1 2 9 0

UNIVERSIDADE D COIMBRA

Miguel António Murtinho Fernandes

A reassessment of the relationship between earnings inequality and growth for the Portuguese economy between 1986 and 2017

Dissertation of the master's in Economics/Economic growth and structural policies of the Faculty of Economics of the University of Coimbra

Supervisor: Maria Adelaide P. Silva Duarte

Supervisor: João Alberto Sousa Andrade

January 2020

Acknowledgements

My gratitude to Professor Maria Adelaide Duarte for the guidance during this semester, for all the valuable suggestions on how to develop this work and for the thorough review of this work. I would also like to express my gratitude to Professor João Sousa Andrade for teaching me how to use the R software, for all the help with this work, especially the econometric part and for the review of this work. I would like to thank my colleague at CeBER, Derick Almeida for his suggestions for this work and timesaving share of information and code. To my family, my brother, father and mother a special thanks for always supporting me. Finally, I would like to thank my friends and colleagues of this master´s course (who I will not name out of fear that I forget someone) for making these last 3 semesters so enjoyable.

Cofinanciado por:

Referência do projeto:

POCI-01-0145-FEDER-029365 | PTDC/EGE-ECO/29365/2017

"WISER Portugal – Welfare Intervention by the State and Economic Resilience in Portugal"

Financiamento:

Projeto cofinanciado pelo COMPETE 2020, Portugal 2020 e União Europeia, através do FEDER e pela Fundação para a Ciência e Tecnologia I.P./MCTES através de fundos nacionais (PIDDAC).

Abstract

The empirical literature that studies the relationship between income inequality and growth has often found conflicting results. While much of it is based on cross-country studies, some authors have called for a country-specific approach, since inequality may influence each country differently. This work focuses on the study of the Portuguese case, from 1986 to 2017, by descriptively analyzing earnings inequality in Portugal and by empirically attempting to determine the sign of the relationship between inequality and growth and its underlying mechanisms. The country's growth trajectory: initially high growth rates, followed by a period of stagnation after 2001, leading up to the sovereign debt crisis and its consequences; and its higher levels of inequality and lower levels of human capital and productivity when compared with other European countries, make it a meaningful case study. The source of data used was *Quadros de Pessoal*, a micro-database comprised of information about millions of private-sector workers, and the online database PORDATA. In order to examine the relationship between inequality and growth a VAR model with 4 variables (Human Capital, Inequality, Investment and GDP) was implemented. From the interpretation of the impulse response functions obtained it was found that inequality has a negative impact on growth, a negative impact on human capital and an initial negative impact on investment, eventually becoming positive. Both human capital and investment have a positive effect on growth. There was also evidence that the "human capital", "savings" and "domestic demand" mechanisms have driven the relationship between inequality and growth in Portugal.

Keywords: Inequality, Growth, VAR, Inequality measures, Portugal.

JEL Codes: *O15, O52*

Resumo

Os resultados da literatura empírica que estuda a relação entre a desigualdade de rendimento e o crescimento, têm sido, muitas vezes, contraditórios. Muito do trabalho empírico é baseado em estudos com vários países, mas alguns autores relevam a necessidade de uma abordagem país a país, pois a desigualdade pode influenciar cada país de maneira diferente. Este trabalho foca-se no estudo do caso português, para o período compreendido entre 1986 e 2017, através de uma análise descritiva da desigualdade salarial em Portugal e através de uma análise empírica com o propósito de determinar o sinal da relação entre crescimento e desigualdade e os mecanismos subjacentes. A trajetória de crescimento do país: inicialmente um crescimento rápido, seguido de um período de estagnação a partir de 2001 e eventualmente chegando à crise da dívida soberana e às suas consequências; e os seus elevados níveis de desigualde e baixos níveis de produtividade e capital humano, fazem de Portugal um importante caso de estudo. A fonte de dados usada é a base de dados *Quadros de Pessoal*, uma base que contém micro-dados sobre milhões de trabalhadores do setor privado e também a PORDATA. A escolha do modelo empírico para analisar a relação entre crescimento e desigualdade recaiu sobre um VAR com quatro variáveis (Capital Humano, Desigualdade, Investimento e PIB). Através da interpretação das funções impulsos resposta, obtidas através do modelo VAR, concluiu-se que a desigualdade tem um impacto negativo no crescimento, um impacto negativo no capital humano e um impacto inicial negativo no investimento, mas que eventualmente se torna positivo. Tanto o capital humano como o investimento mostraram ter um efeito positivo no crescimento. Foi também possível concluir que os mecanismos do "capital humano", da "poupança" e da "procura interna" influenciaram a relação entre desigualdade e crescimento.

Palavras-chave: Desigualdade, Crescimento, VAR, Medidas de desigualdade, Portugal.

Códigos JEL: *O15, O52*

List of Figures

List of Tables

1. Contents

2. Introduction

For a long time now, growth economics has discussed the role of the welfare state in stimulating the growth of an economy. There is no consensus amongst economists, some highlight the potentially distortionary effects of wealth (income) redistribution and others point to the positive externalities that the welfare state can generate. The ways in which the welfare state may affect growth are many and in order to understand this complex issue it is necessary to separate it into narrower questions. One of such questions is how inequality of income affects growth.

Over the last three decades, there has been an extensive research covering the relationship between income inequality and growth and no definitive conclusion has been reached (Cingano 2014).

The aim of this work is to add to this debate by analyzing earnings inequality, its relation to growth and the mechanisms through which this relationship might operate, for Portugal, during the period 1986-2017. The country's growth trajectory: initially high growth rates with real convergence towards the European Union average, followed by a period of stagnation and divergence after 2001, leading up to the sovereign debt crisis, the recession during the crisis and the slow recovery after; coupled with its high levels of inequality, low levels of human capital and productivity when compared to the rest of Europe, make the Portuguese case a meaningful case study.

In order to achieve its goal this work was broadly divided into two parts. The first one covers the implementation of suitable inequality measures from the database *Quadros de Pessoal* and of an encompassing characterization of the evolution of earnings inequality in Portugal for the period under consideration. To accomplish that objective, several statistical measures are computed and displayed in time-series plots, which allow compelling interpretations and a clear visualization of inequality trends. The second part is to build an econometric model capable of providing useful information to determine the signal of the interdependencies between inequality and growth with special focus on human capital. To perform the econometric analysis the choice fell on a VAR (Vector Auto Regressive) model due to its non-restrictive nature and capacity to deal with endogeneity, Sims (1980). A brief analysis of causality was carried out, based on non-parametric estimation and on the methodology proposed by Vinod (2017). The variance decomposition was computed as well as the impulse-response functions. The outputs from the later were crucial to drawing conclusions. It was found that inequality has a negative effect on growth and that there could be multiple channels for that effect, namely the "human capital", "savings" and "domestic demand" mechanisms.

The work is organized in six sections, starting with the introduction, followed by section 3, the literature review, which briefly covers selected literature on inequality and growth, on VAR models applied to inequality and on inequality measures. Section 4 presents the database in the first two subsections and then carries on with the characterization of earnings inequality in Portugal. Section 5 presents the empirical strategy and contains the relevant information regarding the choice and specification of the VAR model. Section 6 presents the results of the variance decomposition and impulse response analysis and the respective economic interpretation. Section 7 concludes.

3. A selected literature review

The literature review is divided into three topics. The first, inequality and growth, contains a brief overview of empirical studies and the theoretical framework in which they operate. The second, VAR models applied to income inequality, focuses on the empirical literature that has used the VAR model to study inequality and growth and the third, income inequality measures, is a concise description of some inequality measures and their characteristics.

3.1 Inequality and growth

There is an abundance of literature concerning the relationship between inequality and growth. The question that researchers have been trying to answer is whether initial inequality has a positive or, instead, a negative influence on growth. The results have not pointed clearly to a specific sign, possibly due to income inequality having different effects in different countries, the lack of quality data for many countries, the estimation methods used and the choices of the inequality index (Cingano, 2014). Furthermore, the economic literature does not agree on the mechanisms through which inequality affects growth, with different authors advancing several hypotheses. Some of the most popular and relevant to this work are:

-The human capital mechanism, which states that higher inequality, coupled with financial market imperfections will lead to a situation where the poor under-invest in human capital

(Galor & Zeira, 1993). For example, a poor individual might not invest in a tertiary education that would have a high return in terms of his future income, due to budget constraints. In this case, higher inequality would lead to lower growth.

-The idea that higher inequality is met by a popular demand for more redistribution and therefore inequality is endogenous to policy making. Assuming redistribution policies have an adverse effect on growth this would mean higher inequality leads to slower growth (Alesina & Rodrik, 1994).

- The possibility that the adoption of new technologies is dependent on a certain amount of domestic demand (Murphy et al., 1989) and inequality can adversely affect domestic demand (Krueger, 2012).

- The "savings" mechanism, which essentially considers that with more inequality, the rich, who have a higher propensity to save, will have a larger share of the wealth and foster the aggregate savings of the economy. According to the neo-classical growth models, growth is directly and positively related to the amount of savings in an economy (Kaldor, 1955). In this case, higher inequality would lead to higher growth.

- The argument that higher inequality is an incentive for the individual to take risks, work harder and investing in education and therefore leads to higher growth.

There have been some studies attempting to test the prevalence of a specific mechanism as the driver of the relationship between inequality and growth, but the literature in this specific aspect is still limited when compared to the literature that aims to estimate the reduced-form relationship between those two variables. (Cingano, 2014; Neves, Afonso & Silva, 2016). In addition to this it has been found by Neves, Afonso and Silva (2016), building on the meta-analysis work of Dominics et al. (2008), that different country contexts can influence which mechanism is the dominant one in each country and that this influence is not time-invariant, potentially changing depending on whether the short-run or the long-run is considered. These conclusions are in consonance with the work of Gobbin and Rayp (2008) that highlighted the need for a country-specific approach, after obtaining significantly different results between countries, with Johansen's cointegration methodology, when studying Belgium, the US and Finland. The above-mentioned issues provide a strong case for studying inequality using a time-series approach as done, for example, by Andrade, Duarte and Simões (2014).

3.2 VAR models applied to income inequality

While most studies try to find the impact inequality has on growth, it is important to note that growth also influences income inequality as evidenced by the Kuznets Curve. The fact that these two variables are inter-dependent is an obstacle to empirical research as it raises a problem of endogeneity and valid exogenous instruments are difficult to find (Durlauf, Johnson & Temple 2005).

Sims (1980) proposal of VAR models solves these problems by allowing interdependencies between the variables, therefore not needing to disentangle between exogenous and endogenous variables.

There are several examples of VAR models being applied to income inequality. Assane and Grammy (2003) used a trivariate VAR model with real GDP, the Gini and a human capital measure to study the US during the period 1960-96 and found that causality ran from growth to inequality with higher growth causing more inequality. Frank (2009) however, found that top income shares causes income growth and weak evidence that income growth causes the top decile share of income. Risso and Carrera (2012) found a positive relationship between inequality and growth for China, corroborated by the bivariate VAR model of Chan, Zhou and Pan (2014). Atems and Jones (2014) made use of a panel VAR using U.S. state-level data and with their baseline bivariate VAR model found that shocks to the Gini index have an adverse impact on the level of income. Andrade, Duarte and Simões (2014) also included a SVAR and Near-Var analysis and found, for Portugal, for the period 1985-2007 "that inequality was detrimental to growth". This study also looked at the transmission mechanisms to determine which were responsible for this negative relationship and found the most likely one to be the idea that more inequality leads to more redistribution and, through distortionary effects, to slower growth.

3.3 Income inequality measures

One of the main empirical problems when studying income inequality is how to measure it. A vast and continuous literature studies the advantages and disadvantages of conventional measures and proposes new measures, ((see, Langel (2012) and (McGregor, Smith & Wills, 2019)). An inequality measure is defined by Cowell and Victoria-Feser (1996) as a member of a class of functions that is defined by a set of essential characteristics. Those characteristics should include Pigou-Dalton transfer sensitivity: a transfer of income from a richer person to a poorer person must reduce the measured inequality Dalton (1920); scale independence: the size of the population does not influence measured inequality; mean independence: if all incomes are multiplied by a scalar (other than 0) measured inequality must remain constant and decomposability: the measured inequality may be separated into fractions of that inequality which are attributed to sub-groups of the original distribution.

The Gini coefficient is arguably the most popular inequality measure and it satisfies most of the properties listed above, with the one downside being that it is not easily decomposable. Other popular inequality measures are often based on the income shares of certain parts of the distribution, examples are the 20:20 ratio that compares the income of the top 20% income share to that of the bottom 20% income share and the Palma-ratio, proposed by Palma (2011), which divides the share of income of the 10% richer by that of the 40% poorer. The merit of this measure compared to the Gini index is that it is more sensible to changes in the top and bottom of the distribution, while the Gini can be oversensitive to changes in the middle of the distribution. Inequality measures based on only certain parts of the distribution, as the two mentioned above, do not fulfill the criteria of the Pigou-Dalton transfer sensitivity.

A useful set of inequality measures that satisfies all the conditions generally considered desirable is the generalized entropy index. This index comes from information theory and was proposed by Theil in 1967. The equation for the generalized entropy index can be written as:

$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\ (\frac{y_i}{\bar{y}})^{\alpha} - 1\right]$, where N is the number of observations, y_i is the income of observation i, $\bar{y}$ is the average income and α is the parameter that determines the weight given to the different parts of the distribution.

From this equation three important inequality measures can be deduced, the Theil’s T, GE(1), Theil’s L, GE(0) and half the squared coefficient of variation, GE(2). The higher α, the higher the sensibility of this index to changes in the top of the distribution.

4. A portrait of Portuguese earnings inequality from 1986 to 2017

This section of the work is divided into three subsections; the first one describes the database *Quadros de Pessoal* used in this work. The second one details the adjustments done to the database in order to improve the reliability of the statistics computed from it. The final subsection presents six figures and the respective interpretation, with the purpose of illustrating the dynamics of earnings inequality in Portugal.

4.1 Overview of *Quadros de Pessoal* database

Quadros de Pessoal database is a systematic compilation of information regarding workers of the private sector resulting from an annual compulsory survey conducted by the Ministry of Solidarity and Social Security (MSSS) where firms provide information about their workers on a variety of items. The information contained in this database includes, for each worker, its base salary, total salary, level of qualifications, job description, among others. Thus, this database works with data at a micro level, making it extremely well suited to be used in the computation of inequality measures.

In this database the base salary is defined as the amount in money and/or goods, before taxes and deductions, payed to the worker on a monthly basis, corresponding to the month of October and to the normal number of hours worked. Total salary is equivalent to the base salary plus all regular bonuses and subsidies and earnings from extra hours worked.

The “workers of the private sector” included in this database are employees working in the private sector, employers who have job in the company, non-payed family workers and active members of a production cooperative. The database does not include all workers that satisfy one of these conditions, since not all companies were surveyed every year, but the number of workers included has been steadily increasing over the years, starting at 1.9 million in 1986 and having over 3 million in 2017. Some of the information included in the database regarding each worker includes, among others, level of education, tenure, sex, region and sector of activity.

The main limitations of this database are its lack of information about public employees and the fact that it was not built for the years 1990 and 2001.

4.2 Adjustments to the database *Quadros de Pessoal*

In order to produce relevant statistics and inequality measures suited to our research, the database had to be adjusted. The following steps were undertaken:

-Every worker whose base salary was below the minimum salary for that year was removed from the database. This measure removed observations that had a salary of 0 and removed potential mistakes that might have happened somewhere in the process of collecting and assembling the data, since no worker can receive an amount below the minimum salary for a full-time job. Another result from this transformation was removing many of the part-time workers, which are not directly comparable to full-time workers.

-The highest 0.5% of total salaries were removed in order to eliminate extreme outliers and mistakes in the data that could taint the analysis.

-The base has nominal salaries, so all salaries were properly deflated (base=2010) as to allow for comparisons between different years.

All these adjustments and the following statistics, computations and figures were made using the R software. Because of the two missing years 1990 and 2001, the time-series produced were interpolated by Kalman smoothing of an automatically generated ARIMA process selected by the AICc criterium, based on the R package "imputeTS" (Moritz S, 2017).

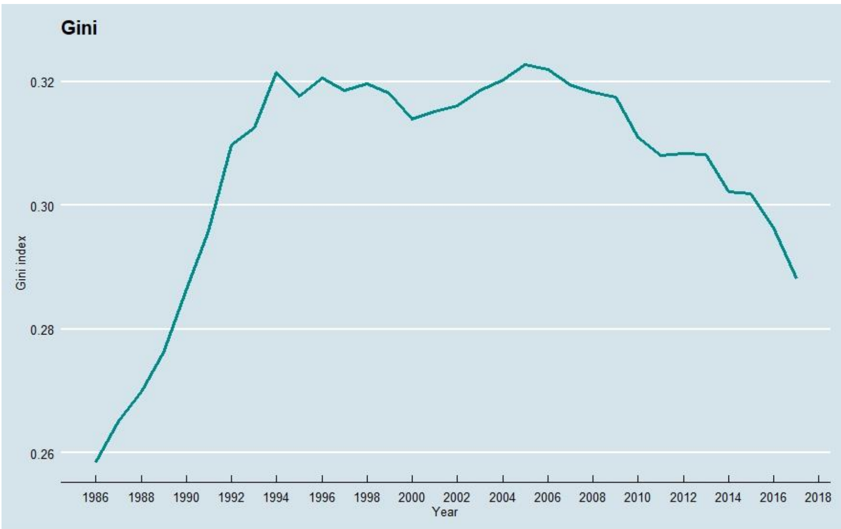
4.3 Earnings Inequality: a descriptive analysis

While inequality measures like the Gini and Theil can be very useful when one wants a synthetic indicator for comparisons between years or countries, the fact that they boil down inequality to a single number means that a lot of relevant information is lost. As such, for studying inequality, several different statistics have to be produced to mitigate the above mentioned problem.

This section aims to present such statistics and interpret them in a way that sheds light on the specificities of the Portuguese earnings distribution. The R package "ineq" (Zeileis, Achim, 2014), was used to compute the inequality metrics.

It makes sense to start by presenting the Gini index, computed from the *Quadros de Pessoal* database, for the period 1986-2017 (this will be the reference period for every analysis unless it is explicitly said otherwise). The Gini index is the most widely used inequality measure and therefore the ideal benchmark with which to compare other statistics that attempt to capture specific characteristics of the earnings distribution. Figure 1 presents the Gini index. The evolution of the index can be divided into three separate periods. The first from 1986 to 1994 is a period of rapid inequality increase, the second from 1995 to 2007 presents inequality fluctuating but not growing or decreasing, finally in the third period, from 2008 to 2017, inequality starts decreasing. The inverted U shape of this curve resembles the Kuznets curve and there is a case to be made that it applies here. Initially inequality is low as the economy is not very developed, then a period of rapid growth leads to increasing inequality, but as the economy converges with more advanced economies the increase in inequality slows down and eventually reverses. Another good argument is that the 2007-08 crisis and the subsequent euro area crisis is responsible for the decrease in inequality amongst active workers. The upper classes being relatively more affected in terms of their salaries and the growing unemployment of this time period affecting mostly low-earning workers, meaning they were removed from statistics computed from the source *Quadros de Pessoal*. The work of Rodrigues, Figueiras and Junqueira (2016) found that individuals at the bottom 10% of the distribution lost the most income out of all deciles of the income distribution, based on an inquiry that was not limited to active workers. This discrepancy in findings suggests that earnings inequality in the labor market might differ from income inequality in society at large, under certain economic circumstances.

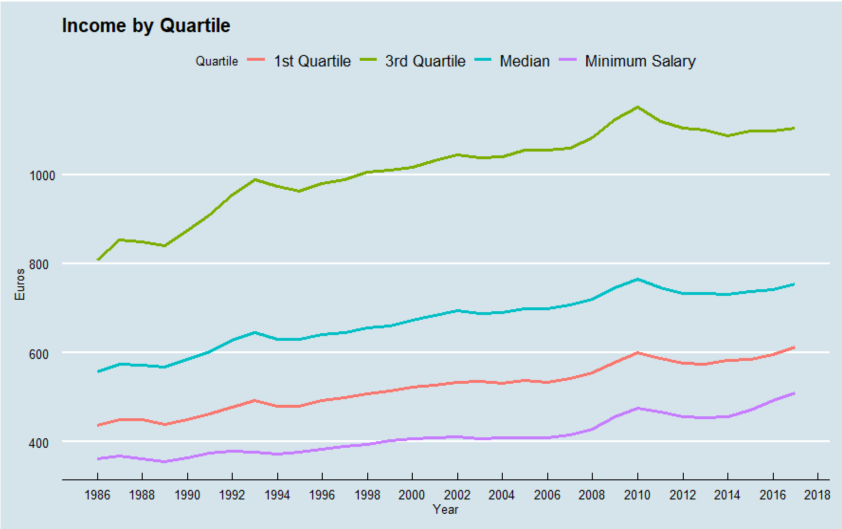
Figure 1: Evolution of the Gini index



Source: author’s own computations based on *Quadros de Pessoal* database.

Figure 2 shows the evolution of total salary by quartiles, as well as the evolution of the minimum salary (was also adjusted by the deflator base 2010). As expected for an earnings distribution the inter-quartile difference grows bigger for each successive quartile. What is noteworthy here is that while the financial crisis and subsequent euro area crisis resulted in losses for every quartile, the 3rd quartile had the steepest drop and the slower recovery. This suggests the crisis might have provoked a decline in inequality by relatively affecting more the upper half of the earnings distribution than the lower half as considered before.

Figure 2: Earnings by Quartile



Source: author's own computations based on *Quadros de Pessoal* database.

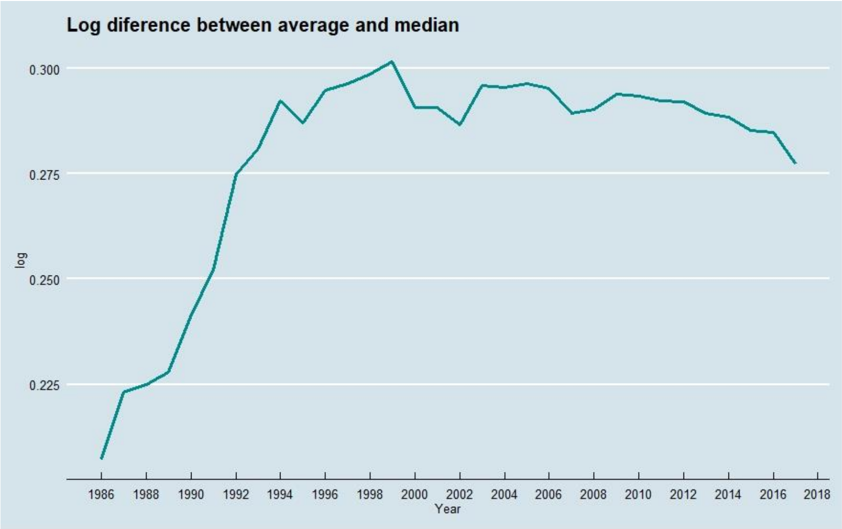
Figure 3 has plotted the difference between the logarithm of the average, $\ln \bar{y}$, and the logarithm of the median, $\tilde{y}$, for each year, which can be written as:

$$d = \ln\frac{\bar{y}}{\tilde{y}}$$

This can be interpreted as an approximation of the relative difference between the average and the median and it can be considered a rudimentary inequality measure.

The fact that this value is significantly positive strongly suggests that the distribution is skewed to the right, as is expected in an earnings distribution. A more interesting takeaway is that when the information of Figures 2 and 3 is put together it becomes clear that up to 1999 the bottom half of earners didn´t reap as many benefits from the economic growth as the upper half.

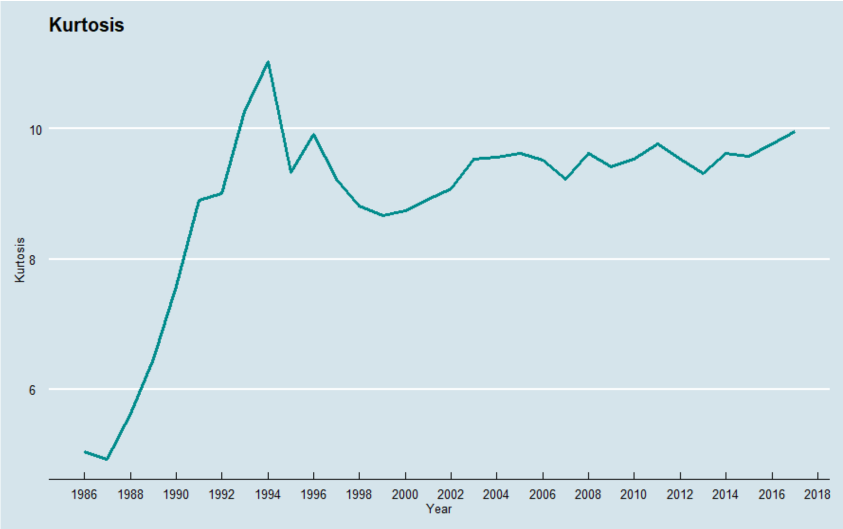
Figure 3: Log difference between average and median



Source: author’s own computations based on *Quadros de Pessoal* database.

The kurtosis is another statistical measure that can be applied in this context and the fact it measures the “tailedness” of the distribution makes it very sensible to changes in the top brackets of earnings. In the case of the Portuguese earnigns inequality, it is not as sensitive to changes in the bottom of the earnings distribution as it might be expected since the lower values are much closer to the average than the higher values. Each individual measure of kurtosis does not tell much about inequality but its evolution over the years does: Figure 4 shows that up to 1995 the kurtosis is increasing, mimicking the Gini index, but then, unlike the Gini, it briefly decreases and then maintains its value, albeit some fluctuations remain visible for the most recent years under analysis. What this suggests is that while the increase in inequality during the period 1986-1994 was driven by increases in the earnings of top earners, the decrease in inequality post 2008 is not driven by an opposite movement in top earners wages. Therefore, it should be a result of a loss of earnings share by the middle and/or upper middle class.

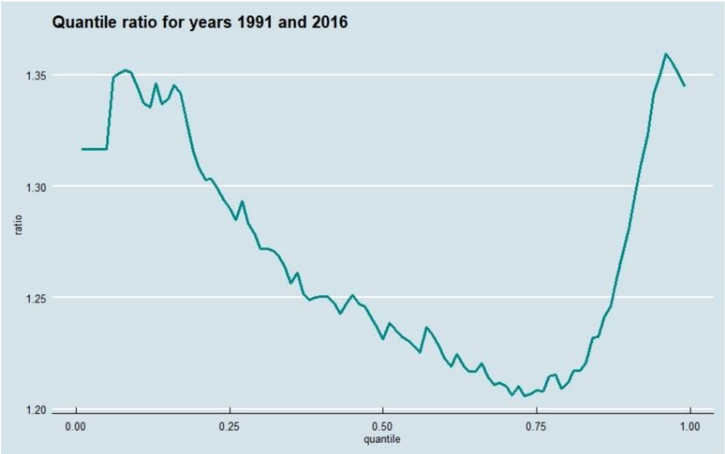
Figure 4: Kurtosis of the earnings distribution



Source: author's own computations based on *Quadros de Pessoal* database.

To corroborate this idea, Figure 5 shows the ratio of all the percentiles, from 1 to 99, for the years 2016 and 1991. These two years were chosen because the Gini index is very similar for both, 0.2964013 and 0.2958507 respectively, but the behavior of inequality should not be the same based on the previous statistics. As such, it should be a result of a loss of earnings share by the middle and/or upper middle class. This graph clearly shows the phenomenon of wage polarization. As expected, wages at all percentiles increased from 1991 to 2016, at the top and bottom of the distribution were the bigger gains, close to 35%, while between the 40th and the 88th percentile occurred the smaller gains, with no more than a 25% increase in salary between the two years.

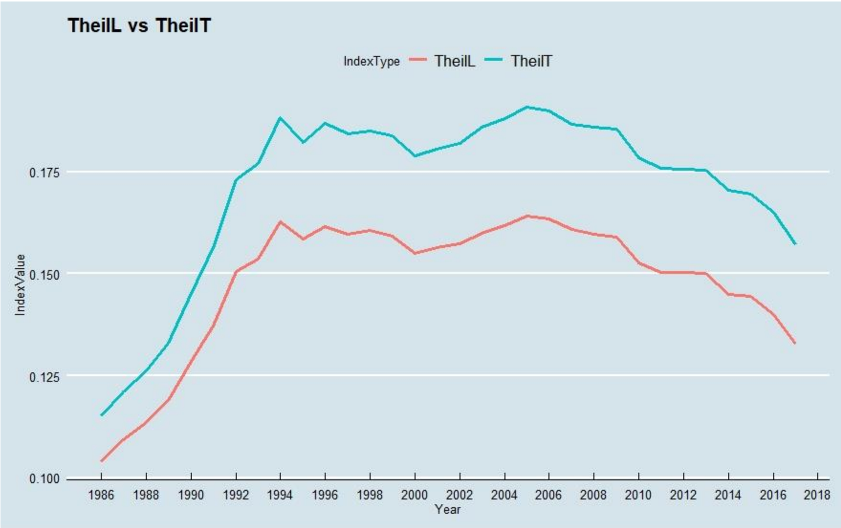
Figure 5: Quantile ratio for year 2016 relative to 1991



Source: author's own computations based on *Quadros de Pessoal* database.

Finally, Figure 6 has Theil's L and Theil's T plotted against each other. Theil's L is more sensitive to changes at the bottom of the distribution and less sensitive to changes at the top compared to Theil's T. Bearing this in mind, the faster growth of Theil´s T in the period of rising inequality indicates that this increase in inequality can most likely be attributed to increases in earnings at the top of the distribution, supporting the conclusions drawn from the previous statistics. This kind of evolution is confirmed today in developed countries, and several causes have been advanced to explain it, (Goos, Manning & Salomons, 2009), Manning (2019) and (Bárány & Siegel, 2019).

Figure 6: A comparison between Theil indexes



Source: author's own computations based on *Quadros de Pessoal* database.

Having analyzed some of the crucial dynamics of Portuguese earnings inequality over the last three decades, the issue at stake onwards becomes whether those dynamics have been favorable to economic growth or a roadblock in the path of economic development. The next section aims to answer this question.

5. Econometric strategy

The economic literature highlights the interdependency between earnings inequality and growth, therefore, a VAR model is considered for our country study of the Portuguese economy, in order to deal with the potential endogeneity of the variables. The final objective of our VAR analysis will be the interpretation of impulse-response functions, which can be used to draw conclusions regarding the effect of inequality on growth as well as the economic mechanisms responsible for this effect. The impulse-response function is the output of a dynamic system, over time, in response to an external change. In the framework of this work, the dynamic system is the VAR(p) model and the external change is an exogenous shock to one of the variables.

The first goal of the empirical strategy is, consequently, to find a suitable specification for the VAR(p) model what requires the accomplishment of several methodological steps. The first one is to choose appropriate variables that can capture information that clarifies both the direction and the mechanisms acting upon the relationship between inequality and growth. The second step is to specify the model. To do that it is necessary to analyse causality and to order the variables according to their degree of exogeneity. The ordering of the variables is relevant because the Cholesky decomposition is used to produce the impulse-response functions. The number of lags is then determined based on the information criteria. The third step is to run diagnostic tests that check the validity of the model, namely the Portmanteau test for residual autocorrelation, the multivariate Jarque-Bera test for normality and the OLS-CUSUM for stability.

Once all the three methodological steps are performed and the model is confirmed to be robust, the variance decomposition and the impulse-response functions are computed and their results presented and analysed in section 6. *Results and discussion* of this work.

All the computations needed were done, as in the previous section, with R software, in particular the functions of the packages “vars” (Pfaff, Bernhard 2008), “aTSA”(Qiu, Debin, 2015) and “generalCorr” (Vinod,2019).

In the following subsections, each aspect of the empirical strategy will be presented in detail.

5.1 The VAR model

The basic form of a VAR(p) model is the following

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

Where y_t is a set of K time-series variables, A_i is a (K × K) coefficient matrix, u_{it} is an unobservable error term and c is the constant.

In our model, $y_t = (H_t, INEQ_t, K_t, Y_t)$, where H is the log of the percentage of workers that have attained at least secondary education, INEQ is the log of GE(2), K is the log of gross fixed capital formation (base = 2016, measured in thousands of million euros) and Y is the log of real GDP(base = 2016, measured in thousands of million euros). H and INEQ were computed from the database *Quadros de Pessoal*, while Y and K were taken from PORDATA.[1]

The purpose of this model is to understand how inequality relates to the GDP, thus Y and INEQ are the two variables that the model must include. The inclusion of a human capital proxy as well as a physical capital proxy serves two purposes. The first is explaining Y since physical capital and human capital are key factors of production and should be considered to have different elasticities in growth models as done by Uzawa (1965) and Lucas (1988). The second purpose is to shed some light into how inequality affects growth, by giving information about which transmission mechanisms are most likely at work.

The inequality measure chosen was the Generalized Entropy Index with $\alpha = 2$, which is equivalent to half the squared coefficient of variation. This measure penalizes relatively more inequality at the top of the distribution. This makes it ideal to test mechanisms related to investment and the mechanism related with the incentive to work harder in order to get higher returns. In the context of the Portuguese economy, it also makes sense to use a measure like the one above mentioned to test the "human capital" mechanism since, for the earlier part of the period under consideration (1986-2017), the workforce had a low percentage of workers with a 12th grade education (only 9% in 1986, up to only 18% in 1997). This suggests that inequality at the bottom should not be as relevant as inequality at the top in order to explain changes in H in the first half of the sample. The fact that the GE(2), although putting relatively more emphasis on the top of the distribution, does not

[1] PORDATA is a certified and official database that collects its data from several different statistical sources. The original source for Y and K is the INE, the Portuguese Nacional Statistical Institute.

ignore the rest, still makes it competent in assessing the "human capital" mechanism, even if for the second half of the sample inequality at the bottom plays a bigger role in determining the human capital level in the workforce.

5.2 VAR model Specification

The PP and KPSS unit root tests were inconclusive regarding the presence of unit roots in data, both not rejecting the null hypothesis of non-stationarity and stationarity respectively, but, from theory, it is expected that at least GDP (Y) and Gross Fixed Capital Formation (K) should have a unit root. Under this scenario, the usual procedure is to differentiate the variables, but, by differentiating them, some of the dynamics of the variables could be lost and so, the variables were kept in levels.

Considering that at least two of the four variables should be I(1) the tests for Granger causality are not applicable in this instance and, therefore, an alternative approach is needed in order to determine causality and a hierarchy of exogeneity among the variables.

A methodology proposed by Vinod (2017), based on the concept of kernel causality, was used.

GMC(Y|X) is the R^2 of the Nadaraya-Watson nonparametric Kernel regression:

$$Y = g(X) + \epsilon$$

Where g(X) is a non-parametric, unspecified function.

Vinod's methodology consists in defining δ as the difference between the generalized measure of correlation (GMC) with one variable as the dependent variable and the GMC with the other variable as dependent: $\delta = GMC(X|Y) - GMC(Y|X)$, and then checking the sign of δ: if δ > 0 then Y is the kernel cause and if δ< 0 then X is the kernel cause.

Using the R package "generalCorr", a matrix of the GMC's is computed with the causes as column names.

Table 1: Matrix of the generalized measure of correlations

	H	**INEQ**	**K**	**Y**
H	1	0.975	0.930	0.950
INEQ	0.996	1	0.761	0.967
K	0.996	0.960	1	0.458
Y	0.999	0.993	0.934	1

Source: author's own computations using the R package "generalCorr".

Notes: H is the human capital proxy, INEQ is the inequality index GE(2), K is the investment proxy, Y is the real GDP. Each value represents a generalized correlation coefficient (GMC). The variable in each column is the cause. $\delta = GMC(X|Y) - GMC(Y|X)$ where Y and X are the causes, respectively, if $\delta > 0$ then Y is the kernel cause and if $\delta < 0$ then X is the kernel cause.

From these results, it can be concluded that H is a kernel cause of INEQ, K and Y, INEQ is a kernel cause of K and Y, and K is a kernel cause of Y. A clear hierarchy of exogeneity is established: H → INEQ → K → Y, from the most exogenous to the least.

Moreover, the R^2 is high for almost every Nadaraya-Watson nonparametric Kernel regression, meaning it is very likely that each of these variables has a causal effect on the others, the potential exception being that Y "does not cause" K, since the GMC value of 0.458 is not as conclusive as the others.

In order to select an appropriate number of lags the Schwarz information criterion (SIC) was used, since it tends to suggest a lower number of lags than the other criterion and it is preferable to have a lower number of lags in the model because there are only 32 observations for each variable. The number of lags selected was one.

Table 2: Lag selection

	1	**2**	**3**	**4**
AIC	-2.828473e+01	-2.857146e+01	-2.919192e+01	-2.937231e+01
HQ	-2.799383e+01	-2.804783e+01	-2.843557e+01	-2.838323e+01
SC	**-2.733316e+01**	-2.685862e+01	-2.671783e+01	-2.613695e+01
FPE	5.282259e-13	4.291504e-13	2.856178e-13	3.816088e-13

Source: author's own computations.

Notes: The criteria are listed in the first column. AIC is the Akaike information criterion, HQ is the Hannan-Quinn criterion, SC is Schwarz criterion and FPE is the final prediction error criterion. The lowest value of the criterion indicates the ideal number of lags. SC was the selected criterion.

Having selected the lags, it is now possible to specify the VAR model as:

$$y_t = c + Ay_{t-1} + u_t$$

with $y_t = (H_t, INEQ_t, K_t, Y_t)$.

5.3 Diagnostic tests

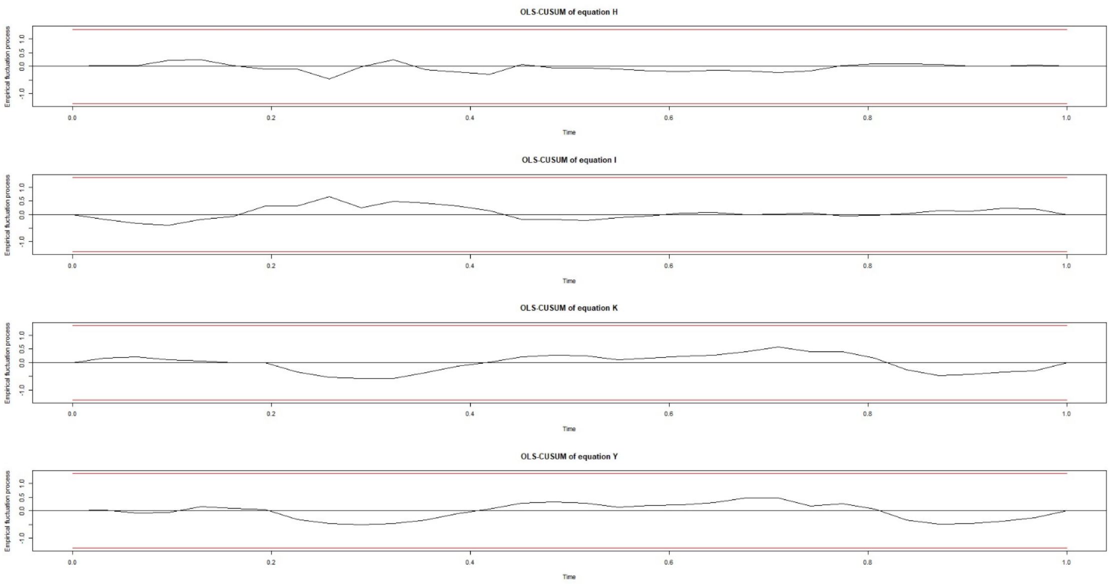
The model was tested for residual autocorrelation, normality and stability.

The Portmanteau test with 10 lags reports a p-value of 0.703, so the null hypothesis of no autocorrelation is not rejected.

The multivariate Jarque-Bera test reports a p-value of 0.1979, so the null hypothesis of normality is not rejected. The "skewness only" part of the test reports a p-value of 0.174 and the "kurtosis only" part of the test reports a p-value of 0.3185.

To test the model's structural stability, Figure 1 shows an OLS-CUSUM test for structural change based on the work of Ploberger and Kramer (1992). The model appears to be structurally stable in terms of parameters evolution.

Figure 7: OLS-CUSUM test



Source: author's own computations.

Moreover, all the characteristic roots have modulus less than 1, which indicates stationarity.

6. VAR results and discussion

6.1 Variance decomposition

With the aim of confirming the forecasting ability of the chosen variables, the variance decomposition 10 years after a shock were computed (Table 2). Each value indicates the amount of the forecast error variance of a variable that can be explained by exogenous shocks to itself or to any of the other variables. The columns represent the variable that receives the exogenous shock and the rows the variable that is affected, consequently each row sums 1 (100%).

Table 3: Variance decomposition (%)

	H	**INEQ**	**K**	**Y**
H	0.4479905	0.0834984	0.0417870	0.4267242
INEQ	0.4938232	0.4037304	0.0821255	0.0203209
K	0.0144885	0.2056899	0.7347249	0.0450966
Y	0.0201200	0.1655092	0.6751175	0.1392533

Source: author's own computations.

Notes: The column names are the variables that receive the shock, row names are the affected variables. The values should be read as percentages, each representing how much of each variable's forecast error variance is explained by shocks to another variable or itself. Each row adds up to 1.

These results show that all the selected variables have an interesting role in the model.

The main takeaways are that human capital is not particularly affected by gross fixed capital formation, being mainly driven by its own inertia and by output. Inequality has a relevant influence on human capital, albeit smaller than that of the output. Inequality is fundamentally dependent on human capital and on its own inertia. It is more affected by the investment than the output, still, these two variables only explain 10% of its variance. An important result is that over 20% of the fixed investment's variance is explained by inequality, reinforcing the need to include these two variables in a parsimonious model like this one. Finally, output is explained mostly by the investment with the remaining 32.5% explained mostly by inequality and the output´s inertia. Human capital accounts for only 2% of output variance.

6.2 Impulse-response analysis

The orthogonal impulse response function indicates how the value of a variable is affected by a one standard deviation shock in another variable or in itself.

Figures 8 to 12 show for 10-step-ahead and one 20-step-ahead impulse response functions selected for their importance in the framework of this work. (See figure 13 in the appendix for all the impulse response functions) Figure 12 is 20-step-ahead rather than 10 in order to capture a change in the sign of its effect that is relevant for the analysis. The Confidence Interval is a 95% bootstrap CI, based on 100 runs.

Figure 8 and Figure 9 show that exogenous shocks on H and K have a positive impact on output (Y), as expected from economic theory, for instance, (Lucas, 1988),(Mankiw, Romer and Weil, 1992) and Jones (1995). A 1% increase in the percentage of workers with at least secondary education leads to a 0.046% increase in output after 10 years. A 1% increase in the formation of gross fixed capital leads to a 0.107% increase in output after 10 years. This is a crucial starting point because most of the transmission mechanisms considered in this work rely on the assumption that both human capital and investment positively influence growth.

Figure 10 shows that an increase in inequality has a negative effect on output, a 1% increase in the inequality index leads to a 0.155% decrease in output after 10 years. This conclusion falls in line with previous literature, for example Andrade, Duarte and Simões (2014). It is now relevant to attempt to understand which mechanisms could be responsible for this.

Figure 11 shows that a shock to inequality has a growing negative impact in human capital. It is logical that this effect grows over the ten year period depicted, up to -0.49% per 1% increase in inequality after 10 years, because there is a time gap between a person's decisions that affect its level of education and that same person entering the labor market. More equality means a higher human capital. The output depicted on Figure 11, therefore, suggests that the "human capital" mechanism is active in the Portuguese economy, meaning that, by decreasing inequality, more people would be able to attain higher levels of education, because parts of the society that were once unable to get secondary education due to budget constraints, might be able to overcome those constraints after a decrease in inequality and, through that, an increase the GDP(Y).

Figure 8: H →Y

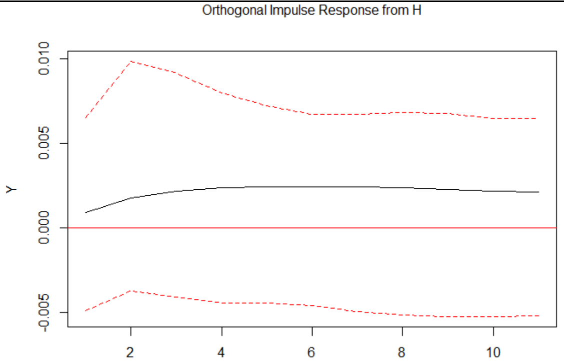


Figure 9: K →Y

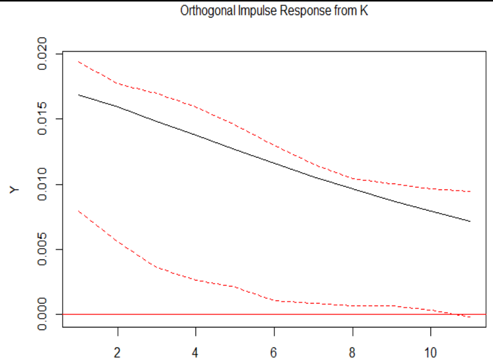


Figure 10: INEQ →Y

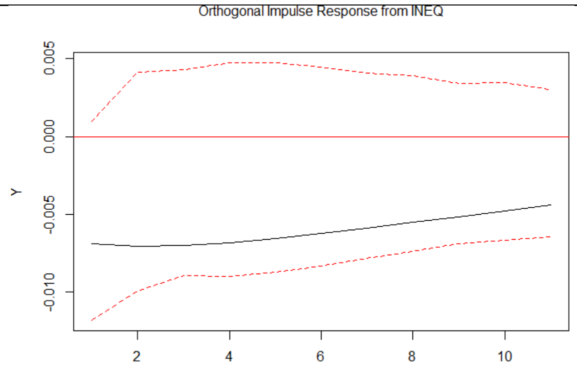


Figure 11: INEQ →H

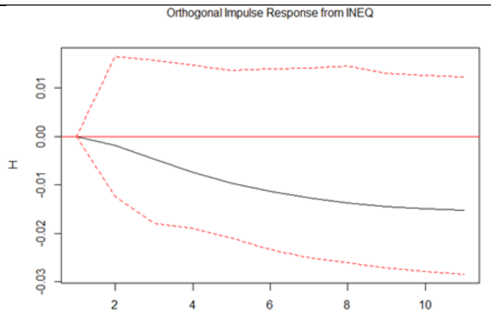
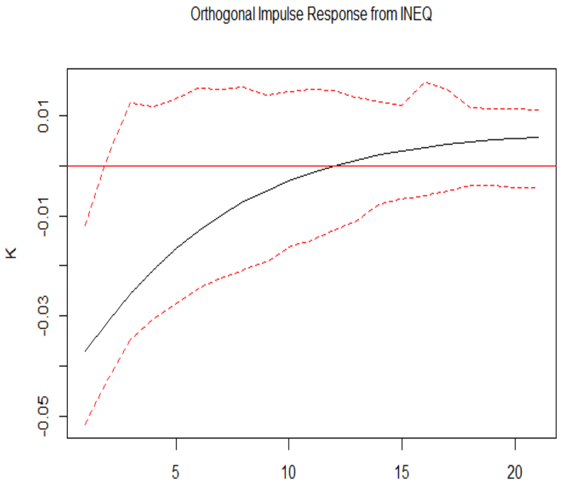


Figure 12: INEQ →Y



Source: author's own computations

Figure 12 is not as straightforward as the other figures since the impact of an increase in inequality on investment starts out as negative but ends up positive after 12 years. An explanation for this could be that two mechanisms of different signs are at work here. On the short run, an increase in inequality causes demand to decrease and because of that, there is less adoption of new technologies by the firms, causing a decrease in investment. On the other hand, with more inequality, aggregate savings are higher; the accumulation of an increased amount of savings leads to lower interest rates, making investment cheaper. After a certain time-delay the “savings” mechanism becomes stronger than the other and inverts the sign of the relationship between inequality and investment. The final effect after 20 years, is that a 1% increase in inequality originates a 0.17 % increase in investment.

A final note on the results: the explanation of the results was based on the expected value of the shocks, however, in most cases (Fig 8, 10, 11,12), the confidence intervals do not exclude the impact being zero.

In order to make the results more robust, a sensitivity analysis was implemented using Theil’s L, Theil’s T and the Gini coefficient as alternative inequality measures. The results in terms of impulse-response functions were very similar and the models passed the diagnostic tests. The impulse-response graphs and the test statistics of these alternative specifications are in the appendix. Because the conclusions remain similar, there is nothing useful to add.

7. Conclusion

This work was based on an analysis of the distribution of earnings for Portugal, in the period 1986-2017, and on an empirical study of the relationship between inequality and growth using a VAR model with four variables (GDP, Inequality, Investment and Human Capital) for the same period.

From the analysis of the earnings distribution the most relevant takeaway are that there are 3 distinct periods in the evolution of earnings inequality, rapid increase, stabilization and decrease, each with their own characteristics, the crucial one being that the early increase in inequality was due, for the most part, to a growth of the larger salaries, but the later decrease was not mainly due to a decrease of those same salaries. The other takeaway is that, in terms of total salary, the middle and upper-middle class workers did not gain as much in terms of salary as the rest of the workers.

From the VAR model, the main conclusions are that inequality negatively impacts growth, both directly and indirectly through Human Capital and Investment, but on the long run the path that runs from Inequality to Investment to Output changes from negative to positive. The most likely mechanisms present in the Portuguese economy are the "human capital" mechanism, the "savings" mechanism and the "domestic demand" mechanism. The choice of the inequality index does not have a noticeable effect on the impulse response general configuration.

Future research, as a follow up to this work, will include studying the cointegration relationships between the variables to potentially estimate a VECM model and simulate the shocks, as well as, estimating a SVAR model with restrictions that allow for a more precise study of the mechanisms.

8. References

Adrián Risso, W., & Sánchez Carrera, E. J. (2012). Inequality and economic growth in China. *Journal of Chinese Economic and Foreign Trade Studies*, 5(2), 80-90. https://doi.org/10.1108/17544401211233453

Alesina, A., & Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal of Economics*, 109(2), 465-490. https://doi.org/10.2307/2118470

Andrade, J. A. S., Duarte, A. P. S., & Simões, M. C. N. (2014). Inequality and growth in Portugal: A time series analysis. *Revista Portuguesa de Estudos Regionais*, 37(1), 29–42.

Assane, D., & Grammy, A. (2003). Institutional framework and economic development: international evidence. *Applied Economics*, 35(17), 1811-1817. https://doi.org/10.1080/0003684032000152862

Atems, B., & Jones, J. (2015). Income inequality and economic growth: a panel VAR approach. *Empirical Economics*, 48(4), 1541-1561. https://doi.org/10.1007/s00181-014-0841-7

Bárány, Z. L., & Siegel, C. (2019). Job polarization, structural transformation and biased technological change. *Travail et Emploi*, 157, 25-44. DOI : 10.4000/travailemploi.8996

Pfaff Bernhard (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. Journal of Statistical Software 27(4). URL http://www.jstatsoft.org/v27/i04/.

Chan, K. S., Zhou, X., & Pan, Z. (2014). The growth and inequality nexus: The case of China. *International Review of Economics & Finance*, 34, 230-236. https://doi.org/10.1016/j.iref.2014.08.004

Cingano, F. (2014), Trends in Income Inequality and its Impact on Economic Growth, *OECD Social, Employment and Migration Working Papers*, No. 163, OECD Publishing, Paris. https://doi.org/10.1787/5jxrjncwxv6j-en.

Cowell, F. A., & Victoria-Feser, M. P. (1996). Robustness properties of inequality measures. *Econometrica: Journal of the Econometric Society*, 64(1), 77-101. DOI: 10.2307/2171925

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119), 348-361. DOI: 10.2307/2223525

De Dominicis, L., Florax, R. J., & De Groot, H. L. (2008). A meta-analysis on the relationship between income inequality and economic growth. *Scottish Journal of Political Economy*, 55(5), 654-682. https://doi.org/10.1111/j.1467-9485.2008.00470.x

Durlauf, S. N., Johnson, P.A & Temple, J. R. W. (2005). Growth Econometrics. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1, Part A (pp. 555-677. Boston, MA: Elsevier- North-Holland. https://doi.org/10.1016/S1574-0684(05)01008-7

Rodrigues, C. F., Figueiras, R., & Junqueira, V. (2016). *Desigualdade de Rendimento e Pobreza em Portugal: As consequências sociais do programa de ajustamento.* Lisbon: Fundação Francisco Manuel dos Santos.

Frank, M. W. (2009). Income inequality, human capital, and income growth: Evidence from a state-level VAR analysis. *Atlantic Economic Journal*, 37(2), 173-185. https://doi.org/10.1007/s11293-009-9172-z

Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *The Review of Economic Studies*, 60(1), 35-52. https://doi.org/10.2307/2297811

Gobbin, N., & Rayp, G. (2008). Different ways of looking at old issues: a time-series approach to inequality and growth. *Applied Economics*, 40(7), 885-895. https://doi.org/10.1080/00036840600771106

Goos, M., Manning, A., & Salomons, A. (2009). Job polarization in Europe. *American Economic Review*, 99(2), 58-63. DOI: 10.1257/aer.99.2.58

Jones, C. I. (1995). R & D-based models of economic growth. *Journal of political Economy*, *103*(4), 759-784. DOI:10.1086/262002

Langel, M. (2012). Measuring inequality in finite population sampling (Doctoral dissertation, Université de Neuchâtel).

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22 (1), 3–42. https://doi.org/10.1016/0304-3932(88)90168-7

Kaldor, N. (1955) Alternative Theories of Distribution, *Review of Economic Studies* 23(2), 83–100. DOI: 10.2307/2296292

Manning, A. (2019). Polarization (s) in Labour Markets: Synthesis and Perspectives. *Travail et emploi*, 1, 13-24.

McGregor, T., Smith, B., & Wills, S. (2019). Measuring inequality. *Oxford Review of Economic Policy*, *35*(3), 368-395. https://doi.org/10.1093/oxrep/grz015

Moritz S, Bartz-Beielstein T (2017). “imputeTS: Time Series Missing Value Imputation in R.” _The R Journal_, *9*(1), 207-218. doi: 10.32614/RJ-2017-009 (URL: https://doi.org/10.32614/RJ-2017-009).

Murphy, K. M., Shleifer, A., & Vishny, R. W. (1989). Industrialization and the big push. *Journal of Political Economy*, 97(5), 1003-1026. https://doi.org/10.1086/261641

Neves, P. C., Afonso, Ó., & Silva, S. T. (2016). A meta-analytic reassessment of the effects of inequality on growth. *World Development*, 78(C), 386-400. https://doi.org/10.1016/j.worlddev.2015.10.038

Palma, J. (2011). Homogeneous middles vs. heterogeneous tails, and the end of the ‘inverted-U’: It's all about the share of the rich. *Development and Change* 42(1), 87-153. https://doi.org/10.1111/j.1467-7660.2011.01694.x

Qiu, Debin (2015). aTSA: Alternative Time Series Analysis. R package version 3.1.2. https://CRAN.R-project.org/package=aTSA

Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, *48*(1), 1-48. doi:10.2307/1912017

Uzawa, H. (1965). Optimum Technical Change in an Aggregative Model of Economic Growth. *International Economic Review*, 6(1), 18–31. doi:10.2307/2525621.

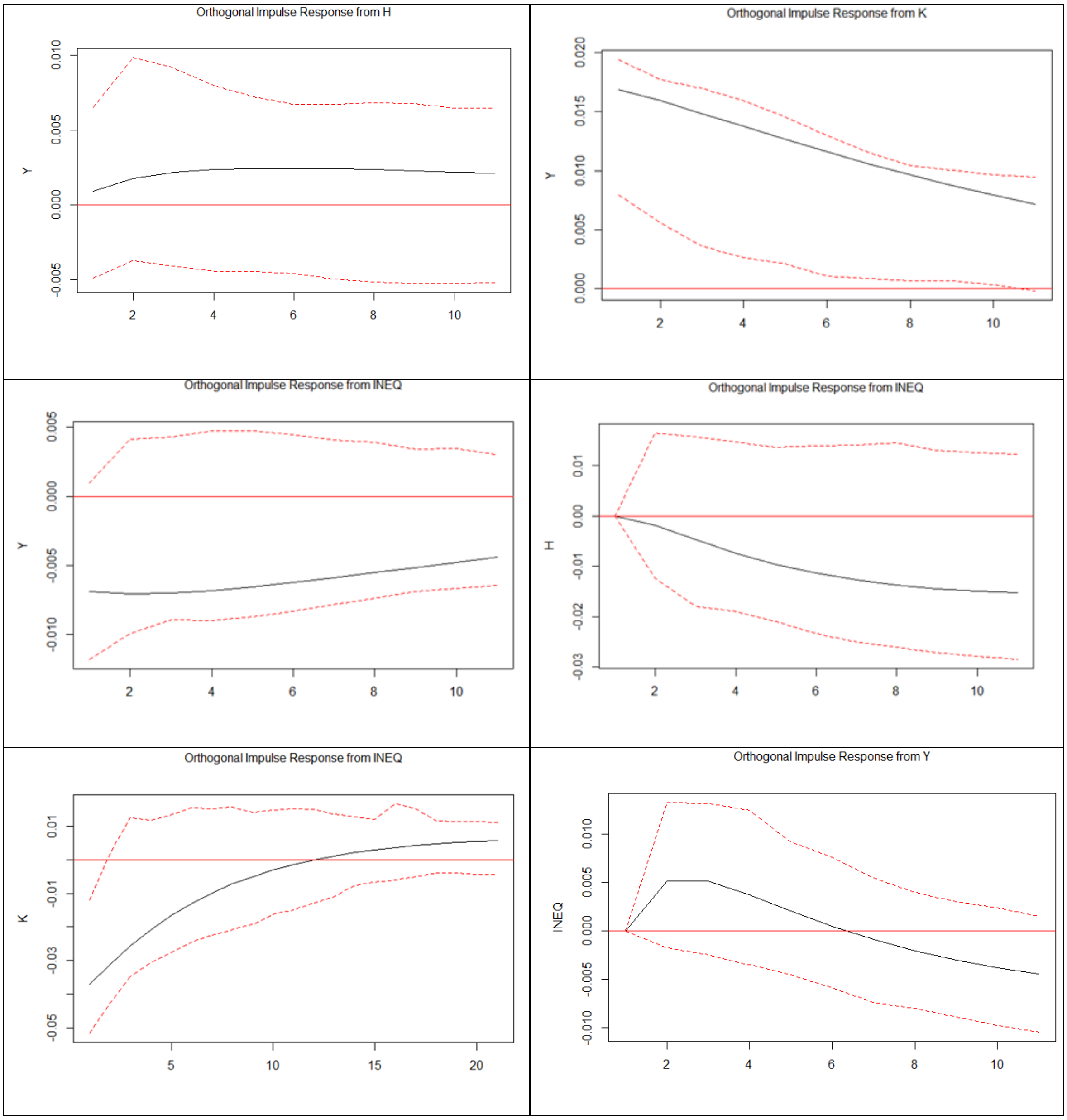
Vinod, H. D. (2017). Generalized correlation and kernel causality with applications in development economics. *Communications in Statistics-Simulation and Computation*, *46*(6), 4513-4534. https://doi.org/10.1080/03610918.2015.1122048

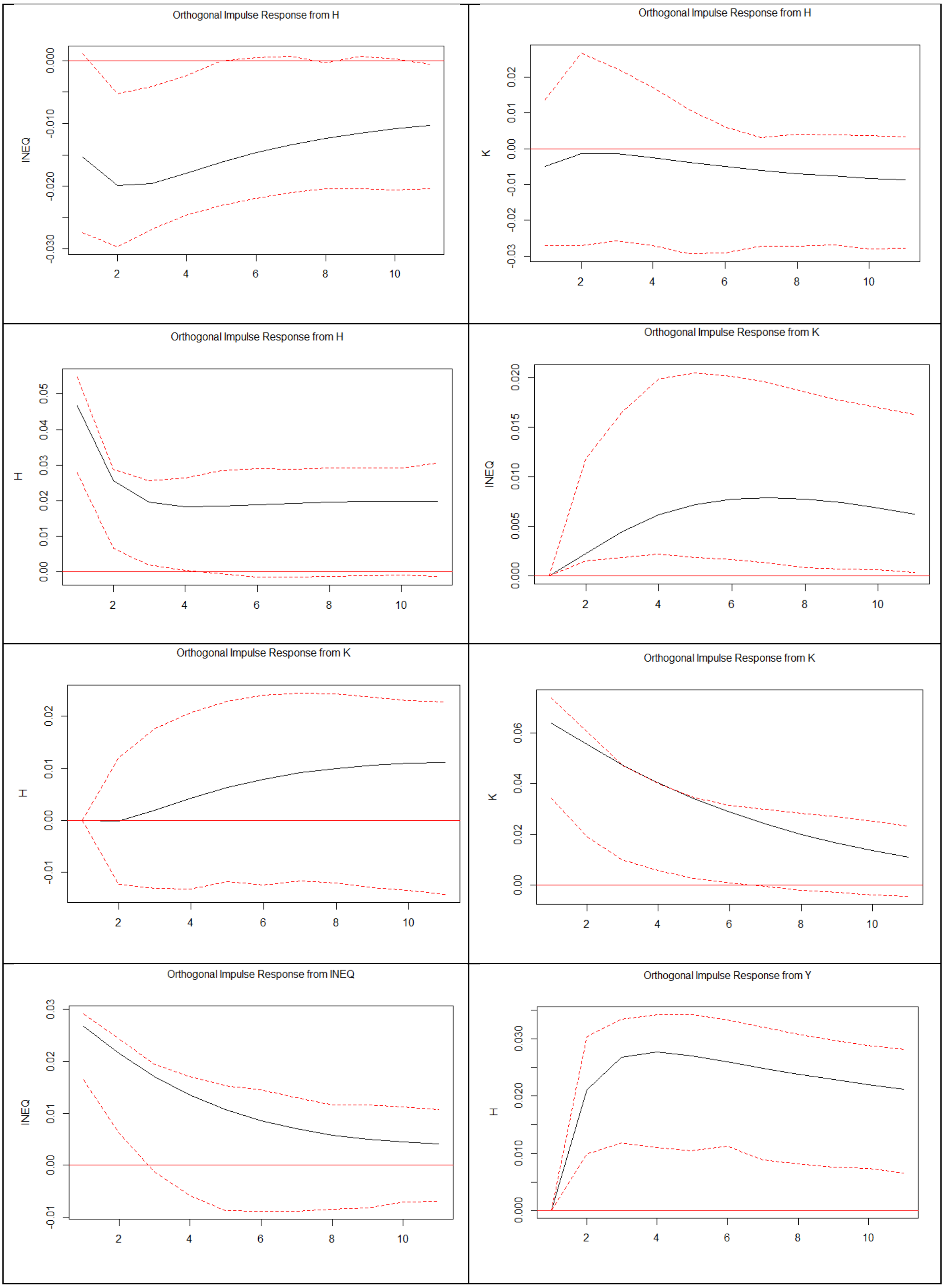
Vinod, H. D. (2019). generalCorr: Generalized Correlations and Plausible Causal Paths. R package version 1.1.5. https://CRAN.R-project.org/package=generalCorr

Zeileis, Achim (2014). ineq: Measuring Inequality, Concentration, and Poverty. R package version 0.2-13. https://CRAN.R-project.org/package=ineq

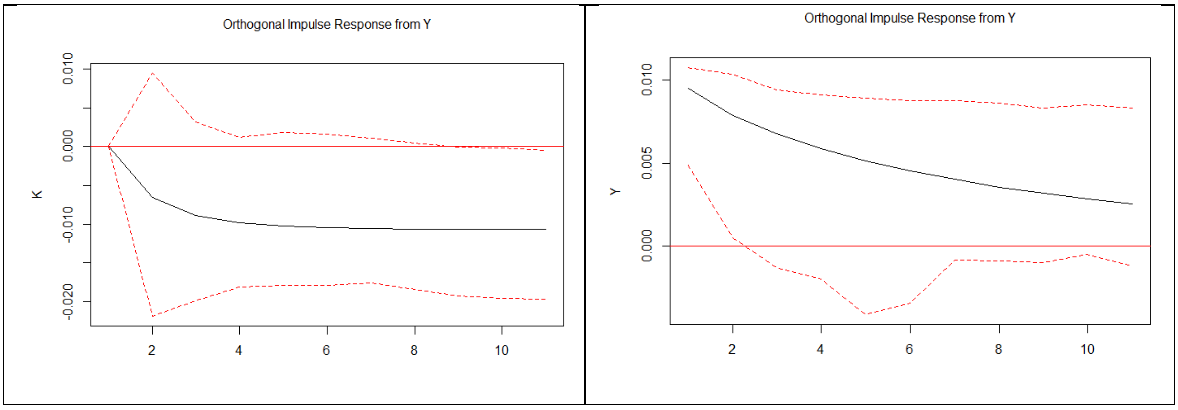
Appendix

Figure 13: Impulse-response functions from the VAR model with GE(2)





Orthogonal Impulse Response from H
INEQ
Orthogonal Impulse Response from H
K
Orthogonal Impulse Response from H
H
Orthogonal Impulse Response from K
INEQ
Orthogonal Impulse Response from K
H
Orthogonal Impulse Response from K
K
Orthogonal Impulse Response from INEQ
INEQ
Orthogonal Impulse Response from Y
H



Source: author's own computations

Table 4: Diagnostic tests for the alternative specifications

	Portmanteau	Multivariate JB	Roots of the characteristic polynomial
INEQ1	p-value = 0.6689	p-value = 0.1468	All characteristic roots have modulus less than 1
INEQ0	p-value = 0.6773	p-value = 0.1354	All characteristic roots have modulus less than 1
INEQG	p-value = 0.7002	p-value = 0.1149	All characteristic roots have modulus less than 1

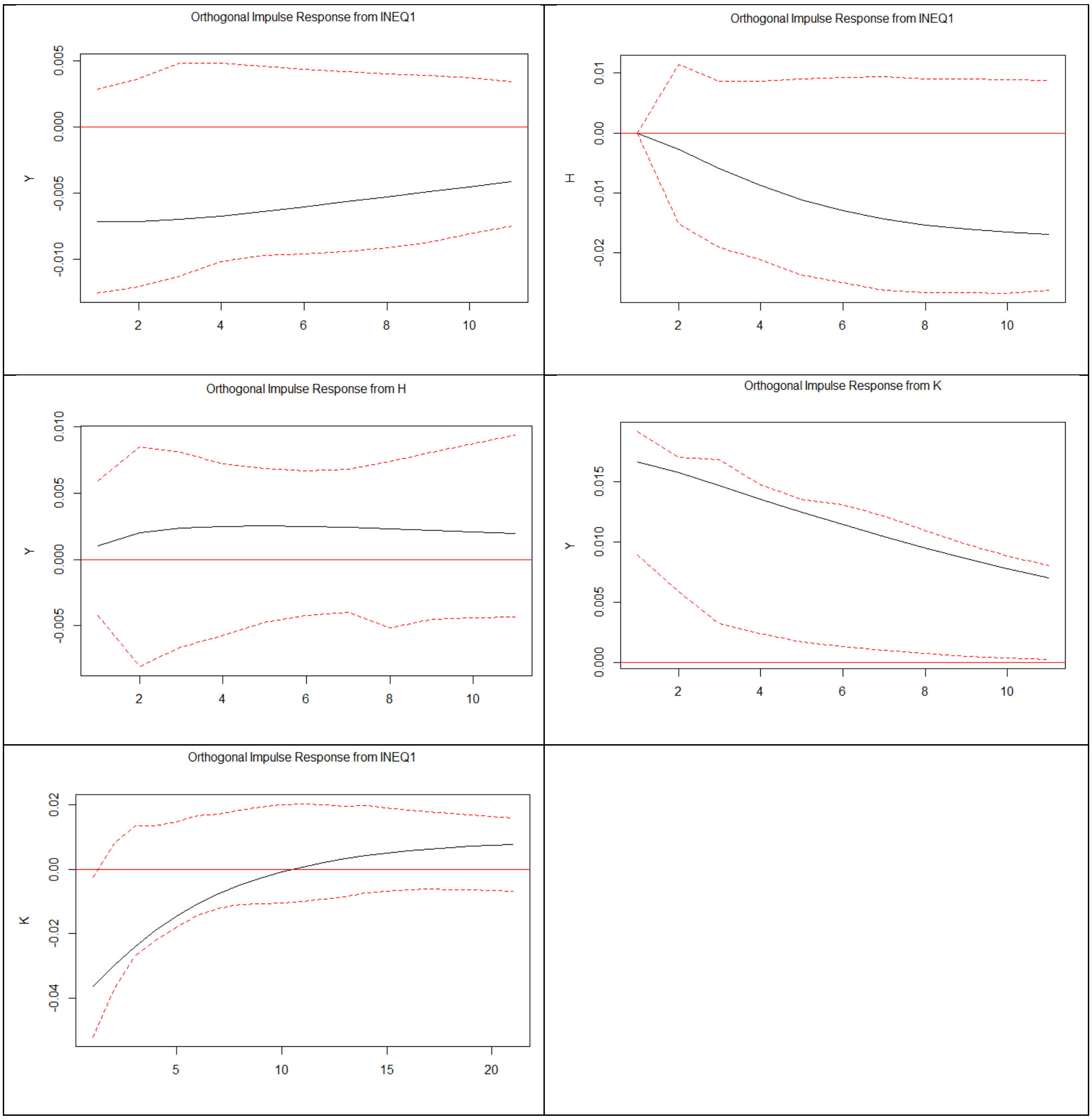
Source: author's own computations

Notes: Null hypothesis of the Portmanteau test: no autocorrelation. Null hypothesis of the Multivariate JB test: normality. P-values over 0.1 do not reject the null hypothesis.

INEQ1 is GE(1), INEQ0 is GE(0) and INEQ is the Gini index. The number of lags used in each of the VARs was 1.

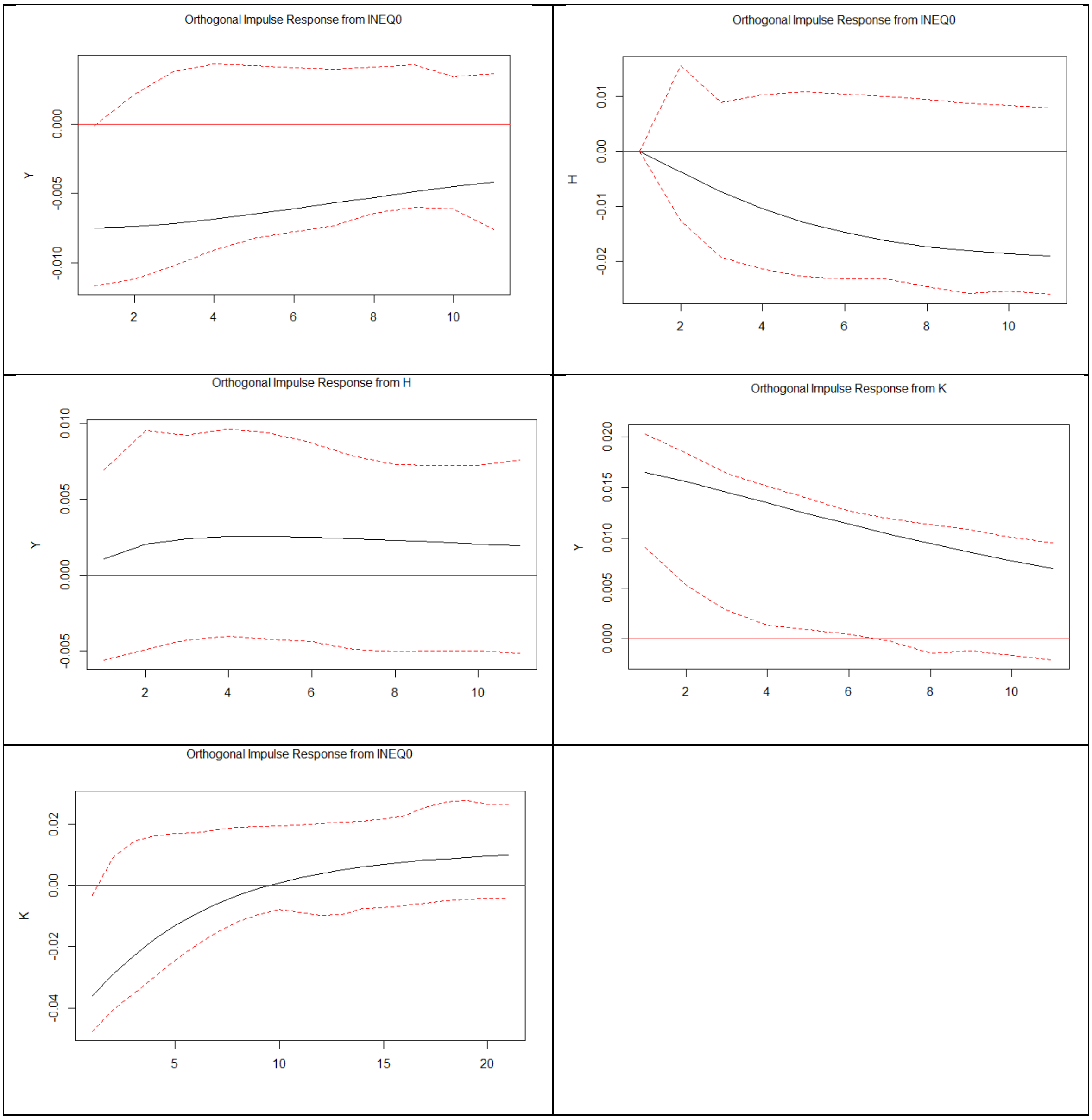
The following Figures contain the same 5 impulse response graphs displayed in the main body of this work, for each alternative specification.

Figure 14: Impulse response functions from the VAR model with GE(1)



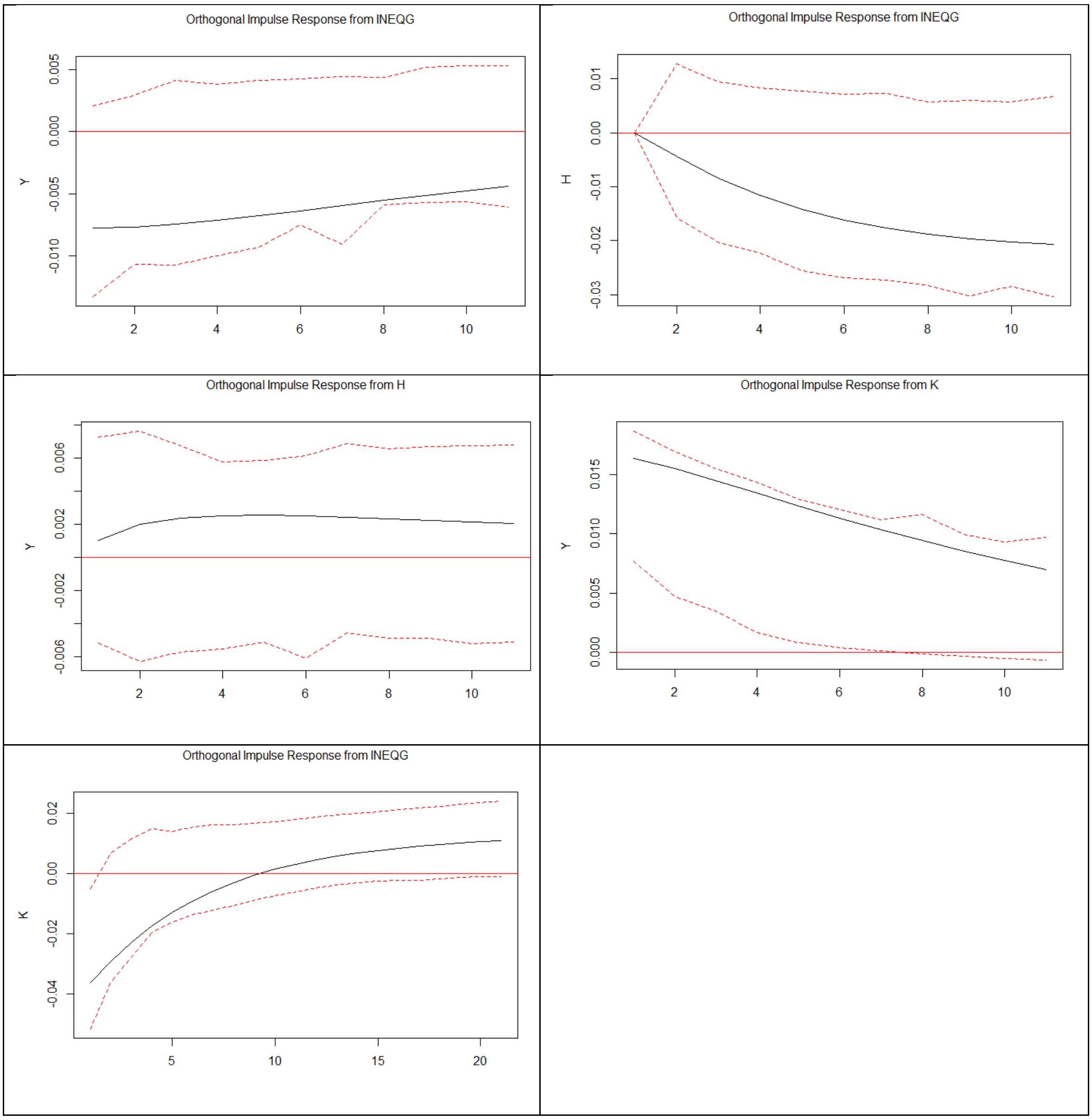
Source: author's own computations

Figure 15: Impulse response functions from the VAR model with GE(0)



Source: author’s own computations

Figure 16: Impulse response functions from the VAR model with the Gini index



Source: author's own computations