ASSESSMENT OF THE RELATIONSHIP BETWEEN INCOME INEQUALITY AND ECONOMIC GROWTH :*

A Panel data analysis across the 32 Federal Entities of México 1960-2002

(Version August 2003)

By Araceli Ortega-Díaz

Abstract

The relationship between income distribution and economic growth has been found to depend on several factors such as capital markets imperfections, moral hazard, indivisibility in investments, and existence of dual economic characteristics. In recent literature the importance of geography has been emphasized in defining this relationship due to it's relevance to trade and openness. The current work assesses how income inequality influences growth estimating a reduced form growth equation. Using dynamic panel data analysis for the 32 States of Mexico with both, urban personal income for grouped data and household income from national surveys, it is found that inequality and growth are positively related. When analysing different periods, two different relationships emerge: 1) a negative influence of inequality on growth in a period of low trade policies, and a positive influence in a period more open trade, when urban personal income is considered, and 2) the relationship is reversed when monetary household income is used. To complete the research, we also estimated a structural form equation taking into account the fiscal effects of inequality on growth, finding that the relationship is positive but requires improvement of the explanatory variables involved.[1]

[1] I would like to thank the participants of the Fourth Summer School: Polarization And Conflict /XX Summer Courses / XIII European Courses in Basque Country University, also the participants of the IESG 26th Annual Conference: Trade Liberalisation, Growth & Poverty at the University of Wales Conference Centre, Gregynog; and the participants of I Mediterranean Summer School in Theoretical and Applied Economics: Inequality, Welfare and Redistribution at the Universitat de les Illes Balears for their invaluable comments. I would like to thank to Professor John Black for his help on language usage corrections in the first version of this paper. My sincere thanks to all academic staff at UNU-WIDER for their excellent comments, specially to Professor Tony Shorrocks for an excellent prompted and discussed in page 24. Any error in this piece of work is my own.

* This paper represents one chapter of the dissertation towards a PhD in Economics "Inequality and Growth in Mexico".

ASSESSMENT OF THE RELATIONSHIP BETWEEN INCOME INEQUALITY AND ECONOMIC GROWTH

(An analysis across the 32 Federal Entities of México 1960-2002)

Section I. Introduction

The relation between economic growth and income distribution is still a controversial topic. When making economic policies, government is interested in economic growth, taking this as a way of increasing economic welfare[2]. But in doing so it faces the problem of increasing income inequality[3], causing a reduction in economic welfare. However, if government targets reducing income inequality as a way of improving welfare[4], economic growth may slow down, leading to welfare loss. This dilemma between income inequality and economic growth has prompted many researchers to explore sources of income inequality, and the channels through which inequality affects economic growth.

On one hand, economic theory suggests that the relation between income distribution and growth differs according to the economic context (market settings). On the other hand, empirics suggest that divergence in results comes from different data quality, period length, omitted variable bias, or even the econometric technique used[5].

---

[2] A kind of trickle down effect like the one described in Aghion and Bolton (1997).

[3] By pushing the skill premium up, creating more inequality like in Aghion, Philippe and Williamson (1998).

[4] Garcia R.,A (1986) provides several definitions of welfare according to different schools of economic thought.

[5] A recent article, Panizza (2002), show that results under panel data estimations do not necessarily turn out to find a positive relationship between economic growth and inequality, as was thought. Meaning that technique does not necessarily influence the results.

But, do results have to be the same for all cases? Does a positive or negative relationship have to hold for every case? Obviously not, and thinking that the relationship has to be the same for every case, is what makes the existing literature look contradictory.

Previous works about income distribution and economic growth have limited their scope, most of them just analyse income distribution with pure times series, at a National level, few of them do in fact analyse a panel for every State using micro data.[6]

Analysing the results of previous literature, we have observed that the reason why the findings vary is that there is not a conceptual framework, which clearly identifies the characteristics of the model we would be interested in analysing under a particular socio-economic scenario. Relationship among countries or within a country, developed or underdeveloped countries, perfect or imperfect capital markets, agents' skilled level, particular characteristics of economic situation (financial crisis, trade openness, fiscal reforms and others), i.e., there is a lack of an agreed starting point.

Therefore, it is not true that income inequality affecting economic growth in a negative (positive) way is the definitive rule; the truth is that income inequality effects on growth will depend on the kind of economy (ies) we are analysing.

[6] In the case of Mexico this can be verify for example in works like the one of Székely (1996) and Esquivel (1999). Both authors explained how they have to correct their time series in order to make them comparable. Although their techniques are not optimal as they involved escalating to National Account (for the first author), they intrinsically show the necessity of providing comparable and State level data sets.

Some researchers have chosen to analyse how economic growth influences income inequality following Kuznets' hypothesis, which states that in the first stages of a country's development, inequality increases as the national economy grows until it reaches a peak. Then, in the later stages of growth, income inequality decreases, this economic path describes an inverted "U" shape. Other researchers have chosen to look at the influence of income inequality on growth without reaching a consensus yet (this literature is vast). A third type of research is on the lines of Quah (1997); he states that it is not that inequality influences growth or vice versa, but that interactions between both have to be analysed simultaneously.

The current work follows the second approach: how income inequality influences economic growth. We focus on analysing this relationship across the 32 States that constitute Mexico. To my knowledge, this kind of work is the first based in that country and contributes to the country case studies of the relationship of income inequality and growth as described by Kanbur (1996).

The present study is structured as follows. Section II defines the economic factors we should take into account when analysing the relationship between inequality and growth. In section III we give a succinct motivation for have chosen Mexico as a country case study. In section IV we define the model to be used, which is mainly that used previously by Forbes (2000), based on Perotti's (1996) reduced form equation, but using panel data methods for the 32 States of Mexico. The fifth section explains the data to be used, sections VI-VIII present the estimation of the model as well as some sensitivity analysis using personal income. In section IX we changed the database from personal income to household income and present the re-estimation of the model. Section X briefly uses Perotti's structural form with a fiscal approach. Section XI concludes.

Section II. Defining a Framework

Through the analysis of several models that provide theoretical as well as empirical findings about the sign of this relationship, we have identified two models as base models, from which other researchers depart and in one way or another relate income distribution and economic growth. 1) Loury's (1981) model of intergenerational income distribution, whose most generalised is found in Mookherjee & Ray (2000). These kinds of models are purely theoretical and applying them requires highly stylised numerical problem solving to find multiple equilibria. 2) Solow (1956) growth model, but modified, such that now we can use a more general endogenous growth model, like the enclosed growth model of Hammond & Rodriguez-Clare (1993).

There is a huge amount of literature of this kind of model, which results, as we explain before, seem to be contradictory. In particular, within these kinds of models, we will use Perotti (1993) and Forbes (2000).

From these models emerge series of variants, changes or mixtures. In first place, looking at the empirical approach, we are interested in knowing how low/high income countries perform through time.

According to Barro & Sala-i-Martin (1992), in a closed economy, where States have similar preferences and technologies, poor states grow faster in per capita terms than rich states, independently of the change in production composition through time. Thus the neoclassical standard model of economic growth with technological progress in a closed economy, will always predict convergence. And as Aghion (1998) points out, the convergence model assumes perfect capital markets, so that convergence results may not hold for developing countries.

A step further is to look at income inequality and the economic performance. For instance, Quah (1997) found that assuming that each country / state has an egalitarian income distribution, their income dynamics across countries / states will show stratification, persistence or convergence. The latter as well as their economic growth will depend on their spatial location, the countries with which they trade, among other factors.

Moreover, if we look at income inequality and income mobility, would inequality matter for economic growth? According to Loury (1981), we should distinguish between social (income) mobility across generations, from that within the same generation. In his model, we can distinguish four combinations of this type:

Table 0. Distribution and Mobility

Income with:	Distribution with:
1. High mobility	Low dispersion.
2. Low mobility	High dispersion.
3. High mobility	High dispersion.
4. Low mobility	Low dispersion.

If dispersion in income in an economy is high but mobility (probability of passing from one income class to another) is also high, then, in the long run, initial income would not matter. Alternatively, if the dispersion is low but mobility is also low then inequality although low will be persistent.

Can we find from here a link between long-term inequality and economic growth? Loury's results arise from the assumption that there is not a loan contract market to invest in human capital, and families within the same generation cannot lend to or borrow from each other.

Galor & Zeira (1993) and Banerjee & Newman (1993) found that under restrictions on borrowing to invest in human/physical capital, income distribution polarizes (into rich and poor for Galor & Zeira) or divides into classes with no mobility out of poverty (Banerjee & Newman), where economic growth will depend on initial conditions, such that the economy can be prosperous or lead to stagnation.

For example, following Aghion & Williamson (1998) and Ljunqvist (1993), a dual economy that among other characteristics has a high number of unskilled agents compared to the number of skilled ones, and where capital markets are imperfect, a high income inequality will lead to a low product[7] level and therefore to a low economic growth. In other words, inequality and growth are negatively related.

[7] Higher income is related to higher effort and a higher probability of success in production. So due to market imperfections, some agents are not able to borrow remaining with low income.

On the other hand in economies where markets are perfect[8], but there is a high concentration of unskilled agents and a high income inequality, it can be attractive for a firm that needs non-specialized labour to invest in those economies. Because markets are perfect, agents invest the same amount and put the same effort, and then can generate more product and economic growth. In other words, inequality and growth are positively related. Whether this relationship holds in the long term is another story.

These are only two examples, but we can define several frameworks to have an a priori idea of what kind of relationship might turn up. An interesting exercise is to analyse what kind of relationship we would get, applying some of the findings to a particular scenario. We do this by using data for the 32 states of Mexico.

Section III. Motivation.

Why the 32 States of Mexico? The answer is easy. We perform the analysis across the 32 states because they exhibit a high inequality in welfare, in GSP[9] growth, in allocation of resources, in shares of government income and expenditure[10]. There are conflicts of redistribution among them: a rich North contrasting with a poor south.

There are several articles that can be found about income inequality and growth in Mexico, but their analysis is at national level (pure time series). Some others analysing convergence in Mexico have regional data but not on inequality.

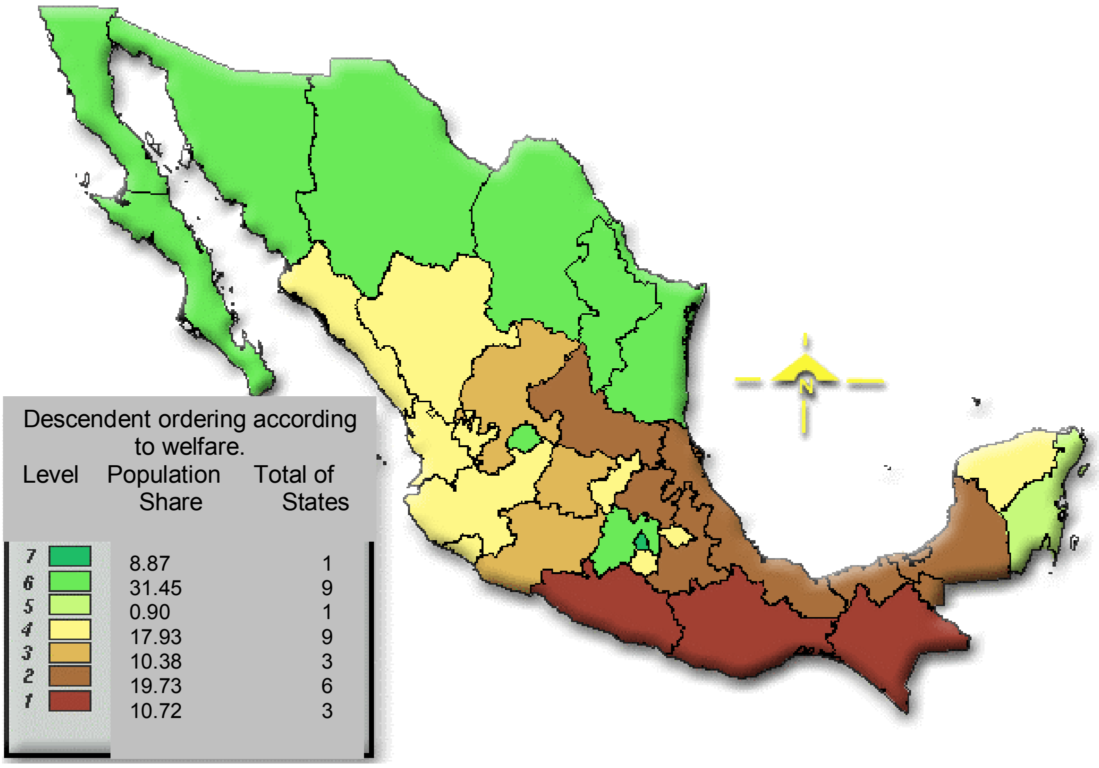
---

[8] All individuals are able to invest the same amount of capital, irrespective of the initial distribution of wealth.

[9] Gross State Product.

[10] As Sempere & Sobarzo (1998) point out, currently in Mexico, there is not a Federal Fiscal policy that relates income with expenditure for each State, a political discretional expenditure has been the main rule.

Figure 1. MEXICO
(Welfare Groups)



Original Map taken from INEGI. Current Map changed by author.

One of the main reasons is that the National Income and Expenditure of Households Survey (ENIGH) by INEGI is available only from 1984 to 2002 (eight surveys in total). If we were to analyse only this period, we would lose some important economic changes. Therefore, the current work uses two surveys: a) personal income from employment ad income surveys from 1960 to 2000, on a decade base and b) household income survey for period 1984-2002[11] using ENIGH. It is original in that it

[11] ENIGH-2002 survey has been released just recently (May 2003), in section IX we are presenting the inequality measure calculated with it, but it cannot be included on the regression, as it ill be seen, we would have to regress 2004 growth on 2002 inequality. We are in year 2003, and 2002 GDP is the most recent estimate at national level.

analyses the relation between income inequality and economic growth at a State or Federal Entity level (see map) and compares the results obtained from each survey.

Section IV. Model

As have been discussed, several economists still doubt if the relation between growth and inequality is inverse (negative); recent work like the one of Forbes (2000) has found that the relation between these two macroeconomic variables is positive.

Works previous to Forbes conclude that economic growth and inequality are negatively correlated: Benabou (1996) provided a large number of articles that concluded so, but only 43% of them are consistent and significant. Kristin Forbes (2000) argues that reading carefully all those articles, they suggest that this relation is not definite and depends on several assumptions and external factors to the models. Even more, according to Forbes, they all present three potential problems in the empirical work performed:

1) Not all results are robust when their sensitivity is analysed by introducing different explanatory variables, or dummy variables (region, sector, etc) sometimes the inequality coefficient becomes insignificant. For example, when Perotti (1996) introduced POP65[12] as one more explanatory variable, the coefficient of inequality decreased significantly[13].

2) a) Problems of measurement error on inequality. It is known that in developing countries, a national household survey is not always available;

[12] POP65 = share of the population aged 65 or more.

[13] We will find the same result when analysing the structural form in section VIII.

therefore, institutions tend to overestimate several figures. One of the measures subject to high measurement error is the income inequality index.

b) Another big problem is the biased caused by the omitted variables in the estimations.

3) Most of the estimations are only of cross section. It is necessary to use panel data methods to detect how changes in inequality affect changes in the growth rate. The latter has the advantage of correcting the bias caused by the omitted variables as well as taking into account problems of measurement error.

In section VI we will use panel data methods in order to overcome these three problems. Although researchers like Quah (1997) do not use panel data, arguing that estimation via panel eliminates heterogeneity and makes it not possible to appreciate the income dynamics, panel data helps to reduce the bias caused by omitted variables but does not control for omitted variable whose value changes over time. In spite of this, estimations are richer in information with panel than with a cross section analysis.

Following Forbes we start analysing a similar equation to Perotti's reduced form equation[14], where we will be looking at influence of income inequality (measured by the GINI coefficient of income) on growth (measured by the Gross State Product per capita growth rate) plus the influence of the human capital variable (measured by

---

14 Perotti reduce equation:

$$Growth_{it} = \beta_1 MID_{i,t-1} + \beta_2 GDP_{i,t-1} + \beta_3 MSE_{i,t-1} + \beta_4 FSE + \beta_5 PPPI_{i,t-1} + e_{it}$$

where MID is a measure of inequality. MSE and FSE are the average years of schooling for male and females respectively. PPPI is the investment cost of investing in a country with respect to USA, in order to capture the socio-political instability of the country. And e is the error term.

years of schooling or by the literacy[15] rate when available) and some dummy variables for each State in order to control for time-invariant omitted variables bias effect and the time dummies to control for aggregate shocks that can not be explained by the explanatory variables. The reduce form equation is:

$$Growth_{it} = \beta_1 Income_{i,t-1} + \beta_2 Inequality_{i,t-1} + \beta_3 Schooling_{t-1} + \alpha_i + \eta_t + u_{it} \quad \textbf{(1)}$$

Where i = 1,…, 32 is the panel variable and, t=1,…,5 is the time variable. α_i are country dummies which can be interpreted as the unobservable individual effect, n_t are period dummies denoting unobserved time effect and u_{it_i} is the remainder stochastic disturbance term.

As an alternative model when data is available, we will use.

$$Growth_{it} = \beta_1 Income_{i,t-1} + \beta_2 Inequality_{i,t-1} + \beta_3 Male\_literacy_{i,t-1} + \beta_4 Female\_literacy_{i,t-1} + \alpha_i + \eta_t + u_{it} \quad (2)$$

We can include more explanatory variables; however we are limited by the regional information considered. Due to income data availability, we are restricted to 1960-1999 having four ten-year-periods. Although data was available for 1960, 1970, 1980, 1990, 1995, 2000, we decided to drop 1995 for two reasons.

[15]Literacy is the percentage of the population aged 10 or more who is able to read and write. We are expecting this coefficient to be positive.

1. Including 1995 would make our ten-year periods non homogeneous and we will not be able to apply the Arellano and Bond Technique.
2. Year 1995 was a year of severe recession for Mexico.[16]

From the explanatory variables included, we expect β_1 to be negative so that results are consistent with the convergence theory. We would expect β_3 and β_4 to be positive[17], according to the theory mentioned in section II explaining that human capital accumulation is beneficial for growth. We are interested in sign of β_2.

Section V. Data[18]

Schooling comes from the Ministry of Public Education (SEP) and is defined as the number of years of schooling.

Literacy will be used as an alternative definition of human capital, which defines the proportion of the population who can read and write, for females and males separately, this data comes from INEGI for sections V-VII and IX. Data comes from ENIGH for section VIII.

Growth: Unfortunately there is not a complete series for the Gross State Product per capita (GSP) at constant prices. Calculation methodology has varied over time, and last publication presents Gross State Product at 1993 constant prices, only from 1993 and onwards. The deflator for Gross National Product at 1993 prices was available so we applied it to the Gross State Product at current prices series previous to 1993

[16] Data come from several sources (see section V).

[17] Barro(1996) mention to have found a negative coefficient for male literacy in some cases.

[18] For more details see Appendix C.

weighted by their respective share over time. Then we weighted it by its corresponding population obtained from INEGI's population census to obtain the Gross State Product per capita (GSP).

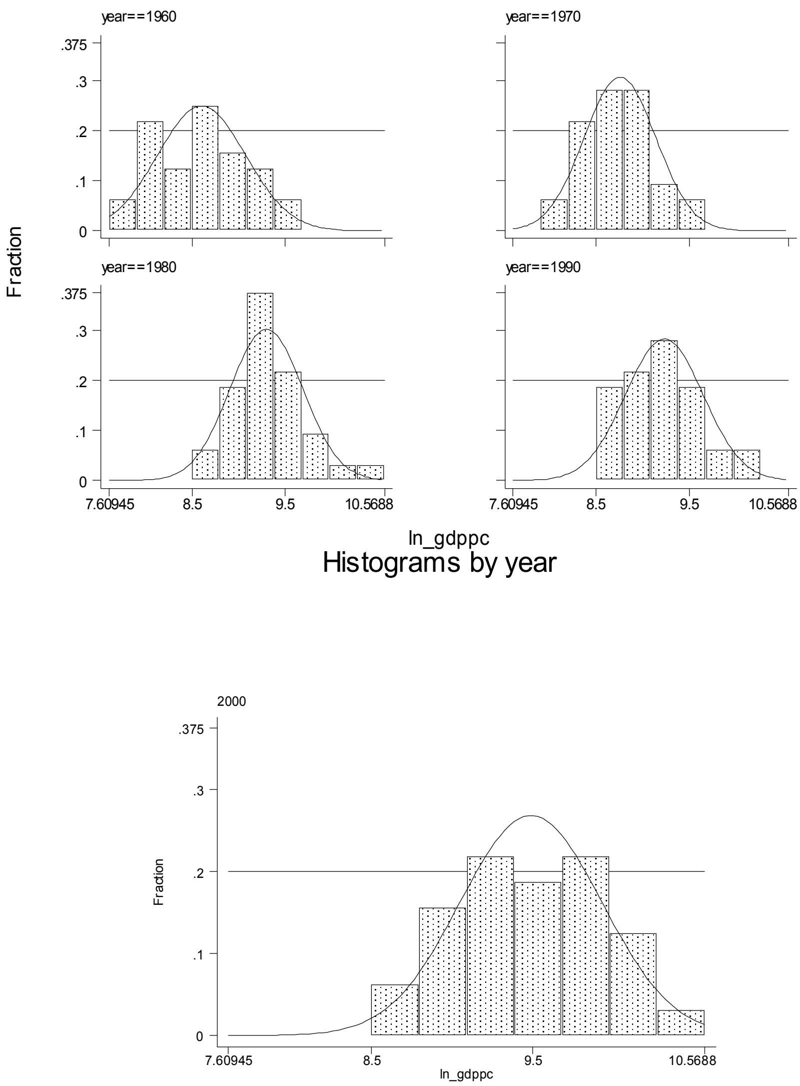
Alternatively, we use GSP series at 1995 constant prices calculated by G. Esquivel (1999) when he faced the same problem on data availability and that he used in his calculations on regional convergence for Mexico.

It has to be emphasized that it is very important to look at the nature of the income and GSP data we are using in the calculation of equation (1). We shall remember that β_1 coefficient is the coefficient of conditional convergence and as we can see in the set of graphs below the natural log(GSP) has been spreading over time, but the way it has spread is what will determine if β_1 will be high, low or not significant, i.e. GSP can spread forming two polarised groups where the GSP of the States within a particular group are converging whereas the GSP between the two polarised groups diverges.

The following is not clearly seen using just histograms like the ones below, but it will be clearer when using kernel estimation.

In our case, we can see that the percentage of States in the middle-income group (around the mean) has decreased, and the percentage of the States below the mean has considerably increased. The problem of using histograms to visualize distributions is that they are bin-width dependent. Changes can be more evident if we use kernel estimation to describe the distribution of income across States. Graphs of kernels provide us with a better look at the changes in the relative concentration of States at each income level. We would later compare these changes over time with β_1 and its consistency with β_2.

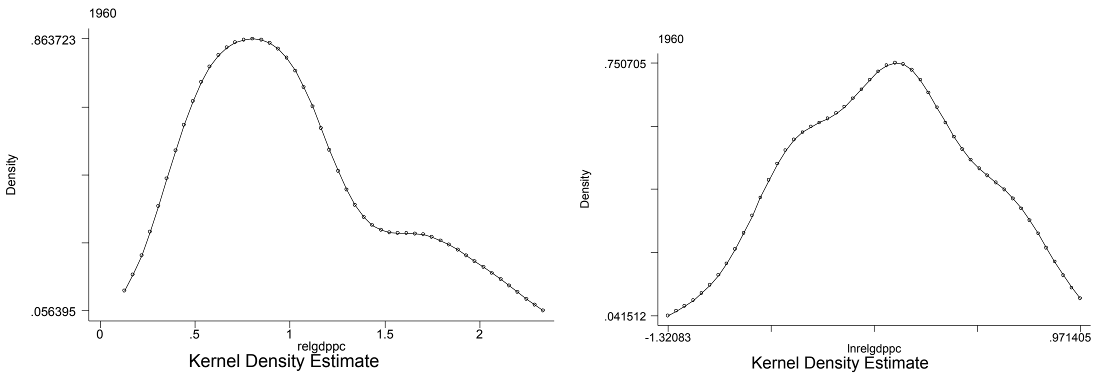
Figure 2. Distribution of the natural log of GSP per capita across 32 States



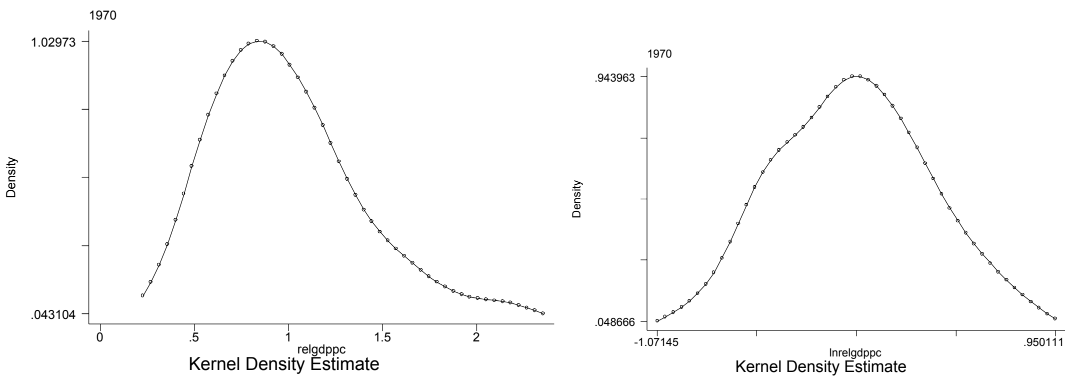
As it is evident in the kernel, the distribution of GSP across states has spread out. It is worth pointing out that the distribution on the natural log of the relative GSP per capita for 2000 exhibits what Quah (1997) describes as a twin peak density. This type of distribution may imply polarization and we will see that divergence instead of convergence is expected for this period.

Figure 3. Densities of Relative GSP per capita across 32 States (left) and densities of natural log relative GSP per capita income across 32 States (right)

1960



1970



1980

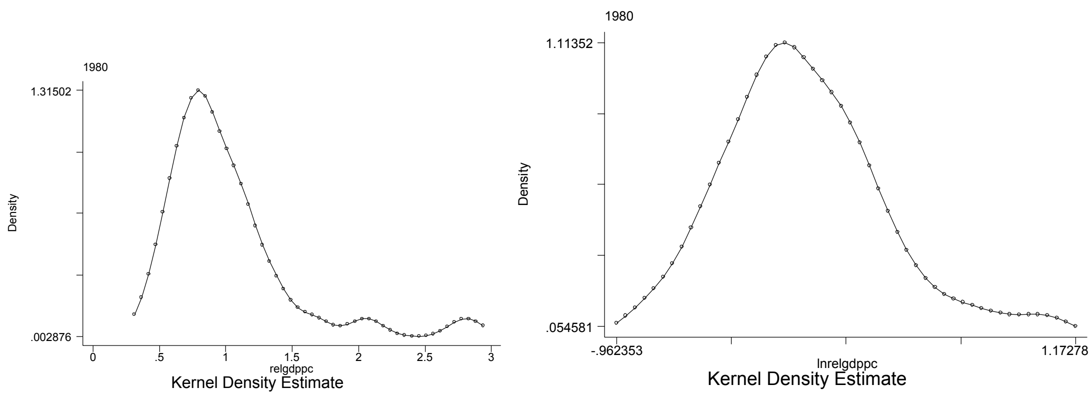
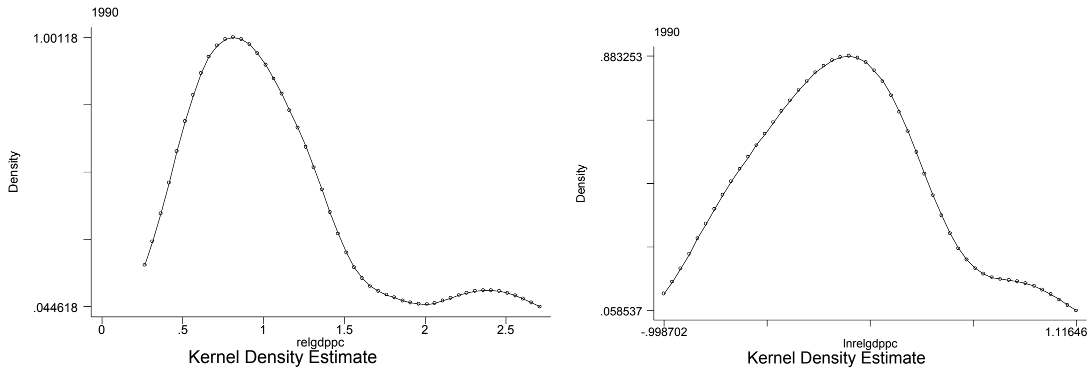
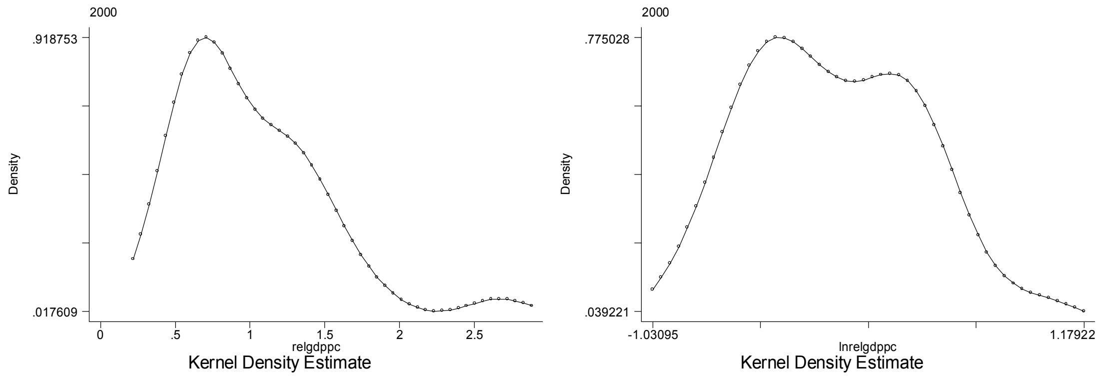


Figure 3 (cont.) Densities of Relative GSP per capita across 32 States (left) and densities of log relative GSP per capita income across 32 States (right)

1990



2000



<u>Inequality</u>: National household surveys are not available as such for each State for the whole period we are analysing 1960-2000; they exist only for the period 1984-2002. Therefore, in the current work, inequality for sections VI to VIII and X was calculated with the Gini coefficient using income per person from the following surveys.

For 1960 we analysed the 1956 Mexican Population Income and Expenditure survey of the Secretary of Economy, the 1958 Mexican Population Income and Expenditure survey of the Secretary of Industry and Trade, and the 1960 Sixteen principal

Mexican Cities' Income and Expenditure survey also from the Secretary of Industry and Trade. All of them available from INEGI's archives.

For 1970 we used the 1968-70 Mexican Population Income and Expenditure survey of the Secretary of Industry.

For 1980 we used the 1977 Income and Expenditure of Household survey of the Secretary of Budgeting and Planning.

For 1990 and 2000 we use National Employment surveys from INEGI.

One of the problems faced was that the number of income groups of data for each survey is different and standardising data loses some of the intergroup inequality we should account for. We used the method of Yitzhaki & Leman to compute the Gini. In the sensitivity analysis section we also present the results using the Gini coefficient calculated POVCAL[19] program.

A second problem is survey comparability. Given that every survey was made using relatively different methodology (mostly 1960 survey), some researchers have escalate to national accounts to correct the problem and be able to use the data. In our case, when trying to escalate to national accounts we obtained Ginis from the magnitude of 70% which were considered as erroneous. Therefore, in the present work, we use the grouped data without escalating to national accounts.

[19] POVCAL was developed by Chen-Datt-Ravallion at the World bank to calculate poverty and inequality measures using grouped data.

For section IX, we observed a similar problem with the Federal Entity representatively of the ENIGH surveys. Using the expansion factors provided in the ENIGH surveys, the Gini coefficients obtained were also from the magnitudes of 70%. Similarly, we decided to use only the Ginis calculated without using expansion factors. Calculation were performed with and without using expansion factors, but only the former are reported in this work, the later are available form author on request.

The summary statistics of the set of data used in sections VI-VIII and X is described below.

Table 1—Summary Statistics for Sections V-VII and IX[20]

Variable	Definition	Source	Year	Mean	Std. Dev.	Min	Max
Schooling	Average years of schooling of the population.	SEP	1960	2.46	0.918	1.0	5.0
			1970	3.19	0.897	1.8	5.8
			1980	4.31	0.955	2.5	7.0
			1990	6.29	1.007	4.2	8.8
			2000	7.53	1.001	5.7	10.2
Income	Ln of Real GSP per capita in 1993 pesos. Correcting with national deflator before 1990.	INEGI	1960	8.60	0.474	7.60	9.46
			1970	8.75	0.385	7.93	9.60
			1980	9.29	0.390	8.56	10.40
			1990	9.23	0.417	8.53	10.16
			2000	9.49	0.439	8.71	10.56
Inequality	Inequality measured by the Gini Coefficient using Leman and Yitzhaki formula.	SE (1960), SIC (1970) & INEGI (1980-2000)	1960	0.387	0.0599	0.20	0.47
			1970	0.430	0.0630	0.32	0.57
			1980	0.456	0.0394	0.40	0.54
			1990	0.379	0.0297	0.34	0.48
			2000	0.413	0.0385	0.34	0.51
Female Literacy	Share of the female population aged over 15 (10) who can read and write.	INEGI	1960	63.99	15.83	34.93	85.589
			1970	73.22	12.05	50.38	87.928
			1980	79.61	10.91	54.94	92.284
			1990	85.09	8.743	62.35	94.529
			2000	88.70	6.906	69.95	96.080
Male Literacy	Share of the male population aged over 15 (10) who can read and write.	INEGI	1960	70.84	12.26	44.87	92.175
			1970	79.06	8.75	59.66	94.318
			1980	85.52	7.21	68.94	96.890
			1990	89.95	5.15	77.52	97.872
			2000	92.11	4.02	82.86	98.260
Income2	Ln of Real GSP per capita In 1995 pesos. Correcting for 2000.	G. Esquivel	1960	9.071	0.445	8.328	10.050
			1970	9.464	0.465	8.562	10.385
			1980	9.770	0.438	8.952	10.653
			1990	9.776	0.444	9.043	10.841
			2000	9.791	0.416	9.016	10.808

[20] When estimating the model in sections V-IX we will use the natural logarithm of the explanatory variables, this has the advantage of providing the elasticity directly from the estimated coefficient. We also verified that the sign of the coefficients is still the same with or without natural logarithm, as well as the significance of the coefficients under both estimations did not alter our results. We will only report the results when using natural logarithm , but the rest of the results are available from the author on request.

For section IX inequality was calculated using the ENIGH (National Survey of Income and Expenditure of the Households) surveys for 1984-2000[21]. For each of the 32 States, we computed the quarterly monetary income which is the result of the sum of six different sources of income: 1) earned income, 2) self-employment income, 3) property rents income, 4) Income from cooperatives, 5) Transferences, 6) Other sources of monetary income. [22]. This data set is more representative of each State compared with the previous set, as it includes a bigger pool of agents from high and low density cities.

The summary statistics for section VIII is provided at the beginning of that section in order to avoid misunderstanding, and with the aim of outlining the difference in income data sources

Section VI. Estimation

Following Forbes (2000), there are three factors to be considered to estimate in the best way equation (1)

1. The relation between the state-specific effect and the regressors.
2. The presence of a lagged endogenous variable (income).
3. The potential endogeneity of other regressors.

[21] ENIGH survey has been release just recently (May 2003), we are presenting the inequality calculated with it in section IX, but it cannot be included on the regression, as it ill be seen, we would have to regress 2004 growth on 2002 inequality. We are in year 2003, and 2002 GDP is the most recent estimate at national level.

[22] For a detailed list of the Sources see ENIGH, or alternatively Garcia R. (1999). For details on the procedure of sorting the data see appendix C.

Previous researchers estimated equation (1) via OLS for a unique long period, without controlling for the three factors above described[23], and they found that the relation between inequality and growth is negative. We will instead use panel data (when observations allow for it). If we use panel data, we have to choose how we will perform the estimations: using fixed effects or random effects estimators. Even though random effect estimation is more efficient, it may not be suitable for our estimation across the 32 States. The estimation of the model is complex given that there is a lagged endogenous variable in the model[24]. Given that growth is defined as the difference in logarithms: $growth_{it} = Income_{it} -_2 Income_{i,t-1}$, where $Income_{it}$ is the logarithm of the income variable for State i at time t. We can visualise the lagged endogenous variable if we rewrite the equation (1) as:

$$Income_{it} - Income_{i,t-1} = \beta_1 Income_{i,t-1} + \beta_2 Inequality_{i,t-1} + \beta_3 Schooling_{i,t-1} + \alpha_i + \eta_t + u_{it} \quad (3)$$

$$Income_{it} = \gamma_1 Income_{i,t-1} + \beta_2 Inequality_{i,t-1} + \beta_3 Schooling_{i,t-1} + \alpha_i + \eta_t + u_{it}. \quad (4)$$

where $\gamma_1 = 1 + \beta_1$.

In matrix notation, it is equivalent to writing:

$$y_{it} = \gamma\, y_{i,t-1} + X'_{i,t-1} B + \alpha_i + \eta_t + u_{it} \quad (5)$$

[23] Perotti did not do it arguing that it was difficult to instrument for the income variable.

[24] It is worth noticing that Barro & Sala-i-Martin never controlled for this kind of effect.

Even if y_{it} and u_{it} are not correlated between them, but t does not go to infinity, estimation with fixed effects or random effects will be inconsistent[25]. Nevertheless our option is to use the GMM estimator of Arellano-Bond which can deliver a consistent estimator for t≥3.

Arellano-Bond Technique

Arellano and Bond method makes use of the generalised method of moments, taking first differences of each variable to eliminate the country-specific effect, and then uses the lag of each variable as instruments. This method not only corrects the bias caused by the lagged endogenous variable, but also allows certain degree of endogeneity to the other regressors.

$$y_{it} - y_{i,t-1} = \gamma \ (y_{i,t-1} - y_{i,t-2}) + (X'_{i,t-1} - X'_{i,t-2})B + (u_{it} - u_{it-1}) \qquad (6)$$

Table 2 shows the results when estimating equation (1) with fixed effects, random effects, and Arellano Bond method, with and without year-dummy variables.

In choosing the best model we have performed a Hausman test to detect that the state-specific effects are not correlated with the regressors. The test shows that Random Effect estimator is always rejected in favour of Fixed Effects. However, given that FE

25 For this estimator to be consistent, it must be satisfied that:

$X'_{i,t-s}$ is predetermined at least for one period: $E(X'_{i,t}\ u_{i,s}) = 0$ for every s>t.

Error cannot be correlated $E(u_{it}u_{it-s}) = 0$ for any s $\geq$ 1. See Appendix A.

See Arellano, M. and Bond S. (1991).

is inconsistent when n is small, like in our case[26], we would stick to the only valid estimation that is GMM estimator of Arellano and Bond (column 5), but will provide the result with FE estimator to have a reference of comparability.

Table 2. Regression Results: Alternative Estimation Techniques

Estimation Method	Fixed Effect (1)	Random Effect (2)	Fixed Effect with Year Dummies (3)	Random effects with Year Dummies (4)	Arellano and Bond (5)	Arellano and Bond With year dummies (6)
Income	- .0968172** (.0111083)	-.0461308** (.0076212)	-.0880741** (.0108487)	- .0334504** (.0067694)	-.3318917** (.1316786)	.0976317 (.1571716)
Inequality	.0053129 (.0195199)	-.0197478 (.0171346)	-.012119 (.01836)	-.0162466 (.0157563)	.395237** (.1920085)	-.0265872 (.2103216)
Schooling	0565433** (.0096834)	.0316067** (.0085654)	-.015377 (.0190816)	.0307657** (.0104152)	-.0724878 (.2437229)	-.0739707 (.2159668)
Dummy 70-80	--	--	.0579678** (.0073311)	.0365021** (.006182)	--	--
Dummy 80-90	--	--	.0506136** (.013916)	-.0145821* (.0077672)	--	-.6428542** (.1029075)
Dummy 90-00	--	--	.0808449** (.0205317)	.0004934** (.009449)	--	-.9056076** (.1519488)
R-squared	0.4725	0.2489	0.7200	0.5680	--	--
States	32	32	32	32	32	32
Observations	128	128	128	128	96	96
Period	1960-2000	1960-2000	1960-2000	1960-2000	1980-2000	1980-2000
Hausman Test	chi2(3)= Prob>chi2=	51.50 0	chi2(6)= Prob>chi2 =	49.75 0.00	--	--
Sargan Test	--	--	--	--	chi2(5)=32.18 Prob>chi2=0.00	chi2(5) = 8.03 Prob>chi2=0.15
A&B acov res 1st	--	--	--	--	z =-2.85 Pr>z = 0.0043z	z = -3.68 Pr > z = 0.0002
A&B acov res 2nd	--	--	--	--	z =-2.40 Pr>z = 0.0162	z = -0.34 Pr > z = 0.7317

Notes: Dependent variable is average annual per capita growth. Standard errors are in parentheses. R-squared is the within R-squared for the fixed effects model and the overall R-squared for random effects. ** significant at 1%. * significant at 5%.

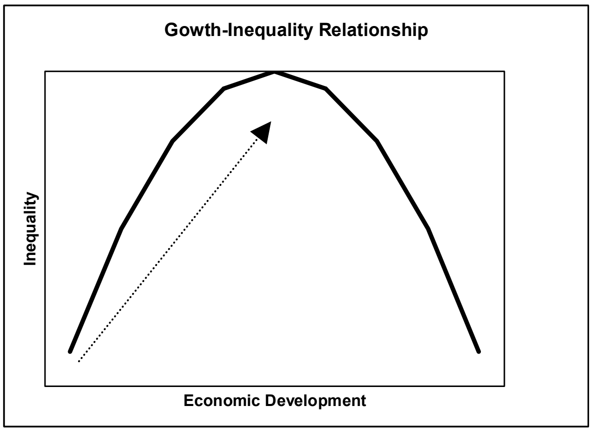
The results show that the relation between inequality and growth is positive for both FE and A&B, but and only significant for Arellano-Bond estimation.

This result implies that in the short run (considering periods of ten years each) positive changes in lagged inequality (i.e. Gini coefficient increasing) are associated with positive changes in natural log GSP (i.e. current GSP growth). Does this provide evidence of the initial part of Kuznets curve? where in the first stages of a country

[26] See Appendix A.

economic development, a positive relationship between inequality and growth is observed.

Figure 4. Kuznets' curve[27]



The arrow shows that our finding refers only to the upward part of the curve.

Kuznets' hypothesis answers the question of how during the stages of economic development of a country inequality changes, but not vice versa, i.e., does not answer the question of how income inequality influences growth.

Therefore, an important question arises, does the positive relationship found between income inequality and growth, means that we have to create more inequality in order to achieve higher rates of growth? [28]

Before answering this question, we have to address the following questions: is it only the method of estimation that makes the relation between growth and inequality differ from other results? Or are there other factors that need to be considered?

[27] See Kuznets (1955).

[28] I would like to thank Professor Tony Shorrocks for prompting at me this question during a UNU/WIDER seminar. This helped me to go back to Income Inequality and Growth literature, and found that indeed this question has not been answered. The related literature get stack into the problem of addressing the relationship, but have not reach the point where they are able to explain it, see Forbes (2000), Banerjee & Duflo (1999), Paniza (2002). We would discuss this matter in the conclusions and deepen into it in the following chapter.

Section VII. Factors that affect the coefficient of inequality.

According to previous works, we expect the coefficient of inequality to be affected by several factors, among them, we considered:

1. Data quality (source).
2. Period's coverage and method of estimation.
3. Different definitions of inequality and literacy.

Using the factor variation technique[29] to identify which factor has the greatest effect on the income inequality coefficient, we have found the following.

1. Data quality can vary depending on the source. We performed FE and A&B estimation using an alternative source for the GSP.

Table 3a. Regression Results: What affects the coefficient on Inequality? Source!

Estimation Method	Fixed Effect (1)	Random Effect (2)	Fixed Effect with Year Dummies (3)	Random Effects with Year Dummies (4)	Arellano and Bond (5)	Arellano and Bond With dummies (6)
Income (G. Esq.)	-.0766066** (.0087347)	-.022848** (.0068259)	-.0868894** (.0100194)	-.0335581** (.006998)	.3383265** (.164285)	.1829996 (.181215)
Inequality	.0047376 (.0154536)	-.0313321 .0156458)	-.0035913 (.0186316)	-.0053471 (.016570)	.0730769 (.2046191)	.0391183 (.2215911)
Schooling	.0188611** (.0076411)	-.003805** (.0079754)	.0069347 (.0194182)	.0414773** (.011566)	. 234679 (.2429086)	.2617124 (.229321)
Dummy 70-80	--	--	.0237293** (.0081935)	-.0074633 (.006435)	--	--
Dummy 80-90	--	--	.018501 (.0138272)	-.0397245** (.008174)	--	-.2860239** (.089866)
Dummy 90-00	--	--	.0247518 (.0207772)	-.0476207** (.010060)	--	-.4377743** (.1870335)
R-squared	0.5835	0.2399	0.6359	0.4153	--	--
States	32	32	32	32	32	32
Observations	128	128	128	128	96	96
Hausman Test	chi2(3) = 82.99	Prob>chi2 = 0	chi2(6) = 30.30	Prob>chi2 = 0	--	--
Period	1960-2000	1960-2000	1960-2000	1960-2000	1980-2000	1980-2000
Sargan Test			--	--	chi2(5) = 8.80 Prob>chi2=0.11	chi2(5) = 5.68 Prob>chi2 = 0.3381
A&B acovres 1st	--	--	--	--	z = -4.84 Pr > z = 0.0000	z = -3.74 Pr > z = 0.0002
A&B acovres2nd	--	--	--	--	z = 1.51 Pr > z = 0.1322	z = 0.22 Pr > z = 0.8291

[29] This consists on estimating the above equation changing the definition of the proxy used for one of the variables in the equation. And repeat the process for all the variables to test how robust results are to variable definitions.

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the within R-squared for the fixed effects model and the overall R-squared for random effects. ** significant at 1%. * significant at 5%.

The results show the same sign for the coefficient of inequality whether using my data or Esquivel's data. Signs coincide across estimation techniques, but the significance of the coefficient has changed. It might be the case that Esquivel's data contains different implicit information, as it takes as a base year 1995, which was a year of severe recession in Mexico. Or it can be due to the fact that NT is small. What is important though, is the fact that for the benchmark estimations (i.e. columns 1 and 5), changes in inequality are positively related to changes in growth, and that the data source does not affect the sign in our estimation.

It is worth to remember that through all this work, the only valid estimation for this type of data set is the one where results were obtained using GMM estimator (A&B)30.

Before moving to the next estimation technique, it is important for many researchers to look at any source of implication outliers in the sample may cause. In our case we have spotted three States with a different behaviour compared to the 29 remaining States: Campeche and Tabasco that are the oil producers, and Chiapas which is among the poorer States. They have been treated different in the literature as seen in Esquivel (1999) and Rodriguez-Pose (1999). The following table summarises the outliers and the estimation treating them as different and excluding them.

[30] The only way where FE will be allowed in a data set with this type of N and T, is if we use in the estimation equation a different definition of economic growth. For example TFP.

Table 3b. Outliers per year and explanatory variable

Period	LnGSP	LnEsq	POVCAL	INEQ	Schooling	Literacy
1960-1970	-		15, 13	15, 13	15	15, 23
1970-1980	Oiltab, Chis	Oiltab, Chis	Oiltab, Chis	Oiltab, Chis	Oiltab, Chis	Oiltab, Chis
1980-1990	Oil, Chis	Oil, Chis	Oil, Chis	Oil, Chis	Oil, Chis	Oil, Chis
1990-2000	oil	Oil	oil	oil	oil	oil

See graph in Appendix B.

Notice that given the nature of the estimation, it is not possible to include a dummy variable for oil producers or for Chiapas, both estimation techniques will eliminate the State dummies. But we can account for the influence they have on our estimation by excluding them and compare the marginal change between the results in table 2 and table 3c. The sign on inequality does not change, but significance slightly increases.

Table 3c. Regression Results : What Affects the coefficient on Inequality? Outliers

Estimation Method	Fixed Effect (Excluding oil) (1)	Arellano and Bond (Excluding oil) (2)	Fixed Effect (Exc. oil & Chis) (3)	Arellano and Bond (Exc. oil & Chis) (4)
Income	-.0922712** (.0111481)	-.5386487** (.1217264)	-.0851842** (.0112286)	-.5882834** (.1238494)
Inequality	.0118467 (.0156453)	.5284371** (.1429985)	.0073691 (.0144092)	.4751712** (.1292132)
Schooling	.0533281** (.008759)	-.046459 (.1777213)	.0508715** (.0086878)	-.0023052 (.158222)
R-squared	0.4599	--	0.4288	--
States	30	30	29	29
Observations	120	90	116	87
Period	1960-2000	1980-2000	1960-2000	1980-2000
Sargan Test	--	chi2(5)= 37.55 Prob>chi2 = 0.00	--	chi2(5) = 40.68 Prob >chi2 = 0.00
A&B acov res 1st	--	z = -1.25 Pr > z = 0.2105	--	z = -0.44 Pr > z = 0.6614
A&B acov res 2nd	--	z = -0.46 Pr > z = 0.6425	--	z = 0.64 Pr > z = 0.5231

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the within R-squared for the fixed effects model. ** significant at 1%. * significant at 5%.

2. A second factor is the length of the periods considered, so we performed several estimations of equation (1) varying the period lengths. As we took one long period as the long-term influence from 1960 to 1999, then equation (1) has to be rewritten as:

$$Growth_{it} = \alpha_0 + \beta_1 Income_{i,t-1} + \beta_2 Inequaltiy_{i,t-1} + \beta_3 Schooling_{i,t-1} + u_i \qquad (7).$$

And we will be using OLS estimation methods.

Long-term relationship. The relationship between growth and inequality turns out to be positive when considering a whole period 1960-2000, for my GSP data source, as well as for Esquivel's data. Even though that, the coefficient is not significant (columns 1 and 2 in table 3b.)

Short-term relationship. Now we divide the 40-year period into three short periods according to the degree of trade openness. In the case of Mexico we will consider the period before Mexico joined the GATT as the Non-Trade period (although we know that trade was taking place), then we will consider the GATT period as the period between joining GATT and signing NAFTA. The last period will be the NAFTA period.

The two first periods (table 3.d, columns 3 and 4) present a negative coefficient but they are not significant. The NAFTA period presents a positive but non-significant coefficient (the significance increase when we drop the outliers). The change in sign across periods suggests that the relationship between inequality and growth has been changing across time, and one of the reasons can be trade openness. We will revise this in section IX when using the households' income surveys. If we contrast this result with Barro's estimations described in Banerjee & Duflo (1999), we can see that Barro found that the relationship between growth and inequality during 1960-1995 is a U shape for poor countries, and positive for Latin-America. Therefore the signs delivered by the estimation (-, - , +) do not contradict Barro for that period, and add the upward slopping for the 1994-2000 period (NAFTA).

Table 3d. Regression Results : What Affects the coefficient on Inequality? Period.

Estimation Method	OLS (40 year) G. Esquivel (1)	OLS (40 year) (2)	OLS (20 year) No trade (3)	OLS (10 year) GATT (4)	OLS (10 year) NAFTA (5)	FE by trade Period (6)	A & Bond by period of trade (7)
Income	-.018995** (.004951)	-.015366** (.003476)	-.017389** (.007583)	-.071957** (.014970)	-.015086** (.007552)	-.072209** (.013802)	.097631 (.157171)
Inequality	.002641 (.007454)	.0035033 (.006234)	-.0030501 (.013599)	-.0350722 (.099857)	.0443563 (.037920)	-.0057232 (.023746)	-.026587 (.21032)
Schooling	.01806** (.005816)	.0143253** (.004370)	-.00077 (.009533)	.1055127** (.039673)	.0669645** (.024888)	.0864551** (.018228)	-.073970 (.215966)
R-squared	0.3510	0.4250	0.3266	0.4897	0.2216	0.5255	--
States	32	32	32	32	32	32	32
Periods	1	1	1	1	1	3	3
Outliers	No change in sgn & sgfc of Ineq	No change in sgn & sgfc of Ineq	No change in sgn & sgfc of Ineq.	No change in sgn & sgfc of Ineq	No change in sgn but incr. sigfc.	Change in Sgn & Sgfc incr.	No change in sgn but incr. sigfc

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the overall R-squared for OLS and within for FE. ** significant at 1%. * significant at 5%.

Using FE and A&B with period dummies (three periods in this case), the FE inequality estimate is positive and highly significant, not so for the second. We can suggest that time length and the period studied affect the relation between inequality and growth.

We turn now to see if changing the human capital variable from schooling to literacy has some effects.

Table 3e. Regression Results: What Affects the coefficient on Inequality? Period and Human Capital Definition.

Estimation Method	OLS (40 year) G.Esquivel (1)	OLS (40 year) (2)	OLS (20 year) No trade (3)	OLS (10 year) GATT (3)	OLS (10 year) NAFTA (4)	FE by trade Period (5)	A & B by period of trade (6)
Income	-.012567** (.0045825)	-.01163** (.003520)	-.01889** (.006807)	-.069103** (.015957)	-.008091** (.007031)	-.072001** (.013334)	.095141 (.154920)
Inequality	.0032256 (.008304)	.0038669 (.007068)	-.0021486 (.01367)	-.1286831 (.110655)	.029681 (.048384)	-.0028819 (.023171)	-.0152256 (.2122082)
Male-Lit	.0291219 (.0253538)	.0157887 (.021670)	.0539208 (.041908)	.2995215 (.214927)	.0872378 (.127927)	-.1777842* (.103729)	-1.099854 (1.017392)
Female-Lit	-.0000895 (.0003034)	.000035 (.000252)	-.0005796 (.000487)	-.0010554 (.001878)	.0002467 (.000973)	.0052026** (.001471)	.0181179 (.0159143)
R-squared	0.2480	0.3105	0.3654	0.4518	0.1281	0.5595	--
States	32	32	32	32	32	32	32
Periods	1	1	1	1	1	3	3
Outliers	No change in sgn & sgfc of Ineq.	No change in sgn & sgfc of Ineq.	No change in sgn & sgfc of Ineq.	No change in sgn, sgfc decr.	No change in sgn but incr. sigfc.	Change in Sgn & Sgfc incr.	Change in Sgn & Sgfc incr.

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the overall R-squared for OLS and within for FE. ** significant at 1%. * significant at 5%.

Table 3e shows that it only affects the sign of the inequality coefficient of the last trade period (column 5). The rest of the inequality coefficients remind with the same sign and significance, suggesting that we need further research for NAFTA period. Particularly interesting are two facts:

1) The human capital variable is non significant in most of the cases. Comparing our results with those of Rodriguez-Pose (1999), about economic convergence, he found that education[31] has a significant and negative effect on growth between 1980-1985, then negative and non significant between 1985-994, and then positive but non significant from 1994-1998. Suggesting that the economy was more natural resource's dependent at the beginning and then became more skilled dependent. In table 3d we find that for Non-Trade period schooling coefficient was negative and then switch to be positive, mildly coinciding with Rodriguez-Pose (1999) findings.

[31] He just states the variable "edu" as education , but do not specify if it is schooling or literacy.

Moreover, when we use literacy, results differ very much, changing sign from period to period for female literacy, and been constant for male literacy. Barro & Lee (1996) have a discussing regarding female and male performance. Perotti's estimations also find this type of switching sign from male to female literacy. We have constructed a composite literacy ratio, regardless of sex, the estimation behave as in table 3d, the coefficient of the compose literacy ratio behaves like the schooling variable but is not always significant. This may suggest that beyond the explanations of Barro & Lee or Perotti's, the change in sign from male to female literacy has to do more with partial correlation between these two variables (moderator & suppressors effects) rather than with explanation about male and female literacy performance. Later, we will contrast these finding with those in section IX.

2) The relation presented with trade periods using OLS and schooling as human capital variable, shows that the sign of the coefficient of inequality varies with trade period:

No trade	GATT	NAFTA
(-)	(-)	(+)

It may suggest that in the early stages of trade openness, inequality was detrimental to growth but as the economy becomes more open, inequality has a positive relation with growth. NAFTA period would be uncertain to interpret as its initial stage coincides with a two years economic crisis that Mexico suffered from December 1994. The results can be interpreted in three ways:

a) If we follow Perotti (1996); when Government implements an economic policy to lessen inequality, say increasing public spending financed with an increase in taxes, then inequality decreases but also growth does[32]. In that sense we can say that within a fiscal framework approach, the channel through which inequality affects growth is public expenditure financed with taxation, so that we explain why the positive relation. And according to Boltvinik's (1999) poverty research on Mexico, in the period 1983-1988 the efforts for fighting poverty and inequality were discontinued (Coplamar and SAM), and new efforts and programs started from the 1988 presidential period that follow (Solidaridad and Progresa (nowadays named Oportunidades)). It may explain why before 1988 the relationship is negative and afterwards positive. (We will analyse this structural form in section X).

b) Another interpretation is to make a change to equation (1) and try to explain how openness affects inequality and growth. (*Aghion & Williamson (1998)*). We can, for instance, incorporate openness as an additional explanatory variable. (This research is in progress).

c) A further interpretation is to follow Quah's (1997) approach about club formation so that rich states are located near rich states and poor near poor. The last approach is developed by looking at seven geographical regions. In next section we will analyse this approach.

[32] In this case incentives for savings also decrease.

Section VIII. Sensitivity Analysis.

Regions relationship: The regional division presented below is taken from Esquivel (1995), we use it trying to see if there is any difference in inequality coefficient across regions as North is rich contrasting with a poor South.

Table 4.a. Regions according to Esquivel (1999)

North	Capital	Golf	Pacific	South	C.North	Centre
B.C.N. Chihuahua Coahuila N. Leon Sonora Tamaulipas	D.F. Mexico	Campeche Quintana Roo Tabasco Veracruz Yucatan	B.C.S. Colima Jalisco Nayarit Sinaloa	Chiapas Guerrero Michoacan Oaxaca	Aguascalientes Durango Guanajuato Queretaro S. Luis Potosí Zacatecas	Hidalgo Morelos Puebla Tlaxcala
6	2	5	5	4	6	4

In these regional groups we can find intrinsic characteristics that make economies differ. North Region for instance, closer to USA, has 6 of the highest product per capita States, and the highest share of foreign direct investment. The capital region includes D.F., which has the highest product per capita (38903 in the year 2000); it is 6 times higher than the poorest state which is Oaxaca. If we were to discuss that product per capita may reflect population concentration in a region, then taking away the per capita considerations, we can see that D.F. product is between 18 and 15 times more than Oaxaca from 1960-2000. So we expect the relation of income inequality and growth to differ for each region.

Table 4.b. Regression Results: What affects the coefficient on Inequality? Region

Geographical Regions	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
North (.3220762)	-0.0759956*	0.0395976	6	24	1960-2000	FE
Capital (.3295079)	0.0326881	0.10534	2	8	1960-2000	FE
C. North(.337028)	0.0414241	0.1309243	6	24	1960-2000	FE
Golf (.3372416)	0.0293938	0.0285472	5	20	1960-2000	FE
Pacific (.3448039)	-0.031164	0.0808259	5	20	1960-2000	FE
South (.3784491)	0.0441505	0.0817078	4	16	1960-2000	FE
Centre (.3980587)	0.0496995**	0.0190213	4	16	1960-2000	FE

Geographical Regions	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
North (.3220762)	-1.162785**	0.2762	6	24	1980-2000	A&B
Capital (.3295079)	0.8142841**	0.1755365	2	8	1980-2000	A&B
C. North(.337028)	0.4543819**	0.2043235	6	24	1980-2000	A&B
Golf (.3372416)	-0.4491927	0.6752689	5	20	1980-2000	A&B
Pacific (.3448039)	0.6604339	0.989981	5	20	1980-2000	A&B
South (.3784491)	0.783074	0.1922	4	16	1980-2000	A&B
Centre (.3980587)	0.573872**		4	16	1980-2000	A&B

**Significant at 1%. * Significant at 5%.

Notes: Initial Gini coefficient is in brackets. GSP coefficient is always negative. Schooling coefficient is always positive but not always significant. North Region (according to A&B) outcome, seems to have a negative relationship, this result is in line with the theory that specifies that richer countries in later stages of their development have a negative relation with income inequality. Geographical regions are ranked by initial inequality (ascendant). Below we present the regions ranked by average inequality across periods.

1. Capital 0.450 2. North 0.453 3. Pacific 0.476	4. Golf 0.494 5. C. North 0.511	6. Centro 0.531 7. South 0 .570

The results in table 4.b. show that in spite of the fact that the coefficient of income inequality is positive in 71% of the cases, it is only significant in 43% of them. One important fact is that the coefficient of inequality is higher with A&B technique; it is also higher in magnitude for the richest region (north and capital) and only negative and significant for the richest region. This could be explained by Aghion & Williamson's (2000) discussion about income inequality having a stimulating effect on growth "there *would be a fundamental trade off between productive efficiency (and/or growth) and social justice, as redistribution would reduce differences in income and wealth, but would also diminish the incentives to accumulate wealth*". This argument is true for the poorest regions, as we have always had problems of social justice, not only in monetary terms but also in racial terms. 57% of the indigenous population is concentrated in those areas. The average schooling years of the population in that region are 5.7 and 6.7 compared with 9.6 in D.F., and their poor access to public services, keep them backward. That may explain the stronger positive coefficient in those regions. Compared to section IX where more data points are available, we will see that the coefficient of inequality is both positive and significant only for south and pacific regions.

Although the relationship seems to be positive in most of the cases, we cannot conclude that the relationship really does not change across regions as data is not enough, as before, NT is too small[33]. Later we will compare these results with the ones in section IX where we will be using ENIGH[34].

In previous literature and work related to inequality and growth, criticisms arise because results are not robust when analysing the sensitivity to the inclusion of new variables or the way data has been grouped, therefore we do perform a sensitivity analysis.

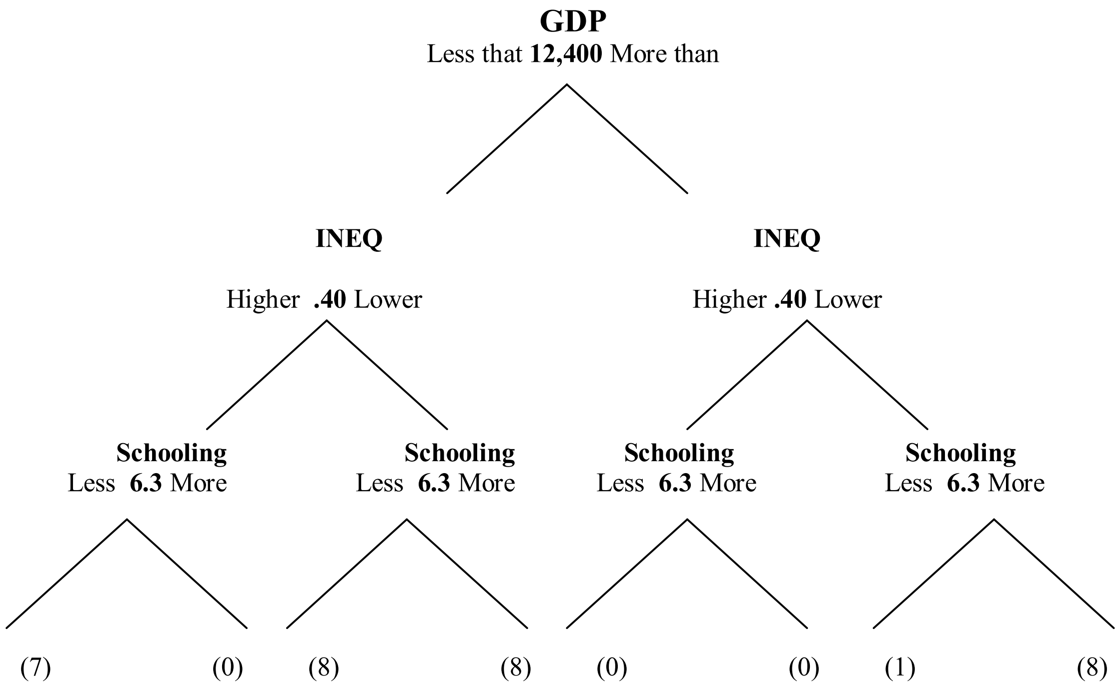
In what follows, income inequality coefficient was always positive through out all the different analyses that were performed.

1. a) It can be argued that the way we previously grouped the 32 States following Esquivel's (1999) geographical regions may not be the adequate to probing theory of coalition or club formation. Therefore we re-estimate the model grouping States according to their income per capita level, then their inequality level and finally their schooling level according to their rounded mean in 1990. Using the following tree diagram[35].

[33] See end of Appendix A.

[34] National Household Income and Expenditure Survey for 1984-1998.

[35] This section is inspired by the tree algorithm technique used in Durlauf & Johnston (1995).



With this technique, we have five groups, which we can use to define our own welfare regions, where region 1 has the lowest welfare and region 5 the highest welfare.

Table 4c. Regional Groups according Tree Welfare Regions

Region 1	Region 2	Region 3	Region 4	Region 5
Chiapas Guerrero Oaxaca Puebla SnL.Potosí Tabasco Yucatán	Durango Guanajuato Hidalgo Michoacán Nayarit Querétaro Veracruz Zacatecas	Aguascalientes Colima Jalisco México Morelos Sinaloa Tamaulipas Tlaxcala	Campeche	Baja California Baja Calif. Sur Coahuila Chihuahua Distrito Federal Nuevo León Quintana Roo Sonora
7	8	8	1	8

Hence, performing the estimation for our two different ways of grouping the States we have the following results.

Table 4d. Sensitivity Analysis: Own Welfare Regions definition.

Regional Groups	Coefficient on INEQ	Standard Error	States	Obs.	Period of growth	Estimation Technique
R1	.0400574	.0769969	7	28	1960-2000	FE
R2	.0329265	.0215983	8	32	1960-2000	FE
R3	.0056755 -	.0302505	8	32	1960-2000	FE
R5	.0405788	.0438244	8	32	1960-2000	FE
R1	.5529425	.8683937	7	21	1980-2000	A&B
R2	.5724293**	.2215417	8	24	1980-2000	A&B
R3	.4109088	.2702965	8	24	1980-2000	A&B
R5	-.3809469	.3395797	8	24	1980-2000	A&B

[a] Region by region FE and A&B estimators give always a positive coefficient for inequality, but only significant for regions 1and 5. Region 2 cannot be compute due to there is only one State in that group. The only consistent and significant in all regressions is the negative sign in coefficient for lagged income.

Again we can find the same pattern as in previous table 4b. The richest region (the ones with the highest welfare) has a negative sign on inequality. But this time results are not significant, so we can not drive a strong conclusion from this table. This pattern will be observed again in table 4e., when we use INEGI's welfare definition. Once more results are significant only in 20% percent of the cases.

It is a well known fact, that significant results arise when there are plenty of time and panel points so that NT goes to infinity and explanatory variables are highly correlated with the dependant variable. In our case, the non-significance results are mainly due to the lack of enough observations. This problem is attenuated when we use households' surveys in section IX.

b) Grouped States using INEGI[36]'s welfare regions.

[36] INEGI stands for National Institute of Statistics, Geography and Informatics in Mexico.

Table 4e. Regional Groups according to INEGI Welfare Regions.

WRegion 1	WRegion 2	WRegion 3	WRegion 4	WRegion 5	WRegion 6	WRegion 7
Chiapas Guerrero Oaxaca	Campeche Hidalgo Puebla Sn Luis P. Tabasco Veracruz	Guanajuato Michoacán Zacatecas	Colima Durango Jalisco Morelos Nayarit Querétaro Sinaloa Tlaxcala Yucatán	Quintana Roo	Aguascalientes B. California B. California Sur Coahuila Chihuahua México Nuevo León Sonora Tamaulipas	D. F.
3	6	3	9	1	9	1

Table 4f. Sensitivity Analysis: INEGI's Welfare Regions definition.

INEGI Regional Welfare Groups,	Coefficient on INEQ	Standard Error	States	Obs.	Period of growth	Estimation Technique
W1	.0874066	.1003845	3	12	1960-2000	FE
W2	.0155403	.0533907	6	24	1960-2000	FE
W3	-.0267263	.117516	3	12	1960-2000	FE
W4	.0448268*	.0235286	9	36	1960-2000	FE
W6	-.0480285	.0313881	9	36	1960-2000	FE
W1	.9098775	1.168124	3	9	1980-2000	A&B
W2	.4555975	.5339787	6	18	1980-2000	A&B
W3	-.8738871	.6859252	3	9	1980-2000	A&B
W4	.6562777**	.2343143	9	27	1980-2000	A&B
W6	-.3536001	.241561	9	27	1980-2000	A&B

[b] Region by region FE and A&B estimators give always a positive coefficient for inequality, but only significant for regions 4 for both FE and A&B . Region 7 cannot be computed due to there is only one State in that group.

2. Estimates for different definitions of inequality.

Finally, as recent literature tries to explain that the relation between income inequality and growth might be quadratic when using GINI coefficient, so that the functional form is not linear[37], this could be an explanation for getting the positive coefficient. We swap the Gini coefficient calculated with Leman-Yitzhaki formula for the Gini calculated with POVCAL.

[37] See Banerjee & Duflo (1999).

Table 4.g. Sensitivity Analysis: inequality definitions.

Inequality definitions	Coefficient on INEQ	Standard Error	States	Obs.	Period of growth	Estimation Technique
20/20 Ratio	.0098957	.0073175	32	128	1960-2000	FE
	.1747568**	.0705838	32	96	1980-2000	A&B
20/20 Ratio no oil	.0079801	.0061773	30	120	1960-2000	FE
	.2226989**	.0570699	30	90	1980-2000	A&B
POVCAL	.0068484	.0244949	32	128	1960-2000	FE
	.3486201	.247042	32	96	1980-2000	A&B
POVCAL No oil	.0297187	.0196563	30	120	1960-2000	FE
	.5779624**	.1810965	30	90	1980-2000	A&B

[v] 20/20 ratio is the income share held by the richest 20 percent of the population, to the share held by the poorest 20 percent.

[v] POVCAL is developed by Chen-Datt-Ravallion at the World Bank to compute poverty and inequality measures, including grouped data

We also drop our definition of inequality based on Gini and used the 20/20 ratio as an alternative measure of inequality. The 20/20 ratio is the quotient between the income of the twenty percent of the richest population to the 20 percent of the poorest, the higher the ratio, the higher the gap between the rich and the poor. Even when we use this new measure of inequality, the estimated inequality coefficient is still positive and very significant when using Arellano and Bond technique.

3. Economic performance and income are highly related, in order to view this interaction we divide our data in income groups.

Results from this table are in line with those observed in the groped data by regions. The richest groups have a negative coefficient on inequality, whereas the poorest have a positive coefficient. Results are non significant in all cases except one.

Table 4.h. Sensitivity Analysis: Income groups definitions.

Income Groups	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
Using INEGI data						
Poor<6037	.0319056	.0267858	17	68	1960-2000	FE
6037≤ Mid <9932	-.0119909 -	.0390737	10	40	1960-2000	FE
Rich≥9932	.0033813	.0494454	5	20	1960-2000	FE
Poor<6037	.5062141**	.2857591	17	51	1980-2000	A&B
6037≤ Mid <9932	-.0480928	.3232534	10	30	1980-2000	A&B
Rich≥9932	.224814	.2132983	5	15	1980-2000	A&B
Using G. Esquivel data						
Poor<9000	.0265088	.019809	17	68	1960-2000	FE
9000≤ Mid <16000	-.0024942	.0397209	10	40	1960-2000	FE
Rich≥16000	-.0548315	.026092	5	20	1960-2000	FE
Poor<9000	.1845697	.265695	17	51	1980-2000	A&B
9000≤ Mid <16000	.0679125	.4807856	10	30	1980-2000	A&B
Rich≥16000	-.1107083	. 4679214	5	15	1980-2000	A&B

[iv] States are categorised base on GSP per capita in 1990. Income is measured in 1993 pesos. A & B estimator is always significant and rejects Sargan null for groups poor and rich, the no autocorrelation is reject for group "poor".

From the two previous sections we can conclude that the relation between inequality and growth depends on the period analysed, because when using FE or A&B with trade period dummies (short run relationship) it can be identified as a positive relation. Using OLS period by period we can identify a negative relation for earlier periods and a positive for the latest, i.e., the relation may depend on the country's stage of development. In addition, we can conclude that the relation between inequality and growth is in most of the cases positive.

These results can be influenced severely by the effect produced by the small amount of data and data quality (personal income instead of households' income) and by omitted variables (taxes and government expenditure[38], public education quality), therefore the next two sections deal with these problems which need further research.

Section IX. Using Household Income.

One of the most severe criticisms from the results in previous sections arises from the fact that we were using personal income surveys and grouped data to calculate the Gini coefficient. Nevertheless, this was the only available income for us to use from 1960 to 1999 on a decade basis. Looking for a more representative variable of income for people among States, we will use the household's income from the National Household Income and Expenditure Survey in Mexico (ENIGH) produced by the National Institute of Statistics, Geography and Informatics in Mexico. There have been nine surveys: 84, 89, 92, 94, 96, 98, 00, 02 which implies we will only be able to analyse the relationship between growth and inequality for the GATT period (1984-1994) and the NAFTA period (1994-2002).[39]

In their sample design, INEGI reports a specific number of households to be surveyed, but not all of them were successfully surveyed. According to the data we obtained from the ENIGH, the table below shows the number of households whose income was available in the survey.

Year	1984	1989	1992	1994	1996	1998	2000	2002
Households	4727	11517	10508	12791	13075	10098	11170	19856

[38] Tabellini points out the importance that government expenditure structure has in fighting inequality and promoting growth.

[39] 2002 survey is not used as the last period in the regression analysis takes into account growth from 2000 to 2002 regress on inequality of 2000.

The survey allows us to add all monetary sources of income[40] for each person and then aggregate each person in households, so that later we directed each household to the State it belongs to, and calculate the Gini coefficient. The survey also allowed us to calculate the literacy rates for males and females.

<u>Data</u>

*GSP (Gross State Product per capita) is still coming from INEGI, and again 1993 is the base year. The alternative GSP coming from Esquivel (1999) cannot be used in this section as dates are completely different from the ones we are dealing with now.

*Human capital explanatory variable is female and male literacy is the percentage of population aged ten or more who can read and write, calculated from the corresponding ENIGH surveys.

*GINI was calculated with ENIGH quarterly monetary income. For a detailed scope of the income distribution of the households see *Appendix B*.

As we clarified before, we performed the estimations with and without using the expansion coefficients provided by the ENIGH, Using the expansion factors lead us to obtaining Gini coefficients of the magnitude of 70% which we considered as not informative, not representative and erroneous. Nevertheless we conducted the estimation for both cases. The table below only reports the estimates without using the expansion factors, the rest of the summary statistic are available on request.

[40] For each of the 32 States, we computed the quarterly monetary income which is the result of the sum of six different sources of income: 1) earned income, 2) self-employment income, 3) property rents income, 4) Income from cooperatives, 5) Transferences, 6) Other sources of monetary income.

Table 5—Summary Statistics for Households Data

Variable	Definition	Source	Year	Mean	Std. Dev.	Min	Max
Income	Ln of Real Gross State Product (GSP) per capita in 1993 pesos. Correcting with national deflator before 1990. * Calculating 2002 using national GDP 2002 and State's share in 2001.	INEGI	1984	2.569559	.4200224	1.927049	4.015779
			1989	2.432083	.4205658	1.725786	3.3789
			1992	2.447025	.4121614	1.779344	3.431072
			1994	2.504636	.4252418	1.820721	3.533127
			1996	2.463371	.4242932	1.780849	3.478531
			1998	2.494974	.4259837	1.770198	3.556282
			2000	2.583899	.4398278	1.848565	3.662692
			2002*	2.543401	.4300736	1.832969	3.642119
Inequality	Inequality measured by the Gini Coefficient.	ENIGH	1984	.4272346	.059682	.2769763	.5229887
			1989	.473915	.0624284	.3461361	.6306674
			1992	.5519121	.0618629	.4319943	.7258033
			1994	.4785259	.0509095	.3745871	.6016473
			1996	.4958984	.0579005	.4215405	.7152544
			1998	.5119454	.0497137	.4192214	.6178948
			2000	.5002526	.0562213	.3781551	.5826821
			2002	.4713063	.0420932	.3726221	.5620028
Female Literacy	Share of the female population aged over 15 (10) who can read and write.	ENIGH	1984	84.54062	9.783308	64.38	98.31
			1989	85.655	8.538247	62.73	97.03
			1992	82.37344	9.878367	60.92	94.9
			1994	83.12219	9.375087	60.05	94.77
			1996	84.66781	7.658279	64.84	95.21
			1998	85.5625	7.93256	69.55	97.9
			2000	86.88799	5.984771	73.75	95.59
			2002	87.18169	7.63206	70.26	97.10
Male Literacy	Share of the male population aged over 15 (10) who can read and write.	ENIGH	1984	91.05375	6.23506	79.38	100
			1989	88.71844	5.846874	78.72	97.05
			1992	86.13969	6.476109	73.97	97.66
			1994	86.93531	6.052723	70.83	97.67
			1996	88.17062	4.632704	76.57	97.37
			1998	87.46031	5.654914	73.09	97.53
			2000	88.49079	4.37435	81.08	98.37

Model and Estimation

We estimate with FE equation (2), as before[41]:

$$Growth_{it} = \beta_1 Income_{i,t-1} + \beta_2 Inequality_{i,t-1} + \beta_3 Male\_literacy_{i,t-1} + \beta_4 Female\_literacy_{i,t-1} + \alpha_i + \eta_t + u_{it}$$

And its equivalent with Arellano and Bond Technique:

$$y_{it} = \gamma\, y_{i,t-1} + X'_{i,t-1} B + \alpha_i + \eta_t + u_{it} \qquad (5)$$

Estimates with FE and A&B show a positive inequality coefficient, whether using year-dummies or not, but only significant when not controlling with year dummies. Using a Hausman test, Random Effects estimators are rejected in favour of Fixed Effects. Appendix A, provides a detail mathematical reason why FE estimation is not

[41] See page 10.

valid due to the presence of an endogenous lag variable, so that the best estimation and valid estimation arises when using Arellano and Bond GMM technique.

Table 6. Regression Results: Alternative Estimation Techniques

Estimation Method	Fixed Effect (1)	Random Effect (2)	Fixed Effect with Year-Dummies (3)	Random effects Year-Dummies (4)	Arellano and Bond (5)	Arellano and Bond Year-dummies (6)
Income	-.162138** (.0150075)	-.023433** (.0079067)	-.144912** (.0104796)	-.038343** (.0073887)	.2844816** (.0571731)	.5404574** (.0682532)
Inequality	.0611511** (.0154486)	.0893637** (.01714)	.0040675 (.011633)	.0235404* (.0143918)	.1370858** (.037846)	.0512578 (.0399755)
Male Literacy	-.166511** (.0662672)	-.0983245 (.0675219)	-.0415821 (.0477453)	-.0014061 (.0556643)	-.3768628** (.1606021)	-.1844717 (.1531613)
Female Literacy	.143724 ** (.0477476)	.1347708** (.0427833)	.0513933 (.0347943)	.0678431* (.0376015)	.2955332** (.1135278)	.1557168 (.1077121)
R-squared	0.4946	0.1438	0.8016	0.4554	--	---
States	32	32	32	32	32	32
Observations	224	224	224	224	192	192
Period	1984-2002	1984-2002	1984-2002	1984-2002	1989-2002	1989-2002
Hausman Test	chi2(4) = Prob>chi2 =	118.26 0.00	chi2(11)= Prob>chi2 =	298.46 0.0000	--	--
Sargan Test	--	--	--	--	chi2(20)=100.4 rob>chi2=0.0	chi2(20)=19.37 Prob>chi2=0.49
A&B acov res 1st	--	--	--	--	z = -5.01 Pr > z = 0.00	z = -3.60 Pr > z = 0.0003
A&B acov res 2nd	--	--	--	--	z = -2.82 Pr >z =0.0048	z = 0.58 Pr>z = 0.5597

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the within R-squared for the fixed effects model and the overall R-squared for random effects. *Non-significant.

Our conclusions on this work will be derived by using A&B like in column 5. In any case we keep reporting the FE estimator just to compare it with the sign and magnitude of the GMM (A&B) estimator.

The main result of this section is that the coefficient on inequality is positive and significant across all periods (1984-2002), which implies that increases in inequality lead to increases in growth.

Nevertheless, as in section VI, results are questionable: “Is it only the method that matters or there are other factors to be considered in the relationship between inequality and growth”. Mimicking Section VII, we start by analysing one factor at the time.

Factors that affect the coefficient of inequality.

A) Period Length.

As before, we estimate the model for different period lengths using OLS, FE and A&B. Overall, the relationship is still positive, but this time the relationship estimated with OLS switches sign for NAFTA and GATT period. This is a puzzle given the observation we have made on page 23, about the possible causes of the relationship being negative during the GATT period and then positive in the NAFTA period.

B) Outliers

Problems arising from outliers in this data set when using panel data methods were not as worrying as in the data set in section VII. The main reason is because the Oil Producers are not playing a role as outlier beyond 1989-1992 growth. So even when we drop the outliers, the coefficient on inequality is still positive and becomes more significant.

Table 7a.OUTLIERS CAUSING PROBLEMS (HOUSEHOLDS DATA)

	Ln	Yit	LiFM	Yit EXP	LiFM EX
84-89	Oil, poor	Oil, poor	Oil, poor	Oil, poor	Oil, poor
89-92	Oilcam, 23, 9	oilcam	oilcam	oilcam	oilcam
92-94	-	-	-	-	12,20,poor
94-96	-	17, 5	-	-	Poor
96-98	-	28	4	-	Poor
98-2000	9	-	-	-	12
2000-2002	-	-	2,9	-	-

*Poor is the id dummy for the State of Chiapas.

Table 7b. Regression Results : What affects the coefficient on Inequality? Period and Outliers

Estimation Method	OLS (15 years)	OLS[A] (10 year) GATT	OLS[A] (5 year) NAFTA	FE (10 years) GATT	FE (5 years) NAFTA	A & Bond* GATT	A & Bond* NAFTA	FE with trade-period dummies	A & Bond trade-period dummies
Income	-.01878** .0083966	- .027913** .0140413	-.0064755 .0051276	- .1305183** .0159127	- .364024**1 .0460518	.4377996** .1130227	-.1746629 .1378475	- .1635538** .0151974	.2538069** .0532829
Inequality	.018667 .0241556	.0297279 .0403944	.0077099 .0167079	.1022064** .0169616	.0521183 .0288924	.1538257* .0743721	.1169376** .0426366	.0601031** .015562	.1780821* .0755034
Male Lit	.0247755 .1009062	.0286115 .0902156	-.0046838 .0465238	.1053729* .0529692	-.2022146 .1252798	-.3339697 .2870416	-.2412882 .1904538	- .1555649** .0685966	-.3717864* .1763379
Female Lit	.0334228 .0539484	.0615593 .1687412	.0390864 .02833	-.1306455 .0763155	.2690098** .0878711	.3183309 .1894263	.1082593 .1567745	.1354335** .0495911	.3048566 * .1381437
R-squared	0.1961	0.1569	0.1327	0.7840	0.4336	--	--	0.4956	--
States	32	32	32	32	32	32	32	32	32
Periods	1	1	1	3	5	2	4	2	2
Dropping Outliers	Sgn chnge Sgfc decr	Sgn chnge Sgfc decr	No effect	Same Sgn Sgfc incr.	No effect	Same Sgn Sgfc decr .	No effect	Same Sgn Sgfc incr.	Same Sgn Sgfc decr .
Dummies Outliers	Sgn chnge Sgfc decr	Sgn chnge Sgfc decr	No Effect	Same Sgn Sgfc incr	No Effect	Same Sgn Sgfc decr.	No Effect	Same Sgn Sgfc incr	Same Sgn Sgfc decr.
Driving the change	OIL	OIL	--	OIL	--	OIL	--	OIL	OIL

Notes: Dependent variable is average annual per capita growth. Standard errors are in parenthesis. R-squared is the overall R-squared for OLS and within for FE.

[A] Period by Period OLS shows a positive coefficient of inequality for periods 1984-89, 1989-92, 1992-94, 1994-96 and negative for periods 1996-98, 1998-1999.

*Significant 5-10%. ** Significant 0-5%.

The problem with outliers arises when estimating the model using OLS technique for the GATT period (1984-1994) due to the severe influence that oil producer estates play in driving the relationship from negative to positive.

It can be seen in table 7b, that the GATT period when using OLS estimation gives a positive coefficient of inequality. Once we include a dummy for the oil producer, or drop them, the relationship changes to negative and significance decrease.

Even though results using OLS estimation are not significant in any of these estimations, we can still be able to say that dropping outliers we obtain exactly the same results as in section VII. Relationship is changing across trade period. The gist of the previous table is:

Table 7c. Period Length Effects

Estimation Technique	WHOLE	GATT	NAFTA
OLS	(+)	(+)	(+)
OLS treating outliers	(-)	(-)	(+)
FE	(+)	(+)	(+)
A&B	(+)	(+)	(+)

C) Data availability.

Given that the gap (number of years) between one survey and another is not the same for the first two surveys in comparison with the rest of them whose availability is every two years, we performed the analysis dropping 1984 (first survey), then dropping 1984 and 1989 (second survey). In all cases, using FE and A&B techniques, the coefficient of inequality was positive and significant. Meaning then that from 1984 until now the relationship between income inequality and growth has remained positive.

Sensitivity Analysis.

A) Regions.

Table 8. Regression Results : What affects the coefficient on Inequality? Region

Geographical Regions	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
North	0.104832**	.0445662	6	42	1984-2002	FE
Capital	.0860682	.0807005	2	14	1984-2002	FE
Golf	-.0068577	.0468655	5	35	1984-2002	FE
Pacific	.0493436	.025864	5	35	1984-2002	FE
South	.0502914	.0424864	4	28	1984-2002	FE
C. North	.0565872	.0377702	6	42	1984-2002	FE
Centre	.0871629	.0477207	4	28	1984-2002	FE
North	.0979573	.0794194	6	36	1989-2002	A&B
Capital	.2138542	.2216171	2	12	1989-2002	A&B
Golf	.0370503	.1363271	5	30	1989-2002	A&B
Pacific	.119299*	.0579346	5	30	1989-2002	A&B
South	.1498383 *	.0755402	4	24	1989-2002	A&B
C. North	.08881	.068347	6	36	1989-2002	A&B
Centre	.1119644	.0847619	4	24	1989-2002	A&B

Income coefficient is always negative and significant, literacy coefficient is always positive but not always significant.

As Table 8 shows, we have grouped the data according to their geographical zone defined by Esquivel (1999). In table 9 by INEGI and then we also compute our own geographical zones according to the tree diagram shown on page 28 for the base year 1990 and for the base year 1984 to make it comparable. The results show that income inequality and economic growth are positively correlated in all cases. Once more due to the groups formation that leads into few observations for each group, NT becomes small and results are non significant.

Table 9. Sensitivity Analysis: Regions and inequality definitions

	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
Standard Analysis Whole Sample						
Whole Sample	.0611511 **	.0154486	32	224	1984-2002	FE
	.1370858 **	.037846	32	192	1989-2002	A&B
INEGI Regional Welfare Groups						
W1	.0233719	.0505956	3	21	1984-2002	FE
W2	.1018888 **	.0347111	6	42	1984-2002	FE
W3	.0204677	.0505393	3	21	1984-2002	FE
W4	.0240185	.0298068	9	63	1984-2002	FE
W6	.0800286 **	.0283777	9	63	1984-2002	FE
W1	.1546196	.0941024	3	18	1989-2002	A&B
W2	.2074687 **	.0741402	6	36	1989-2002	A&B
W3	.0414458	.0745795	3	18	1989-2002	A&B
W4	.0275678	.0630692	9	54	1989-2002	A&B
W6	.0929115	.0610038	9	54	1989-2002	A&B
Regional Groups (for 1990 classif.)						
R1	.0315152	.0351329	7	49	1984-2002	FE
R2	.0510049	.0279804	8	56	1984-2002	FE
R3	.0629348*	.0286169	8	56	1984-2002	FE
R5	.0704772*	.0352204	8	56	1984-2002	FE
R1	.1282237	.069092	7	42	1989-2002	A&B
R2	.1204493*	.0577911	8	48	1989-2002	A&B
R3	.0759185	.0637662	8	48	1989-2002	A&B
R5	.1319166	.0701356**	8	48	1989-2002	A&B
Regional Groups (for 1984 classific.)						
NR1	.0351983	.0271184	7	49	1984-2002	FE
NR2	-.0643638	.0743809	3	21	1984-2002	FE
NR3	.1500355**	.0548765	3	21	1984-2002	FE
NR4	.0322828	.0331687	5	35	1984-2002	FE
NR6	.0896044*	.0470269	7	49	1984-2002	FE
NR7	.0785425*	.0321012	6	42	1984-2002	FE

	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
NR1	.1232764 *	.053389	7	42	1989-2002	A&B
NR2	-.0770881	.147043	3	18	1989-2002	A&B
NR3	.2317824**	.0842537	3	18	1989-2002	A&B
NR4	-.0418675	.0740838	5	30	1989-2002	A&B
NR6	.0788825	.0838852	7	42	1989-2002	A&B
NR7	.1759866*	.079673	6	36	1984-2002	A&B

Notice that this time the first significant contrast that we observe in tables 8 and 9, compare to the results obtained for the different region in section VIII is that the coefficient of inequality is positive for the richest region. Tables 8 and 9 can be summarized into:

Table 10. Summary of Regional Effect

	Esquivel	INEGI	1990	1984
FE	(+)	(+)	(+)	(+)
A&B	(+)	(+)	(+)	(+)

B) Income groups.

We also group the data according to their level of income, and re-estimate equation (2). Results show that inequality coefficient is still positive and highly significant for the poor and medium income groups, non-significant for the rich. This is a major change from table 4h., where the coefficient was only positive for the poorest states. Table 4h, considers all the period from (1960-2000) while table 11, with a largest set of observation, explains only what happen to this coefficient form 1984-2002. This fact give us more support to say that the relationship between income inequality and economic is in fact changing across time.

Table 11. Income Effect

	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
Income Groups						
Poor<13330	.0535485**	.0194024	17	119	1984-2002	FE
13330≤Mid <19800	.0652917*	.0284456	11	77	1984-2002	FE
Rich≥19800	.0841509	.0589067	4	28	1984-2002	FE
					1984-2002	
Poor<13330	.088719**	.0392196	17	102	1989-2002	A&B
13330≤Mid <19800	.1247725*	.0697542	11	66	1989-2002	A&B
Rich≥19800	.2121862	.1326261	4	24	1989-2002	A&B

C) Inequality Definition.

Because the use of GINI coefficient, as a measure of inequality, has been considered to have a non linear relationship with growth, we perform the analysis using the 20/20 ratio as an alternative measure for inequality. This measure is defined as the income share held by the richest 20 percent of the population to the share held by the poorest 20 percent. Doing that, results do not change, the sign of the relationship is still positive and highly significant.

Table12. Sensitivity Analysis: Income Inequality Measure Effect

	Coefficient on INEQ	Standard Error	States	Obs	Period of growth	Estimation Technique
Inequality						
definitions	.0357901**	.0069601	32	224	1984-2002	FE
20/20 Ratio	.0565163**	.0146711	32	192	1989-2002	A&B

20/20 ratio is the income share held by the richest 20 percent of the population, to the share held by the poorest 20 percent

Results using the Households Income and Expenditure Survey (ENIGH) deliver in more than 90% of the cases a positive coefficient for inequality. This coefficient becomes negative only when analysing the GATT period using OLS estimator treating oil producers as different. Therefore, comparing this section with the results using personal income in sections V-VIII, we have found that the relationship between economic growth and income inequality is positive, and only changes sign when we analyse with OLS the trade period of NAFTA and GATT.

Section X. Structural Form.

Most of the models of inequality and growth up to 1996 use a reduced form estimation where they add the income distribution variable as one more explanatory variable in a standard economic growth regression. Perotti (1996) suggests that it is not enough to estimate the growth equation in its reduced form; but it is necessary to look for the channel through which inequality influences economic growth and estimate this relation following three steps. The first step is to decide which approach we will follow: fiscal policy, political instability, investment in human capital with borrowing constraints, or joint decisions on fertility and education. Once we have decided the approach, the second step consists in identifying the channels through which inequality affects growth, using these channels as instrumental variables. The third step is to estimate the growth equations once we have used the instruments. An example of Perotti's reduced form equation would be:

$$Growth_{it} = \beta_1 GSP_{i,t-1} + \beta_2 EXP_{i,t-1} + \beta_3 Male\_literacy_{i,t-1} + \beta_4 Female\_literacy + z_i + e_{it}$$

$$EXP_{it-1} = \gamma_1 Inequality_{i,t-1} + \gamma_2 POP65_{i,t-1} + \gamma_3 Deprivation_{i,t-1} + v_i + u_{it}$$

where z_i, v_t are state effects and u_{it}, e_{it} are the error term.

- The reduced form will be: Growth increases when equality (inequality) increases (decreases).

- Perotti three steps approach in a Fiscal framework would be:

Step 1: Growth increases when distortionary taxation decreases.

Step 2: Government redistributive expenditure and then distortional taxation decreases when equality (inequality) increases (decreases).

Step 3: Growth increases when equality (inequality) increases (decreases).

Therefore, the Fiscal approach at a National level requires us not only to perform an ordinary OLS regression but also to apply two-stage instrumentation that may be more accurate according to Perotti.

The original aim of the current section was to implement the structural form for the 32 states of Mexico, for the 6 periods from 1960-2002. But due to data availability in government expenditure variables among states, we are implementing the fiscal approach for the period 1989-2002, using household survey.

As a measure of re-distributive government expenditure we will use government expenditure of the counties, aggregated by State. We use this measure, as it is the one easily available from INEGI. Government expenditure is the result of adding expenditure on administrative issues, construction and public fostering, transferences, debt, disposable, third parties, and other expenses. From all of this, it is considered that only the expenditure in construction and public fostering is the one that plays a major redistribution role. So we re-estimate the model using this measure instead of the total government expenditure.

In addition, we use the share of the population who is more than 65 years old due to the fact that they require more expenditure on health care. If government expenditure is effective, we would also expect that the deprivation index[42] would be reduced.

According to Perotti's theory, we would be expecting to find a negative coefficient between taxation and growth given the distortionary effects of taxation, and a positive coefficient between income inequality and demand for redistribution. Our results would differ from those of Perotti, in the sense that expenditure is in States is not financed directly form taxation. Income to use as government expenditure for each state comes from the federal government by formula, such that poorer estates are not self financing their own expenditures, (see Cayeros, 1995). Richer States may gather around 90% of their own government expenditure, but poorer estates may just gather 20% of it. It is a well know fact that the Mexican Government has been fighting in the last years Reform its Federal system of taxation. Therefore even if Federal government increases taxation, that relationship is not directly linked with an increase in government expenditure by States.

1. The estimation shows that the more unequal the distribution of income is, the lower the higher total government expenditure is.

2. Column (1) shows that the higher the total government expenditure is, the higher the growth will be, this result together with the previous one make the overall relation between inequality and growth is positive and very significant.

[42] Sempere-Sobarzo (1998).

This implies that results in the current section are consistent with the findings in all previous sections. Inequality and growth are positively correlated. In this case via government expenditure.

One interesting fact is that the higher the share of the population aged 65 or more, the higher the government expenditure will be. As Perotti argues, this happens because older people need more social care, social security, so that government expenditure should be higher.

Literacy coefficients still showing the partial correlation effect between them and they are not significant.

Table 13. Structural Estimation. Fiscal Approach

Estimation Method	(Growth) FE	(EXP) FE	(Growth) RE	(EXP) RE
Income	-.5159453** (0.0555587)		- .0282027** (0.0117427)	
Inequality		.367482 (0.230145)		2.238682** 0.4408971)
F-Literacy	.1037198 (0.0881061)		.0757281 (0.0617059)	
M-Literacy	-.3332556** (0.1206781)		-.0738521 (0.0980025)	
EXP	.0536399** (0.0069648)		.0310512** (0.0085235)	
Pop65		4.539486** (0.2816646)		1.500914** (0.3020056)
Depriva		--		- .0152619** (0.1198809)
Obs	341	341	341	341
Hausman	chi2(4) = 71.81			
Period	1989-2001	1989-2001	1989-2001	1989-2001

Standard errors are in brackets.

Growth represents the average growth rate 1989-2002.

Schooling, F-Literacy, M-Literacy and Inequality are all for 1989-2000.

EXP is government expenditure in 1989-2000.

EXP is government expenditure in public construction 1989-2000.

Pop65 is the share of population of 65 years or more in 1980, 90, 97, 2000.

Depriva is the deprivation index for 1990, calculated by Sempere & Sobarzo (1998)

Inequality comes from the gini using ENIGH 1989-2000.

Estimation is repeated but using Government expenditure in construction and public fostering only. The relationship is still positive and significant.

Table 14 Structural Estimation. Fiscal Approach

Estimation Method	(Growth) FE	(EXP-Contr) FE	(Growth) RE	(EXP-Const) RE
Income	-.020144* (0.010233)		-.0201449* (0.0102339)	
Inequality		.3241877 (0.324919)		2.433465** (0.5138075)
F-Literacy	-.0179362 (0.0942073)		.0468938 (0.0592744)	
M-Literacy	.0468938 (0.0592744)		-.0179362 (0.0942073)	
EXP-Cntr	.0594348** (0.0092255)		.0169082** (0.0066872)	
Pop65		3.95058** (0.407827)		1.960546** (0.3437326)
Depriva		--		.1400234 (0.1354987)
Obs	341	341	341	341
Hausman	chi2(4) = 47.89			
Period	1989-2001	1989-2001	1989-2001	1989-2001

A sensible explanation of why are economic growth and government expenditure positively related, is the fact that government expenditure enters in the equation of national accounts via all the growth enhancing expenditure like construction. However, only 20% of it is spent in that account, the remaining is devoted to debt payments and public administration.

A simple correlation analysis will show us that expenditure in construction is positively correlated to Growth whereas transferences expenses and other expenses are negatively correlated.

Table 15 . Government Expenditure Correlations Coefficients

t-1	Growth(t)	Total Expenditure(t-1)
Total Expenditure	0.1423 (0.0085)	1
Administration	0.0365 (0.4829)	0.9716 (0.0000)
Construction	0.0040 (0.9382)	0.8984 (0.0000)
Debt	0.0522 (0.3178)	0.6927 (0.0000)
Disponibilidades	0.0428 (0.4186)	0.7759 (0.0000)
Third Parties	0.0207 (0.7553)	0.5419 (0.0000)
Other	-0.0808 (0.6829)	---- (1.0000)
Transferences	-0.3285 (0.0000)	0.7885 (0.0000)

* Standard Errors are in brackets.

Section XI. Conclusions and Possible Extensions

Results coming from this work have to be treated with reasonable caution due to the limited amount of data we have. In this work, we have found that analysing personal income[43] using dynamic panel data methods for 1960-2000, the short term relationship between income inequality and economic growth is positive. The analysis did not reveal a different behaviour from that of household income for 1984-2002[44].

The puzzle arises when we used OLS regressions for different trade periods as we found that inequality is negatively correlated for Non-Trade and GATT periods in Mexico, and positively related to growth in the NAFTA; for both data sets. In particular, when households income is used, GATT period deliver a positive coefficient of income inequality, but if we account for the effects of oil producer

[43] We calculated the income Gini coefficient with grouped data in this case.

[44] We calculated the income Gini coefficient with micro data in this case.

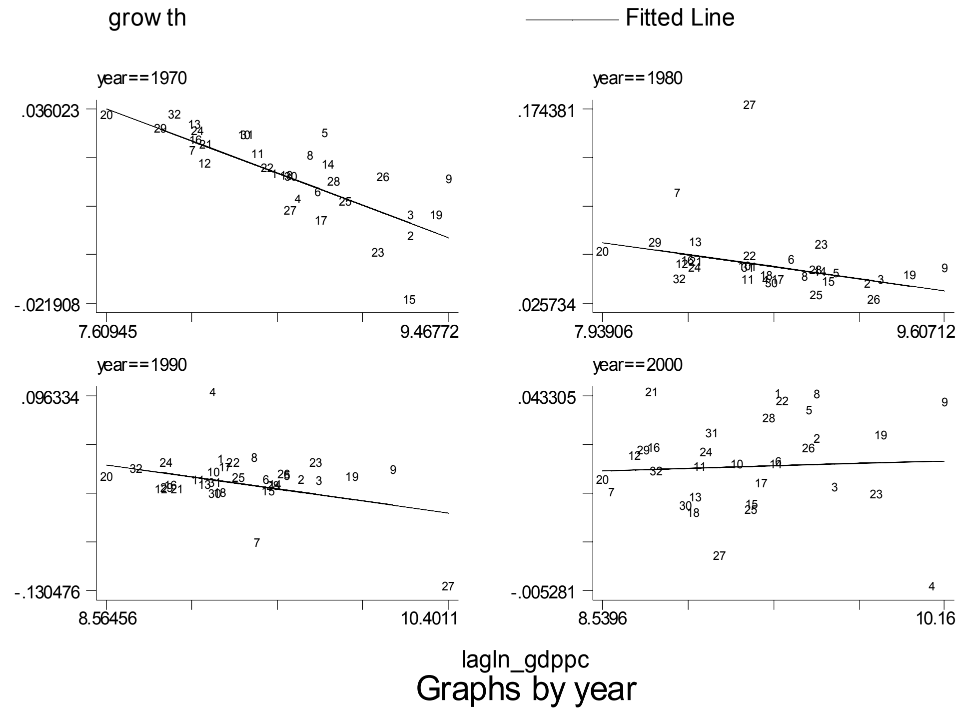
States GATT period coefficient of inequality becomes negative. In general, for both data sets using OLS, results suggest that as the Mexican economy becomes more open, the relation between growth and inequality is changing[45]. As we can see the relationship depends on the period we are analysing.

We should notice that personal income survey provides the monetary perception of the occupied people (except the ones that will start a job) in terms of current minimum wage in the quarter that the survey was performed. Whereas household income is the quarterly monetary income, which is the result of adding six different sources of income: 1) earned income, 2) self-employment income, 3) property rents income, 4) Income from cooperatives, 5) Transferences, and 6) Other sources of monetary income.

An economic shock will affect personal income faster that it will affect household income. Economic shocks may have a lagged effect in household income, say the crisis in 1994, the impact in personal incomes due the 1994 crisis was strong enough to make inequality reflect a severe increase (+ change). Despite the crisis of 1994 there was economic growth from 1990-2000 (+ change). These two facts explain the positive relation between these two variables for NAFTA period (1990-2000) and may imply that inequality grew faster than economic growth due to the financial crisis of 1994 and slow positive benefit of the economic integration.

[45] Which does not contradict Barro results, described in Banerjee and Dufflo(1999).

Figure 2. Growth for the periods used by the Grouped Data (1960-2000)



Our results are also consistent with convergence theory of previous works on Mexico (Rodriguez-Pose, 1999) because the coefficient of lagged GSP[46] is changing with time. It is negative from (1960-1990) and positive for (1990-2000). See figure 2.

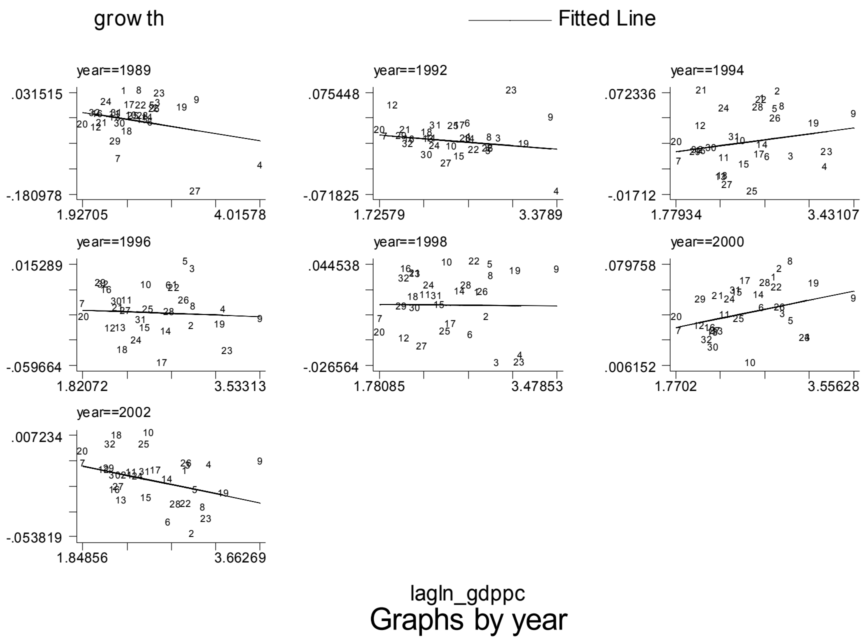
When the GATT period 1980-1990 is subdivided in three periods 1984-1989, 1989-1992 and 1992-1994, we can observed that the relationship changes to positive in the last period (see figure 3). This change is heavily influenced for the GSP results of year 1994, given the Mexican economic crisis in December 1994. Then when NAFTA is subdivided in four periods, we can observed that period 1994-1996 and 1996-1998 show neither convergence nor divergence. 1994-1996 is mostly influence by the 1994 crisis, and 1996-1998 receives lagged influence from the 1994 crisis and the first effect from trade openness, (NAFTA at seven, 2001). The 1998-2000 period exhibits a positive coefficient, which again implies that as the economy becomes

[46] Gross State Product.

more open, the driving forces of growth are such that states are diverging. This coincides with a positive relationship between growth and income inequality for that period. Finally, period 2000-2002 is again showing convergence[47].

Finally, Perotti's structural form, with a fiscal approach shows that the relation between inequality and growth is always positive. Results remain the same when we introduce deprivation as a proxy for well-being of the population in each State.

Figure 3. Growth for the periods used by the Household Income Set (1984-2002)



<u>Caveats</u>

Does this imply that we have to create more income inequality to achieve a higher rate of growth? What kind of policy implication it has? Is income inequality beneficial for growth as the economy becomes more open?

[47] Any comment about this period is still under research as 2002 GSP has been calculated using preliminary data, final GSP for 2002 will be available by the beginning of year 2004

Kuznets hypothesis is more suitable to describe the relationship between economic development and income inequality, within a context of the process of development and the role of the institutions. In the first stages of development of a country inequality increases as a results of skill level bias of the economic process, that leave behind the unskilled and embrace the skilled. However, once institutions are settle down, and fiscal and labour regulations laws passed, where income redistribution and unemployment benefits emerge, in addition to labour training (see Boot & Bryan, 2002), the poor are able to catch up with the middle class and the rich, closing the gaps among them, lessen income inequality. As a consequence, in the later stages of development of a country, inequality decreases.

However in our context we are looking for the reasons why income inequality will increase economic growth? We need to define what the determinants of inequality are and what the sources of economic growth are. We would do so in next chapter.

We also need to address the questions: why is convergence changing? If it is true, that inequality sign changes for GATT and NAFTA periods, does Kuznets hypothesis apply? Are there other factors that we should be considering, such as the stages of economic development?. Is there any real relation between the human capital variable, the income distribution and economic growth?[48].

Further research is performed in subsequent chapters looking at growth accounting factors, sources of growth and determinants of income inequality.

[48] This is analysed in a subsequent paper.

Bibliography

Arellano, M. and Bond S. (1991), " Some test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". Review of Economic Studies, 58, pp. 277-297.

Aghion, Philippe and Bolton, Patrick (1997), "A theory of Trickle-Down Growth and Development", Review of Economic Studies, 64, 151-172.

Aghion, Philippe and Williamson, Jeffrey, "Growth, Inequality and Globalization", Cambridge Press. Edition 1998.

Alesina, Alberto and Rodrik, Dani, (1994) " Distributive Politics and Economic Growth", Quarterly Journal of Economics, 465-490.

Banerjee, Abhijit, and Newman, Andrew (1993), " Occupational Choice and the process of development", Journal of Political Economics vol1, no.2, 274- 298.

Banerjee, Abhijit, and Duflo, Esther (1999), "Inequality and Growth: What can the data say?" Working paper MIT, 1-32.

Barro, Robert, and Sala-I-Martin, Xavier(1992), "Convergence", Journal of Political Economy, vol. 100. no. 21.

Barro, Robert J. and Lee, Jong Wha (1996), " International Measures of Schooling Years and Schooling Quality", AEA Papers and Proceedings, Vol 86 No. 218-223.

Barro, Robert J. and Lee, Jong Wha (2001), " Schooling Quality in Cross-Section of Countries", Economica, Vol 68 No. 272. 465-469.

Benabou, Roland (1996), "Inequality and Growth", NBER , 11-72.

Boltvinik, Julio and Hernandez Laos (1999), "Pobreza y Distribución del Ingreso en México". Siglo Veintiuno Editores. 354pp.

Chotikapanich, Duangkamon and Griffiths, William (2001), "On Calculation Of the Extended Gini Coefficient", Review of Income and Wealth. 541pp.

Díaz Cayeros, Alberto (1995), "Desarrollo Económico e Inequidad Regional en México", México, Editorial Porrua.

Esquivel, Gerardo (1999), "Convergencia Regional en México, 1940-1995", Cuaderno de trabajo de El Colegio de México, no. IX-99.

Forbes, Kristin (2000), "A reassessment of the relationship between inequality and growth", American economic review.

Galor, Oded and Zeira, Joseph (1993)" Income Distribution and Macroeconomics". Review of Economic Studies, 60,35-52.

García R., Adalberto, “La desigualdad Económica”, El Colegio de México. Primera Edición 1986.

Garcia R.,Ma. Elena, “Polarization of Income: The Mexican Case”. December 1999. Working paper at the Young Economist Conference. Oxford 2000.

Hammond, Peter and Rodriguez-Clare, A. (1993), “On endogenizing long-run growth “, Scandinavian Journal of Economics, 391-425.

INEGI. ENIGH (National Survey of Income and Expenditure of the Households) surveys for 1984,1989,1992,1994, 1996, 1998, 2000 and 2002.

Kanbur, Ravi (1996). “Income Distribution and Development”. Paper prepared for the forthcoming North-Holland Handbook on Income Distribution. pp 1-41.

Kuznets, Simon (1955). “"Economic Growth and Income Inequality", American Economic Review, 45, 1-28.

Ljungqvist, Lars (1993) “ Economic Underdevelopment. The case for a missing market for human capital” , Journal of development Economics 40 , 219-239

Loury, Glenn (1981), “Intergenerational Transfers and the distribution of earnings”, Econometrica, vol. 49, no.4. , 843-867.

Mookherjee, Dilip and Ray, Debraj (1998), "Persistent Inequality and Endogenous Investment Thresholds", MIMEO. working paper Boston University .

Nafta at Seven, (2001), "It's Impact on workers in all three regions". Briefing Paper. Economic Policy Institute. Washington D.C.

Paniza, Ugo (2002), "Income Inequality and Economic Growth: Evidence from American Data". Journal of Economic Growth, 7. 25-41.

Perotti, Roberto (1996), "Growth, Income Distribution, and Democracy: What the data Say.", Journal of Economic Growth, 149-187.

Quah, Danny (1997), "Empirics for Growth and distribution: Stratification, Polarization, and Convergence Clubs". Journal of Economic Growth, 2: 27-59.

Rodriguez-Pose, Andres and Sanchez-Reaza, Javier (1999), "Economic Impact of Trade and Liberalization on Mexico's Regional Growth", Working Paper in the Department of Geography and Environment, LSE.

Sempere, Jaime and Sobarzo, Horacio (1998), "Federalismo Fiscal en México". El Colegio de México, 227pp.

Secretaria de Economía, "Ingresos y Egresos de la Población de México (Investigación por Muestreo Octubre 1956)", Secretaria de Industria y Comercio, “Ingresos y Egresos de la Población de México (Investigación por Muestreo Julio 1958)”. Marzo 1960.

Secretaria de Industria y Comercio, “La 16 Ciudades Principales de la Republica Mexicana, Ingresos y Egresos Familiares. Agosto-Diciembre de 1960.

Secretaria de Industria y Comercio, “Ingresos y Egresos de la Población de México (1969-1970)”. Tomos I-IV.

Secretaria de Programación y Presupuesto, “Encuesta de ingresos y gastos de los hogares 1977, primera observación gasto seminal y mensual”.

Székely, Miguel (1996), “Explaining Changes in Poverty: Some Methodology and Its Application to Mexico”. Documento de trabajo de El Colegio de México, no. V-1996.

Appendix A. Panel Data Models with Lagged Dependant Variables

1) Base Model

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + X'_{i,t-1} B + u_{it} \tag{1}$$

Note:

$$\Rightarrow \Delta y_{it} = \alpha_i + (1-\lambda) y_{i,t-1} + X'_{i,t-1} B + u_{it} \tag{2}$$

2) Estimation: Invalid

We cannot validly apply either: usual FE estimator; nor usual RE estimator.

a) Problem with FE

FE estimator is implemented as LSDV and can be obtained by regressing.

$(y_{i,t-1} - \overline{y}_i)$ on $(y_{i,t-1} - \overline{y}_{i,-})$ and $(X_{i,t-1} - \overline{X}_{i,-})$

where:

$$\overline{y}_i = (T-1)^{-1} \sum_{t=2}^{T} y_{i,t} \quad \text{and} \quad \overline{y}_{i,-} = (T-1)^{-1} \sum_{t=2}^{T} y_{i,t-1}$$

$$\overline{X}_{i,-} = (T-1)^{-1} \sum_{t=2}^{T} X_{i,t-1}$$

Observe that

$$\overline{y}_i = \alpha_i + \lambda \overline{y}_{i,-1} + \overline{X}'_{i,-1} B + \overline{u}_i$$

Hence

$$(y_{i,t} - \overline{y}_i) = \lambda (y_{i,t-1} - \overline{y}_{i,-1}) + (X_{i,t-1} - \overline{X}'_{i,-1} B) + (u_{i,t} - \overline{u}_i)$$

But then $(y_{i,t} - \overline{y}_i)$ and $(u_{i,t} - \overline{u}_i)$ are correlated (due to the presence of $\overline{y}_{i,-}$ and $\overline{u}_i$).

Note : FE estimator of (2) suffers from an identical problem.

b) Problem with RE

RE estimator is FGLS using covariance matrix for $\upsilon_{i,t} = \alpha_i + u_{i,t} - \eta$,

where $\eta = E(\alpha_i)$.

$$Cov(\eta_{i,t}, \eta_{i,s}) = \begin{cases} \sigma_\alpha^2 + \sigma_u^2 & i = j \quad and \quad t = s \\ \sigma_\alpha^2 & i = j \quad and \quad t \neq s \\ 0 & otherwise \end{cases}$$

But the problem is that $y_{i,t-1}$ and $\upsilon_{i,t}$ are correlated due to the presence of α_i in both. Then conventional RE estimator is invalid.

Observer

$$y_i = \alpha_i j + \lambda y_{i,-1} + X'_{i,-1} B + u_i$$

$$y_i = \lambda y_{i,-1} + X'_{i,-1} B + \upsilon_i + \eta j$$

where $Var(\upsilon_i) = \sigma_\alpha^2 jj + \sigma_u^2 I_{T-1} = \Omega(\sigma_\alpha^2, \sigma_u^2)$

GLS:

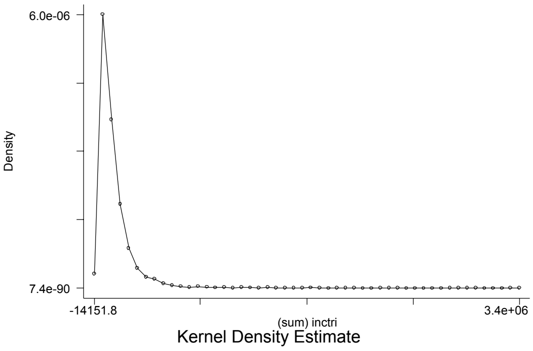
$$\min_{\alpha,\beta,\eta} y_i = \sum (y_i - \lambda y_{i,\_} \eta j - X'_{i,-1} B)' \Omega^{-1} (y_i - \lambda y_{i,\_} \eta j - X'_{i,-1} B) \lambda j$$

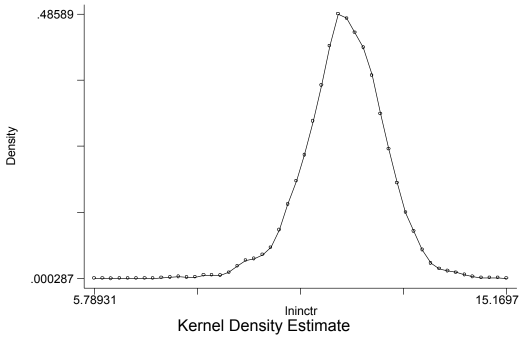
And the problem is that $p\lim \frac{1}{n(T-1)} \sum y_{i,\_}' \Omega^{-1} \upsilon_i \neq 0$

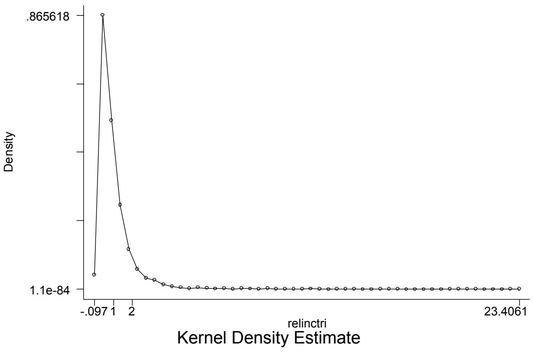
Appendix B . Gaussian Kernels

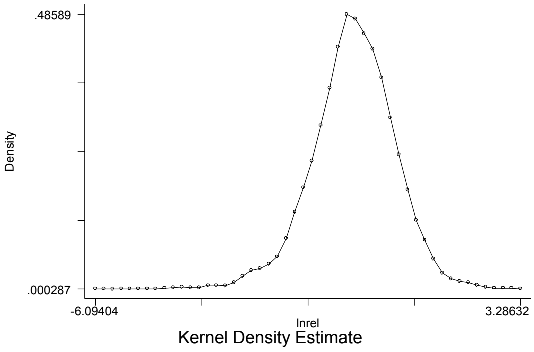
The following Gaussian kernels were calculated using the quarterly monetary income of the Households reported by INEGI in ENIGH surveys. Relative income for each household, is its income relative to the whole sample mean income.

4724 Households 1984



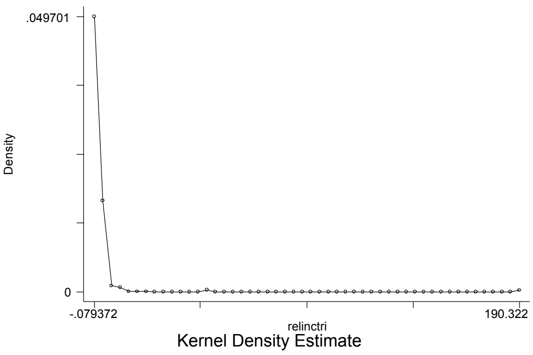


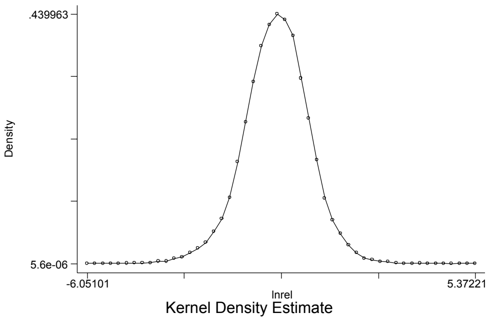




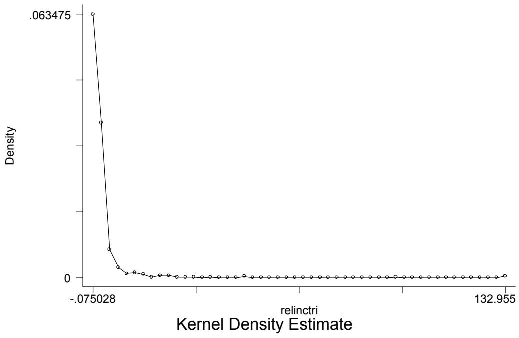
Note. Kernel plotting for the income (first figure) and relative income (third figure) are the same, given that the distribution is exactly the same. This argument also holds when we plot the natural log of the income and the natural log of the relative income. For space purposes we will only plot the figures three and four for each year.

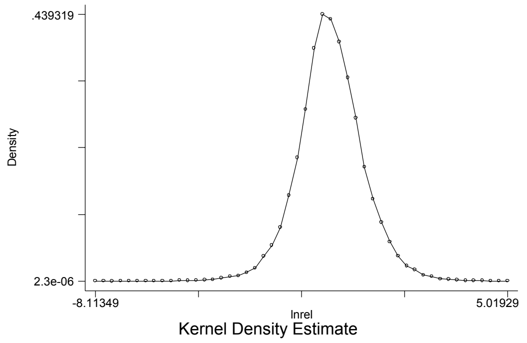
11517 Households 1989



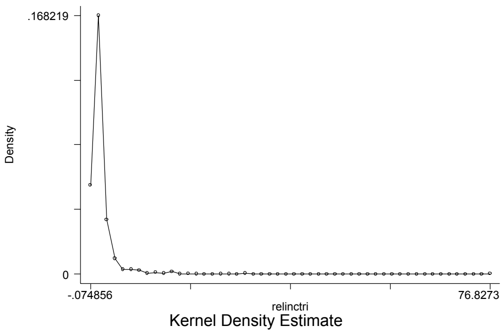


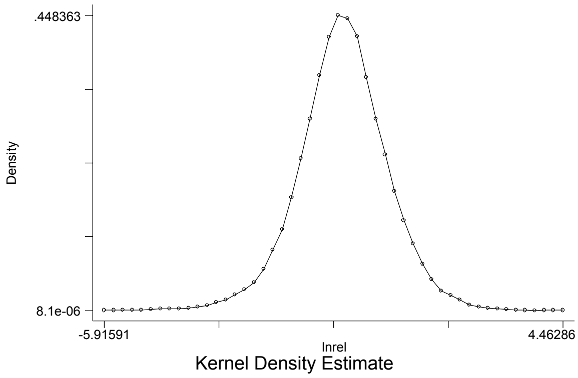
10508 Households 1992



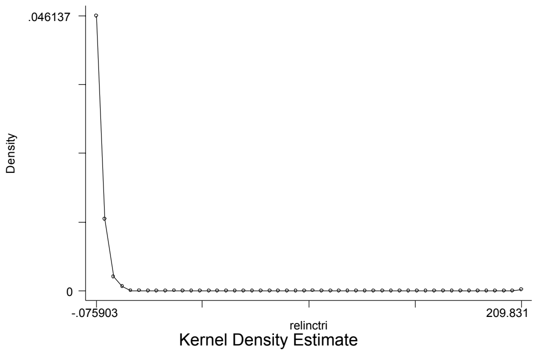


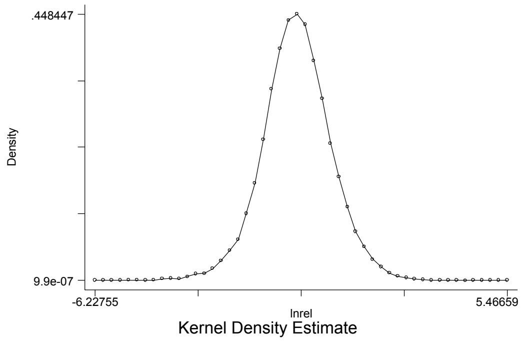
12791 Households 1994



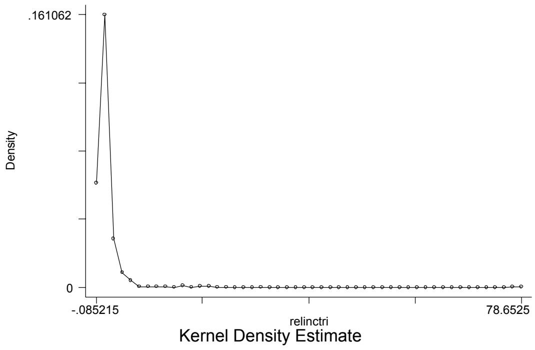


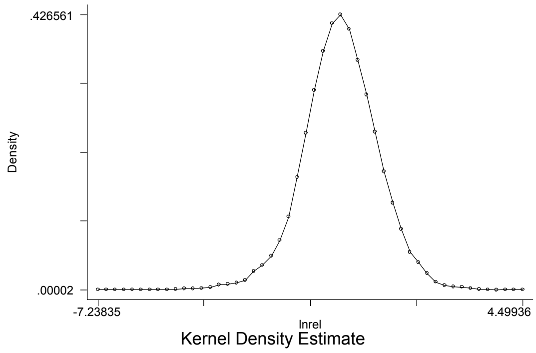
13075 Households 1996



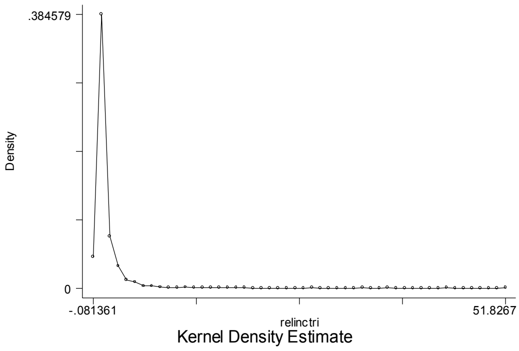


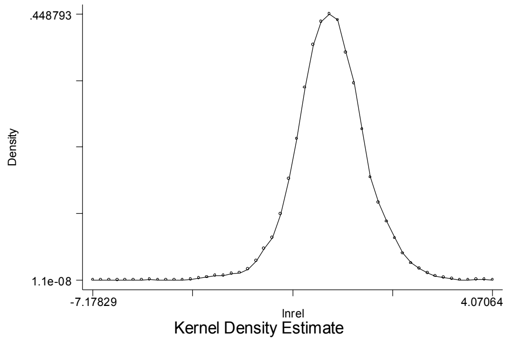
10098 Households 1998



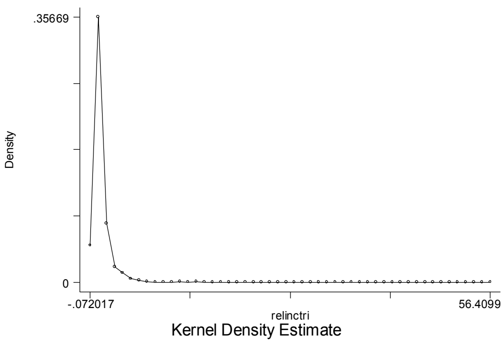


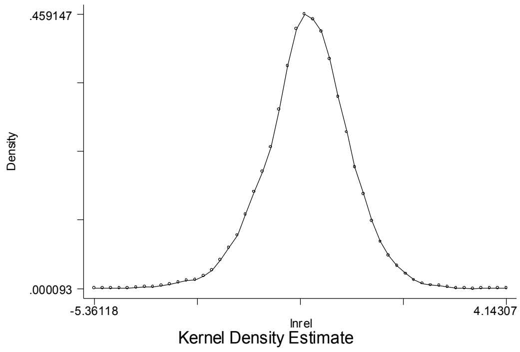
10089[49] Households 2000





17137 Households 2002





Note: Given that the number of household with income 10 times or more the mean is around 3% of the whole sample, we decided to drop those observations, and recalculate the kernel. Again, the kernel did not show much valuable information about the distribution of income given that the number of household above 1 time the mean was less than 9% altogether, we decided to drop the observations whose income is above 1 times the mean of the original sample. In all cases, we will specify the percentage of the observations that have been dropped. Dropping observations was only for plotting the kernel graph, we do not drop observations for estimation purposes.

[49] INEGI determined a sample size of 11170 households, but in the data set there are only 10089 that were included. The same applies for 2002, the estimated sample size was 19856 but we only were obtained .

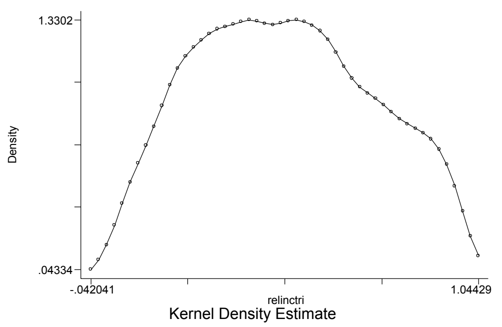
Gaussian Kernel for 1984
(Households with income below 1 times the mean)

Total Households= 4724

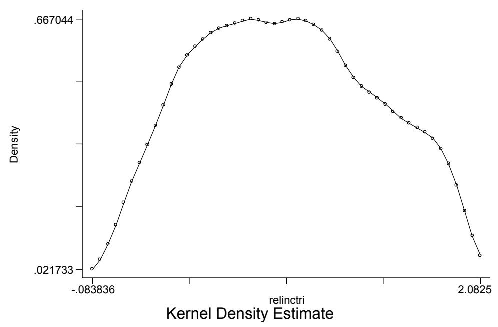
Households above the 10 times the mean income = 11
Households above the 5 times the mean income = 57
Households above the 2 times the mean income = 483
Households above the 1 times the mean income = 1548.

Drop all the Households with income greater than 1 time the mean income (33% dropped).
Original gini= .46497562
New gini= .28525216

Relative Income (original Data)



Relative Income (Remaining Data)



Note. Given that the kernel density plots for data below one times the mean, then 5 times the means and so on, are the same, i.e. distribution doesn't change, just boundary changes. For space purposes we will only plot the data remaining after dropping the observations that are 1 times below the mean.

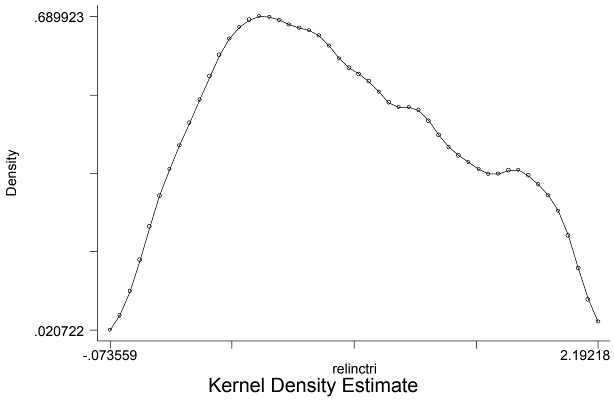
Gaussian Kernel for 1989
(Households with income below 1 times the mean)

Total Households= 11517

Households above the 10 times the mean income = 37
Households above the 5 times the mean income = 174
Households above the 2 times the mean income = 1117
Households above the 1 times the mean income = 3492.

Drop all the Households with income greater than 1 time the mean income (30.32% dropped).
Original gini= . .50836066
New gini= .30470994

Relative Income (Remaining Data) 1989



Gaussian Kernel for 1992
(Households with income below 1 times the mean)

Total Households= 10508

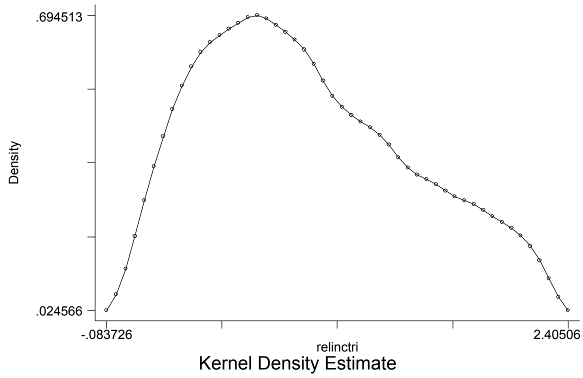
Households above the 10 times the mean income = 68
Households above the 5 times the mean income = 221

Households above the 2 times the mean income = 994

Households above the 1 times the mean income = 2658.

Drop all the Households with income greater than 1 time the mean income (25.29% dropped).
Original gini= . . .57674314
New gini= .32135624

Relative Income (Remaining Data) 1992



Gaussian Kernel for 1994
(Households with income below 1 times the mean)

Total Households= 12791

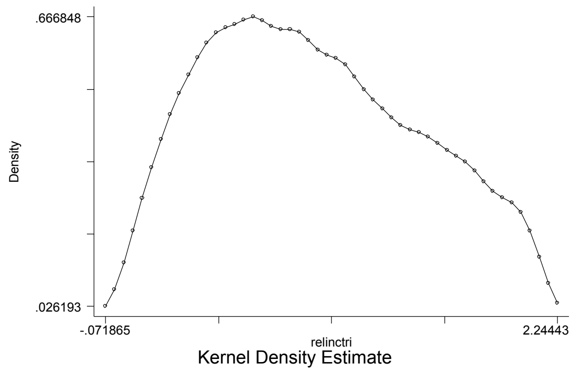
Households above the 10 times the mean income = 52
Households above the 5 times the mean income = 238

Households above the 2 times the mean income = 1334

Households above the 1 times the mean income = 3664.

Drop all the Households with income greater than 1 time the mean income (28.64% dropped).
Original gini= .5262996
New gini= . .30967255

Relative Income (Remaining Data) 1994



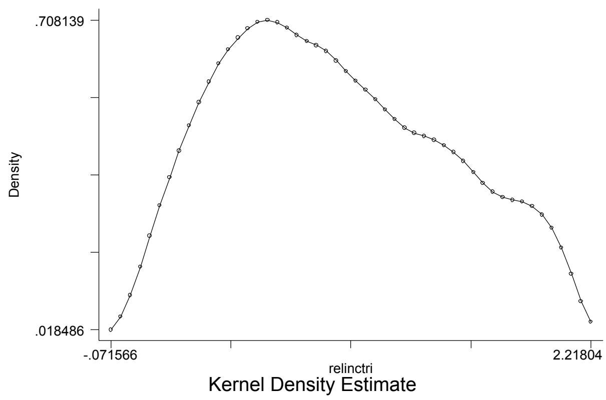
Gaussian Kernel for 1996
(Households with income below 1 times the mean)

Total Households= 13075
Households above the 10 times the mean income = 44
Households above the 5 times the mean income = 227

Households above the 2 times the mean income = 1340
Households above the 1 times the mean income = 3746.

Drop all the Households with income greater than 1 time the mean income (28.65% dropped).
Original gini= .52132036
New gini= .3022854

Relative Income (Remaining Data) 1996



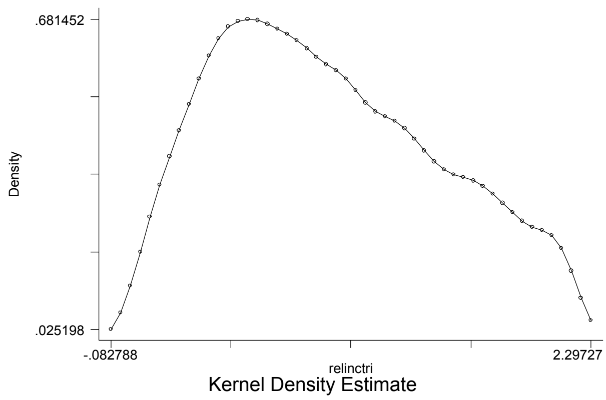
Gaussian Kernel for 1998
(Households with income below 1 times the mean)

Total Households= 10098
Households above the 10 times the mean income = 41
Households above the 5 times the mean income = 187

Households above the 2 times the mean income = 1065
Households above the 1 times the mean income = 2905.

Drop all the Households with income greater than 1 time the mean income (28.65% dropped).
Original gini= .53304116
New gini= .31742147

Relative Income (Remaining Data) 1998



Gaussian Kernel for 2000
(Households with income below 1 times the mean)

Total Households= 10089.

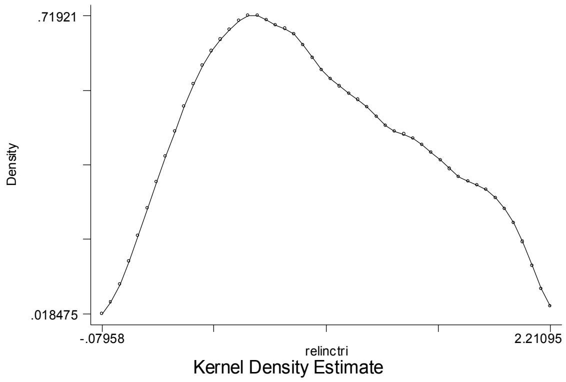
Households above the 10 times the mean income = 36
Households above the 5 times the mean income = 192

Households above the 2 times the mean income = 1040

Households above the 1 times the mean income = 2934.

Drop all the Households with income greater than 1 time the mean income (29.08% dropped).
Original gini= .51651525
New gini= .35631384

Relative Income (Remaining Data) 2000



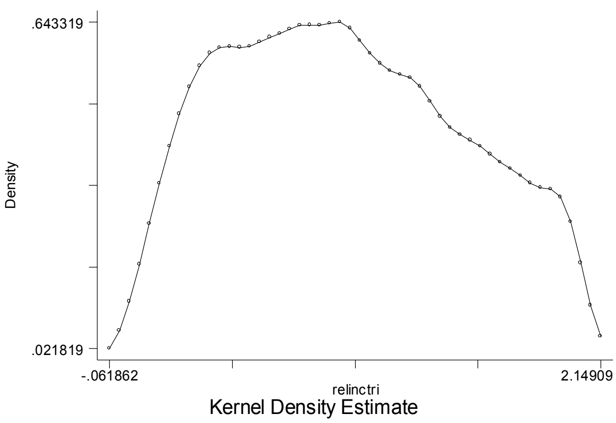
Gaussian Kernel for 2002
(Households with income below 1 times the mean)

Total Households= 17137

Households above the 10 times the mean income = 45
Households above the 5 times the mean income = 260
Households above the 2 times the mean income = 1840
Households above the 1 times the mean income = 5348.

Drop all the Households with income greater than 1 time the mean income (31.20% dropped).
Original gini= .49205247
New gini= .31766094

Relative Income (Remaining Data) 2002



COMPARING ENIGH SAMPLES

Year	Sample Size	Households Income>10	Households Income>5	Households Income>2	Households Income>1
1984	4724	11 0.23%	57 1.21%	483 10.22%	1548 32.77%
1989	11517	37 0.32%	174 1.51%	1117 9.70%	3492 30.32%
1992	10508	68 0.65%	221 2.10%	994 9.46%	2658 25.30%
1994	12791	52 0.41%	238 1.86%	1334 10.46%	3664 28.65%
1996	13075	44 0.34%	227 1.74%	1340 10.25%	3746 28.65%
1998	10098	41 0.41%	187 1.85%	1065 10.55%	2905 28.77%
2000	10089	36 0.35%	192 1.94%	1040 10.30%	2934 29.08%
2002	17137	45 0.26%	260 1.51%	1840 10.73%	5348 31.20%

Household Income>10 gives the number (and percentage) of households that are above ten times the mean income.