Tsallis statistics in the income distribution of Brazil

A.D. Soares
National Commission for Nuclear Energy - CNEN

Marcelo B. Ribeiro
Universidade Federal do Rio de Janeiro – UFRJ

N.J. Moura Jr.
Brazilian Institute for Geography and Statistics - IBGE
Econophysics modeling of income distribution

• Usually by a two fitting functions:
  - Pareto power-law for the rich (~ 1%);
  - other functions (exponential, Gompertz, log-normal, Gamma, etc) for the overwhelming majority (~ 99%);
• This reinforces a common sociological view that societies are basically structured by two classes;
• What about a 3rd class, generically called “middle class”, of unclear proportion and unknown dynamics?
• What would it be the possible dynamics and interplay of such possible three-classes?
• Possibilities:
  1) include a third segment with another yet unknown function, possibly complicating the modeling considerably;
  2) Fit the data with a single function and see if some new feature comes out.
• This single function must come from some physically plausible model which allows further dynamical features to be studied (not a simply a function that fits the data).
Tsallis Distribution (TD)

- Not simply a function that fits the data;
- Could indicate some physical process, possibly applicable to income dynamics;
- Two parameters only;
- E.P. Borges and J.C. Ferrero already fitted income data with TD;
- Borges (2004): two TDs, 2\textsuperscript{nd} and 3\textsuperscript{rd} segment
- Ferrero (2005, 2011): just one year for several countries;

\textbf{PROPOSAL:}

- \textit{fit the whole annual Brazilian income data from 1978 to 2014 with a single two-parameters function}
Tsallis functions

• q-exponential and q-logarithm:

\[ \ln_q x ≡ \frac{x^{(1-q)} - 1}{1 - q}, \]
\[ e_q^x ≡ [1 + (1 - q)x]^{1/(1-q)}. \]

• Proposed complementary cumulative income distribution function:

\[ F(x) = A e_q^{-Bx}, \]

• Linearized

\[ \ln_q \left[ \frac{F(x)}{100} \right] = -B x. \]

• Fit data to find q and B for the whole distribution;
Data fitting

- Data distribution:
  - every year produces \( n \) observed income values \( x_i \) (\( i=1,\ldots,n \))
- Corresponds to \( F_i=F(x_i) \)
- Finding \( q \) for specific yearly dataset \( \{F_i, x_i\} \)
- Assuming: \( q \) lies in the interval \([-10,+10]\)
- Ranging within the interval in steps \( \Delta q=0.003 \) (rough uncertainty)
- Generate \( m \) values \( q_j (j=1,\ldots,m) \)
- For every \( q_j \) fit a straight line to find corresponding \( B_j \)
- Each year produces \( m \) quantities \( \{q_j, B_j\} \)
- Residue for each pair \( q_j, B_j \)

\[
R_j(q_j) = \sum_{i=1}^{n} \left[ \ln q_j \left( \frac{F_i}{100} \right) + B_j x_i \right].
\]

- Optimal pair \( q_j, B_j \) \( \iff \) minimum \( R_j \) \( \iff \) maximum \( \frac{d^2R_j}{dq_j^2} \)
Results (1)

- $1.24 \leq q \leq 1.51$
- $0.42 \leq B \leq 4.6$
- Both parameters vary periodically in time
- Period $\sim 3.5$ years
- Ups and downs seem correlated
- This is similar to the periodic oscillation observed for the Pareto index by other authors
- Apparently higher $q$ means higher Gini (Ferrero 2005)

<table>
<thead>
<tr>
<th>Year</th>
<th>$q(\pm 0.003)$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td>1.397</td>
<td>0.484 ± 0.065</td>
</tr>
<tr>
<td>1979</td>
<td>1.322</td>
<td>0.421 ± 0.067</td>
</tr>
<tr>
<td>1981</td>
<td>1.235</td>
<td>0.994 ± 0.034</td>
</tr>
<tr>
<td>1982</td>
<td>1.418</td>
<td>2.462 ± 0.088</td>
</tr>
<tr>
<td>1983</td>
<td>1.238</td>
<td>0.838 ± 0.050</td>
</tr>
<tr>
<td>1984</td>
<td>1.253</td>
<td>1.124 ± 0.034</td>
</tr>
<tr>
<td>1985</td>
<td>1.241</td>
<td>0.779 ± 0.039</td>
</tr>
<tr>
<td>1986</td>
<td>1.382</td>
<td>1.234 ± 0.112</td>
</tr>
<tr>
<td>1987</td>
<td>1.424</td>
<td>2.133 ± 0.095</td>
</tr>
<tr>
<td>1988</td>
<td>1.247</td>
<td>0.838 ± 0.044</td>
</tr>
<tr>
<td>1989</td>
<td>1.397</td>
<td>1.342 ± 0.058</td>
</tr>
<tr>
<td>1990</td>
<td>1.490</td>
<td>3.312 ± 0.155</td>
</tr>
<tr>
<td>1992</td>
<td>1.415</td>
<td>1.737 ± 0.113</td>
</tr>
<tr>
<td>1993</td>
<td>1.397</td>
<td>1.564 ± 0.070</td>
</tr>
<tr>
<td>1995</td>
<td>1.244</td>
<td>0.846 ± 0.039</td>
</tr>
<tr>
<td>1996</td>
<td>1.238</td>
<td>0.799 ± 0.045</td>
</tr>
<tr>
<td>1997</td>
<td>1.361</td>
<td>1.627 ± 0.075</td>
</tr>
<tr>
<td>1998</td>
<td>1.328</td>
<td>1.369 ± 0.033</td>
</tr>
<tr>
<td>1999</td>
<td>1.301</td>
<td>1.250 ± 0.043</td>
</tr>
<tr>
<td>2001</td>
<td>1.187</td>
<td>0.549 ± 0.055</td>
</tr>
<tr>
<td>2002</td>
<td>1.352</td>
<td>1.676 ± 0.059</td>
</tr>
<tr>
<td>2003</td>
<td>1.292</td>
<td>1.229 ± 0.038</td>
</tr>
<tr>
<td>2004</td>
<td>1.292</td>
<td>1.216 ± 0.045</td>
</tr>
<tr>
<td>2005</td>
<td>1.382</td>
<td>1.910 ± 0.108</td>
</tr>
<tr>
<td>2006</td>
<td>1.229</td>
<td>1.032 ± 0.047</td>
</tr>
<tr>
<td>2007</td>
<td>1.349</td>
<td>1.414 ± 0.077</td>
</tr>
<tr>
<td>2008</td>
<td>1.313</td>
<td>1.298 ± 0.060</td>
</tr>
<tr>
<td>2009</td>
<td>1.511</td>
<td>4.551 ± 0.271</td>
</tr>
<tr>
<td>2011</td>
<td>1.379</td>
<td>2.033 ± 0.081</td>
</tr>
<tr>
<td>2012</td>
<td>1.538</td>
<td>3.089 ± 0.335</td>
</tr>
<tr>
<td>2013</td>
<td>1.265</td>
<td>0.998 ± 0.068</td>
</tr>
<tr>
<td>2014</td>
<td>1.265</td>
<td>0.958 ± 0.067</td>
</tr>
</tbody>
</table>
Results (2)

- Simplest correlation between $q$ and $B$ (unrelated to time): linearity

$$B = aq + b,$$ having the following fitted parameters: $a = 4.86 \pm 0.88$ and $b = -5.13 \pm 1.13$. 
Results (3)

- $q$ and $B$ also cycle in time
Results (4)

- Still the cycling behavior of $q$ and $B$ in time
In 3D: helix like evolution of parameters (mostly clockwise)
Oscillatory nature of income (1)

- Just some years as example, but ALL years behave similarly
- Complementary cumulative distribution oscillates as the income increases

\[
\ln_q[F(x)/F(0)] = -0.8378 \times x \\
q = 1.238
\]

\[
\ln_q[F(x)/F(0)] = -1.1243 \times x \\
q = 1.253
\]
Oscillatory nature of whole income (2)

\[ \ln_q [F(x)/F(0)] = -1.3423 \times x \]
\[ q = 1.397 \]

\[ \ln_q [F(x)/F(0)] = -3.3121 \times x \]
\[ q = 1.490 \]

\[ \ln_q [F(x)/F(0)] = -0.9977 \times x \]
\[ q = 1.265 \]

\[ \ln_q [F(x)/F(0)] = -0.9575 \times x \]
\[ q = 1.265 \]
Oscillatory nature of whole income (3)

- Already present in previous income studies, but apparently went unnoticed
- Just few examples below

Dragulescu & Yakovenko (2000)

Moura Jr & Ribeiro (2009)

Clementi et al. (2008)
Conclusions

- Results show a apparent nontrivial dynamics in the income distribution, both in time and with income value.

  - **In time**: helical like evolution, also identified by Moura Jr. & Ribeiro (2013 - arXiv:1301.1090) by means of an entirely different approach (Goodwin growth-cycle model).

  - **In income**: log-periodic oscillations, especially in the tail (Pareto segment) of the distribution.

  - Tsallis distribution behaves as exponential at low income and power-law at high incomes.

  - Log-periodic oscillations may indicate an intermediate nontrivial dynamics of unknown origin (middle class that inflates and deflates?).

  - These oscillations may be described by allowing $q$ to become complex.

- **Poster**: Abreu et al “The complex Tsallis parameter and its impact in econophysics”

END
Mainstream economics

- Dominated by the study of equilibrium states:
  - states of static equilibrium: “natural” prices, demand and supply plots;
  - states of proportional and balanced growth: quantities increase uniformly with time;
  - linear models.
- In general, dynamical behavior cannot be understood by starting from systems in equilibrium states;
- In systems like weather, the important and interesting features are essentially dynamic, not small perturbations around some equilibrium state;
- These systems are generally nonlinear;
- The most striking example of a truly dynamic phenomena in economics are **trade cycles**;
- Trade cycles discuss economic quantities as “aggregated” or “macroeconomic” in nature, like total output, investment, consumption, stocks, etc;
Sketch of some modern economic trade cycles models

- **Hicks model:**
  - oscillations of aggregated quantities lead to their ceiling and floor levels;
  - income has a ceiling level.

- **Frisch-type models:**
  - economy is basically stable;
  - in the absence of “shocks”, oscillations die down and system reaches equilibrium;
  - no ceiling or floor levels.

- **Goodwin model:**
  - no full employment;
  - distributive shares of output going to capital and labor;
  - investment is tied to profits;
  - wages are related to the fraction of workers unemployment.

In the spirits of (econo)physics, it matters less the basic assumptions of the models, but more the differential equations produced by them and how these equations can be tied to measurable quantities and then empirically tested.
Marx's qualitative and empirically descriptive model of the macroeconomic dynamics of capitalist systems

- **Karl Marx** (Capital, vol 1, book 1; 1867):
  - capitalist production grows on cycles of booms and busts;
  - during a boom, profits increase and unemployment decreases (shortage of manpower);
  - a boom is followed by a bust: less unemployment reduces profits, causing then higher unemployment;
  - smaller salaries increase profit margin, renewed investment and a new boom starts...
  - ...followed by another bust, and so on...

- **A century** later **Richard Goodwin** (1967) proposed a mathematical model that tried to capture the essence of this dynamics.
The Goodwin model

- Marx's qualitative dynamics is represented by a modified Lotka-Volterra predator-prey system of 1st order ODEs;
- The number of predators and preys are replaced by two variables, $u$ and $v$:
  
  $u \rightarrow$ worker's share of total production $\leftrightarrow$ capitalist's profit margin
  
  $v \rightarrow$ employment rate $\leftrightarrow$ total share of those marginalized by the production, the unemployed

- To build the model, Goodwin advanced a series of economic hypotheses linking capital, output, total labor, output labor ratio, population, average and total wage, employment rate, profit level and investment;

- The model translated these hypotheses into an ODE system with various parameters whose signs are also fixed.
The Goodwin model (part 2)

Goodwin model is a Lotka-Volterra predator-prey like system of 2 ODEs:

\[ \begin{align*}
\dot{u} &= -a_1 + b_1 v \\
\dot{v} &= a_2 - b_2 u
\end{align*} \]

- 1\text{st} equation: positive slope
- 2\text{nd} equation: negative slope
- The model also has a fixed center \((u_c, v_c)\)

\[ \begin{align*}
u_c &= ca_2, \quad c > 0; \\
v_c &= \frac{a_1}{h}, \quad h > 0
\end{align*} \]
The Goodwin model (part 3)

- Goodwin model has clockwise orbits with an unique center in the *u-v* phase plane;
- Variables have a predator-prey like time evolution;
- Model is unstable to a change in its parameters, but the single center remains.
Empirical evidence

- Since its proposal, several *theoretical* developments were advanced by economists;
- The Goodwin model gained a dedicated group of supporters;
- However, after 47 years very few *empirical* studies were carried out trying to *test* its validity with real data;
- The limited empirical results range from partial qualitative acceptance to total quantitative rejection;
- Partial qualitative acceptance motivated this study;
- It is based on a different approach to analyze data, inspired by recent efforts made by econophysicists on the problem of characterizing income distribution;
- Income distribution functions are used to characterize the model's variables $u$ and $v$. 
Testing the Goodwin model with Brazilian data

- Individual income distribution can be modeled by the Gompertz-Pareto distribution (GPD);
- Gompertz curve (double exponential) + Pareto power law
- Complementary cumulative distribution for average income $x$

\[
F(x) = \begin{cases} 
G(x) = e^{e^{(A-Bx)}}, & (0 \leq x < x_t), \\
P(x) = (x_t)^{\alpha} e^{e^{(A-Bx)}} x^{-\alpha}, & (x_t \leq x \leq \infty),
\end{cases} \quad \text{(Gompertz)}
\]

\[
P(x) = (x_t)^{\alpha} e^{e^{(A-Bx)}} x^{-\alpha}, \quad (x_t \leq x \leq \infty), \quad \text{(Pareto)}
\]

- $x_t$ – transition income value
- $\alpha$ – Pareto index
- $B$ – Gompertz parameter
- $A$ – Gompertz boundary condition

\[A = \ln (\ln 100) = 1.5272.\]

- Availability of Brazilian income data and previous studies with that database made this work possible.
The Gompertz-Pareto distribution (GPD) and Brazilian individual income data

- Essential results stemming from recent studies:
  
  

  1% are “the rich” (Pareto) and 99% are “the rest” (Gompertz);

- GPD is a good approximation for highly polarized (high Gini coefficient) income distributions, like the Brazilian one:

- “Middle class” is represented by the exponential approximation of the upper part of the Gompertzian component (consistent with Dragulescu and Yakovenko 2001);

- Detailed characterization of “middle class” is work in progress;

- GPD is a tool to partition the income distribution in segments capable of characterizing the Goodwin variables $u$ and $v$;
GPD and the Goodwin model

- Gompertzian segment characterizes $u$;
- Unemployment is characterized as a lower limit income threshold value;
- Unemployment share is obtained from the income distribution, and not from official unemployment statistics (drop long term unemployment), but percentages are in general agreement for the last 15 years;
- Employment rate $v = (100\% - \text{unemployment share})$;
- Once $u$ and $v$ are yearly defined, their time derivatives, $\dot{u}$ and $\dot{v}$ can be obtained numerically and straight line fitting is used to ascertain the validity of the economic hypotheses of the model;
- Time evolution of these variables can also be studied.
Time evolution of the Goodwin variables in Brazil 1981-2009

- Approximate cycling behavior with 4-year periods;
- Variables have phase difference of about 2 years;
- Short term cycles.
Brazilian $u-v$ phase plane: 1981-2009

- Clockwise cycles, but no single center;
- Two cycling regions, “center” appears to move to the upper region of the plane;
- Qualitative agreement (clockwise cycles), but quantitative disagreement (no single center) with the original model.
Phase plane evolution (↑) of the Brazilian macroeconomic system

Closer look at phase plane shows that the data can be divided in two distinct regions:

(A) 1981 to 1994; (B) 1995 to 2009

Outliers: 1986, 1990
Tentative interpretation of the divided Brazilian $u$-$v$ phase plane: 1981-1994 (left) and 1995-2009 (right)

- 1994 event: abrupt end of Brazilian hyperinflation;
- Inflation can be considered as an additional tax on labor;
- This may have triggered the system to move into a new position in the phase plane, where employment rate, i.e., Gompertzian (labor) component share, is higher;
- Earlier, failed, attempts to control hyperinflation: 1986 and 1990 (outlier points 6 and 10);
- Runaway inflation started in 1982, so the system may have been in another region in the phase plane before that.
Time evolution of the $u$-$v$ phase plane

- Previous results in a tri-dimensional plot;
- Points 1 and 2 seem to be the transition from an unspecified earlier region;
- Points 3 to 14 correspond to the hyperinflationary period in the Brazilian economy, abruptly finished in 1995;
- Projection (left vertical plane) also seems to indicate that different regions correspond to different inflationary (income) regimes in Brazil.
Temporal variation of employment rate and workers' share

- Numerical evaluation of the variables' derivatives;
- Straight line fitting to determine the parameters of the model observationally;
- Points show important dispersion;

Observational results imply slopes opposite to what the model is supposed to obey, according to the economic hypotheses.

\[ A_1 = 0.17 \pm 0.06, B_1 = -0.0019 \pm 0.0006 \]  
\[ A_2 = -0.52 \pm 0.22, B_2 = 0.006 \pm 0.003 \]
Effect on the economic hypotheses

• All underlying economic hypotheses might be affected in one way or another by the opposed than expected parameter behavior;
• But, this occurs in non obvious ways
• However, results suggest that, more strongly:
  - $B_1 < 0$ affects assumption of wage rate rises with employment; labour force and productivity grow exponentially
  - $B_2 < 0$ affects wage rate = capital / output
Conclusions

- The Goodwin model agrees only qualitatively with the data: it fails on the quantitative front, though this is not a new conclusion;

- It may, however, be useful as a starting point for an improved model showing better data agreement;

- Much theoretical work done on this model by economists have been focused on its economic hypotheses;

- Empirical studies indicate that these hypotheses may, perhaps, not be valid, or need to be modified;

- Perhaps, it is more fruitful to focus on its empirical validation, rather than speculate about hypotheses which may not be substantiated by real data;

- Macroeconomics can be approached at the macro level without any need of the so-called “microeconomic foundations of individual's rationality” and, so, no need to talk about “agent's behavior”, representative or not;

- As the focus is on average individual income evolution, this work can be viewed under an economical distributive dynamics perspective;

- The GPD and its exponential approximation provide a useful tool for proposing dynamical models of economic systems;

- It would be very interesting to see similar works carried out using income data of other countries.

Conclusions