Income inequality regression models with applications

Adrian Mehic

STAM01: Master's thesis in Statistics (15 ECTS)

Lund University
School of Economics and Management
Supervisor: Krzysztof Podgórski
3 May 2017

Abstract

This thesis addresses three topics in income inequality. First, a cross-sectional dataset of 30 countries is used to investigate the causes of between-country differences in income inequality. Secondly, a panel of the same countries is used to examine the drivers behind the change in inequality between 1985 and 2013. This part of the thesis utilizes dynamic regression models for panel data, and compares the estimates from the two most common dynamic panel models. The problem of endogenity of explanatory variables is addressed, and possible solutions to this problem are discussed, with an emphasis on techniques based on the generalized method of moments (GMM). The main findings are that the trade-to-GDP ratio, the industrial employment share and the political color of the government are the most important explanatory variables of income inequality over time. The final question addressed in the thesis concerns cross-country income inequality convergence over time. Notably, inequality is shown to converge at slightly slower rate than reported by previous studies.

Contents

1 Introduction

In the wake of the late-2000s financial and economic crisis, the topic of income inequality has again become prominent in the national political debate in many countries. Income inequality, technological change, as well as a perceived increase in distance between ordinary people and the "establishment" are generally considered to be important factors in explaining the rise of new political and social movements on both sides of the traditional left-right spectrum. For example, the slogan of the Occupy Wall Street movement, "We are the 99%", refers to the bottom 99% of the income distribution as compared to the top 1%. The 2015-16 presidential campaign of U.S. senator Bernie Sanders was highly focused on highlighting the perceived increase in societal income disparity. This development has been mirrored in Europe, where left-wing political parties such as *Podemos* in Spain and *Syriza* in Greece have gained momentum. Moreover, the election of Donald Trump as President of the United States, as well as the UK vote to leave the European Union ("Brexit") are considered by many to be at least partially related to economic globalization and the loss of high-paid industrial jobs in certain parts of these countries. Additionally, the financial sector, considering its rapid expansion in virtually all nations of the world, has frequently been scapegoated for increased income inequality.

The purpose of this thesis is threefold. Firstly, to examine which factors affect the difference in income inequality between countries. Secondly, to assess why some countries have experienced an increase in income inequality, while others have not. Finally, the thesis will examine if there has been convergence in income inequality between countries. That is, if inequality is decreasing in countries where it is high, and conversely, increasing in countries with low income inequality. Specifically, in answering the first question, a cross-sectional dataset on 30 countries for the year 2013 is used, whereas a panel of the same countries between 1985 and 2013 is used for answering the second and third questions.

In the thesis, models based on the so-called *generalized method of moments* (Hansen 1982) will be emphasized in order to answer the questions formulated above. Hence, a rigorous introduction to this important parameter estimation technique will be given.

The thesis is structured in the following way. The problem of measuring income inequality is discussed in Section 2. In Section 3, a methodological background to the theories and models employed in the analysis of income inequality are presented, including a literature review. Section 4 outlines theoretical models for static and dynamic regression, with an emphasis on the generalized method of moments technique, which forms the basis for both the cross-sectional and panel data analyses. In Section 5, the data is presented and the regression models described in Section 4 are implemented to the empirical data. Finally, in Section 6, the results obtained in the previous parts are analyzed and discussed.

2 Fundamentals of income analysis

Income can be viewed both as a discrete and as a continuous variable. Assuming first that income is discrete, let $\mathbf{y} = (y_1, ..., y_N)'$ be the vector of incomes, where $0 \leq y_1 \leq y_2 \leq \cdots \leq y_N$. It is well-known that the mean income may be defined as $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$. Also, $F(y) = (\#y_i : y_i = y)/N$ denotes the cumulative probability.

Income may also be seen as a continuous variable. Denoting the sample space by Ω, it is possible to define the random variable income as a measurable function, $Y : \Omega \to \mathbb{R}_+$. The corresponding probability density function (p.d.f.), $f(y)$ describes how the income is distributed on $\mathbb{R}_+$. The cumulative density function (c.d.f.) is denoted $F(y)$, and the cumulative probability is $F(y) = \int_0^y f(s)ds$. The mean income is $\mu = \int_0^\infty yf(y)dy$. It is usually easier, from a mathematical perspective, to view income as a continuous variable.

2.1 Lorenz curves

The cumulative income distribution can be graphically represented by the so-called *Lorenz curve* (Lorenz 1905). It displays on the x-axis the cumulative share of the population and on the y-axis the corresponding cumulative income. Hence, if income is uniformly distributed, the slope of Lorenz curve is equal to the diagonal of the unit square. An example of this would be if 50% of the population would have exactly 50% of income. Mathematically, let ξ_p denote the p:th quantile of $f(y)$; $p \in [0, 1]$. Also $\xi_p = F^{-1}(p) = \inf\{y \in \mathbb{R}_+ : F(y) \geq p\}$. Let now

$$Y_p = Y \mathbb{1}_{(Y \leq \xi_p)} \tag{1}$$

Then, the Lorenz curve is defined as

$$L(p) = \frac{E(Y_p)}{E(Y)} \tag{2}$$

This may equivalently be expressed as

$$L(p) = \frac{1}{\mu}\int_0^p F^{-1}(y)dy \tag{3}$$

From (3) it is follows that

$$L'(p) = \frac{dL(p)}{dp} = \frac{F^{-1}(p)}{\mu} \geq 0$$

Also,

$$
\begin{aligned}
L''(p) = \frac{d^2L(p)}{dp^2} &= \frac{1}{\mu}\left[\frac{d}{dp}F^{-1}(p)\right] \\
&= \frac{1}{\mu}\left[\frac{1}{F'(F^{-1}(p))}\right] \\
&= \frac{1}{\mu}\left[\frac{1}{f(\xi_p)}\right] \\
&\geq 0
\end{aligned}
$$

where the second-to-last equality follows from the inverse function theorem. Since $L'(p) \geq 0$ and $L''(p) \geq 0$ for all p, the Lorenz curve is an increasing, monotonic and convex function that is differentiable on [0,1]. The convexity of the Lorenz curve makes intuitive sense since the rate at which incomes are added to curve as $p \to 1$ is increasing. In addition, convexity here implies that the curve is always below the diagonal, and if the Lorenz curve would touch the diagonal at one point, then it would have to be equal to the diagonal over the entire range.

2.2 The Gini coefficient

The most common measure of income inequality is the so-called *Gini coefficient*, or *Gini index* (Gini 1912). It will be denoted as G throughout this thesis. It can be interpreted and derived in many different ways, and its simplicity has made it popular especially amongst economists. There are several other measures of income disparity; this topic is, however, beyond the scope of this paper. The Gini coefficient is named after the Italian statistician Corrado Gini (1884-1965), who developed the measure in the early 20th century.

2.2.1 Definition

There are several ways of viewing the Gini coefficient. The simplest approach is geometrical and is based on the Lorenz curve. If income in a certain population were evenly divided; i.e. if everyone had the same income, the Lorenz curve would be a straight line and $L(p) = p$. Thus, one may examine the deviation of the empirical Lorenz curve from the 45 degree line. Since the Lorenz curve has percentages on both the x- and y-axes, the area of the unit square in which the Lorenz curve is located is equal to one. Thus, each of the two triangles that are formed by constructing the 45 degree line has an area of 1/2. Viewed from this geometrical perspective, the Gini coefficient is calculated as

$$G = \frac{A}{A+B} \tag{4}$$

where A is the area between the line of perfect equality and the Lorenz curve, and B is the area between the Lorenz curve and the line of perfect inequality.

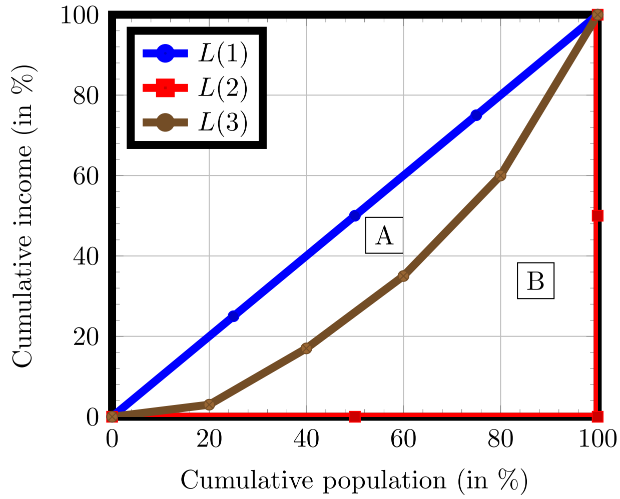

Figure 1: Three different Lorenz curves. The blue curve, $L(1)$, corresponds to perfect equality of incomes ($G = 0$). The red one, L(2), represents perfect inequality ($G = 1$). The brown curve, $L(3)$, corresponds to $G = 0.34$. The areas A and B correspond to $L(3)$.

2.2.2 Interpretation

From the definition, it follows that $G \in [0, 1]$. There are two extreme cases, $G = 0$ and $G = 1$. If all individuals in the population have the same income, $G = 0$. This is a state of perfect income equality. On the other hand, $G = 1$ in the limiting case (for N large) when all income is concentrated at one point. The latter case, representing a situation with perfect inequality, can be viewed as a bimodal distribution with two modes, $y = 0$ and $y = y_{max}$. In this case, only one person in the population attains y_{max}, while the $N-1$ remaining individuals have $y = 0$. This is obviously a situation with extreme income inequality.

Figure 1 shows three different Lorenz curves. The blue curve represents perfect equality of incomes, i.e. $G = 0$, the red curve corresponds to $G = 1$, where only one person holds all income, while the brown curve represents the empirically more realistic situation where $G = 0.34$. The areas A and B in Figure 1 correspond to this final situation.

Proposition 1. The Gini coefficient is can be expressed by the formula

$$\begin{aligned} G &= 2\int_0^1 [p - L(p)]\, dp \\ &= 1 - 2\int_0^1 L(p)dp \end{aligned} \tag{5}$$

The above statement is of importance, as it relates the Gini coefficient to the Lorenz curve directly, without having to define the areas A and B.

2.2.3 Estimation - covariance approach

This section considers the practical calculation of the Gini coefficient, introducing two distinct estimation methods, yielding the estimates $\hat{G}_1$ and $\hat{G}_2$, respectively. First, the Gini coefficient can be seen as the covariance between Y and the empirical c.d.f. $\hat{F}_y$. In order to see this, substitute (3) into (5). This yields

$$\begin{aligned} G &= 1 - \frac{2}{\mu}\int_0^1 \int_0^{F^{-1}(p)} ydF(y)dp \\ &= 1 - \frac{2}{\mu}\int_0^1 \int_0^{x} ydF(y)dF(x) \\ &= 1 - \frac{2}{\mu}\int_0^\infty \overbrace{\int_y^\infty dF(x)}^{=F(\infty)-F(y)} ydF(y) \\ &= 1 - \frac{2}{\mu}\int_0^\infty yS(y)dF(y) \\ &= 1 - \frac{2}{\mu}\int_0^\infty [1-F(y)]ydF(y) \\ &= 1 - \frac{2}{\mu}\int_0^\infty ydF(y) + \frac{2}{\mu}\int_0^\infty yF(y)dF(y)dy \\ &= 1 - \frac{2}{\mu}\overbrace{\int_0^\infty yf(y)dy}^{=\mu} + \frac{2}{\mu}\int_0^\infty yF(y)f(y)dy \\ &= \frac{2}{\mu}\int_0^\infty yF(y)f(y)dy - 1 \end{aligned}$$

The plug-in estimator of G is clearly

$$\hat{G}_1 = \frac{2}{\hat{\mu}}\int_0^\infty y\hat{F}_N(y)\hat{f}_N(y)dy - 1 \tag{6}$$

where $\hat{f}_N(y)$ is the kernel density estimator and $\hat{F}_N(y)$ is the empirical c.d.f. It is possible apply the Lindeberg-Lévy central limit theorem to (6); it can be shown that $\sqrt{N}(\hat{G}-G) \xrightarrow{L} \mathcal{N}(0, \mathbb{V}_{\hat{G}})$, where $\mathbb{V}_{\hat{G}}$ is the asymptotic variance of $\hat{G}$; see e.g. Davidson (2009) for further details.

Since $\mathbb{E}[Y] = \mu$, $\mathbb{E}[F(Y)] = \frac{1}{2}$ and $\mathbb{E}[YF(Y)] - \mathbb{E}[Y]\mathbb{E}[F(Y)] = \mathbb{C}[Y, F(Y)]$, the expression for the Gini coefficient can be written as

$$\begin{aligned} G &= \frac{2}{\mu}\left[\int_0^\infty yf(y)F(y)dy - \frac{\mu}{2}\right] \\ &= \frac{2}{\mu}\{\mathbb{E}[YF(Y)] - \mathbb{E}[Y]\mathbb{E}[F(Y)]\} \\ &= \frac{2}{\mu}\mathbb{C}[Y, F(Y)] \end{aligned} \tag{7}$$

Hence, the Gini coefficient is proportional to the covariance between income and the cumulative distribution of income. However, it is possible to estimate G without having to estimate the density of income. In fact, it is not even necessary to assume that income has a density, which would be the case for discrete models. Let $y_{(i)}, i = 1\dots, N$ denote the order statistics, R_i denote the rank of the observations, $\bar{R}$ denote the mean rank, and $\overline{\hat{F}_N(\boldsymbol{y})}$ be the mean of $\hat{F}_N(\boldsymbol{y}) = (\hat{F}_N(y_1), \dots, \hat{F}_N(y_N))$. Now, equation (7) can be estimated as

$$\begin{aligned} \hat{G}_2 &= \frac{2}{\bar{y}N}\sum_{i=1}^{N}(y_i - \bar{y})(\hat{F}_N(y_i) - \overline{\hat{F}_N(\boldsymbol{y})}) \\ &= \frac{2}{\bar{y}}\left[\frac{1}{N}\sum_{i=1}^{N} y_{(i)}\hat{F}_N(y_{(i)}) - \bar{y}\overline{\hat{F}_N(\boldsymbol{y})}\right] \\ &= \frac{2}{\bar{y}N}\left(\frac{1}{N}\sum_{i=1}^{N} y_i R_i - \bar{y}\bar{R}\right) \end{aligned} \tag{8}$$

Thus, from (8), it is clear that it suffices to know the rank of incomes in addition to the observations $\{y_i\}$, in order to estimate the Gini coefficient.

2.2.4 Estimation - expected gains approach

Another approach to the Gini coefficient is the so-called expected gains method. It was first proposed by Pyatt (1976). Assume that one person from the sample and one income from $\mathbf{y} = (y_1, \dots, y_N)'$ is selected at random, say y_j. The individual selected is then offered to move to that income level; the individual's current income is denoted by y_i. Note that $\Pr(y_i = y_j) = 1/N$. The average expected payoff, or expected gain, for all individuals in the sample is

$$\mathbb{E}[gain|i \to j] = \frac{1}{N}\sum_{j=1}^{N}\max(0, y_j - y_i),\ \forall i \tag{9}$$

Clearly, an individual would not switch incomes if $y_j < y_i$, and if $y_j = y_i$, the individual would be indifferent between moving and staying. In order to derive the Gini coefficient, it is assumed that income is divided into k groups of equal size. This setting is slightly more realistic and eases calculations considerably.

Now, let $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)'$ be the vector of mean incomes, where each element of $\boldsymbol{\mu}$ is the mean income of group i. The proportion of individuals in group i is denoted by p_i. It holds that $\sum_{i=1}^{k} p_i = 1$. The p_i:s are stacked into $\mathbf{p} = (p_1, ..., p_k)'$. Furthermore, let $\mathbf{E}$ be a k x k matrix where the element E_{ij} is

$$E_{ij} = \max(0, \mu_j - \mu_i)$$

which implies that $\mathrm{tr}(\boldsymbol{E}) = 0$. Using this notation, the Gini index can then be calculated as

$$\hat{G}_3 = (\boldsymbol{\mu}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{E}\mathbf{p} \tag{10}$$

3 Methodology

3.1 Background

This section briefly discusses some different aspects of income inequality in a non-technical fashion. It also provides some insight into previous empirical research on the matter.

3.2 Empirical topics in income inequality

3.2.1 General outline on covariates affecting income inequality

One of the most comprehensive studies is due to Jaumotte et al. (2008). In that paper, the Gini coefficient is used as the dependent variable in a panel model of 51 countries, both advanced and developing; the time period covered being 1981-2003. The findings suggest that the most significant variables for explaining change in income inequality are credit to the private sector, the ratio of inward FDI (Foreign Direct Investment) to GDP and the industrial employment share. The first two variables have a positive correlation with income inequality; that is, an increase in one of the two causes an increase in Gini, while industrial employment share is negatively associated with Gini. Furthermore, the export-to-GDP ratio as well as the inverse of the tariff rate are both significantly negatively correlated with Gini, which suggests that trade globalization is associated with lower income inequality (ibid.).

3.2.2 Globalization variables

When it comes to trade globalization, most studies have focused on developing countries, rather than on advanced economies. Chakrabarti (2000) uses a panel of both high- and low-income countries, and finds that an increase in trade-to-GDP significantly reduces income inequality. Other papers, for example Aradhyula et al. (2007) have reached the opposite conclusion. Consequently, more empirical studies are likely required in order to over-bridge the current research gap regarding the effect of trade flow increases on societal income disparity.

As mentioned before, another globalization variable is industry's share of total employment. It has been decreasing in most developed economies as a consequence of globalization; lower tariff rates means that it is cheaper to export from low-wage countries, and hence production has shifted from high-wage countries to low-wage countries. On the other hand, if it is true that trade-to-GDP-ratio is negatively correlated with income inequality, there exists a globalization paradox, at least in countries where globalization has caused a decline in the relative importance of the industrial sector. Consequently, it is not *a priori* clear which of the two effects dominates. It is also likely that the share of the labor force employed in the agricultural sector is important in explaining cross-country income inequality, as agricultural jobs are often comparatively poorly paid. However, agricultural employment is likely not a suitable variable in explaining the change in inequality over time, since the share of labor force employed in the agricultural sector is declining, and has been doing so for over a century, in virtually all countries of the world.

3.2.3 Domestic variables

Apart from globalization variables, such as the trade-to-GDP-ratio and the industrial employment share, inequality of incomes is likely affected by domestic policy, such as regulatory changes, which can be difficult to quantify. However, it is reasonable to believe that left-wing parties exhibit a greater willingness to reduce income inequality, while right-wing parties are more focused on economic growth. Empirical research give support to the claim that growth rates tend to increase during the tenure of right-wing governments (Aidt et al. 2016). Given this result, an interesting question would therefore be if income inequality as measured by Gini is decreasing at times of left-leaning governments, or equivalently, increasing during right-wing governments. To the best of this author's knowledge, the effect of political ideology on income inequality as measured by Gini has never been empirically investigated.

Government actions, regardless of its political color, can also be a contributing factor in reducing or increasing inequality. For example, Dabla-Norris et al. (2015) has shown that increased government spending over time reduces income inequality. Additionally, it is reasonable to believe that countries with higher government expenditure have lower values of Gini than countries with lower levels of government expenditure. Government-funded social programs usually target the bottom earners in a society; hence, higher government spending should, to some extent, equalize income distribution. However, empirical research has not found a definite evidence on this matter. For example, cross-country studies by Lundberg and Squire (2003), de Mello and Tiongson (2006) and Jauch and Watzka (2012) have found that government spending has in fact exacerbated income inequality. A country-specific study using only data from Brazil points in the same direction (Clements 1997). One plausible reason may be that increased education spending, which is often a great proportion of government expenditure, does not benefit the low income earners, since it is often directed at universities and other institutions of higher education. In many countries, the working class does not have access to tertiary education, and hence, government spending on education can in fact worsen inequality.

3.2.4 Financial globalization

The financial sector has grown considerably in size in virtually all countries. For example, in the United States, the financial sector contributed to 7.2% of total GDP in 2015, compared to 4.8% in 1980 (Bureau of Economic Analysis 2017). More importantly, empirical research (cf. Philippon and Reshef 2012) has shown that workers in the financial sector tend to earn more than workers in other sector with the same education level. For top executives, this premium can be as high as 250% (ibid.). Moreover, higher dividends and greater capital gains on the stock market will disproportionally benefit those with large holdings of liquid assets, further increasing income inequality. However, while it is true that wealthy individuals tend to invest more in stock market than low-income earners, since the early 1980s, the relative importance of risky assets in the savings portfolio of the broad middle-class in many Western countries has increased (Keister 2000, p. 70-71). The emergence of Internet-based platforms for asset trading has further exacerbated this trend. Instead, credit expansion is likely a better predictor for income inequality than stock market returns, since credit is even more unequally distributed than income (Denk and Cournède 2015).

There have been quite a few empirical studies on the relationship between income inequality and credit expansion. Focusing on middle- and high-income countries, Beck et al. (2007) and Clarke et al. (2006) have found that private credit decreases inequality as measured by Gini. Other studies, for example Gimet and Lagoarde-Segot (2011), Fournier and Koske (2013), and the aforementioned Jaumotte et al. (2008) and Jauch and Watzka (2012) have obtained the opposite result.

3.2.5 Between-country convergence

Solow (1956) and Swan (1956) independently claimed that per capita income will converge between countries; this is an important feature of the eponymous *Solow-Swan neoclassical growth model*. Basically, this implies that poorer countries should grow faster than richer ones, until their economies have reached the same per capita income. However, according to the Solow-Swan model, there can be full convergence only if all countries have access to the same technology. Since this is an unrealistic assumption, there will only be partial convergence in practice. Several empirical studies have tried to estimate the speed of growth convergence. A highly-cited paper by Barro and Sala-i-Martin (1991) put the rate at 2% per year; this has given rise to the "two percent convergence rule" (Burda and Wyplosz 2013, p. 85). Decades after the development of the Solow-Swan model, it was shown that convergence of income is limited not only to the mean, but that it can also be applied to the variance of incomes (Bénabou 1996). However, convergence of income inequality has received considerably less attention in the literature. Bénabou (1996) estimates the average rate of convergence at -3.9%, Ravallion (2003) at -2.8%, while Bleaney and Nishiyama (2003) put the figure at approximately -1.3%, and Dhongde and Miao (2013) at -2.4%. Hence, the average convergence speed seems to be hovering around 1-3% on a yearly basis.

Additionally, both for inequality convergence and growth convergence, the speed of convergence is decreasing according to most studies (Abreu et al. 2005). This means that the cross-country gap seems to be closing at a decreasing pace.

3.3 Problems in analyzing income inequality

As with any problem involving empirical analysis, there are a number of issues arising when analysing income inequality. Often, data from national statistics agencies is used. Hence, definitions on some variables might differ between countries. This is particularly true for the Gini coefficient. For example, in some studies, the Gini coefficient net of tax is used; in other papers, the gross measure is used. Moreover, there are several different organizations that provide Gini data (Milanovic and Squire 2005).

To remedy the problem of different data sources, Solt (2016) has developed the so-called *Standardized World Income Inequality Database* (SWIID), with the purpose of standardizing Gini observations from different sources. To briefly summarize the SWIID approach, it uses data from several databases, wherein one of the databases, namely the Luxembourg Income Study (LIS), is chosen as the "reference" data, to which all other series are harmonized. However, the LIS has relatively few data points, and hence, requires imputation. By using loess regression on the data points of each of the other series used in constructing the harmonization, it is possible to predict the missing points using the coefficients from those regressions. For each point estimate, the SWIID uses five-year moving average smoothing, with twice as much weight on the estimate for the current year. To further account for the uncertainty associated with the imputation, each of the predicted variables is re-generated 1,000 times using Monte Carlo simulations, of which 100 are reported in the database. Then, it is possible to calculate the mean of these 100 observations, in order to obtain point estimates and the corresponding confidence intervals for each year. In the empirical section of this thesis, the SWIID will be used as the data source.

Another problem arises when analyzing the trade-to-GDP ratio. There is a reason to believe that the trade-to-GDP ratio, if used as an explanatory variable for income inequality, may be *endogenous*, that is, correlated with the error term. In the trade-to-GDP case, the reason for endogenity is likely that richer countries tend to trade more, for reasons other than trade openness (Frankel and Romer 1999). Hence, Frankel and Romer (1999) construct an alternative measure of the trade share, based on geographical variables such as proximity between trading partners. Geographical factors are likely not affected by variables omitted from the model, and given this, the constructed trade share is highly correlated with the true trade share. Hence, it is reasonable to use this variable in lieu of the actual trade share. A variable that is used instead of an endogenous variable in this fashion, is called an *instrument* for that variable. In

Section 4.1, endogenity and instruments will be discussed from a more technical perspective.

Additionally, there are a number of statistical issues that arise when introducing time-dependence, mainly concerning the stationarity of panel models. These will be discussed in Section 4.2.3.

4 Regression models for income inequality

In this thesis, the goal is to study which variables affect income inequality today, as well as how the change over time in certain variables has affected income inequality. Section 4.1 presents the theory behind the time-invariant regression models, while Section 4.2 deals with the time-dependent models.

4.1 Static linear regression models

Assume the standard regression model

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta} + u_i \tag{11}$$

for $i = 1, \dots, N$, where y_i and u_i are scalars, $\boldsymbol{x}_i$ is $1 \times K$ and $\boldsymbol{\beta}$ is $K \times 1$. Note that the above expression defines a generic regression model, so y_i does not need to be related to income or income inequality. Stacking the $\boldsymbol{x}_i$:s into one matrix gives $\boldsymbol{X}$, which is $N \times K$. Moreover, let $\boldsymbol{y} = (y_1, \dots, y_N)'$. The expression for the least squares estimate of $\boldsymbol{\beta}$ is then

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{12}$$

An estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is said to be *asymptotically consistent* if $\lim \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ as $N \to \infty$. By rewriting (12) and applying Slutsky's theorem, it is possible to show that the limit condition stated above is satisfied if and only if $\mathbb{E}[\boldsymbol{x}_i | u_i] = \boldsymbol{0}$ so that $\mathbb{E}[\boldsymbol{x}_i u_i] = \boldsymbol{0}$. In practice, this means that an observation $\mathbf{x}_i$ must not be correlated with the random error term u_i. However, in econometrics, it is often the case that there is some sort of dependence between one or several of the independent variables and the error term. An independent variable that is correlated with the error term is said to be *endogenous*, while a variable uncorrelated with the error term is called *exogenous.* In the case of endogenity, $\hat{\boldsymbol{\beta}}$ is not a consistent estimator of $\boldsymbol{\beta}$. To remedy this problem, the well-known concept of *instrumental variables* will be briefly reviewed.

4.1.1 Instrumental variables estimation

Following White (1982), introduce a sequence of independent, not necessarily identically distributed random $L \times 1$ vectors $\{\boldsymbol{z}_i\}$ that for all i satisfy the properties $\mathbb{E}[|u_i^2|^{1+\delta}] < \Delta$, $\mathbb{E}[|\boldsymbol{z}_{il} u_i|^{1+\delta}] < \Delta$ and $\mathbb{E}[|\boldsymbol{z}_{il} \boldsymbol{x}_{ik}^{1+\delta}|] < \Delta$ for some finite constants $\delta \in \mathbb{R}_+$ and $\Delta \in \mathbb{R}_+$, where $(k = 1 \dots K)$ and $(l = 1 \dots L)$, as well as $\mathbb{E}[\boldsymbol{z}_i u_i] = \boldsymbol{0}$ for all i. The last assumption means that there must not be any correlation between $\boldsymbol{z}_i$ and the random error term u_i. A sequence $\{\boldsymbol{z}_i\}$ that satisfies these conditions is called a sequence of *instrumental variables vectors*. Stacking the N instrumental variables vectors $\boldsymbol{z}_i$, gives the $N \times L$ matrix $\boldsymbol{Z}$. Note that the matrix $\boldsymbol{Z}$ contains both the exogenous variables in $\boldsymbol{X}$ as well as the instrumental variables for the endogenous variables in $\boldsymbol{X}$. Now, the instrumental variables estimator $\boldsymbol{\beta}^{\star}_{IV}$ of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^{\star}_{IV} = (\hat{\boldsymbol{P}}'_N \boldsymbol{Z}'\boldsymbol{X})^{-1} \hat{\boldsymbol{P}}'_N \boldsymbol{Z}'\boldsymbol{y} \tag{13}$$

where $\hat{\boldsymbol{P}}_N$ is an $L \times K$ projection matrix. It can be shown, for example in the appendix of White (1982), that $\sqrt{N}\boldsymbol{\Omega}_N^{-1/2}(\boldsymbol{\beta}^{\star}_{IV} - \boldsymbol{\beta}) \xrightarrow{L} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$, where $\boldsymbol{\Omega}_N$ is the asymptotic covariance matrix of $\boldsymbol{\beta}^{\star}_{IV}$. Depending on the choice of $\boldsymbol{P}_N$, different estimators of $\boldsymbol{\beta}$ are obtained. Setting $\boldsymbol{P}_N = \boldsymbol{I}$, yields

$$\hat{\boldsymbol{\beta}}_{IV} = (\boldsymbol{Z}'\boldsymbol{X})^{-1} \hat{\boldsymbol{Z}}'\boldsymbol{y} \tag{14}$$

which is the instrumental variables version of the LS estimator (12). This estimator only works if $L = K$. If, however, $L > K$, set $\boldsymbol{P}_N = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$ and the corresponding estimator (15) is called the *two-stage least-squares* estimator, $\hat{\boldsymbol{\beta}}_{2SLS}$. Some algebra yields the compact form for the analytical expression,

$$\hat{\boldsymbol{\beta}}_{2SLS} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \tag{15}$$

The name ”two-stage least squares” is because the procedure can be done by two separate regressions: First, LS regression of $\boldsymbol{X}$ on $\boldsymbol{Z}$, to obtain the fitted values $\hat{\boldsymbol{X}}$, and then a second LS regression of $\hat{\boldsymbol{X}}$ on $\boldsymbol{y}$ to obtain $\hat{\boldsymbol{\beta}}_{2SLS}$. Later in this thesis, it will be shown that this is also the so-called *generalized method of moments* (GMM) estimator of $\boldsymbol{\beta}$.

4.1.2 GMM estimation

The problem is to estimate the unknown parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ denotes the set of all possible parameter vectors. In order to do this, the i.i.d. sample $\{(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_N, \boldsymbol{x}_N)\}$ is drawn. Throughout this thesis, it is assumed that $\{(y_i, \boldsymbol{x}_i), i \geq 1\}$ was generated by a stationary and ergodic stochastic process. An estimator $\hat{\boldsymbol{\theta}}_M$ of $\boldsymbol{\theta}$ of the type

$$\hat{\boldsymbol{\theta}}_M = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\omega}(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}) \tag{16}$$

where $\boldsymbol{\omega}(y_i, x_i, \boldsymbol{\theta})$ is a continuous and twice differentiable function, is known as an *M-estimator* (van der Waart 1998, p. 41). Two special cases of M-estimators are the maximum likelihood estimator, where $\boldsymbol{\omega}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) = \log\left[\boldsymbol{f}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\right]$ and the least squares (LS) estimator, in which $\boldsymbol{\omega}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) = -\left[\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\right]^2$.

Another class of estimators are the *Z-estimators*, where ”Z” means ”zero” (ibid.). Assuming that the true parameter value $\boldsymbol{\theta}$ is the unique solution to the *population moment restriction*

$$\boldsymbol{f}(\boldsymbol{\theta}) \equiv \mathbb{E}[\boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta})] = \boldsymbol{0} \tag{17}$$

then a Z-estimator $\hat{\boldsymbol{\theta}}_Z$ of $\boldsymbol{\theta}$ is an estimator that solves the corresponding system of equations

$$\boldsymbol{g}_N(\boldsymbol{\theta}) \equiv \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{f}(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}) = \mathbf{0} \tag{18}$$

Naturally, there is no guarantee that a solution to (18) exists. Hence, in practice, it is more common to minimize the squared norm of this function, that is, to solve the optimization problem

$$\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{g}_N(\boldsymbol{\theta})\|_W^2$$

where the squared Euclidean norm $\|\|^2$ has been replaced by the squared weighted norm $\|\|_W^2$, defined as $\|\boldsymbol{x}\|_W^2 = \boldsymbol{x'Wx}$, where $\boldsymbol{W}$ is a positive definite weighting matrix. The estimator that solves the minimization problem defined above is known as the *generalized method of moments estimator* (GMM) of $\boldsymbol{\theta}$, and can be defined as

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta}) \tag{19}$$

where $\boldsymbol{W}_N$ is some positive definite weighting matrix that converges to in probability to $\boldsymbol{W}$ (Hansen 1982); the choice of $\boldsymbol{W}_N$ for the IV regression case will be discussed in Section 4.1.4. Looking at the definition of Z estimators, it is clear that the maximum likelihood method can be seen as a special case of the GMM. In the ML case, the moment restriction is that the score function must be equal to zero. Four important asymptotic theorems related to the GMM follow below.

Theorem 1. *The estimator $\hat{\boldsymbol{\theta}}_{GMM}$ is a consistent estimator of the unknown parameter vector $\boldsymbol{\theta}$.*

Proof. See Appendix 1 of this thesis. ■

Theorem 2. *Under the assumptions above, $\sqrt{N}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})$ is asymptotically normal with mean zero and covariance matrix $\boldsymbol{\Omega}$.*

where

$$\boldsymbol{\Omega} = \left[(\boldsymbol{G'WG})^{-1}\boldsymbol{G'W\Upsilon W'G}(\boldsymbol{G'W'G})^{-1}\right] \tag{20}$$

and

$$\boldsymbol{G} = \mathbb{E}\left[\frac{\partial \boldsymbol{f}(y_i, \boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right]$$

$$\boldsymbol{\Upsilon} = \mathbb{E}\left[\boldsymbol{f}(y_i, \boldsymbol{\theta})\boldsymbol{f}(y_i, \boldsymbol{\theta})'\right]$$

Proof. Se Appendix 2 of this thesis. ■

However, setting $\boldsymbol{W} = \boldsymbol{\Upsilon}^{-1}$, equation (20) simplifies to

$$\boldsymbol{\Omega}^{\star} = (\boldsymbol{G'\Upsilon^{-1}G})^{-1} \tag{21}$$

Theorem 3. *$\boldsymbol{\Omega}^{\star}$ has the smallest possible asymptotic variance amongst all weight matrices.*

Proof. See Appendix 3 of this thesis. ■

As a consequence of the result established above, using the inverse of the covariance matrix of the sample moment vector as a weighting matrix minimizes the asymptotic variance matrix $\boldsymbol{\Omega}$. Consequently, $\boldsymbol{\Omega}^\star$ is called the *optimal weight matrix*.

Theorem 4. *The 2SLS estimator is efficient in the class of all IV estimators for instruments linear in* $\boldsymbol{Z}$.

Proof. Omitted, see Wooldridge (2002), p. 96-97. ■

Hence, GMM estimators are consistent and asymptotically normal. In addition, Theorem 4 says that the GMM estimator $\hat{\boldsymbol{\beta}}_{2SLS}$ of $\boldsymbol{\beta}$ is efficient.

4.1.3 Advantages and disadvantages of the GMM

The main advantage of the GMM framework is that it does not require the distribution of the disturbances to be known. Assuming normality of the errors, it would be possible to estimate the parameters using maximum likelihood. However, if the normality assumption fails, the resulting parameter estimates are be inconsistent (Wooldridge 2002, p. 385). In economics and finance, the distribution of the data usually has heavier tails than does the normal distribution. Two examples of heavy tailed data frequently encountered in practice are income distributions, covered in this thesis, and asset returns in finance. The obvious drawback of the GMM is the efficiency issue described above; hence, in cases where it is easy to specify the full model including the distribution of the disturbances, the maximum likelihood method is usually preferred.

4.1.4 GMM estimation for instrumental variables

For instrumental variables regression, the objective function is

$$\boldsymbol{g}_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) = \frac{1}{N}(\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta})$$

The weight matrix is

$$\boldsymbol{W}_N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_i\boldsymbol{z}_i' = \frac{1}{N}\boldsymbol{Z}'\boldsymbol{Z}$$

for which it holds (according to the weak law of large numbers) that $\boldsymbol{W}_N \xrightarrow{P} \boldsymbol{W}$. Then

$$Q(\boldsymbol{\beta}) = \frac{1}{N}(\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta})$$

The partial derivative with respect to $\boldsymbol{\beta}$ of the above expression is

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{2}{N}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta}) \tag{22}$$

Setting (22) to $\mathbf{0}$ and solving for $\hat{\boldsymbol{\beta}}_{2SLS}$, yields

$$\begin{aligned}
&\frac{2}{N}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{y} - \mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}}_{2SLS}) = \mathbf{0} \\
&\iff \boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_{2SLS} = \mathbf{0} \\
&\iff \hat{\boldsymbol{\beta}}_{2SLS} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \qquad (23)
\end{aligned}$$

given that the matrix $\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$ is nonsingular. Note that (23) is equal to the previously presented expression (15).

4.1.5 Covariance matrix of the 2SLS under homoscedasticity

Assuming that the variance of the error terms $\{u_i\}_i^N$ given $\boldsymbol{z}_i$ is equal for all i, that is, $\mathbb{E}\left[u_i^2|\boldsymbol{z}_i\right] = \sigma^2$, where a consistent estimator of σ^2 is $\hat{\sigma}^2 = N^{-1}\sum_{i=1}^N \hat{u}_i^2$ $= N^{-1}\sum_{i=1}^N (y_i - \boldsymbol{x}_i\hat{\boldsymbol{\beta}})^2$, the estimator of $\boldsymbol{\Omega}$ under homoscedasticity is

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}} &= \mathbb{E}\left[(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta})'\right] \qquad (24) \\
&= \hat{\sigma}^2\left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})\right]^{-1} \\
&= \hat{\sigma}^2\left[\boldsymbol{X}'\boldsymbol{Z}\overbrace{(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{Z}'\boldsymbol{Z})^{-1}}^{=\boldsymbol{I}}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{X})\right]^{-1} \\
&= \hat{\sigma}^2\left\{\left[\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]'\left[\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]\right\}^{-1} \\
&= \hat{\sigma}^2(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}) \qquad (25)
\end{aligned}$$

where $\hat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$.

4.1.6 Covariance matrix of the 2SLS under heteroscedasticity

Relaxing the assumption of homoscedasticity of the errors, and using the same definition of $\hat{\mathbf{X}}$ as above, it is possible to write the 2SLS form of (20) as

$$\begin{aligned}
\hat{\boldsymbol{\Omega}} = &(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\left(\sum_{i=1}^N \hat{u_i^2}\hat{\boldsymbol{x}}_{\boldsymbol{i}}\hat{\boldsymbol{x}}'_{\boldsymbol{i}}\right) \times \\
&(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1} \\
= &(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\left(\sum_{i=1}^N \hat{u_i^2}\hat{\boldsymbol{x}}_{\boldsymbol{i}}\hat{\boldsymbol{x}}'_{\boldsymbol{i}}\right)(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \qquad (26)
\end{aligned}$$

The heteroscedasticty-adjusted covariance matrix (26) is known as the *Eicker-Huber-White covariance matrix*.

4.1.7 Testing for endogenity

There are several tests for endogenity, see for example Holmquist et al. (2011). The most commonly used is the Durbin-Wu-Hausman test (Durbin 1954, Wu 1973, Hausman 1978), which has the test statistic

$$(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{LS})' \boldsymbol{D}^{+} (\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{LS}) \tag{27}$$

where $\boldsymbol{D} = \boldsymbol{V}_{IV} - \boldsymbol{V}_{LS}$, and $\boldsymbol{D}^{+}$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{D}$. Under the null hypothesis, both estimators are consistent, but $\hat{\boldsymbol{\beta}}_{LS}$ is more efficient. Under the alternative hypothesis, $\hat{\boldsymbol{\beta}}_{IV}$ is consistent whereas $\hat{\boldsymbol{\beta}}_{LS}$ is inconsistent. The distribution of the test statistic under the null is asymptotically chi-squared with number of degrees of freedom equal to the rank of $\boldsymbol{D}$.

4.1.8 Inferences about $\boldsymbol{\beta}$

The results described in previous sections can be used for testing hypotheses about $\boldsymbol{\beta}$. Consider for example the linear hypotheses

$$H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$$

versus

$$H_1 : \boldsymbol{R\beta} \neq \boldsymbol{r}$$

where $\boldsymbol{R}$ is $q \times K$ and $\boldsymbol{r}$ is $q \times 1$. The test statistic

$$N(\boldsymbol{R}\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{r})'(\boldsymbol{R}\hat{\boldsymbol{W}}_N \boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{r}) \tag{28}$$

is due to Wald (1943). The distribution of (28) is asymptotically chi-squared with q degrees of freedom, which means that it is possible to use the Wald test in practice when testing the significance of $\boldsymbol{\beta}$.

4.1.9 Testing for instrument validity

If there are more instruments than endogenous regressors, it is possible to test the validity of the instruments by using the Sargan (1958) test, which has test statistic

$$J = \frac{\hat{\boldsymbol{u}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\hat{\boldsymbol{u}}}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/N} \xrightarrow{L} \chi^2_{L-K}(\alpha) \tag{29}$$

where $\boldsymbol{u}$ is the vector of residuals. As an alternative to using (29) directly, the Sargan test can be done in two steps: First, estimate the residuals from the regression, then regress the estimated residuals $\hat{\boldsymbol{u}}$ on all exogenous variables. It is well-known that the resulting R^2 value multiplied by the sample size N is asymptotically chi-square distributed. The null hypothesis is that the instruments utilized in the regression are uncorrelated with the error term. If at least one of the instruments is in fact endogenous, the null hypothesis is rejected at the significance level α. The Sargan test is also known as the *Sargan-Hansen test*, since Hansen (1982) showed that it could be applied in a GMM setting.

4.1.10 Weak instruments

An instrument is said to be *weak* if it is only weakly correlated with the endogenous regressor. This can cause serious problems when estimating a model with endogenous variables, since $\boldsymbol{\beta}^{\star}_{IV}$ will be inconsistent and biased, and its asymptotic distribution will not be normal (Staiger and Stock 1997). In fact, it is possible to show that under weak instruments, both the IV and LS estimators will be biased, but the bias of the instrumental variables estimator will be larger. Let $\mathcal{W}_{\text{bias}}$ be a set of instrumental variables matrices such that $\mathcal{W}_{\text{bias}} = \{\boldsymbol{Z} : B \geq b\}$, where B is a measure of the relative bias of the instrumental variables model relative to the least squares (cf. Stock and Yogo 2005), and b is the desired maximal bias of the IV estimator compared to the LS. Let also $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. A test of the null hypothesis

$$H_0 : \boldsymbol{Z} \in \mathcal{W}_{\text{bias}}$$

against

$$H_1 : \boldsymbol{Z} \notin \mathcal{W}_{\text{bias}}$$

was proposed by Cragg and Donald (1993) and is defined by

$$\text{C} = \text{mineval}(\boldsymbol{\Lambda}) \tag{30}$$

where mineval() denotes the minimum eigenvalue, and

$$\boldsymbol{\Lambda} = (\hat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{X}'\boldsymbol{Z}'\boldsymbol{Z}^{-1}\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\Sigma}}^{-1/2})/L \tag{31}$$

and

$$\hat{\boldsymbol{\Sigma}} = (\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X})/(N-K) \tag{32}$$

Hence, failure to reject the null indicates the presence of weak instruments. It is possible to show that the test statistic follows the noncentral chi-squared distribution; the critical values for the 2SLS estimation method are given in Stock and Yogo (2005). These depend on the number of endogenous regressors, the number of instrumental variables and the allowed maximal bias of the 2SLS estimator relative to the LS estimator, b.

4.2 Capturing the time effect - dynamic panel data models

This section of the thesis considers models for analysing the change in income inequality between countries over time. This requires the use of panel data. In economics, it is often the case that a lag of the dependent variable is used as a regressor; such a model is called a *dynamic panel model*. The two most frequently encountered models for estimating equations of the dynamic-panel type are the *first-differenced GMM estimator* and the *system GMM estimator*.

4.2.1 First-differenced GMM

Consider first the so-called *first-differenced GMM estimator*, introduced by Arellano and Bond (1991). The model is built upon the AR(1) time series model

$$y_t = \phi y_{t-1} + u_t \tag{33}$$

for time $t = 1, \ldots, T$ and $|\phi| < 1$. This model can be extended by using a sample of N individuals, so that $i = 1, \ldots, N$. This gives the model

$$y_{it} = \phi y_{i,t-1} + u_{it} \tag{34}$$

where the disturbance term u_{it} can be divided into η_i, the error specific to each individual, and the *idiosyncratic error*, ν_{it}, that varies both across individuals and over time. In this thesis, y_{it} is the income inequality of country i at time t, and η_i is the country-specific error. Since y_{it} depends on η_i, and η_i is the same for all time periods, the explanatory variable $y_{i,t-1}$ also depends on η_i. Hence, the model described by (34) is endogenous. A further assumption is that $\mathbb{E}[\nu_{it}] = 0$ and that $\mathbb{E}[\nu_{it}\nu_{is}] = 0$ for all $s \neq t$.

Augmenting the model described above by adding $(K-1)$ independent explanatory variables, the model can be written as

$$y_{it} = \phi y_{i,t-1} + \boldsymbol{\beta}' \boldsymbol{x}_{it} + u_{it} \tag{35}$$

where the assumption of strict exogenity of the explanatory variables normally encountered in panel models is relaxed. This allows for $\mathbb{E}[\boldsymbol{x}_{it}\nu_{is}] \neq \mathbf{0}$ for $s < t$ and $\mathbb{E}[\boldsymbol{x}_{it}\nu_{is}] = \mathbf{0}$ for $s \geq t$. It was shown by Nickell (1981) that, because endogenity, the standard fixed-effects panel model is biased when using the lagged dependent variable as an explanatory variable, even as $N \to \infty$. This potentially serious problem can be solved in two steps. First, take the first difference of equation (35), which yields

$$y_{it} - y_{i,t-1} = \phi(y_{i,t-1} - y_{i,t-2}) + \boldsymbol{\beta}'(\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}) + u_{it} - u_{i,t-1} \tag{36}$$

This expression can be written more compactly as

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \boldsymbol{\beta}' \Delta \boldsymbol{x}_{it} + \Delta u_{it} \tag{37}$$

In equation (37) above, $\Delta y_{i,t-1}$ is correlated with Δu_{it}. This is because the term $y_{i,t-1}$, which is in $\Delta y_{i,t-1}$, contains $u_{i,t-1}$, which is also in Δu_{it}. To solve this problem, Anderson and Hsiao (1981) suggest that lags of the dependent variable are used as instruments. For $\Delta y_{i,t-1}$, the suggestion is to use $\Delta y_{i,t-2}$ that is uncorrelated with $\Delta u_{i,t}$ given that the errors are serially uncorrelated. Expanding this approach, Holtz-Eakin et al. (1988) propose the utilization of all available lags as instruments. For example, for $t = 3$, the model can be written

$$y_{i3} - y_{i2} = \phi(y_{i2} - y_{i1}) + (\boldsymbol{x}'_{i3} - \boldsymbol{x}'_{i2})\boldsymbol{\beta} + u_{i3} - u_{i2} \tag{38}$$

and the available instruments are $y_{i1}, \boldsymbol{x}'_{i1}$ and $\boldsymbol{x}'_{i2}$. For $t = 4$, it is easy to see that the instruments available are $y_{i1}, y_{i2}, \boldsymbol{x}'_{i1}, \boldsymbol{x}'_{i2}$ and $\boldsymbol{x}'_{i3}$. The number of available lags is highest for the time period closest to the final time T, that is, for $t = T$. For this period, the model can be written as

$$y_{i,T} - y_{i,T-1} = \phi(y_{i,T-1} - y_{i,T-2}) + (\boldsymbol{x}'_{i,T} - \boldsymbol{x}'_{i,T-1})\boldsymbol{\beta} + u_{i,T} - u_{i,T-1} \tag{39}$$

for which the instruments $y_{i1}, y_{i2}, \ldots, y_{i,T-2}$ and $\boldsymbol{x}'_{i1}, \boldsymbol{x}'_{i2}, \ldots, \boldsymbol{x}'_{i,T-1}$ are available. Thus, the instrument matrix $\boldsymbol{Z}_i$ is the block diagonal matrix $\boldsymbol{Z}_i = \text{diag}([y_{i1},\ \boldsymbol{x}'_{i1},\ \boldsymbol{x}'_{i2}],\ [y_{i1},\ y_{i2},\ \boldsymbol{x}'_{i1},\ \boldsymbol{x}'_{i2},\ \boldsymbol{x}'_{i3}],\ \ldots,\ [y_{i1},\ \ldots,\ y_{i,T-2},\ \boldsymbol{x}'_{i1},\ \ldots, \boldsymbol{x}'_{i,T-1}])$. Even for relatively small T, there will be more instruments than parameters. Hence, a GMM approach is appropriate. A possible estimator of $\boldsymbol{\psi} = (\phi, \boldsymbol{\beta}')'$ would thus be the solution to the GMM optimization problem

$$\hat{\boldsymbol{\psi}}_{AB} = \arg\min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left(\frac{1}{N}\sum_{i=1}^{N} \Delta\boldsymbol{u}'_i\boldsymbol{Z}'_i\right) \boldsymbol{A}_N \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i\Delta\boldsymbol{u}_i\right) \tag{40}$$

where $\boldsymbol{\Psi}$ is the set of all possible parameter vectors. This estimator was first proposed by Arellano and Bond (1991). The estimate of the weighting matrix $\boldsymbol{A}_N$ is

$$\hat{\boldsymbol{A}}_N = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}'_i\boldsymbol{H}\boldsymbol{Z}_i\right) \tag{41}$$

where $\boldsymbol{H}$ is the tridiagonal $(T-1)\times(T-1)$ matrix

$$\boldsymbol{H} = \begin{pmatrix} 2 & -1 & 0 & \ldots & 0 \\ -1 & 2 & -1 & \ldots & 0 \\ 0 & -1 & 2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix} \tag{42}$$

Define also

$$\tilde{\boldsymbol{X}}_i = \begin{pmatrix} y_{i2} - y_{i1} & \boldsymbol{x}'_{i3} - \boldsymbol{x}'_{i2} \\ y_{i3} - y_{i2} & \boldsymbol{x}'_{i4} - \boldsymbol{x}'_{i3} \\ \vdots & \vdots \\ y_{i,T-1} - y_{i,T-2} & \boldsymbol{x}'_{i,T} - \boldsymbol{x}'_{i,T-1} \end{pmatrix} \tag{43}$$

and

$$\boldsymbol{\Delta y}_i = \begin{pmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ \vdots \\ y_{i,T-1} - y_{i,T-2} \end{pmatrix} \tag{44}$$

Using this, a closed-form solution the the optimization problem in (40) can be written as

$$\hat{\boldsymbol{\psi}}_{AB} = (\tilde{\boldsymbol{X}}'\boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}'\Delta\boldsymbol{y} \tag{45}$$

where $\boldsymbol{\Delta y} = \boldsymbol{I}_N \otimes \boldsymbol{\Delta y}'_i$, $\tilde{\boldsymbol{X}} = \boldsymbol{I}_N \otimes \tilde{\boldsymbol{X}}'_i$, $\boldsymbol{Z} = \boldsymbol{I}_N \otimes \boldsymbol{Z}'_i$, and $\otimes$ denotes the Kronecker product.

Since the number of available instruments grows very quickly as T increases, most statistical software enable the user to limit the instruments used in the model, so that the number of instruments does not exceed the total observations.

While it is not required that $\mathbb{E}\left[\Delta u_{it}\Delta u_{i,t-1}\right] = 0$, since the u_{it}:s are first differences of serially uncorrelated errors, it can be shown that the consistency of the model relies heavily on the assumption that $\mathbb{E}\left[\Delta u_{it}\Delta u_{i,t-2}\right] = 0$. With this in mind, Arellano-Bond (1991) have developed a test for second-order serial autocorrelation. The test statistic is

$$m = \frac{\Delta\hat{\boldsymbol{u}}_{-2}\Delta\hat{\boldsymbol{u}}_{\star}}{(\Delta\hat{\boldsymbol{u}}'\Delta\hat{\boldsymbol{u}})^{1/2}} \xrightarrow{L} N(0,1) \tag{46}$$

where $\Delta\hat{\boldsymbol{u}}$ is the vector of first-differenced residuals from the regression, $\Delta\hat{\boldsymbol{u}}_{-2}$ is the vector of the first-differenced residuals lagged twice, and $\Delta\hat{\boldsymbol{u}}_{\star}$ is the vector of trimmed first-differenced residuals to match $\Delta\hat{\boldsymbol{u}}_{-2}$.

4.2.2 System GMM

The Arellano-Bond model has been shown to preform poorly in the presence of weak instruments (Blundell and Bond 1998). In the dynamic panel setting, the weak instruments issue arises when the absolute value of the autoregressive parameter ϕ in (35) is close to unity (ibid.). Moreover, Soto (2009) has shown that the Arellano-Bond estimator underestimates the true value of $\boldsymbol{\beta}$ when N is low. Hence, an alternative approach when estimating a dynamic panel data model with endogenous variables was developed by Arellano and Bover (1995) and Blundell and Bond (1998). This model is an extension of the Arellano-Bond approach described in Section 4.2.1. The model is again

$$y_{it} = \phi y_{i,t-1} + \boldsymbol{\beta}'\boldsymbol{x}_{it} + u_{it} \tag{47}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. The Blundell-Bond model uses lagged differences as extra moment conditions. For example, for $t = T$, the model is

$$y_{iT} = \phi y_{i,T-1} + \boldsymbol{x}'_{iT}\boldsymbol{\beta} + u_{iT} \tag{48}$$

and the available extra instruments are $\Delta y_{i1}, \dots, \Delta y_{i,T-2}$ and $\Delta\boldsymbol{x}'_{i1}, \dots, \Delta\boldsymbol{x}'_{i,T-1}$. The instrument matrix can then be written as $\boldsymbol{U}_i = \text{diag}(\boldsymbol{Z}_i, \tilde{\boldsymbol{Z}}_i)$, where $\boldsymbol{Z}_i$ is as defined in the Arellano-Bond model, and $\tilde{\boldsymbol{Z}}_i$ is a block diagonal matrix defined by $\tilde{\boldsymbol{Z}}_i = \text{diag}([\Delta y_{i2}, \Delta\boldsymbol{x}'_{i2}, \Delta\boldsymbol{x}'_{i3}], [\Delta y_{i2}, \Delta y_{i3}, \Delta\boldsymbol{x}'_{i2}, \Delta\boldsymbol{x}'_{i3}, \Delta\boldsymbol{x}'_{i4}], \dots, [\Delta y_{i2}, \dots, \Delta y_{i,T-2}, \Delta\boldsymbol{x}'_{i2}, \dots, \Delta\boldsymbol{x}'_{i,T-1}])$. Further, let $\boldsymbol{U} = \boldsymbol{I}_N \otimes \boldsymbol{U}'_i$, and the weighting matrix is now $\boldsymbol{A} = \text{diag}(\boldsymbol{A}, \boldsymbol{I}_{T-1})$. Let also

$$\tilde{\boldsymbol{X}}_i = \begin{pmatrix} y_{i2} - y_{i1} & \boldsymbol{x}'_{i3} - \boldsymbol{x}'_{i2} \\ y_{i3} - y_{i2} & \boldsymbol{x}'_{i4} - \boldsymbol{x}'_{i3} \\ \vdots & \vdots \\ y_{i,T-1} - y_{i,T-2} & \boldsymbol{x}'_{i,T} - \boldsymbol{x}'_{i,T-1} \\ y_{i3} & \boldsymbol{x}'_{i2} \\ \vdots & \vdots \\ y_{iT} & \boldsymbol{x}'_{iT} \end{pmatrix} \tag{49}$$

and

$$\boldsymbol{\Delta y}_i^\star = \begin{pmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ \vdots \\ y_{i,T-1} - y_{i,T-2} \\ y_{i3} \\ \vdots \\ y_{iT} \end{pmatrix} \tag{50}$$

Then, the corresponding Blundell-Bond estimator of $\boldsymbol{\psi}$ is

$$\hat{\boldsymbol{\psi}}_{BB} = (\tilde{\boldsymbol{X}}'\boldsymbol{U}\tilde{\boldsymbol{A}}\boldsymbol{U}'\tilde{\boldsymbol{X}})^{-1}\boldsymbol{X}'\boldsymbol{U}\tilde{\boldsymbol{A}}\boldsymbol{U}'\boldsymbol{\Delta y}^\star \tag{51}$$

where $\tilde{\boldsymbol{X}} = \boldsymbol{I}_N \otimes \tilde{\boldsymbol{X}}_i'$ and $\boldsymbol{\Delta y}^\star = \boldsymbol{I}_N \otimes \boldsymbol{\Delta y}^{\star\prime}_i$. This model is also known as the *system GMM*. Since this model does not use first differences of the dependent variable, it is not appropriate to use in situations when $\phi \geq 1$. This slightly limits the usability of the system GMM, as compared to the Arellano-Bond model, which does not require stationarity; cf. Madsen (2008). Another issue is that while consistent, asymptotically normal and asymptotically efficient, both the Arellano-Bond and Blundell-Bond GMM have an asymptotic bias that is $\mathcal{O}(\sqrt{T/N})$. This is, obviously, a weakness of the model. As $N/T \to \infty$, the asymptotic bias disappears (Alvarez and Arellano 2003).

There are a number of additional approaches using even more moment restrictions besides the Arellano-Bond and Blundell-Bond models; see for example Cameron and Trivedi (2005, p. 766). However, in this thesis, the number of models is limited to the two most common.

4.2.3 Panel unit root tests

As mentioned before, the Blundell-Bond model is more robust when the autoregressive parameter ϕ is close to unity. Hence, unit root tests are of great importance in panel settings.

By the mid-1990s, several studies had found that conventional univariate unit root tests, such as the Dickey-Fuller test, have low power in panel settings (cf. Oh 1996). To remedy this problem, several tests of unit roots in panel data have emerged. Letting the model of interest be

$$y_{it} = \phi_i y_{i,t-1} + u_{it} \tag{52}$$

a unit root is present if $\phi = 1$. Im, Pesaran and Shin (2003) propose a test of the hypotheses

$$H_0 : \phi_i = 1 \ \forall i$$

against

$$H_1 : \phi_i < 1, i = 1 \dots N_1 \text{ and } \phi_i = 1, i = (N_1 + 1) \dots N$$

that is, in the Im-Pesaran-Shin test, rejecting the null hypothesis means that N_1 of the N individual time series are stationary, while the others have unit roots.

Another panel unit root test of this type is due to Hadri (2000). The difference between the tests is that the Hadri test statistic is based on the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, while the Im-Pesaran-Shin test is based on the augmented Dickey-Fuller (ADF) test.

5 Empirical analysis

The focus of the thesis will now shift towards the empirical analysis of income inequality. The empirical analysis consists of three parts. First, a cross-section of countries is used in order to construct a multiple linear regression model with the Gini coefficient as the dependent variable. Then, a dynamic panel model is fitted to the data in order to analyze what factors have influenced income inequality over time. Finally, in the third section, the focus is on inequality convergence.

The statistical software used in the empirical analysis are *R*, version 3.2.2, *EViews*, version 9.5, and *Matlab*, version R2016a.

5.1 Data description

5.1.1 Dependent variable

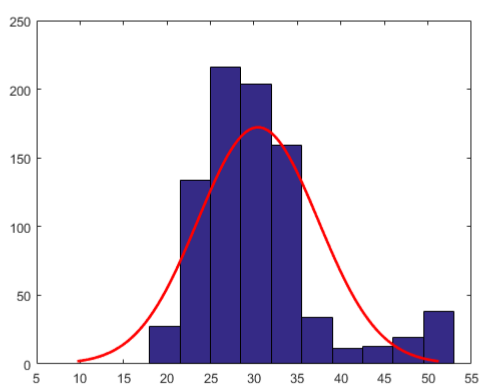

Figure 2: Histogram of the values of the Gini coefficient from 1985 to 2013 for all countries. The red curve is the normal distribution.

The dataset consists of 30 countries listed in Table 1. Also in Table 1, the earliest and most recent value of the Gini coefficient (net of tax) for each country is presented, together with the average yearly change in Gini for each country. For most of the countries, the time range is between 1985 and 2013; the exceptions being Cyprus, Iceland and Romania, for which the time series start slightly later, respectively. The countries for which Gini has decreased during this period are Brazil, Chile, France, Greece, Ireland, Norway, Switzerland and Turkey. This implies that income inequality has increased in all other countries.

The total number of observations on the dependent variable is 840. Figure 2 shows the histogram of all 840 observations with the normal distribution superimposed. Considering the heavy tails of the distribution in Figure 2, the data is clearly non-Gaussian.

Figure 3 highlights the trend in Gini between 1985 and 2013 for seven countries: Brazil, Turkey, the United States, the United Kingdom, Poland, Germany and Sweden. The countries with the highest level of income inequality in 1985 - Brazil and Turkey - have seen a decrease in Gini over the 28-year period, while the other five countries have experienced an increase in Gini. Hence, it seems reasonable to believe that there has been at least some degree of convergence in income inequality during the sample period. This is further supported by Table 1, from which it is apparent that income inequality has increased in most countries. However, with the exception of the United States, inequality has decreased in all countries that in the year 1985 had a Gini value of over 32.

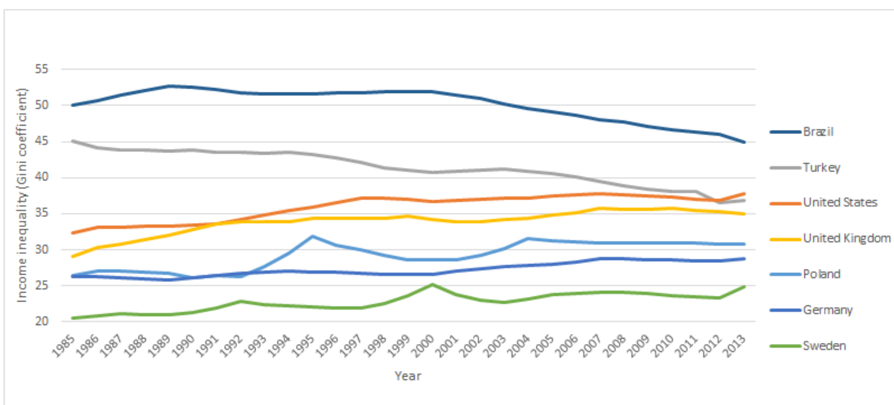

Figure 3: The values of the Gini coefficient for Brazil, Turkey, the United States, the United Kingdom, Poland, Germany and Sweden between 1985 and 2013.

The Gini data comes from the Standardized World Income Inequality Database (*SWIID*); the statistical methodology used in constructing the *SWIIID* was described in detail in Section 3.3.

5.1.2 Independent variables

Table 2 below presents the explanatory variables used in the models. The ratio of government expenditures to GDP, the ratio of trade (imports and exports) to GDP, and the ratio of domestic credit to the private sector to GDP are all used in both the static and dynamic models. The credit variable measures the total credit given by domestic banks and other financial institutions to domestic firms, both public and private.

Country	**Gini, 1985**	**Gini, 2013**	**Avg. ann. change, %**
Australia	29.20	30.90	0.21
Austria	25.00	29.00	0.57
Belgium	23.00	23.00	0.00
Brazil	50.10	45.00	-0.36
Canada	28.50	31.30	0.35
Chile	50.50	47.20	-0.23
Czech Republic	19.50	24.60	0.93
Cyprus	22.10 (1989)	32.40	1.66
Denmark	25.20	26.80	0.23
Finland	20.50	25.40	0.85
France	32.60	30.10	-0.27
Germany	26.30	28.80	0.34
Greece	33.30	33.00	-0.03
Iceland	22.60 (1992)	24.60	0.42
Ireland	32.20	29.00	-0.35
Italy	31.00	31.60	0.07
Japan	25.60	29.10	0.49
Hungary	22.10	28.30	1.00
Netherlands	24.00	25.00	0.15
New Zealand	26.70	32.80	0.82
Norway	23.10	22.80	-0.05
Poland	26.40	30.70	0.58
Portugal	26.60	34.20	1.02
Romania	19.50 (1989)	32.30	2.74
Spain	28.50	33.90	0.68
Sweden	20.50	24.90	0.77
Switzerland	30.20	28.50	-0.20
Turkey	45.00	36.80	-0.65
United Kingdom	29.10	35.00	0.72
United States	32.40	37.80	0.60

Table 1: The values of the Gini coefficient (net of tax) in 1985 and 2013 for the 30 countries used in the empirical analysis, together with the average annual change during this time period.

Variable	Type	Used in
Government-expenditure-to-GDP ratio	continuous	both models
Trade-to-GDP ratio	continuous	both models
Dom. credit-to-private sector GDP ratio	continuous	both models
Agricultural employment share	dichotomous	static only
Industrial employment share	continuous	dynamic only
Country population	dichotomous	static only
Country government	dichotomous	dynamic only

Table 2: Summary of the independent variables used in static and dynamic regression models.

There are two labor market variables, one agricultural employment dummy, and one continuous variable that measures the proportion of the labor force employed in the industrial sector. The first variable is used only in the static model, while the final is used only in the dynamic model. From the discussion in Section 3.2.3, it is reasonable to believe that agricultural employment, albeit being useful when analyzing cross-country income inequality, is unsuitable in the dynamic model. Instead, the labor market variable in the dynamic model is the industrial employment share. In the static model, the final covariate is a size dummy; more populous countries are assumed to have greater income diversity. The dynamic model utilizes a dichotomous variable taking the value 0 when the country is governed by a left-wing parties, and 1 when it is ruled by a right-wing government. The logic behind this variable will be discussed in greater detail in Section 5.3.1.

The static model includes the constructed trade share defined in Section 3.3 as an instrument for the possible endogenous trade-to-GDP-ratio. Since the Frankel and Romer paper uses data from 1985, the measure has been corrected by multiplying the constructed trade shares by the ratios of actual trade-to-GDP ratios in 2013 to the trade-to-GDP ratios in 1985.

The World Bank provides the data on government expenditure (World Bank 2017a), trade share (World Bank 2017b) and domestic credit-to-GDP-ratio (World Bank 2017c), while the data on the two labor market variables comes from the International Labour Organization (ILO 2017). Finally, the CIA World Factbook (Central Intelligence Agency 2017) is used for the data on population and government.

Some descriptive statistics of the data are summarized in Appendix A4. The correlation between the constructed trade share and the actual trade share is of great importance for the validity of the model. In the dataset used in this thesis, it is 0.91.

5.2 Static model

5.2.1 Model specification and main results

The model is

$$\log(y_i) = \beta_0 + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) + \beta_3 \log(x_{3i}) + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i \quad (53)$$

for $i = 1 \dots 30$, where y_i is the income inequality of country i in the year 2013, x_1 is government expenditure as share of GDP, x_2 is the sum of exports and imports divided by GDP, x_3 is the domestic credit to the private sector as share of GDP, and the dummy variables x_4 and x_5 are defined as

$$x_4 = \begin{cases} 1 & \text{if employment in agriculture} \geq 5\% \text{ of total employment} \\ 0 & \text{otherwise} \end{cases}$$

$$x_5 = \begin{cases} 1 & \text{if country population is} \geq 15 \text{ million} \\ 0 & \text{otherwise} \end{cases}$$

The corrected constructed trade share described in Section 5.1 is the sole instrument for the endogenous variable. Table 3 shows the results from the static model. The column denoted by (1) gives the result assuming homoscedasticity, while the column denoted by (2) gives the estimates when using the Eicker-Huber-White estimator. Finally, column (3) presents the LS estimates for comparison. It is possible to conclude that the logarithms of government expenditure and trade-to-GDP-ratio are both significant variables for explaining the cross-country differences in the logarithm of Gini. Both of these have negative coefficients, that is, an increase in any of these decreases inequality. The dummy variable *high level of agricultural employment*, x_4, is also highly significant. The size dummy x_5 is significant at the 10% level for the heteroscedasticity-adjusted instrumental variables and LS models, whereas it is insignificant in the non-adjusted instrumental variables model. Both of the dummy variables have positive coefficients. The sole insignificant variable according to all three models is the logarithm of the domestic credit to the private sector.

Moreover, the p-value of the Durbin-Wu-Hausman test indicates that trade-to-GDP is endogenous. Additionally, the Wald and F-tests show that the joint parameter vector $\hat{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$ is also highly significant. The Cragg-Donald weak instruments test indicates that the instruments in the model are strong. This result was fairly expected given the high correlation between the actual and constructed trade shares. Finally, the R^2 value is well above 70%, implying that the model explains almost three-quarters of the cross-country variation in income inequality.

5.2.2 Model diagnostics

Figure 4 above shows the quantile-quantile (Q-Q) plot and the plot of the residuals versus the fitted values for both models; above, the 2SLS model and below, the LS model. The residuals-versus-fitted-plot does not look too worrying. In

	(1)	(2)	(3)
Intercept, $\hat{\beta}_0$	$4.38^{\star\star\star}$	$4.38^{\star\star\star\star}$	$4.11^{\star\star\star}$
	(8.65)	(7.61)	(8.32)
log(Gov.exp./GDP), $\hat{\beta}_1$	$-0.19^{\star}$	$-0.19^{\star}$	$-0.18^{\star}$
	(-2.72)	(-2.47)	(-2.42)
log(Trade/GDP), $\hat{\beta}_2$	$-0.15^{\star}$	$-0.15^{\star}$	$-0.12^{\star}$
	(-3.07)	(-2.35)	(-2.24)
log(Domestic credit), $\hat{\beta}_3$	0.03	0.03	0.05
	(0.59)	(0.65)	(1.05)
High level of agric. emp., $\hat{\beta}_4$	$0.16^{\star\star}$	$0.16^{\star\star}$	$0.17^{\star\star}$
	(3.71)	(3.28)	(3.61)
Population above 15 mil., $\hat{\beta}_5$	0.08	0.08	$0.10^{\diamond}$
	(1.47)	(1.38)	(2.02)
Cragg-Donald p-value	$< 0.001^{\star\star\star}$	$< 0.001^{\star\star\star}$	
Durbin-Wu-Hausman p-value	$< 0.001^{\star\star\star}$	$< 0.001^{\star\star\star}$	
Wald p-value	$< 0.001^{\star\star\star}$	$< 0.001^{\star\star\star}$	
F-test p-value			$< 0.001^{\star\star\star}$
Kolmogorov-Smirnov p-value	0.58	0.58	0.62
Breusch-Pagan p-value	0.59	0.59	0.48
R^2	73.26%	73.26%	72.93%

Table 3: The results of the three regressions. Column (1) corresponds to instrumental variable regression with heteroscedastic covariance matrix, column (2) is instrumental variables regression assuming homoscedasticty of the errors, and column (3) are the least squares estimates ignoring possible endogenity. In brackets the corresponding t-statistics. Achieved level of significance: $\star\star\star = 0.001, \star\star = 0.01, \star = 0.05, \diamond = 0.10$.

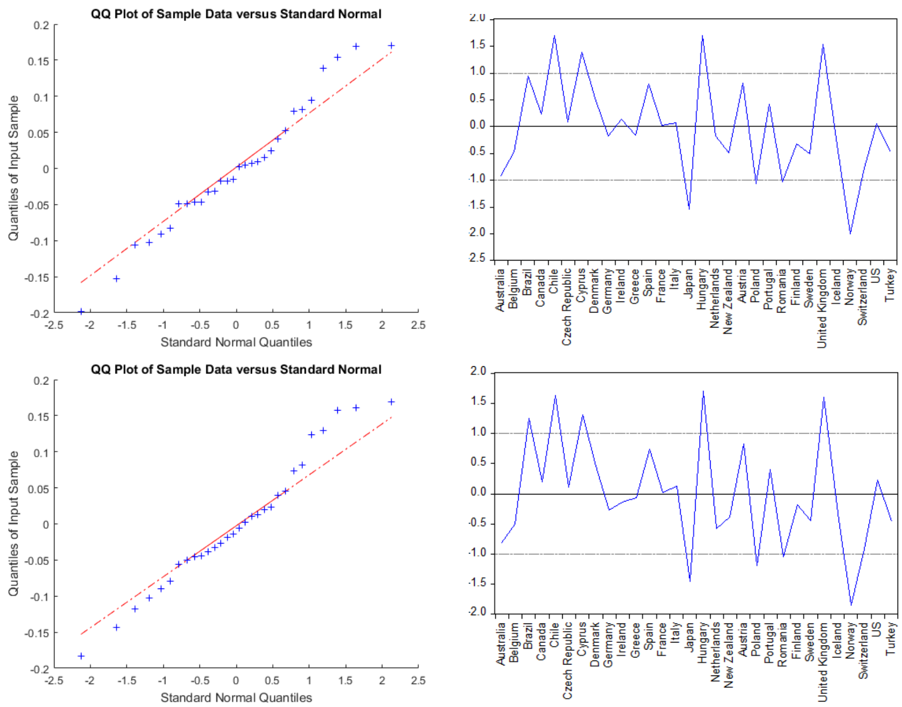

Figure 4: The top figures show the Q-Q plot and residual versus fitted values for the 2SLS. The bottom figures correspond to the least squares model.

Table 3, the value of the standard Breusch-Pagan (Breusch and Pagan 1979) residuals test is presented. This indicates that is not possible to reject the null hypothesis of residual homoscedasticity.

Since the sample size is relatively small ($N = 30$) and the data is heavy-tailed, at least some of the residuals should be far away from the straight line. However, the log-transformation seems to have partially remedied this problem, and the Q-Q plot looks quite decent. Moreover, the Kolmogorov-Smirnov test (Kolmogorov 1933, Smirnov 1948) gives p-values of 0.62 for the LS model and 0.58 for the IV model. Hence, it is not possible to reject the null hypothesis of normality of the residuals. Hence, both models seem acceptable from a residuals perspective.

5.3 Dynamic model

5.3.1 Model specification

The model is

$$\log y_{it} = \phi \log (y_{it-1}) + \boldsymbol{\beta}' \boldsymbol{x}_{it} + u_{it} \tag{54}$$

for and $i = 1, \ldots, 30$ and $t = 1985, \ldots, 2013$. As before the dependent variable is income inequality, x_1 is the logarithm of government expenditure as share of GDP, x_2 is the logartihm of the sum of exports and imports divided by GDP, and x_3 is the logarithm of domestic credit to the private sector as share of GDP. However, x_4 is now industry's share of total employment. x_5 is a dichotomous variable defined as

$$x_5 = \begin{cases} 1 & \text{if a country is governed by a right-wing government} \\ 0 & \text{if a country is governed by a left-wing government} \end{cases}$$

In this slightly simplified approach, a government is either classified as right-wing or left-wing. Governments consisting of solely socialist or social-democratic parties are left-wing; conversely, if conservative or Christian democrat parties form the government, it is classified as right-wing. Coalition governments are given either a 0 or a 1 depending on the main party of that particular government. For example, the current (April 2017) German government is classified as right-wing, since the largest party is the Christian democratic CDU/CSU, even though it forms a coalition with the social-democratic SPD. There are four countries in the dataset with presidential systems: Brazil, Chile, Cyprus and the United States. This means that the head of government is also the head of state. In these cases, the value of x_5 is set depending on the political party of the president. In the US case, $x_5 = 1$ when a Republican is president, and $x_5 = 0$ under Democratic presidents. Additionally, the variable x_5 is lagged one year. This is because it usually takes a certain amount of time until a new political regime has fully implemented its economic policy. If an election leading to a change in government is held in the last quarter of a given year, x_5 is lagged two calendar years.

Test	P-value
Im-Pesaran-Shin	<0.001
Hadri	0.0013

Table 4: Panel unit root tests for the 1985-2013 Gini data.

5.3.2 Results of the dynamic model

First, the tests described in Section 4.2.3 are calculated for the dependent variable. Both the Hadri and Im-Pesaran-Shin tests indicate that there are some stationary time series in the data. The results of the dynamic model are as presented in Table 5. The middle column of the table gives the Arellano-Bond (first-differenced GMM) estimates, while the rightmost column shows the Blundell-Bond (system GMM) estimates. The time period analyzed is 1985 to 2013, and the countries are the same as in the static model. Table 5 shows that the autoregressive parameter is highly significant and over 0.85 in value in the Blundell-Bond model, and almost 0.80 in the Arellano-Bond model. The explanatory variables *trade-to-GDP-ratio*, *industry's share of total employment* and the dummy *right-wing government* are all highly significant according to both models. The signs of the coefficients for the industrial employment and trade variables are negative, whereas it is positive for the government dummy. That is, a decrease in either trade-to-GDP-ratio or employment share in the industrial sector will be associated with an increase in inequality, assuming that the other variables are kept constant. Similarly, a change from a left-wing to a right-wing government should increase income inequality. The numerical effect of such a change is an increase in Gini by between 0.4 and 0.5 percentage points. When it comes to the credit variable, it is positive but not close to being significant in the Blundell-Bond model, whereas it is significant on the 10% level and negative in the Arellano-Bond model. A similar conclusion can be reached regarding the government expenditure variable, which is significant in the Arellano-Bond model and insignificant in the Blundell-Bond setting.

5.3.3 Model evaluation

Regarding the parameter values, the absolute values of the β_i:s are higher for all significant variables using the Blundell-Bond model. This does not come as a surprise, considering the discussion in Section 4.2.2. The value of the autoregressive parameter is also higher in the system GMM. Also, the p-value of the Sargan-Hansen indicates that it is not possible to reject the null hypothesis of instrument validity. Finally, the Arellano-Bond test for AR(2) serial correlation in that model indicates that there is no serial correlation of significance in the AR(2) term.

Moving on to residual analysis, Figure 5 shows the histogram of the residuals as well as the plot of the standardized residuals against the fitted values for each of the models. Table 6 shows the values of the third and fourth moments for the residuals from both models. The kurtosis is much higher than 3, which is

	(1)	(2)
AR parameter, $\hat{\phi}$	$0.778^{\star\star\star}$	$0.852^{\star\star\star}$
	(48.06)	(51.67)
log(Gov.exp./GDP), $\hat{\beta}_2$	$-0.0183^{\star\star\star}$	-0.000895
	(-7.66)	(-0.31)
log(Trade/GDP), $\hat{\beta}_1$	$-0.0365^{\star}$	$-0.0627^{\star\star\star}$
	(-5.24)	(-12.89)
log(Domestic credit), $\hat{\beta}_4$	$-0.00569^{\diamond}$	0.00379
	(-1.84)	(1.22)
Industry's share of total employment, $\hat{\beta}_5$	$-0.0608^{\star}$	$-0.0935^{\star\star\star}$
	(-2.57)	(-4.02)
Right-wing government, $\hat{\beta}_3$	$0.00424^{\star\star\star}$	$0.00484^{\star\star}$
	(6.19)	(3.33)
Sargan-Hansen p-value	0.56	0.37
Arellano-Bond p-value	0.40	
Kolmogorov-Smirnov p-value	$< 0.001^{\star\star\star}$	$< 0.001^{\star\star\star\star}$

Table 5: Coefficient estimates and in brackets, the t-statistics, from the dynamic panel models. In column (1), the results using the Arellano-Bond (difference GMM) model, in column (2) the results of the Blundell-Bond (system GMM) model. Achieved level of significance: $\star\star\star = 0.001, \star\star = 0.01, \star = 0.05, \diamond = 0.10$.

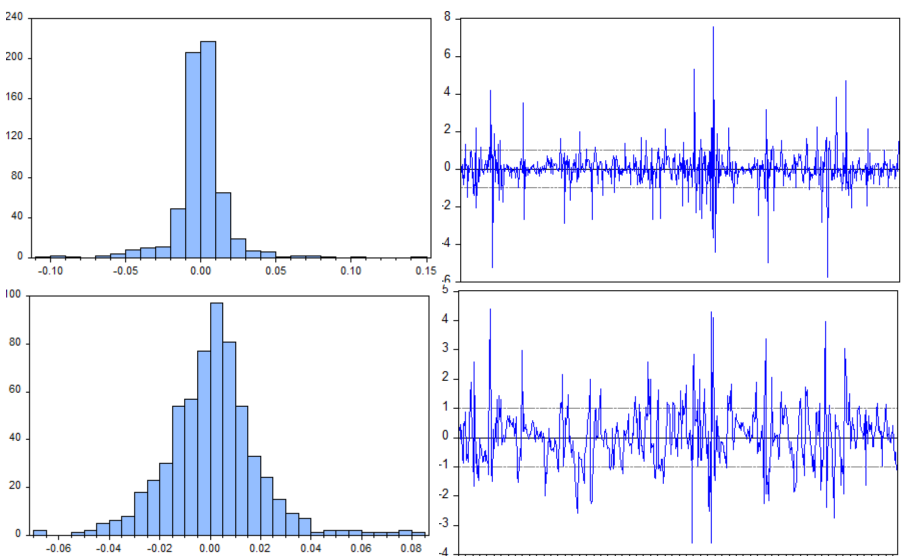

Figure 5: The top left exhibit shows the histogram from the Arellano-Bond model, while the bottom left exhibit is the histogram corresponding to the Blundell-Bond model. The two figures on the right are the corresponding plots of residuals versus fitted values.

Measure	Estimate, AB model	Estimate, BB model
Skewness	0.25	0.40
Kurtosis	15.97	5.50

Table 6: Sample skewness and kurtosis for the residual plots from the dynamic regression.

the value for the normal distribution, for both models. This is especially true for the difference GMM, for which the kurtosis is almost 16. However, since the Sargan p-value indicates that the instruments are valid, the non-gaussianity of the residuals is not of immense importance.

Considering the residuals-versus-fitted values plots in the right-hand side of Figure 5, the Arellano-Bond model has more outlier values compared to the Blundell-Bond model. For example, one observation has a residual equal to almost eight units, which is a relatively large figure, considering that Gini values usually fall in the range of 20 to 40. This is an additional drawback of the AB model.

5.4 Analysis of convergence

The model for analyzing τ-period convergence is

$$\frac{1}{\tau}\log\left(\frac{Gini_{i,T}}{Gini_{t,T-\tau}}\right) = \alpha + \beta\log(Gini_{i,T-\tau}) + u_i \tag{55}$$

where $i = 1, \ldots, N$, $T = 1985, \ldots, 2013$ and $\tau = 1, \ldots, 28$. The model is the same as in Bleaney and Nishiyama (2003) and Dhongde and Miao (2013). Note that the model described by (55) is a standard linear regression, that is, the GMM technique is not used in this case. Set $\tau \in \{5, 10, 15, 20, 25, 28\}$; hence, five-year increments are used. The slightly non-standard value $\tau = 28$ is included to illustrate the average convergence over the entire sample period. The average yearly convergence $\hat{\beta}$ is given for each τ and for different starting years T in Table 7. From this table, it is clear that the average yearly convergence of income inequality is has been slightly less than 2%. For the entire sample period, the average yearly convergence is 1.28 %.

T/τ	5	10	15	20	25	28
1985	-0.0192	-0.0169	-0.0181	-0.0157	-0.0128	-0.0128
	(-1.8867)	(-1.9689)	(-3.1961)	(-4.1271)	(-4.8723)	(-4.9107)
1990	-0.0134	-0.0181	-0.0151	-0.0125		
	(-0.9624)	(-2.4579)	(-3.5954)	(-4.2968)		
1995	-0.0291	-0.0196	-0.0164			
	(-2.5279)	(-4.0983)	(-4.2501)			
2000	-0.0185	-0.0156				
	(-2.2708)	(-2.592)				
2005	-0.0140					
	(-1.3945)					

Table 7: The estimates of the average yearly convergence β for different starting years T and time periods τ. In brackets the corresponding t-statistics for β.

There is a slight tendency of decreasing absolute values of the β_i:s towards the end of the sample period; this is consistent with the literature as described in Section 3.2.5. Not surprisingly, the average convergence rates are less volatile for higher τ, due to the sample period being longer.

6 Conclusion

This thesis has attempted to answer three questions. What factors affect income inequality between countries today? What factors explain the change in inequality over time? And finally, has there been convergence of inequality between countries? The empirical analysis gives strong support for the hypothesis that countries with low levels of agricultural employment have lower inequality, and vice versa. Additionally, countries with high government expenditures have lower levels of inequality, which is consistent with the hypothesis that extensive social safety nets serve to flatten the income distribution. Moreover, a high trade openness as measured by the trade-to-GDP-ratio is also associated with lower income inequality.

Looking back on the period from 1985 to 2013, the income inequality has increased, with some exceptions, in most of the countries sampled. Two variables that are significant in explaining this development are the trade-to-GDP ratio and the relative share of industry employment in the national labor markets. Both of these are negatively associated with income inequality. However, most countries that have seen a rise in one of these variables, have experienced a decrease in the other, giving rise to a globalization paradox. Also, the political color of the government is important in explaining inequality; periods of left-wing government are associated with decreasing levels of income inequality and vice versa. The importance of government political color in explaining income inequality is a new contribution to the existing literature.

Furthermore, the thesis has included a discussion on the performance of the of the most common GMM-based dynamic panel regression models − the Arellano-Bond and Blundell-Bond models. The analysis has confirmed previous empirical research on the matter, that is, that the Blundell-Bond model yields higher parameter estimates, both of the independent regressors, but also of the autoregressive term. Further analysis of the results of the dynamic models shows that the residuals from the Blundell-Bond estimates are more close to a normal distribution than the residuals from the Arellano-Bond model. This result, together with previous theoretical studies on the performance of the two models in Blundell and Bond (1998) and Soto (2009), indicates that the Blundell-Bond model is more robust when the value of the autoregressive parameter is close to unity, as was the case in this thesis.

The analysis of cross-country convergence of income inequality points toward an annual yearly convergence of around 1.5%, which is slightly lower than reported by other studies. However, empirical research on this issue has been relatively scarce. An interesting question for future analysis would be to further investigate the robustness of the two dynamic models under different settings. In particular, the effect of non-stationarity of the time series in the dynamic model is a potential research question.

References

[1] Abreu, M., de Groot, H.L.F. and Florax, R.J.G.M. (2005). A Meta-Analysis of Beta Convergence. *Tinbergen Discussion Paper* 2005-001/3. Amsterdam-Rotterdam: Tinbergen Institute.

[2] Aidt, T.S, Castro, V, Martins, R. (2016). Shades of red and blue: Political ideology and sustainable development. *University of Cambridge Working Paper Economics* 1635.

[3] Alvarez, J. and Arellano, M. (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica* 71 (4) 1121-1159.

[4] Amemiya, T. (1973). Regression Analysis when the Dependent Variable Is Truncated Normal. *Econometrica* 41 (6) 997-1016.

[5] Anderson, T.W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76 (375) 598-606.

[6] Aradhyula, S., Rahman, T. and Seenivasan, K. (2007). Impact of International Trade on Income and Income Inequality. *Working paper, University of Arizona*.

[7] Arellano, M. and Bond, S.R. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58 (2) 277—297.

[8] Arellano, M. and Bover, O. (1995). Another Look at the Instrumental Variables Estimation of Error Components Models. *Journal of Econometrics* 68 (1) 29—52.

[9] Barro, R.J. and Sala-i-Martin, X. (1991). Convergence across States and Regions. *Brookings Papers on Economic Activity* 22 (1) 107-182.

[10] Beck, T., Demirgüç-Kunt, A. and Levine, R. (2007). Finance, Inequality and the Poor. *Journal of Economic Growth* 12 (1) 27-49.

[11] Bénabou, R. (1996). Inequality and growth, in Bernanke B.S. and Rotemberg J. J. (eds), *NBER Macroeconomics Annual* 11, 11–73.

[12] Bleaney, M. and Nishiyama, A. (2003). Convergence in income inequality: differences between advanced and developing countries. *Economics Bulletin* 4 (22) 1-10.

[13] Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1) 115-143.

[14] Breusch, T. S. and Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica* 47 (5) 1287–1294.

[15] Burda, M. and Wyplosz, C. (2013). *Macroeconomics: A European Text*. Sixth edition. Oxford: Oxford University Press.

[16] Bureau of Economic Analysis (2017). *Gross-Domestic-Product-(GDP)-by-Industry Data* [online]. Available at https://www.bea.gov/industry/gdpbyind_data.htm. [Accessed 4 February 2017].

[17] Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

[18] Central Intelligence Agency (2017). *The World Factbook* [online]. Available at: https://www.cia.gov/library/publications/the-world-factbook. [Accessed 4 February 2017].

[19] Chakrabarti, A. (2000). Does Trade Cause Inequality? *Journal of Economic Development* 25 (2) 1-21.

[20] Clarke, G. R. G, Xu, L.C., and Zou, H.(2006). Finance and Income Inequality: What Do the Data Tell Us?. *Southern Economic Journal* 72(3) 578-596.

[21] Clements, B. (1997). The Real Plan, poverty, and income distribution in Brazil. *Finance and Development* 34 (3) 44-46.

[22] Cragg, J.G. and Donald, S.G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory* 9 (2) 222-240.

[23] Dabla-Norris, E., Kochhar, K., Suphaphiphat, N., Ricka, F. and Tsounta, E. (2015). Causes and Consequences of Income Inequality: A Global Perspective. *IMF Staff Discussion Note*. Washington, D,C.: IMF.

[24] Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics* 150 (1) 30-40.

[25] Denk, O. and Cournède, B. (2015). Finance and Income Inequality in OECD Countries. *OECD Economics Department Working Papers* 1224. Paris: OECD Publishing.

[26] Dhongde, S. and Miao, X. (2013). Cross-Country Convergence in Income Inequality. *ECINEQ Working paper*. 2013-290. Verona: Society for the Study of Income Inequality, Università di Verona.

[27] Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute* 22 (1/3) 23–32.

[28] Fournier, J.-M. and Koske, I. (2013). The Determinants of Earnings Inequality: Evidence from Quantile Regressions. *OECD Journal: Economic Studies* 2012/1 7-36. Paris: OECD Publishing.

[29] Frankel, J.A. and Romer, D. (1999). Does Trade Cause Growth? *American Economic Review* 89 (3) 379-399.

[30] Gimet, C. and Lagoarde-Segot, T. (2011). A Closer Look at Financial Development and Income Distribution. *Journal of Banking & Finance* 35 (7) 1698-1713.

[31] Gini, C. (1912). *Variabilitá e Mutabilitá: contributo allo studio delle distribuzioni e delle relazioni statistiche.* Bologna: Tipografia di Paolo Cuppini.

[32] Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3 (2) 148-161.

[33] Hansen, L.P. (1982). Large Sample Properties of Generalized Methods of Moments Estimators. *Econometrica* 50 (4) 1029-1054.

[34] Hausman, J.A. (1978). Specification Tests in Econometrics. *Econometrica* 46 (6) 1251-1271.

[35] Holmquist B., Nasrin, S. and Djurfeldt, G. (2011). Power of tests for endogenity in explanatory binary variables in multivariate linear regression models. *Working paper, Lund University.*

[36] Holtz-Eakin, D., Newey, W. and Rosen, H. (1988). Estimating Vector Autoregressions with Panel Data. *Econometrica* 56 (6) 1371-1395.

[37] ILO (2017). *Share of employment by sector and sex, globally and by region and country* [online]. Available at: http://www.ilo.org/global/research/global-reports/global-employment-trends/2014/WCMS_234879/lang–en/index.htm. [Accessed 4 February 2017].

[38] Im K.S., Pesaran M.H. and Shin Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Economics* 115 (1) 53-74.

[39] Jauch, S. and Watzka, S. (2012). Financial development and income inequality: a panel data approach. *CESifo Working Papers* 3687. Munich: CESifo.

[40] Jaumotte, F., Lall, S., Papageorgiou, C. (2008). Rising Income Inequality: Technology, or Trade and Financial Globalization? *Working paper, IMF* 08/185. Washington, D,C.: IMF.

[41] Keister, L.A. (2000). *Wealth in America: Trends in Wealth Inequality.* Cambridge: Cambridge University Press.

[42] Kolmogorov, A (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4 (1) 83–91.

[43] Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9 (70) 209-219.

[44] Lundberg, M. and Squire, L. (2003). The Simultaneous Evolution of Growth and Inequality. *The Economic Journal* 113 (487) 326-344.

[45] Madsen, E. (2008). GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. *Working paper, University of Copenhagen.*

[46] de Mello, L. and Tiongson E.R. (2006). Income Inequality and Redistributive Government Spending. *Public Finance Review* 34 (3) 285-305.

[47] Milanovic, B. and Squire, L. (2005): Does Tariff Liberalization Increase Wage Inequality? Some Empirical Evidence. *World Bank Policy Research Working Paper* 3571.

[48] Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49 (6) 1417-1426.

[49] Oh, K. (1996). Purchasing power parity and unit root tests using panel data. *Journal of International Money and Finance* 15 (3) 405-418.

[50] Ravallion, M. (2003). Inequality convergence. *Economics Letters* 80 (3) 351–356.

[51] Pesaran, M.H. (2015). *Time Series and Panel Data Econometrics.* Oxford: Oxford University Press.

[52] Philippon, T. and Reshef, A. (2012). Wages and Human Capital in the U.S. Financial Industry: 1909−2006. *The Quarterly Journal of Economics* 127 (4) 1551-1609.

[53] Pyatt, G. (1976). On the interpretation and Disaggregation of Gini Coefficients. *The Economic Journal* 86 (342) 243-255.

[54] Sargan, J.D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica* 26 (3) 393-415.

[55] Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* 19 (2) 279–281.

[56] Solow, R.M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70 (1) 65-94.

[57] Solt, F. (2016). The Standardized World Income Inequality Database. *Social Science Quarterly* 97 (5) 1267-1281. SWIID Version 5.1, July 2016.

[58] Soto, M. (2009). System GMM estimation with a small sample. *Barcelona Economics Working Paper Series* 395.

[59] Staiger, D. and Stock, J.H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* 65 (3) 557-586.

[60] Stock, J., and Yogo, M. (2005). Testing for Weak Instruments In Linear IV Regression in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. Andrews and J. H. Stock. Cambridge: Cambridge University Press, 80–108.

[61] Swan, T.W. (1956). Economic Growth and Capital Accumulation. *Economic Record* 32 (2) 334-361.

[62] van der Waart, A.W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

[63] Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of American Mathematical Society* 54 (3) 426-482.

[64] World Bank (2017a). *General government final consumption expenditure (% of GDP).* Available at: http://data.worldbank.org/indicator/NE.CON.GOVT.ZS. [Accessed 4 February 2017].

[65] World Bank (2017b). *Trade (% of GDP)* [online]. Available at: http://data.worldbank.org/indicator/NE.TRD.GNFS.ZS. [Accessed 4 February 2017].

[66] World Bank (2017c).*Domestic credit to private sector (% of GDP)* [online]. Available at: http://data.worldbank.org/indicator/FS.AST.PRVT.GD.ZS. [Accessed 4 February 2017].

[67] White, H. (1982). Instrumental Variables Regression with Independent Observations. *Econometrica* 50 (2) 483-499.

[68] Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

[69] Wu, D.-M. (1973). Alternative Tests of Independence between Stochastic Regressors and Disturbances. *Econometrica* 41 (4) 733-750.

[70] Yohimoto, H. (2008). Generalized Method of Moment (GMM) Estimator. *ECON 203C: System Models.* UCLA, unpublished.

Appendix

Proof of Theorem 1

Theorem 1. *The estimator $\hat{\boldsymbol{\theta}}_{GMM}$ is a consistent estimator of the unknown parameter vector $\boldsymbol{\theta}$.*

Proof. The proof is based on Yoshimoto (2008). In order to show consistency of the GMM estimator, it first needs to be shown that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_N(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})|$ converges in probability to $\mathbf{0}$. To simplify notation, let $\{\boldsymbol{w}_i\} = \{y_i, \boldsymbol{x}_i\}$. Expanding terms and using the triangle inequality together with some standard algebra, yields

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_N(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta})) - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta}) - \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] \\
&\quad + \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] - \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] \\
&\quad + \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_n(\boldsymbol{\theta}) - \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] - \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]'\boldsymbol{W}\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N \left\{\boldsymbol{g}_N(\boldsymbol{\theta}) - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})\right\}] \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'(\boldsymbol{W}_N - \boldsymbol{W})\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})] \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\left\{\boldsymbol{g}_N(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta}])'\right\}\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N| \overbrace{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta}) - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]}^{\xrightarrow{P} 0} \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta})'| \overbrace{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |(\boldsymbol{W}_N - \boldsymbol{W})|}^{\xrightarrow{P} 0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]| \\
&\quad + \overbrace{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{g}_N(\boldsymbol{\theta}) - \mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta}])|}^{\xrightarrow{P} 0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{W}_N\mathbb{E}[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})]|
\end{aligned}$$

Since all three terms of the final equality converge in probability to $\mathbf{0}$, it holds that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_N(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \xrightarrow{P} 0 \tag{56}$$

Finally, by Lemma 3 of Amemiya (1973), the result in (56) implies that

$$\hat{\boldsymbol{\theta}}_{GMMM} \xrightarrow{P} \boldsymbol{\theta}$$

which shows that $\hat{\boldsymbol{\theta}}_{GMM}$ is indeed a consistent estimator of the unknown parameter vector $\boldsymbol{\theta}$. ■

Proof of Theorem 2

Theorem 2. *The quantity* $\sqrt{N}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})$ *is asymptotically normal with mean zero and covariance matrix* $\boldsymbol{\Omega}$*.*

Proof. The below proof follows Yoshimoto (2008), albeit slightly modified. Since $\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta})$, take the partial derivative of the quadratic form in order to find the argument of the minimum. This yields

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \left[\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta})\right] \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \left[\boldsymbol{g}_N(\boldsymbol{\theta})'\right] \frac{\partial}{\partial \boldsymbol{g}_N(\boldsymbol{\theta})} \left[\boldsymbol{g}_N(\boldsymbol{\theta})'\boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta})\right] \\
&= 2\frac{\partial \boldsymbol{g}_N(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta}) \qquad (57)
\end{aligned}$$

Set (57) equal to $\mathbf{0}$:

$$\frac{\partial \boldsymbol{g}_M(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM}) = \mathbf{0} \qquad (58)$$

Now, according to the mean value theorem it is possible to write $\boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})$ as

$$\boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM}) = \boldsymbol{g}N(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) \qquad (59)$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}_{GMM}$ and $\boldsymbol{\theta}$. Substitute (59) into (58) and obtain

$$\begin{aligned}
&\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N \left[\boldsymbol{g}_N(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})\right] = \mathbf{0} \\
\iff &\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) = \mathbf{0} \\
\iff &\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) = \frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta}) \\
\iff &(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) = \left[-\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]^{-1} \frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\boldsymbol{g}_N(\boldsymbol{\theta})
\end{aligned}$$

Multiplying by $\sqrt{N}$ on both sides gives

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) = \left[-\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N \frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]^{-1} \frac{\partial g_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_N\sqrt{N}\boldsymbol{g}_N(\boldsymbol{\theta})$$

Given that the sequence of observations $\{\boldsymbol{w}_i\}$ are i.i.d., $\mathbb{E}\left[\boldsymbol{f}(\boldsymbol{w}_i, \boldsymbol{\theta})\right] < \infty$ and $\mathbb{V}\left[\boldsymbol{f}(\boldsymbol{w}_i, \boldsymbol{\theta})\right] < \infty$, it is possible to apply the Lindeberg-Lévy CLT to $\sqrt{N}\boldsymbol{g}_N(\boldsymbol{\theta})$, which yields

$$\sqrt{N}\boldsymbol{g}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta}) \xrightarrow{P} \mathcal{N}\left\{\mathbf{0}, \mathbb{E}\left[\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})'\right]\right\} \tag{60}$$

Now, by utilizing first the sum rule of differentiation and then the weak law of large numbers, it holds that

$$\begin{aligned}
\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}')}{\partial \boldsymbol{\theta}'} &= \frac{\partial}{\partial \boldsymbol{\theta}'}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})\right] \\
&= \frac{1}{N}\left\{\frac{\partial}{\partial \boldsymbol{\theta}'}\left[\sum_{i=1}^{nN}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})\right]\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \xrightarrow{P} E\left[\frac{\partial \boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]
\end{aligned}$$

A similar exercise can be done in order to show the convergences of the partials of $\boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'$ and $\boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})$. Since it is known from the weak law of large numbers that $\boldsymbol{W}_N \xrightarrow{P} \boldsymbol{W}$, applying the continuous mapping and Slutsky theorems yields

$$\begin{aligned}
&\left[-\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}}\boldsymbol{W}_N\frac{\partial \boldsymbol{g}_N(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]^{-1}\frac{\partial \boldsymbol{g}_N(\hat{\boldsymbol{\theta}}_{GMM})'}{\partial \boldsymbol{\theta}}\boldsymbol{W}_N\sqrt{N}\boldsymbol{g}_N(\boldsymbol{\theta}_{\mathbf{0}}) \\
&\xrightarrow{P}\left\{-\mathbb{E}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right]\boldsymbol{W}\mathbb{E}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]^{-1}\right\}\mathbb{E}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{w}_i,\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right]\boldsymbol{W}
\end{aligned}$$

And thus, using the result in (60),

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) \xrightarrow{P} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$$

where $\boldsymbol{\Omega}$ is as defined in the main body of the text. ■

Proof of Theorem 3

Theorem 3. $\boldsymbol{\Omega}^\star$ *has the smallest possible asymptotic variance amongst all weight matrices.*

Proof. Following Pesaran (2015, p. 232), let $\boldsymbol{\Omega}^\star = (\boldsymbol{G}'\boldsymbol{\Upsilon}^{-\mathbf{1}}\boldsymbol{G})^{-1}$ and $\boldsymbol{\Omega} = (\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{\Upsilon}\boldsymbol{W}'\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}'\boldsymbol{G})^{-1}$ as before. The goal is to show that $\boldsymbol{\Omega}-\boldsymbol{\Omega}^\star$ is positive semi-definite (p.s.d). This is equivalent to $(\boldsymbol{\Omega}^\star)^{-1} - \boldsymbol{\Omega}^{-1}$ being p.s.d. First, utilize the Cholesky decomposition $\boldsymbol{\Upsilon} = \boldsymbol{C}\boldsymbol{C}'$ and define $\boldsymbol{R} = \boldsymbol{C}^{-1}\boldsymbol{G}$, $\boldsymbol{B} = \boldsymbol{C}'\boldsymbol{W}'$ and $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{G}$. Now,

$$
\begin{aligned}
(\boldsymbol{\Omega}^\star)^{-1} - \boldsymbol{\Omega}^{-1} &= (\boldsymbol{G}'\boldsymbol{\Upsilon}^{-\mathbf{1}}\boldsymbol{G}) - \left[(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{\Upsilon}\boldsymbol{W}'\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}'\boldsymbol{G})^{-1}\right]^{-1} \\
&= (\boldsymbol{G}'\boldsymbol{C}^{-1\prime}\boldsymbol{C}^{-1}\boldsymbol{G}) - (\boldsymbol{G}'\boldsymbol{W}'\boldsymbol{G})(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\boldsymbol{C}'\boldsymbol{W}'\boldsymbol{G})^{-1}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}) \\
&= (\boldsymbol{G}'\boldsymbol{C}^{-1\prime}\boldsymbol{C}^{-1}\boldsymbol{G}) - (\boldsymbol{G}'\overbrace{\boldsymbol{C}^{-1\prime}\boldsymbol{C}'}^{=\boldsymbol{I}}\boldsymbol{W}'\boldsymbol{G})(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\boldsymbol{C}'\boldsymbol{W}'\boldsymbol{G})^{-1} \\
&\times (\boldsymbol{G}'\boldsymbol{W}\overbrace{\boldsymbol{C}^{-1}\boldsymbol{C}}^{=\boldsymbol{I}}\boldsymbol{G}) \\
&= \left[\boldsymbol{G}'\boldsymbol{C}^{-1\prime} - \boldsymbol{G}'\overbrace{\boldsymbol{C}^{-1\prime}\boldsymbol{C}'}^{=\boldsymbol{I}}\boldsymbol{W}'\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\boldsymbol{C}'\boldsymbol{W}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\right] \\
&\times \boldsymbol{C}^{-1}\boldsymbol{G} \\
&= \boldsymbol{G}'\boldsymbol{C}^{-1\prime}\left[\boldsymbol{I} - \boldsymbol{W}'\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\boldsymbol{C}'\boldsymbol{W}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{C}\right]\boldsymbol{C}^{-1}\boldsymbol{G} \\
&= \boldsymbol{R}'\left[\boldsymbol{I} - \boldsymbol{B}\boldsymbol{C}(\boldsymbol{G}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{G})^{=-1}\boldsymbol{G}'\boldsymbol{B}'\right]\boldsymbol{R} \\
&= \boldsymbol{R}'\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]\boldsymbol{R}
\end{aligned}
$$

Now, the final step is to show that $\boldsymbol{R}'\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]\boldsymbol{R}$ is p.s.d. It suffices to show that $\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]$ is p.s.d., because for any p.s.d. matrix $\boldsymbol{A}$, the linear combination $\boldsymbol{q}'\boldsymbol{A}\boldsymbol{q}$ is p.s.d. as well.
Multiplying $\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]$ with itself and expanding terms, yields

$$
\begin{aligned}
&\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right] \\
&= \boldsymbol{I} - 2\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L}\boldsymbol{L}')^{-1}\boldsymbol{L}' + \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L}\boldsymbol{L}')^{-1}\overbrace{\boldsymbol{L}'\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}}^{=\boldsymbol{I}}\boldsymbol{L} \\
&= \boldsymbol{I} - 2\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}' + \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{=1}\boldsymbol{L}' \\
&= \boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'
\end{aligned}
$$

Since $\left[\boldsymbol{I} - \boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\right]$ is idempotent, it is p.s.d., and hence $(\boldsymbol{\Omega}^\star)^{-1} - \boldsymbol{\Omega}^{-1}$ is p.s.d. This completes the proof of Theorem 3. ■

Summary statistics

Tables 6 and 7 show the correlation matrix and data summary for the entire sample, 1985-2013. Similarly, tables 8 and 9 show the correlation matrix and the data summary for the subsample used in the static model, i.e. for the year 2013 only.

	Gini	Gov.exp.	Trade/GDP	Credit	Ind. share
Gini	1	-0.58	-0.43	-0.11	-0.20
Gov.exp.	-0.58	1	0.26	0.01	0.01
Trade/GDP	-0.43	0.26	1	0.00	0.04
Credit	-0.11	0.01	-0.32	1	-0.32
Ind. share	-0.20	0.00	0.04	-0.32	1

Table 8: Correlation matrix for the entire dataset.

Variable	**min**	**Q1**	**mean**	**Q3**	**max**
Gini	18.98	25.79	30.46	33.00	52.65
Gov.exp.	4.58	14.63	16.69	19.39	26.70
Trade/GDP	14.39	47.21	69.50	84.29	197.22
Credit	7.09	49.01	89.38	117.49	312.15
Industry's share of emp.	15.78	22.40	26.23	29.60	44.29
Right-wing govt. $\geq$ 15 mil.	0	0	0.56	1	1

Table 9: The descriptive statistics for the dataset used dynamic model.

Variable	**min**	**Q1**	**mean**	**Q3**	**max**
Gini	22.80	26.80	30.62	32.80	47.20
Gov.exp.	7.16	16.86	18.32	20.06	25.23
Trade/GDP	25.60	60.70	87.89	103.30	194.00
Const. trade share.	4.01	14.71	27.69	43.05	56.56
Credit	33.86	81.67	115.25	142.33	253.57
High agric. emp.	0	0	0.28	1	1
Population $\geq$ 15 mil.	0	0	0.48	1	1

Table 10: The descriptive statistics for the year 2013 only.

	Gini	Gov.exp.	Trade/GDP	Cons. trd. sh.	Credit
Gini	1	-0.53	-0.53	-0.58	0.05
Gov.exp.	-0.53	1	0.06	0.04	-0.06
Trade/GDP	-0.53	0.06	1	0.91	-0.24
Cons. trd. sh.	-0.58	0.04	0.91	1	-0.25
Credit	0.05	-0.06	-0.24	-0.25	1

Table 11: Correlation matrix for the 2013 subsample.