Working Paper No. 259

Constructing Long and Dense Time-Series of Inequality Using the Theil Index

Pedro Conceição
Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin
Pedroc@uts.cc.utexas.edu

James K. Galbraith
The Jerome Levy Economics Institute
Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin
Galbraith@mail.utexas.edu

INTRODUCTION

Most empirical work on inequality uses measures that are based on household surveys. These aim to provide a comprehensive overview of income inequalities, covering all social strata and comparable both through time and between countries. Gini coefficients are the index mostly commonly computed from these sources, though various quintile ratios are also frequently deployed.

Deininger and Squire (1996) have compiled an impressive data set of available Gini and quintile measures of inequality. Yet, the limitations of this data for studies of the evolution of inequality through time are evident from Table 1, which shows the number of data points in a 26 year period (1970-1995) for those countries for which more than three data points are available. Only four countries show data for virtually every year, and most do not have data for even half of the years. And these gaps are irreparable. There is no way to construct Gini coefficients for countries and years for which adequate household sample surveys were never conducted in the first place.

Table 1- Number of Data Points in the High Quality Deininguer and Squire (1996) Data-set Between 1970 and 1995 (only countries with more than 3 data points are shown).

Australia	8	France	4	Norway	7	Taiwan	23
Bangladesh	8	Germany	5	Pakistan	8	Thailand	6
Brazil	14	Hungary	7	Panama	4	UK	22
Bulgaria	24	India	12	Peru	4	USA	22
Canada	17	Indonesia	9	Philippines	4	Venezuela	9
China	12	Italy	15	Poland	17		
Colombia	7	Japan	16	Portugal	4		
Costa Rica	8	Korea, R.	7	Singapore	6		
Cote d'Ivoire	5	Mexico	5	Spain	7		
Denmark	4	Netherlands	12	Sri Lanka	6		
Finland	10	New Zealand	12	Sweden	14		

Fortunately, the decomposability properties of the Theil measure make it possible in part to repair this gap, albeit in most cases only for the limited span of the manufacturing economy. In particular, one can compute between-group measure of inequality (T' hereafter) across industrial sectors, as delineated by national or international industrial classification schemes. Data on industrial wages, earnings and employment are very easily found. The data are also reasonably reliable; there is little reason to suspect that they are faked in any systematic way that would affect a Theil measure. Where gross errors do occasionally enter into the recording, the regularity and hierarchical structure of the data sets often means that these can be detected.

2- THEIL'S INEQUALITY MEASURE

Henri Theil (1967) first noted the possibility of using Claude Shannon's (1948) information theory to produce measures of income inequality. Shannon's theory was motivated by the need to measure the value of information. Shannon argued that the more unexpected an event is, the higher the yield of information it would produce. To formalize this idea, Shannon proposed to measure the information content of an event as a decreasing function of the probability of its occurrence. Adding some axiomatic principles, most importantly that independent events should yield information corresponding to the sum of the individual events' information, Shannon chose the logarithm of the inverse of the probability as the way to translate probabilities into information. The logarithm allows the decomposition of the multiplicative probabilities into additive information content.

If we have a set of n events, one of which we are certain is going to occur, and each with a probability x_i of occurring, then $\sum_{i=1}^{n} x_i = 1$ and the expected information content is given by Shannon's measure:

[1] $$H=\sum_{i=1}^{n} x_i \log\frac{1}{x_i}$$

The information content is zero when one of the events has probability 1; we draw no information from the occurrence of an event we are sure is going to happen. The information content is maximum when $x_i=\frac{1}{n}, i=1,\dots,n;$; in this case $H = \log n$. In other words, maximum information is derived from the occurrence of one event in a context of maximum uncertainty. To borrow from thermodynamics, maximum information is derived from a state of maximum disorder, or maximum entropy. This is the reason why entropy is used as a synonym of expected information.

Theil was attracted to information theory because it might lead toward a general partitioning theory. Beyond dividing certainty (probability 1) into various uncertain probabilities, information theory presented an opportunity to devise measures for the way in which some set is divided into subsets. Theil considered it natural to apply information theory to the partitioning of overall income throughout the taxpayers of a country. If we were to apply Shannon's measure directly to individual shares of income, we would have a measure of equality (recall that the maximum of Shannon's measure occurs when all the shares are equal). Therefore, Theil proposed to subtract Shannon's measure from log n , leading to his well-known measure of *in*equality:

[2] $$T=\frac{1}{n}\sum_{i=1}^{n} r_i.\log r_i$$

Where r_i is the ratio between individual income (y_i) and average income (μ_Y): $r_i=\frac{y_i}{\mu_Y},\ \mu_Y=\frac{\sum_{i=1}^{n} y_i}{n}.$. The value of the Theil index (T index) is a monotonically increasing measure of inequality in the distribution of income, bounded by $T\in[0,\log n]$.

Theil argues that the fact that *T* does not have an upper bound but depends always on population size is desirable. Consider a society with only two individuals in which one earns all the income. In this case, *T* = log 2. Next, consider another society in which all the income is again concentrated in one person, but the overall population is now one thousand. In this case, *T* = log 1000, a much higher value as desired in a much more unequal society. Consider now a different situation: if the division of income in this larger society were in the same proportion as in the first (half of the population having all the income), then we would have again *T* = log 2 for the larger society, as is to be expected. In general, Theil showed that $T - \log\frac{1}{\theta}$, in which *q* is the proportion of the population having all the income (1/2 in our last example). This is independent of the size of the population.

Theil's measure has all of the desirable properties of an inequality measure: it is symmetric (invariance under permutations of individuals), replication invariant (independent of population replications), mean independent (invariant under scalar multiplication of income), and satisfies the Pigou-Dalton property (inequality increases as a result of a regressive transfer). It is also Lorenz-consistent, meaning that it agrees with the quasi-ordering that can be derived from comparing Lorenz curves.

An important characteristic of entropy-based indexes such as the Theil index is that they are decomposable. If individuals are grouped in a mutually exclusive, completely exhaustive way, overall inequality can be separated into a between-group component and a within-group component. If we consider that the population is divided into *m* groups, g_1, g_2, ..., g_m, each with n_j individuals, *j* =*1* , ..., *m* , then the decomposition takes the self-similar form of a fractal:

$$[3] \quad \begin{cases} T = \sum_{j=1}^{m} p_j R_j \log R_j + \sum_{j=1}^{m} p_j R_j T_j \\ T_j = \frac{1}{n_j} \sum_{i \in g_j} r_i \log r_i \end{cases}$$

The population proportion in each group is represented by $p_j = \frac{n_j}{n}$ and the ratio of average group income to overall average income by $R_j - \frac{\mu_j}{\mu_Y}.$.

There are several reasons why it may be of interest to have a decomposable measure of inequality. One might be interested in analyzing the functional distribution of income according to some criterion that divides the overall population into groups. Examples are race, gender (both of which were explored by Theil in 1967), education level, economic sector, age, to name a few. Another reason might be associated with geography (different regions, like, say, states or countries, which were explored also by Theil in 1967). Another possibility is study differences in urban vs. rural populations. Yet another reason may be related to the differentiation of sources of income.

A further important motivation, again recognized by Theil himself, is associated with data. Data on income is often reported in income brackets, which do not give information on what is the distribution of income *within* the income bracket. Theil explored how the decomposition properties of the *T* index might help in devising measures of inequality not based on percentiles.

In this paper we go beyond these efforts, to explore the use of *T* to construct long and dense time-series of inequality measures from industrial data. We are interested in looking at the time evolution of inequality, to allow the study of the processes that drive and determine changes in inequality. Problems associated with data availability and with the choice of the instruments to measure inequality have hindered the possibility of constructing long, dense-time series, as we saw above.

Clearly, $T^{/}$ constructed across industrial sectors yields an incomplete picture of inequality at each point in time.; $T^{/}$ is not a substitute for T. But, we argue, the potential for constructing long and dense time series outweighs this disadvantage. The question is, to what extent do changes in $T^{/}$ measure changes in T? Can we use the change in $T^{/}$ as a proxy for the evolution of inequality in the larger distribution from which $T^{/}$ is computed?. Section 3 quantifies the "information loss" when one uses $T^{/}$ instead of T . In Section 4 we discuss procedures for isolating income-change from population-shift effects, so that we can reduce the range of uncertainty associated with inferring the change in T from the change in $T^{/}$. Finally, in Section 5, we provide some empirical illustrations using data for Brazil.

3- GOING FROM T TO $T^{/}$: WHAT IS INCLUDED, AND WHAT IS LEFT OUT?

The between-group component of Theil's T can be computed from wage or earnings data aggregated by industrial sectors in a very large number of countries. All that is required are measures of total payrolls or the wage bill, and measures of employment or hours, for consistently organized industrial categories. However, inequality overall depends also on the inequality *within* each group, and on the change in population shares across groups as time passes. Therefore, it is of interest to examine these two effects and to assess how large their impact on inequality may be. In this section we discuss how to account for the left-out inequality associated with the unobservable within-group inequality. We leave for the next section a discussion of how to isolate population effects.

We are primarily interested in a dynamic analysis of inequality, in how inequality changes over time. Therefore, we focus on the behavior of rates of change.

From [3] we can compute the change in inequality over time:

$$\dot{T} = \sum_{j=1}^{m}\left[\left(R_j \log R_j + R_j T_j\right).\dot{p}_j + \left(p_j \log R_j + p_j + p_j T_j\right).\dot{R}_j + p_j R_j \dot{T}_j\right] \quad [4]$$

and also the change over time exclusively associated with the between group component:

$$\dot{T}' = \sum_{j=1}^{m}\left[R_j \log R_j \dot{p}_j + \left(p_j \log R_j + p_j\right).R_j\right] \quad [5]$$

This means that [4] can be written as:

$$\dot{T} = \dot{T}' + \sum_{j=1}^{m}\left[\frac{d\left(R_j.p_j\right)}{dt}T_j + R_j p_j \dot{T}_j\right] \quad [6]$$

From [6], the only unobservable components are T_j and $\dot{T}_j$. Therefore, we can measure the first term on the right hand of [6], but we cannot measure the second term. This non-measurable component corresponds to the change unaccounted for by the between group component change, which is given by [5]. However, we can state conditions under which the effect of the within group change is likely to be small.

The within group change is the time derivative of the product $R_j.p_j.T_j$, and therefore, with broad generality, within group changes will be small whenever changes in $R_j.p_j.T_j$ are small. Since

$$R_j = \frac{1}{p_j}.\frac{Y_j}{Y} \qquad [7]$$

where Y_j is group j 's total income, the product $R_j.p_j.T_j$ can be reduced to $(Y_j/Y).T_j$, which is independent of p_j. Finally, we can do the following simplification

$$\frac{d(R_j.p_j)}{dt}T_j + R_j p_j \dot{T}_j = \frac{Y_j}{Y}\left[\dot{T}_j + T_j(g_j - g)\right] \qquad [8]$$

where $\dot{Y}/Y$ and $g_j = \dot{Y}_j/Y_j$.

Two features are immediately apparent from [8]. First, the within group Theil change is independent of the employment structure, and depends exclusively on the **relative levels of average wages** and on the **relative rates of wage change**, besides the unobservable T_j and $\dot{T}_j$. Secondly, there are two contributions to the within group change, one related with the obvious group endogenous change in the distribution over time, T_j, and a second reflecting the effect of relative change in the wage structure. The term $(g_j - g)$ can be understood as the relative growth rate of average wages for group j.

How large is each of the terms given by [8]? We know that Y_j/Y is between zero and one. Moreover, if the number of groups, m, is relatively large, then Y_j/Y is likely to be small. Consequently, the impact of Y_j/Y on the expression is always be to reduce the effect of the within group component on the time variation of the Theil index as a whole.

If $g_j \sim g$, for every j, then the effect associated with structural wage change is low. There is no reason why each group's rate of wage changes should be equal and close to g, but there is a trade-off. Since g is the overall growth rate, if one or a few group rates of growth are higher than g, then it must be that the remaining are lower. Therefore, the coefficients are likely to cancel out on average, and the overall contribution of relative wage change to within-group inequality is likely to be small.

The remaining problem is that we do not know the levels T_j, the extent of inequality within each group. Since the changes in T_j are also unknown, there is not much that one can do with generality from this point on. Nonetheless, we can estimate the **maximum** impact of this unobservable effect. We know that T_j^{MAX})t)=log[n_j(t)], that is, the maximum value of the inequality within each group is equal to the log of the population in that group. It is much more difficult to determine a maximum for the rate of change of the within group inequality. In principle, $\dot{T}_j$ could be almost infinite, since we could go from any distribution of income to a situation under which all the income in the group is concentrated in one individual. However, this is not very likely to happen instantaneously. If we move from the continuous analysis to a discrete analysis across time, the highest change occurs when inequality moves from zero to log (n_j) from t to $t+1$, or from log (n_j) to zero from one period to the other. Note that there is a duality here: the highest possible change implies that at one of the periods t or $t+1$ within group inequality is zero, and the impact of the level component is, therefore, zero. For example, when the within group inequality jumps from virtually zero to T_j^{MAX})t)=log[n_j(t)], the $T_j(t)$ terms would have contributed very little to overall inequality before the jump.

Taking first the issue of the maximum inequality level within group, consider that the upper bound for the summation on the right hand side of [8] occurs when *all* groups have their maximum level of inequality (all the group's income in one individual). And though this is an unlikely and unstable situation, we will assume that all groups will remain with their maximum level of inequality. In this situation $\dot{T}_j = \frac{\dot{p}_j}{p_j}$. Introducing this last expression and T_j^{MAX})t)=log[n_j(t)] in [6], we get that when within group inequality is kept at its maximum level, the maximum impact of the unobservable component is given by:

[9] $$d^{MAX}_{within} = \sum_{j=1}^{m} \left\{ \frac{Y_j}{Y} \left[\frac{\dot{p}_j}{p_j} + \left(g_j - g\right).\log n_j \right] \right\}$$

From a formal point of view it is worthwhile to note the dependency of this expression on the rate of change of the employment structure. The term in square brackets dependss only on rates of change in employment and wages, with the weight for the relative wage growth being now given by the log of the groups' population. All the variables in [9] are observable, but expression [9] still cannot be used empirically, since it contains differential terms. Expressing the growth rates by logarithmic differences, as is standard practice, we obtain an expression that can be used with discrete data:

[10] $$d^{MAX}_{within}(t,t+1) = \sum_{j=1}^{m} \left\{ \frac{Y_j(t+1)}{Y(t+1)} \left[\log\left[\frac{p_j(t+1)}{p_j(t)} \right] + \log\left[\frac{Y_j(t+1)}{Y(t+1)} . \frac{Y(t)}{Y_j(t)} \right] . \log n_j(t+1) \right] \right\}$$

We can thus express the maximum changes in the within group Theil exclusively with observable variables. First, we will assume that from *t* or *t+1* the within group inequality will jump, for every group, from 0 to the maximum level. In this case, again expressing the growth rates by logarithmic differences:

[11] $$\Delta^{MAX}_{within}(t,t+1) = \sum_{j=1}^{m} \left\{ \frac{Y_j(t+1)}{Y(t+1)} . \log\left[n_j(t+1)\right] . \left[1 + \log\left[\frac{Y_j(t+1)}{Y(t+1)} . \frac{Y(t)}{Y_j(t)} \right] \right] \right\}$$

The maximum change in the opposite direction, from the maximum level of within inequality to zero, is given by

[12] $$\Delta^{MAX}_{within}(t,t+1) = -\sum_{j=1}^{m} \left\{ \frac{Y_j(t+1)}{Y(t+1)} . \log\left[n_j(t)\right] \right\}$$

Expressions [9] through [11] include only measurable variables, and can be computed from data on industrial wages. However, one must bear in mind that these estimates are certainly highly exaggerated. We are assuming the unrealistic situation under which eiither all income in each group is concentrated in a single individual, or else the change between consecutive periods goes from one extreme to the other of possible within group Theil values, that is from zero to log (n_j).

How big are changes in within-group inequality likely to prove in practice? It is possible to consider this issue by reflecting on the nature of industrial classification schemes. Consider first the extreme, and once again unrealistic, case where industrial classifications had no intrinsic economic meaning, but are simply a random classification system whose only virtue, for our purposes, is that each factory retains its identifying label through time and is therefore recorded in the same category every year. In this case, the fractal character of the distribution, and of the Theil index, assures us that the change in T^{j} is very close the change in T. This result is obvious from the assumption of randomness:. Changes within groups must also be happening across groups; there is no basis for within-group inequalities to be changing at a different rate from between group inequalities, other than random differences which are likely in any event to be offsetting once groups are added together.

Now consider the more realistic case where industrial classification schemes actually do, to some imperfect extent, distinguish between qualitatively differing types of economic activity. The effect of this is again obvious. Relative to the random-taxonomy case, some within group and unobservable variations must move to the between group part of the expression, where they can be observed. Why? Because industries now mean something, and if they mean anything at all, the effect must be to impose a measure of homogeneity on entities classified together, and a measure of distinctiveness to entities classified as being in different groups.

Pursuing this line of thought further, consider that "industries" are in fact collections of similar factories, which operate from one year to the next with labor forces, internal wage structures, managerial hierarchies and technologies that change fairly little. It seems clear that while within-group inequalities are likely to be large relative to differences between group averages, the internal rigidity of industrial structure tends to assure that changes in within group inequalities in an industrial classification will be small relative to changes between groups. Therefore, a measure of the change in T' is likely to be a robust estimate of the change in T, so long as changes in employment structures and the distribution of the workforce across categories are not too large.

4- Separating out employment effects

We now turn to the effect of changes in employment structures on T'. Our strategy relies on Theil's own hypothetical question: what would have happened to inequality if employment shares had not changed? Taking the beginning of the time-series as a starting point, then, if employment shares do not change, equation [5] simplifies to:

$$\dot{T}_F{}' = \sum_{j=1}^{m} p_j \left(\log R_j + 1\right).\dot{R}_j \qquad [13]$$

What are the implications of a fixed employment structure for our estimates of the maximum change in the within group component of T? We established that the within group component, and its changes over time, are independent of the employment structure. However, when estimating the maximum impact due to high levels of within group inequality, expression [9] shows a dependency on employment changes. With the assumption of a fixed employment structure, [9] turns to:

$$d^{max}_{within,F} = \sum_{j=1}^{m} \left[\frac{Y_j}{Y}.\left(g_j - g\right).\log n_j\right] \qquad [14]$$

Expression [14] can easily be turned into a discrete form, amenable to empirical use, using the same procedure discussed above when moving from equation [9] to [10].

In short, isolating the effect of changes in employment structure on the Theil index is entirely straightforward, and requires only observable data. In the next section we illustrate the practical effect of changing employment structures on inequality in the case of Brazil.

5- Empirical Application to the Case of Brazil

We will now compute T' for the case of Brazil using monthly data on wages and employment for 17 industrial sectors. Data are monthly for the period 1976-1995, leading to the long and dense, continuous time-series of Figure 1.

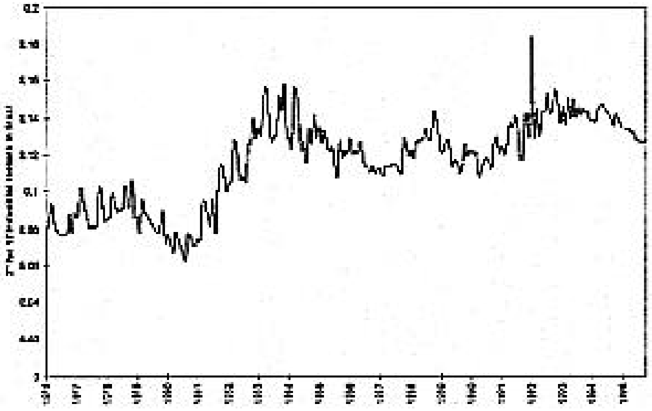

Figure 1- Monthly $T^{/}$ for 17 Industrial Sectors in Brazil 1976-1996.

In Figure 2 we plot the series after smoothing, using a 12 month moving average. Also plotted here are the "high-quality" Gini coefficients from Deininger and Squire's (1996) data set for Brazil.

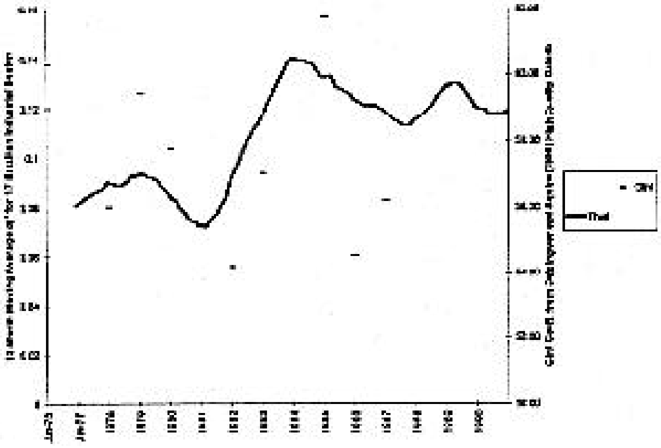

Figure 2- Smoothed $T^{/}$ Series for 17 Industrial Sectors in Brazil 1976-1996 and Available High Quality Gini Coefficients.

To determine how much of the change in inequality overall has not been accounted for, we begin with an informal discussion of the structure of the data. First, consider the wage structure. The proportion of total wages held by each industrial group is depicted in Figure 3, which shows the evolution of each industry wage's share. The highest positive rate of change is 7% and the highest negative rate of change is -4.5%. Average rates of change, both negative and positive, are just over 1%. The mechanical sector has the highest share, which reaches just over 20%. Steel and transportation have shares above 10%, but the remaining fourteen sectors have shares below 8%.

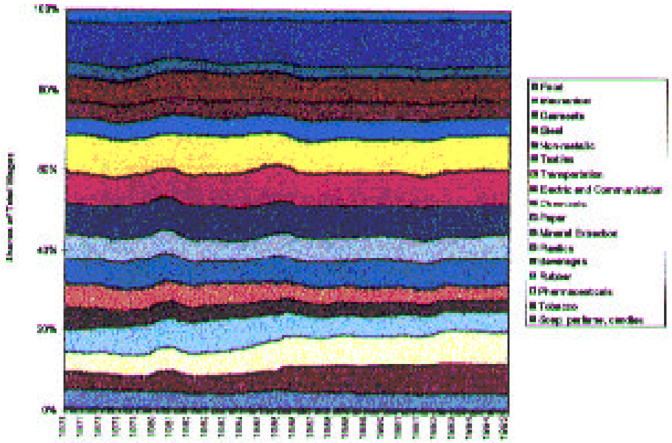

Figure 3- Smoothed Shares of Wages for 17 Industrial Sectors in Brazil.

Figure 4 shows the evolution of the employment structure. Despite being irrelevant to the estimate of the unobservable impact given by [8], which is independent from employment, this effect is important for the maximum potential impact estimate given by [9]. Four groups have consistently more than 10% of employment, and one, food, reaches a high of 19% in 1992 and 1993. The next four industries in employment share, non-metallic, textiles, transportation, and electric/communications, have shares that oscillate between 5% and under 10%. More importantly, since [9] depends only on changes in employment, Figure 4 shows that there are no sharp transitions in population shares. In fact, the highest positive rate of change is 3% and the highest negative -6% (average growth rates, both positive and negative, are about 1%). From this we infer that changing population shares will rarely affect T by very much.

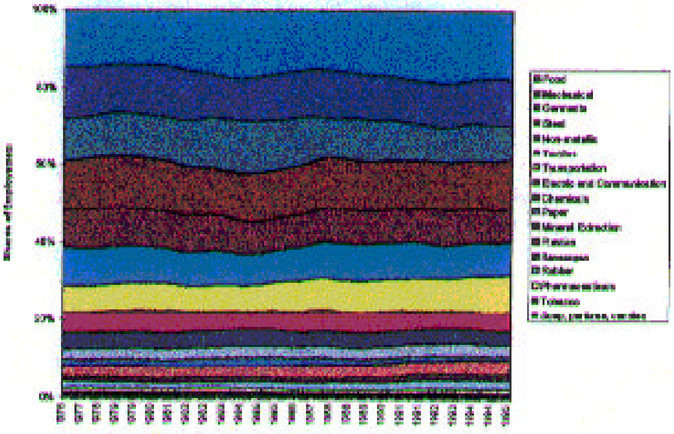

Figure 4- Smoothed Shares of Employment for 17 Industrial Sectors in Brazil.

Given what we know about the changes of the shares of wages and employment, we can explore what the outcomes of [10] through [12] are likely to be. Since the structure of wages changes little from one instant in time to the other, the following approximations are valid:

$$\frac{Y_j(t+1)}{Y(t+1)} \cdot \frac{Y(t)}{Y_j(t)} = \left(\frac{Y_j(t+1)}{Y(t+1)}\right) \Big/ \left(\frac{Yj(t)}{Y(t)}\right) \approx 1 \Rightarrow \log\left[\frac{Y_j(t+1)}{Y(t+1)} \cdot \frac{Y(t)}{Y_j(t)}\right] \approx 0 \qquad [15]$$

Therefore, expression [10] can be approximated by:

[16] $$d^{MAX}_{within}(t,t+1) \approx \sum_{j=1}^{m}\left\{\frac{Y_j(t+1)}{Y(t+1)}\log\left[\frac{p_j(t+1)}{p_j(t)}\right]\right\}$$

and expression [11] by:

[17] $$\Delta^{MAX^+}_{within}(t,t+1) = \sum_{j=1}^{m}\left\{\frac{Y_j(t+1)}{Y(t+1)}.\log\left[n_j(t+1)\right].\right\}$$

Expression [12] is unaffected by approximation [15]. Note, however, that [17] is almost symmetric with [12]; whenever $n_j(t) \approx n_j(t+1),$, we should expect $\Delta^{MAX^+}_{within}(t,t+1) \approx -\Delta^{MAX^-}_{within}(t,t+1).$.

In both [17] and [12] log[n_j (·)] tends to smooth the differences across industries. Taking the extreme cases, employment in the food industries is around 500,000 and in the soap and perfume industries around 20,000; when logs are taken, the values are 13 for food and 10 for soap and perfume. Likewise, and in an even more dramatic way, the log smoothes the changes in employment within industries across time, which means that the expressions [17] and [12] are almost constant, since we also saw that the change over time of the wage shares was smooth. Therefore, there is no need to compute a time series for either $\Delta^{MAX^+}_{within}(t,t+1)$ or $\Delta^{MAX^-}_{within}(t,t+1)$. Their values are likely to be almost constant over time. Furthermore, a time series of either [17] or [12] makes no sense, since, for example, the Theil cannot jump from zero to the maximum from *t* to *t+1* , and then again from zero to the maximum from *t+1* to *t+2* . What would be meaningful would be one time-series in which [12] and [17] alternate from one period to the next. But since we know that the values are almost constant over time, we might as well compute averages over time.

Table 2 shows the average values for $\Delta^{MAX^+}_{within}(t,t+1)$ and $\Delta^{MAX^-}_{within}(t,t+1)$, in which the computation was made using the exact formulas [11] and [12], and the averages are of the monthly values over the entire period of time under consideration.

Table 2- Maximum Impact of the Maximum Changes over Time of Within Industry Inequality

	$\Delta^{MAX^+}_{within}(t,t+1)$	$\Delta^{MAX^-}_{within}(t,t+1)$
Average (n=226)	12.51906	-12.52010
Standard Deviation	0.11943	0.11883

Table 2 confirms that [11] and [12] are symmetrical, and also, that their change over time is low (note the low levels of the standard deviation). Clearly, these values are much above the changes in between group Theil that we observe, which never surpass 0.00437 and are never below 0.00270. But it is also clear that the values of Table 2 are highly unlikely to correspond to a real evolution of the Theil index.

From expression [10], or its simplified form [16], it is not possible to make any further simplifications. In fact, the "strong" term log[n_j (·)] is almost washed out in [16]. In this case, then, it is of interest to know what is the time evolution of [10], and to compare it with the observed between industry Theil. Figure 5 compares what would be the maximum change in of the overall Theil index if the levels of the within industry Theil would remain at their maximum level.

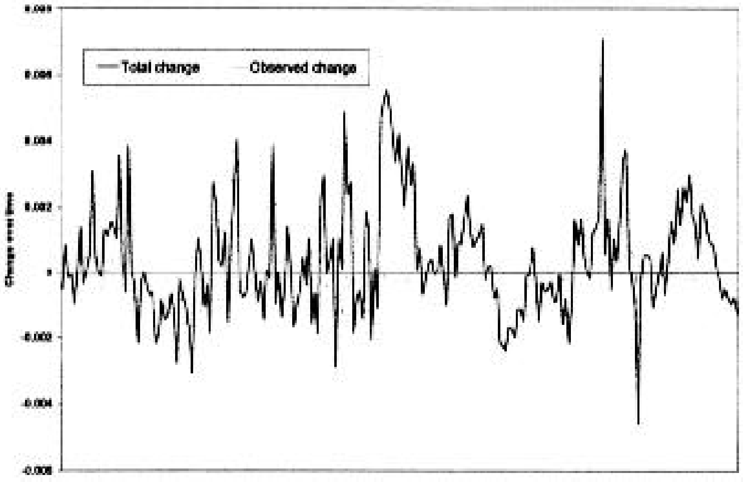

Figure 5- Comparison of the Observed between Industry change and the Maximum Theil change when the Levels of Within Group Theil are at their Maximum.

Again, discrepancies exist, namely when the observed change is of a different sign of the maximum possible change, but the overall pattern of evolution of the two series looks quite similar.

In essence, we have built a framework of analysis to account for the "measure of our ignorance" whenever only the between group component of a measure of inequality is used. Our aim conclusions, the aim was to show how this framework could be implemented, and to illustrate this with real data. Further refinements of our analysis should include explicit consideration of more plausible within group distributions, rather than the extreme cases we considered, which are unlikely to ever be found in practice.

From a conceptual point of view, how important is the within group component in practice? We believe that when the underlying data set is drawn from industrial classification schemes, the answer will generally be "not very important." Industrial classification schemes, after all, are designed to group together entities that are comprised of firms engaged in similar lines of work, and firms, like all bureaucracies, tend to maintain their internal relative pay structures comparatively stable from one period to the next. Thus, the within-group variation of inequality will never approach the extreme case in which all the income moves from equal distribution to concentration in a single individual, the example of Michael Milken in the last years of Drexel Burnham Lambert to the contrary notwithstanding. For this reason, we remain convinced that in practice the effect of the unobservable component on the evolution of T will generally be very small. So long as the group structure is sufficiently disaggregated so that changing population shares and wage shares are not likely to dominate, the movement of T¢ will closely approximate the movement of T. And it is obvious that as one moves toward a finer classification scheme, T¢ must necessarily converge toward T.

To complete the analysis proposed in section 4, we must isolate the population effects. This we will do by fixing the employment structure to the beginning of the period. This way we will compute a $T_{F¢}$, which gives the evolution of inequality under the hypothetical situation of no changes in the structure of employment. Figure 6 presents the results.

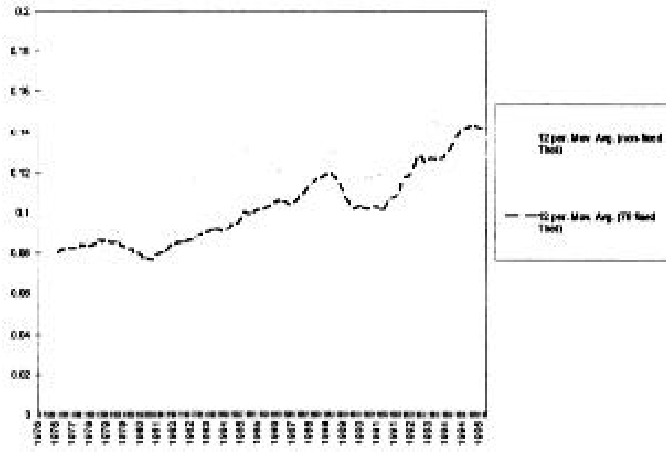

Figure 6- Fixed and Variable $T^{/}$ **for Brazil.**

Figure 6 shows the power of this simple procedure. Changes in $T_F^{/}$ follow changes in $T^{/}$ during most of the period under analysis, showing that wage changes have driven most of the dynamic behavior of inequality. However, between 1982 and 1986 there is a clear discrepancy. $T_F^{/}$ rose steadily, at what looks like a constant rate of change, but our measured $T^{/}$ rose much more sharply between 1982 and 1984, and even decreased between 1984 and 1986. From then on, the changes are very similar. Therefore, we can argue that during the 1982-1986 period, but particularly between 1984 and 1986, changes in $T^{/}$ were driven by changes in the employment structure, rather than by changes in wages.

We have shown how it is possible to study the dynamics of inequality using Theil indexes calculated from industrial wage data. The structure of the data required is extremely simple, basically, only employment and wages or earnings by sector are required. The wide availability of such data for many countries over long periods of time opens new possibilities for the analysis of inequality dynamics.

REFERENCES

Calmon, P., Conceição, P., Galbraith, J. K., Garza Cantu, V., Hibert, A. (forthcoming). "The Evolution of Industrial Wage Inequality in Mexico and Brazil: A Comparative Study", *Review of Development Economics*.

Deininger, K., Squire, L. (1996). "New Ways of Looking at Old Issues: Inequality and Growth", mimeo, The World Bank.

Shannon, C. E. (1948). "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27: 379-423.

Theil, H. (1967). *Economics and Information Theory*. Chicago: Rand McNally and Company.