta}\right] = \theta m_t^{\delta} \text{ for all } t$$

where $\theta := \lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x) - h\right]$. Finally, solving for $\sigma^2(m_t)$ from the previous equation we obtain

$$\sigma^2(x) = h x^{\delta} + \theta x^{\delta-1}. \tag{A.13}$$

Finally, substituting A.12 and A.13 into A.8 confirms that that $\mu(x) = \gamma$. Thus, in any solution, the drift must be $\gamma$ . The diffusion coefficient, in the other hand, can either have the form A.10 or A.13, but A.10 is a particular case of A.13.

**Zipf's Plot Several Year 1840 - 1990**
**100 Largest Cities and Urban Places**




Population of 100 Largest Cities and Other Urban Places in the U. S. 1790 TO 1990, U.S. Census Bureau.
For 1990 we use Metropolitan Area Population, Census Bureau.