Gibrat's law, Zipf's law and Cointegration

Aurélie Lalanne and Martin Zumpe

GREThA, UMR CNRS 5113, University of Bordeaux, France

2015

Gibrat's law, Zipf's law and Cointegration

Aurélie LALANNE

&

Martin ZUMPE

GREThA, UMR CNRS 5113, University of Bordeaux, France

Abstract:

This paper examines the methods to detect the nature of the urban growth processes. It seems that cointegration testing enables to disentangle two versions of Gibrat's law: a first one with growth shocks that are iid across time and cities (implying convergence of the city-size distribution towards Zipf's law), and an alternative one with growth shocks that are only iid over time (implying conservation of the initial structure of the city size distribution).

Keywords: Zipf's law, Gibrat's law, Cointegration tests, unit root tests, urban growth, urban system

JEL Codes: R00, R11, C41, O40

Corresponding author:
Aurelie Lalanne
GREThA UMR CNRS 5113
University of Bordeaux
France

aurelie.lalanne@u-bordeaux.fr

1. **Introduction**

Many papers dealing with urban growth theory try to test Gibrat's law (Eeckhout, 2004, Gonzáles-Val, 2013). According to Gabaix (1999), this law is actually a random growth process that allows explaining one of the most important statistical regularities in urban economics: Zipf's law (Krugman, 1996).

Gibrat's law is a stochastic process in which city *i*'s share of national urban population in period *t*, noted $P_{i,t}$, is that in period *t-1* multiplied by $\gamma_{i,t}$, where $\gamma_{i,t}$ is identically and independently distributed (*iid*) across cities and time (Gabaix, 1999):

$$P_{i,t} = \gamma_{i,t} P_{i,t-1} \qquad (1)$$

To test the empirical relevance of Gibrat's law, Clark and Stabler (1991) recommend making use of unit root testing, and the literature widely agrees on this point (see among others Sharma, 2003). The basic Dickey-Fuller unit root test can be formulated as follows:

$$\Delta ln\, P_{i,t} = \beta_i\, ln\, P_{i,t-1} + \varepsilon_{i,t} \qquad (2)$$

If Gibrat's law is verified, we have the non-stationary unit root process $\Delta ln\, P_{i,t} = ln\, \gamma_{i,t}$, so the Dickey-Fuller test should give an estimated value $\hat{\beta}_i = 0$. The presence of a unit root implies that urban growth depends only on exogenous shocks ($\gamma_{i,t}$) without any restoring force.

However, the presence of a unit root is also consistent with the alternative unit root process $\Delta ln\, P_{i,t} = ln\, \gamma_t$, corresponding to another formalization of Gibrat's law of independence between growth rates and city sizes

$$P_{i,t} = \bar{\gamma}_t P_{i,t-1,} \qquad (3)$$

where $\bar{\gamma}_t$ is *iid* over time, but not across cities (so the growth shocks are collinear in the cross-section). Note that urban systems are generally characterized by the existence of several cities belonging to a same city-type, such as administrative cities, touristic cities, mining cities, etc. One should thus expect that cities belonging to the same city-type are affected by similar random growth shocks, implying that in these subsamples, urban growth should look something like the collinear process (3). Against this background, it should be mentioned that Gabaix's proof of convergence of a Gibrat process towards a Zipf distribution has been established for process (1), but it is not clear whether it still holds for the collinear process (3).

The fact that unit root testing does not enable to distinguish between the Gabaix formalization of Gibrat's law (1), and the alternative process (3), leaves a gap in the empirical understanding of Zipf's law. In order to disentangle between these two processes, we propose the use of cointegration tests.[1]

The remainder of this paper is organized as follows. Section 2 proves that cointegration is inconsistent with the Gabaix process (1), but not with the collinear process (3). Section 3 proves that process (3) does not give rise to convergence towards a Zipf distribution. Section 4 concludes.

2. The inconsistency of cointegration with Gibrat's law "à la Gabaix"

Take the Gabaix (1999) formalization of Gibrat's law (1) with a $\gamma_{i,t}$ distribution $f(\gamma)$ characterized by $E(\gamma) = \mu_\gamma$ and $Var(\gamma) = \sigma_\gamma^2$ verifying $|\mu_\gamma| < \infty$ and $0 < \sigma_\gamma^2 < \infty$. Taking natural logarithms, we get equation

$$lnP_{i,t} = ln\,\gamma_{i,t} + lnP_{i,t-1}, \qquad (4)$$

which is evidently integrated of order 1. Remark that for realistic annual city growth rates, $ln\,\gamma_{i,t}$ is well defined, because the growth factor $\gamma_{i,t}$ is positive. In empirical applications, we have necessarily an initial observation $P_{i,0}$, so we can rewrite equation (1) as follows:

$$P_{i,t} = \gamma_{i,t} \times \gamma_{i,t-1} \times \ldots \times \gamma_{i,1} \times P_{i,0}. \qquad (5)$$

In the same way, we obtain for city *j*:

$$P_{j,t} = \gamma_{j,t} \times \gamma_{j,t-1} \times \ldots \times \gamma_{j,1} \times P_{j,0}. \qquad (6)$$

Now recall that cointegration between two *I(1)*-variables means that there is some linear combination of these variables which is *I(0)*. So we have to find a way to link equations (5) and (6) in a manner that enables us to formulate a linear combination of $\ln P_{i,t}$ and $\ln P_{j,t}$. A general way of doing that is to raise expressions (5) and (6) to powers δ_1 and δ_2, with $[\delta_1 \delta_2]' \neq [0\ 0]'$, to divide the powered equation (5) by the powered equation (6), and then to take natural logarithms. We get the cointegration equation

$$x_{ij,t} = \delta_1 \ln P_{i,t} - \delta_2 \ln P_{j,t} - \alpha_{ij,0} \qquad (7)$$

[1] The well-known low power of cointegration tests does not preclude this empirical use. Due to the fact that unit root tests and cointegration tests are similarly affected by low power, the usual methods of dealing with low unit root power can be applied to cointegration testing: *i)* counterchecking of non-rejections of a unit root by means of stationarity tests such as the KPSS-test (Kwiatkowski et al. , 1992) and *ii)* recourse to panel tests proposed by Levin et al. (2002) and Im et al. (1995).

where $\alpha_{ij,0} = \delta_1 \ln P_{i,0} - \delta_2 \ln P_{j,0}$ has a natural interpretation as the difference between the logs of initial population levels of cities i and j, and with $x_{ij,t} = (\delta_1 \ln \gamma_{i,t} - \delta_2 \ln \gamma_{j,t}) + (\delta_1 \ln \gamma_{i,t-1} - \delta_2 \ln \gamma_{j,t-1}) + \cdots + (\delta_1 \ln \gamma_{i,1} - \delta_2 \ln \gamma_{j,1})$.

Now define the process $\{\omega_{ij,t}\}$:

$$\omega_{ij,t} = \delta_1 \ln \gamma_{i,t} - \delta_2 \ln \gamma_{j,t} \quad . \tag{8}$$

$\omega_{ij,t}$ is a linear combination of (log transformed) *iid* processes, so it is itself *iid*, with mean $E(\omega_{ij,t}) = \mu_\omega$ and variance $Var(\omega_{ij,t}) = \sigma_\omega^2$ verifying $| \mu_\omega | < \infty$ and $0 < \sigma_\omega^2 < \infty$. We can now rewrite $x_{ij,t}$ as follows:

$$x_{ij,t} = \omega_{ij,t} + \omega_{ij,t-1} + \cdots + \omega_{ij,1} \tag{9}$$

implying $E(x_{ij,t}) = t \times \mu_\omega$ and $Var(x_{ij,t}) = t \times \sigma_\omega^2$. Recall that integration of order 0 requires that the first two theoretical moments are finite and independent of time. The only vectors $[\delta_1 \delta_2]'$ which assure time independence of $E(x_{ij,t})$ are those verifying $\delta_1 = \delta_2$. But this vector choice leaves unchanged the variance, we still have $Var(x_{ij,t}) = t \times \sigma_\omega^2$. By consequence, $x_{ij,t}$ is not integrated of order 0, implying that $\ln P_{i,t}$ and $\ln P_{j,t}$ cannot be cointegrated.

For the collinear process (3), we get exactly the opposite result. Proceeding in the same was as above, we find:

$$x_{ij,t} = [\delta_1 - \delta_2] \times [\ln \bar{\gamma}_t + \ln \bar{\gamma}_{t-1} + \cdots + \ln \bar{\gamma}_1]. \tag{10}$$

By choosing $\delta_1 = \delta_2$, we obtain the degenerate random variable $x_{ij,t} = 0$ for which time independency of first and second moments is trivially verified.

3. Gibrat's law and convergence to Zipf's law

The formal proof of convergence of Gibrat's growth to a Zipf's distribution is based on city i's share of national urban population: $S_{i,t} = \frac{P_{i,t}}{\sum_i P_{i,t}}$ (Gabaix, 1999). Gibrat's law thus writes as follows

$$S_{i,t} = \gamma_{i,t} S_{i,t-1}. \tag{11}$$

Gabaix shows that the tail distribution of city sizes $G_t(S) = P(S_{i,t} > S)$ converges to the Zipf distribution which is characterized by $P(S_{i,t} > S) = a\, S^{-1}$, for some parameter a and over a large range of sizes S.

Equation (11) can be rewritten

$$P_{i,t} = \gamma_{i,t}\ \bar{\gamma}_t\ \ P_{i,t-1} \tag{12}$$

with $\bar{\gamma}_t = \frac{\sum_i P_{i,t}}{\sum_i P_{i,t-1}}$ and allows to break down the Gibrat random growth process into overall urban growth shocks $\bar{y}_t$ and city-i specific growth shocks $\gamma_{i,t}$.

Importantly, one cannot relax the assumption of a growth process based on shocks that are *iid* across time **and** cities. Suppose in fact growth shocks which are *iid* across time but collinear in the cross section, impacting growth of cities i and j in the same sense (i.e. a given shock cannot simultaneously lead to an increase of i's and to a decrease of j's population). In this instance, equation (12) transforms to the collinear process (3), where $\bar{y}_t$ is *iid* across time and has some convenient density distribution $f(\bar{\gamma})$.[2] Process (3) implies the following expression for city i's population share in period $t + n$:

$$S_{i,t+n} = \frac{\bar{\gamma}_{t+n} \times \ldots \times \bar{\gamma}_{t+2} \times \bar{\gamma}_{t+1} \times P_{i,t}}{\bar{\gamma}_{t+n} \times \ldots \times \bar{\gamma}_{t+2} \times \bar{\gamma}_{t+1} \times \sum_i P_{i,t}} = S_{i,t} \tag{13}$$

Equation (13) highlights that there is no convergence towards a Zipf distribution, because the initial distribution of city size shares is perfectly conserved over time. So we can conclude that the assumption of growth shocks that are *iid* across time **and** cities cannot be relaxed in the Gabaix (1999) proof.

4. Discussion and conclusion

Most papers applying time series methods on the analysis of urban growth focus on unit root testing in order to prove the validity of random growth à la Gibrat. By contrast, cointegration testing is scarcely used in this literature: Chen et al. (2013) highlight cointegrated growth of a minority of Chinese cities sharing important location-specific characteristics (same region, same resource endowment etc.); Sharma (2003) finds cointegration between the growth of the summed population of a set of 100 major Indian cities and the population growth of most of the individual cities of this set (89%). While these contributions reveal the existence of cointegration schemes in urban growth series, they do not formalize the logical relationships between cointegration, Gibrat's law and Zipf's law.

Our paper aims at filling this gap. In fact, we show that (unit root testable) random growth may correspond to two versions of Gibrat's law, with diametrically opposed

[2] Formally, $\gamma_{i,t}$ becomes a time-invariant collinearity coefficient $\varphi_i = 1\ \forall\ i$.

implications for cointegration and convergence behavior. The well-known Gabaix formalization (1) establishes that growth shocks are *iid* across time and cities, ensuring convergence of the city-size distribution towards Zipf's law (Gabaix, 1999); we prove that this process is inconsistent with cointegrated city growth.[3] The second version of Gibrat's law (3) is characterized by growth shocks that are *iid* across time, but collinear in the cross-section; process (3) is consistent with cointegration, but it does not converge towards Zipf's law.

In spite of their technical similarity[4], unit root tests and cointegration tests should thus be regarded as complementary tools, likely to provide guidance on the precise nature of urban growth and to give a better empirical understanding of Zipf's law.[5]

References

Chen Z., Fu S., and Zhang D., 2013, Searching for the parallel growth of cities in China. Urban Studies 50, 2118-2135

Clark, S., & Stabler, J., 1991. Gibrat's law and the growth of Canadian cities. Urban Studies 28, 635-639

Eeckhout, J., 2004. Gibrat's law for (all) cities. The American Economic Review 94, 1429-1451

Gabaix, X., 1999. Zipf's law for cities: an explanation. The Quaterly Journal of Economics 114, 739-767

Giesen, K., & Südekum, J., 2011. Zipf's law for cities in the regions and the country. Journal of Economic Geography 11, 667-686.

Gonzáles-Val, R., Lanaspa, L., & Sanz-Gracia, F., 2013. Gibrat's law for cities, growth regressions and sample size. Economics Letters 118, 367-369

Krugman, P. , 1996, Confronting the mystery of urban hierarchy. Journal of the Japanese and the International Economies 10, 399-418.

[3] This inconsistency result also holds for Sharma's approach of testing for cointegration between the growth of individual cities and overall urban growth (*see* appendix 1).

[4] Cointegration testing is unit root testing applied on the residuals of a linear combination of time series.

[5] For a similar purpose, Giesen and Südekum (2011) use non-parametric methods.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54, 159-178.
Levin, A., Lin, C. F., and Chu, C. J., 2002, Unit root tests in panel data: asymptotic and finite properties. Journal of Econometrics 108, 1-24.
Im, K., Pesaran, M., and Shin, Y., 2003, Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53-74.
Sharma, S., 2003. Persistence and stability in city growth. Journal of Urban Economics 53, 300-320.

Appendix 1

The proof presented in section 2 can be extended to Sharma's approach of testing for cointegration between the natural logarithms of each individual city *i*'s size and the sum of city sizes across all *i*, $P_{all,t} = \sum_i P_{i,t}$. If each *i* grows according to equation (1), the growth of $P_{all,t}$ is given by

$$P_{all,t} = \gamma_{all,t} P_{all,t-1}, \qquad \text{(A)}$$

with $\gamma_{all,t} = \frac{\sum_i P_{i,t}}{\sum_i P_{i,t-1}}$. Proceeding in the same way as in section 2, we get

$$x_{i,all,t} = \omega_{i,all,t} + \omega_{i,all,t-1} + \cdots + \omega_{i,all,1} \qquad \text{(B)}$$

with $\omega_{i,all,t} = \delta_1 \ln \gamma_{i,t} - \delta_2 \ln \gamma_{all,t}$, characterized by $E(\omega_{i,all,t}) = \mu_{\varpi}$ and $Var(\omega_{i,all,t}) = \sigma^2_{\varpi}$, verifying $|\mu_{\varpi}| < \infty$ and $0 < \sigma^2_{\varpi} < \infty$. The two first moments of $x_{i,all,t}$ are $E(x_{i,all,t}) = t \times \mu_{\varpi}$ and $Var(x_{i,all,t}) = t \times \sigma^2_{\varpi} + 2\sum_{k=0}^{t-2}\sum_{l=-k+1}^{t-1} Cov\,(\omega_{i,all,t+k}, \omega_{i,all,t-l})$. Time independency of $Var(x_{i,all,t})$ requires that $Cov\,(\omega_{i,all,t}, \omega_{i,all,t-1})$ = -0.5 σ^2_{ϖ}, but this is not a general property of $x_{i,all,t}$. It is for example not verified for standard density distributions (normal, uniform, lognormal etc), implying that $\ln P_{i,t}$ and $\ln P_{all,t}$ are not cointegrated.