Universitat Pompeu Fabra Barcelona

Department of Economics and Business

Economic Working Paper Series Working Paper No. 1741

Real-Time inequality and the welfare state in motion: Evidence from COVID-19 in Spain

Oriol Aspachs, Ruben Durante, Alberto Graziano, Josep Mestres, Jose G. Montalvo and Marta Reynal-Querol

September 2020

Real-Time Inequality and the Welfare State in Motion: Evidence from COVID-19 in Spain *

Oriol Aspachs† Ruben Durante‡ Alberto Graziano†

Josep Mestres† Jose G. Montalvo§ Marta Reynal-Querol‡

Abstract

Most official economic statistics have a relatively low frequency. The measures of inequality, in particular, are not only produced with low frequency but also with significant lags. This poses an important challenge for policymakers in their objective to mitigate the effects of a rapidly moving epidemic as the COVID-19. We propose a methodology for tracking the evolution of income inequality in the aftermath of the COVID-19 pandemic using high-frequency, high-quality microdata from bank-records. Using this approach we study the evolution of inequality since the beginning of the COVID-19 pandemic, and its effect on different groups of the population. First, we show that the payroll data managed by banks are an extremely useful source of information to detect, timely and accurately, changes in the distribution of wages. Our data replicate very closely the distribution of wages from the official wage surveys. Second, we show that, in absence of public benefits schemes, inequality would have increased dramatically. The impact of the crisis on inequality is explained mostly by its effect on low-wage workers. Pre-benefits wage inequality has increased significantly among foreign-born individuals, and regions that have a heavy economic dependence on touristic activities. Finally, we show that the public benefits activated soon after the beginning of the pandemic have substantially mitigated the impact of the COVID-19 crisis on inequality.

Keywords: inequality, COVID-19, administrative data, high frequency

JEL Classification: C81, D63, E24, J31

---

*We want to thank Miguel Angel Barcia for his helpful suggestions. Daniele Alimonti provided excellent research assistance.

†Caixabank Research

‡ICREA-UPF, IPEG and BarcelonaGSE

§UPF, IPEG and BarcelonaGSE

1 Introduction

The spread of COVID-19 has taken a huge toll on economic activity around the globe. The IMF expects that global GDP will pencil in the highest contraction on record in 2020, and that it will be way deeper than the one experienced during the global financial crisis (0% in 2009 vs –5% in 2020). In the Euro area, the GDP declined by 3.8% in the first quarter of 2020 and is expected to shrink by 8.3% in 2020 according to the central scenario that the ECB presented in June this year (-12.6% in the severe scenario). The Bank of Spain predicts that the Spanish GDP will shrink even further, between -11.6% and -15.1% depending on the scenario. In fact, given the importance for the Spanish economy of the sectors most affected by the lock-down, and the stringency and the persistence of the restrictions to mobility implemented, it is highly likely that Spain will struggle more than other countries to recover from the COVID-19 recession. The Bank of Spain estimates that the GVA of Trade, transport, accommodation and food declined by -71% in Spain during the lock-down (-44% in the Euro Area), and the weight of these sectors in Spain is 23.8% (19.0% in the Euro area). The expected lackluster recovery of the Spanish economy is reflected in the medium term forecasts of most institutions. The central scenario of the Bank of Spain assumes that the Spanish economy will remain well below pre-crisis levels in 2022 (-1.6%), while the ECB expects the recovery of Euro area to be more advanced at this stage, with GDP reaching pre-crisis levels at the end of 2022.

The evolution of the labor market also reflects the impact of the lock-down measures. In Spain, between the 15th of March and the 31st of April, close to one million workers lost their job (4.4% of the active population). In addition, 4.6 million people was forced to move under different forms of temporary lay-off schemes. Hence, in just a month and a half, 22% of the active population had either lost their job or was under temporary leave schemes. The concerns on the impact on inequality have mounted. The new economic backlash might produce a surge in income inequality right at a time when the large inequalities caused by the financial crisis were returning to their 2008 levels.

Most governments around the world have tackled the economic consequences of the pandemic using a combination of family income support and credit facilities for firms and self-employed workers. Family income support tries to stabilize the income of the families, reducing the impact of the crisis on consumption and inequality. The recovery of consumption is fundamental for the shape of the economic recovery (V, U, L, W) since consumption represents the largest component of demand. In Spain, household's consumption represents 57% of GDP. In addition, a higher level of inequality could erode the social fabric and, ultimately, it could call into question the democratic system and promote populism, in a situation where the pandemic is already giving some support to autocratic and illiberal political views. This is one of the most significant challenges for democratic societies in the years after the pandemic.

However, there is a high degree of uncertainty on the effect of the policy responses and the appropriateness of the total amount of public support to the economy. Unfortunately, most of the official indicators and macroeconomic statistics on the effect of the measures to help the families and the economy have a very low frequency (quarterly or yearly), and are produced with long delays (especially the measures of inequality). This is an important challenge for policymakers in their efforts to tailor their responses to "flatten the recession curve" (Gourinchas, 2020).

One characteristic effect of pandemics is its impact on inequality.[1] However, inequality measures are obtained with long lags. We propose an innovative methodology to track inequality, using data at high frequency, and the evolution of the impact of public subsidies on inequality. Having a granular (by gender, geography, age, etc.) and frequently updated view of inequality helps tailoring policies to contain the increase of inequality in general, and by subgroups of the population. Our proposal is related with new international initiatives to track in real time (or with high frequency indicators) the evolution of economic activity. Cicala (2020) uses electricity usage from the European grid to proxy the evolution of economic activity since the consumption of electricity is highly correlated with its entrepreneurial usage. Bick & Blandin (2020) resort to the a Real Time Population Survey (RPS), following the structure of the government survey (CPS) to construct high frequency estimates of employment, hours worked and earnings. Chen et al. (2020) shows that these high-frequency measures of energy consumption and hours worked are strongly correlated with the mobility indicators from the Google Community Mobility Report. Chetty et al. (2020) have built an economic tracker to measure economic activity at high-frequency in the US. They use anonymized economic information from private companies to measure consumer spending (card-based transactions from Affinity Solutions); change in small business open (business making transactions on a given day from Womply); time spent at work (GPS data provided by Google); and hours worked at small business (provided by Homebase).

In Spain, since the outbreak of the pandemic, CaixaBank Research has published weekly notes analysing credit cards usage at POS, online transactions and cash withdrawals to estimate the impact of the confinement measures on consumption. Carvalho et al. (2020) also uses credit card transactions reported by the POS of BBVA (1,300 million transactions) to measure the evolution of consumption by categories. This type of information is also routinely exploited by other banks like Abanca (Observatorio Abanca by IESIDE) or the

[1] For instance Wade (2020) or Scheidel (2017).

Banc Sabadell (Pulso).[2]

S. R. Baker et al. (2020) goes a step further and analyze the effect of income support on consumption. Following a recent literature that examines, using data from Personal Finance Websites, the effect of income on consumption at daily frequency.[3]

However, the new literature on high frequency and real-time data has not proposed a methodology to construct an indicator of inequality that provides a high frequency view of its evolution. In this paper we propose a new source of high quality payroll administrative data, and a methodology, to calculate inequality indicators at high frequency.

2 Methodology

As in many other countries, the Spanish government approved several measures to support income and revenues of workers and companies to mitigate the impact of the COVID-19 crisis. One of the most important measures adopted in Spain to support employment and salaries are the ERTEs, a mechanism to provide income to workers temporarily (or partially) layed off from work. ERTEs also aim at keeping the match employer-employee in the aftermath of exogenous and temporary shocks. But it is important to evaluate their effectiveness. In fact, the implementation of ERTEs has been surrounded by complaints about delays on the transfers and errors on the administrative procedures.[4]

Researchers have recently started to analyze the impact of the economic stimulus packages to maintain employment and wages. Very recently Autor et al. (2020) have analyze the Paycheck Protection Program of the US (PPP), a program of guaranteed loans for small business to support employment and wages. Using anonymized and aggregated payroll administrative data from ADP (one of the largest providers of cloud-based human resources management services), and the Small Business Administration, they conclude that PPP increased employment approximately by 3.25%. Our objective is to use high-quality microdata to evaluate the evolution of wage inequality in Spain at high frequency. Without this type of data the impact of COVID-19 on financial vulnerability by groups of population can only be estimated indirectly using sectoral

[2]The Bank of Spain has also used POS to track expenditure during the pandemic (Gonzalez et al. (2020)).

[3]For instance Gelman et al. (2014), S. Baker (2018), Olafsson & Pagel (2018) or Kueng (2018).

[4]The problem of management of the large number of public measures adopted to mitigate the impact of the COVID19 on the economy is not specific of Spain. Public agencies in charge of managing these programs are overwhelmed by the size and complexity of the different relief programs. For instance, there are also reports in the US on the reception by large corporations of large amounts of financial aid directed to small and medium size firms. There is also evidence that the Paycheck Protection Program has been disproportionately directed to areas scarcely affected by the COVID19.

information of workers by characteristics (gender, age, etc.) like the recent work of Alvargonzález et al. (2020).

However, proper inequality indices are produced with important lags. While monetary policy has many high frequency indicators available to check its performance (short term interest rates, multiple indicators for credit and financial markets tensions, etc.), fiscal policy does not have the advantage of being able to rely on high frequency data. This is particularly troublesome to evaluate the impact of fiscal policies to mitigate the effect of the COVID-19 pandemic and, in particular, to fight against the increase of inequality. For instance, the latest indicators of wage inequality in Spain, produced by the National Institute of Statistics, were presented at the end of June of 2020, but refer to the situation in 2018. This is not unusual in this type of indicators. The World Bank reports in July of 2020 the 2018 Gini indices of 10.6% of the countries, and for 25% of the countries the latest Gini index refers to 2017. But even when we only consider advanced economies, the lags in the production of inequality indicators are very significant. Eurostat reports only 40% of the Gini indices for 2019 for countries of the EU. Therefore, inequality measures are produced with a significant lag. However, many programs are designed to have a direct impact on the distribution of income. In fact, public transfers are an important political issue and attention has been given to evaluate their effect on the distribution of income, both before and after policy implementation.

Since the outbreak of the coronavirus pandemic researchers have developed online surveys to track the evolution of the labor market in several countries. Bick & Blandin (2020) construct a Real-Time Population Survey (RPS) that follows closely the Current Population Survey of the BLS of US. Adams-Prassi et al. (2020) presents also a real time labor survey for the UK, US and Germany. This is an interesting approach but relies on online surveys, which required access to Internet, and produce relatively small, and potentially biased, samples. Our approach is different and relies on large administrative data from financial institutions like recent research by Gelman et al. (2014) or S. R. Baker et al. (2020). Cajner et al. (2020) uses information by ADP, a large international provider of payroll processing, benefit management and tax services.

In this paper we show the utility, and validity, of bank's information to produce high frequency indicators that can help to track the evolution of wage inequality at high frequency. For this purpose we use aggregated and anonymized data[5] from the second largest Spanish financial institution (CaixaBank) on payroll and

[5]The treatment of information uses anonymized data to generate groups of individuals who are extracted from the company's Data Warehouse without any field identifying the clients.

benefits, for more than 3 million wage earners[6]. The computation of inequality in Spain cannot rely on administrative data and that is the reason why, until now, it has been based on surveys. The data of the tax authority is compiled yearly, and the Social Security information, although in principle is updated every month, does not have information on salaries but on the contribution base which is capped.

Our methodological proposal aims at tracking the evolution of wage inequality and the impact of public support on its attenuation. The information of banks is well suited for this type of exercise since wages and subsidies are very well identified in their information systems. For this purpose we will analyze the monthly change in wages by level of salary in the reference period. The evolution of these distributions (over time and across wage intervals) will allow following the impact of the crisis on the distribution of salaries over time and across the wage distribution. We also provide a methodology to calculate monthly Gini indices and Lorentz curves, before and after accounting for public benefits, to analyze if the schemes to support workers temporarily out of the labor market (including potentially the future system of minimum income guarantee) is working well enough to keep inequality at a low level.

Obviously, it is interesting to analyze not only the evolution of total inequality but also the effect of the COVID-19 crisis on inequality for different subgroups of the population. The information systems of banks allow calculating inequality for different criteria like gender, age, region, or country of birth[7]. Using this information we calculate the monthly evolution of gender inequality. We also calculate the monthly evolution of intergenerational inequality. Economic research shows that young people in transition to the labor market have a worse labor market experience if they access the labor market in the middle of a recession. Therefore, this is a particularly important group of the population. We also analyze the evolution of wage inequality by regions (CCAA).[8]

3 Data

In this section we present the criteria to construct the sample and discuss its representativeness with respect to the population of Spanish workers.

[6] At the end of 2019 Caixabank had 27,1% of the payroll receipts in the financial sector. In Spain, differently from other countries like the US, the payment of the salary using checks is a very rare event. Almost all the payments of wages and salaries use direct deposits.

[7] We leave the effect of the sector of activity for future research.

[8] In future work we will also distinguish between rural and urban areas.

3.1 Sample construction

We use data on monthly payrolls and unemployment benefits paid in personal bank accounts in CaixaBank. CaixaBank is the second largest financial institution in Spain and has the highest market share in payrolls paid in a bank account in the country (27.1%).

Our base sample comes from anonymized personal bank accounts and includes all payroll income and labor market related benefits received in that account. We restrict our sample to accounts with either only one account holder or with multiple account co-holders but only one employer paying-in wages. This way, we ensure that payrolls or transfers recorded correspond to only one individual and avoid recording multiple payrolls or transfers from multiple account holders.[9]. In addition, we exclude from the sample those individuals who died during our period of study or who did not use the bank account for their usual financial transactions during the period.[10] Finally, to ensure some stability on the sample of individuals studied, we require observing either wages or public benefits during two months (that is, in December 2019 and in January 2020) prior to the period of study (starting in February 2020). No other filters are applied, in particular we do not restrict our sample by income level, neither minimum nor maximum. We will only do so in the following section for comparability reasons with other public datasets to show the representativeness of our sample.

This way, our base sample includes individuals aged 16-64 who received either wages or unemployment benefits in December of 2019 and January of 2020. We follow those individuals in the following months starting in February 2020 (lockdown measures started on March 14th 2020 when the state of alarm was declared in Spain) until May 2020 (the lockdown measures started to be lifted that month) and we observe wage income and unemployment benefits received or the absence of any of those. Salaries in month t are defined as all wage payments received in a particular bank account from the 16th of that month until the 15th of the following month.[11] Public benefits paid by the Social Security in month t cover the same time period, and include those workers in furlough schemes (known as ERTE in Spain).[12]

[9] Around 5% of individual bank accounts had multiple holders receiving two or more payrolls or subsidies and were excluded from the sample.

[10] In particular, we identify each month those clients who are actively using their bank account and perform at least two non-automatic transactions (payments, etc) during the last two months. Those represent around 0.7% of the sample.

[11] In Spain, employees' payments are usually deposited monthly, towards the end of the month, but they can also be paid fortnightly, weekly, etc. and they are very rarely paid in cheques.

[12] ERTE refers to "Expediente de Regulación de Empleo" and is a temporary layoff program in which workers remain affiliated to the Social Security but receive benefits while they are not working, or they are working a lower number of hours than in regular times). In addition, we know for each client her gender, age, region and country of birth.

3.2 Sample representativeness

Since our main source of data is related with holding a bank account it is important to start analyzing the level of financial inclusion in Spain. The data of the Global Findex, the index of financial inclusion of the World Bank, shows that 97.6% of Spanish people over 15 years old holds a banking account.[13] There is no gender difference in holding bank accounts in the Spanish case. Men and women have exactly the same proportion.

Given our data source we restrict our sample to people working or getting some kind of subsidy related with their job market participation. We exclude the self-employed from the sample since it is difficult to calculate their monthly income from bank account information. However, it is important to note that the proportion of wage earners among the Spanish working population was 84.4% in the first quarter of 2020 (Labor Force Survey of Spain EPA).[14]

Since most of the individuals in the sample are workers, to analyze the representativeness of the sample in terms of the distribution of salaries we compare our data with the data of the Encuesta de Estructura Salarial (EES, Spanish Wage survey). Therefore, we consider the individuals in our sample who were working in February of 2020. We should notice that the wages received by workers in their accounts are net of tax withholds and Social Security. In order to compare our data with the EES we need to calculate the distribution of net salaries. For that purpose we have transformed the gross salaries of the EES in net salaries by subtracting social insurance payments and taxes withheld.

Figure 1 shows the distribution of the monthly wages of our sample, corresponding to February of 2020, compared with the distribution of monthly net salaries[15] for the Spanish Wage Survey. Unfortunately the latest microdata available for the quadrennial survey on the distribution of wages corresponds to 2014.[16] Since there is a time difference with our data we have adjusted the figure moving the whole distribution by the increase in the average salary between 2014 and 2018 (the last available EES). Figure 1 compares both distributions. We can see that the histogram of the net wages of our sample is very well adjusted by the density estimation of the adjusted distribution of net salaries in the official wage survey.

[13] The average in high income countries is 93.7%.

[14] The relevance of wages as the main source of income can also be seen in the similarity of the inequality measures using income or gross wages. For instance, for the last period for which both measures are available, income inequality in Spain, measured by the Gini index, was 34.5 while wage inequality was 34.3. The average absolute difference in the Gini index for income and wages over the period 2008-2016 was 0.3.

[15] Base salary plus extraordinary payments plus payment for extra hours minus tax withholds minus Social Security contributions.

[16] The microdata for the EES 2018 are not yet available at the time of writing this version.

[Include Figure 1 here]

Since the distributions are so similar it is not surprising to see that the quantile ratios used regularly to describe inequality are very similar in both distributions as shown in Table 1.

Table 1: Inequality: basic quantile ratios (net wages)

	Our sample (CBK)	EES 2014
P90/P10	4.24	4.12
P90/P50	1.88	1.87
P10/P50	0.44	0.45
P75/P25	1.85	1.83

Notes - EES stands for Encuesta de Estructura Salarial (Spanish Wage Survey).

To run an additional check on the representativeness of our sample we compare the distribution of individuals by gender and age with other sources. Table 2 summarizes the comparisons. In our sample 54% of the individuals are male. This compares satisfactorily with the 52% of males in the sample of the EES of 2014. In order to compare with more recent estimates, columns 3 and 4 include the proportions of males among employees in the Labor Force Survey of the last quarter of 2019 and first quarter of 2020.We want to compare with February of 2020, which is in the middle of the first quarter. The proportion of males is identical to the one in the EES of 2014 and very close to the one in our sample. With respect to age we also find that the proportions in our sample are very similar to the age distribution of workers. These distributions of characteristics confirm the appropriateness of our sample.

4 Results

In this section we present the basic results of the paper. One of the objectives of the paper is to analyze the role of public benefits on inequality. For this reason, all the exercises present two situations: pre-public benefits and post-benefits. The pre-benefits scenario considers the wages of the individuals in the reference sample before including unemployment benefits and furlough schemes. The post-benefits scenario considers unemployment benefits, subsidies and furlough schemes perceived by the individuals of the reference sample.

Table 2: Check our data versus labor surveys

	Our sample (CBK)	EES 2014	EPA4T19	EPA1T20
N	3,028,204	209,473	$\approx 200,000$	$\approx 200,000$
Gender				
Male	0.54	0.52	0.52	0.52
Female	0.46	0.48	0.48	0.48
Age				
≤ 19	0.01	0.00	0.008	0.007
20-29	0.18	0.12	0.145	0.142
30-39	0.25	0.31	0.246	0.243
40-49	0.28	0.30	0.305	0.304
50-59	0.21	0.21	0.233	0.237
≥ 60	0.07	0.05	0.060	0.063

Notes - EES stands for Encuesta de Estructura Salarial (Spanish Wages Survey); EPA4T19 refers to the sample of employees in the Spanish Labor Survey (EPA) in the last quarter of 2019; EPA1T2020 refers to the sample of employees in the Spanish Labor Survey in the first quarter of 2020.

4.1 Changes in the distribution of wages

Figure 2 shows the changes in wages from the period before the beginning of the lockdown (February) to April for different levels of wages in the reference period (February) before accounting for public schemes. We include not only the distribution of the change in 2020 but also in 2019.[17] In the x-axis we show the percentage change in salaries while in the y-axis we include the distribution. To simplify the figure we include only three levels of initial wages, measured in February of 2020, and the average for all the individual in the sample of reference. We consider five representative levels of wages. The interval from 900 to 1,000 euros includes the 25 percentile of the distribution of wages. The median is included in the interval from 1,200 to 1,300 euros. The interval from 1,700 to 1,800 euros includes the 75 percentile. The interval from 2,900 euros to 3,000 euros includes to the 95 percentile. Finally, the interval 4,700 to 4,800 represents to top 1%. The comparison of the changes between April and February in 2019 and 2020 show several interesting facts. First, in general there is a reduction in the proportion of salaries that are unchanged between both months in 2020 versus 2019. Second, we observe that the movement to the category no income is much larger in 2020 than in 2019. Finally, Figure 2 shows a large increase in the probability mass around a large reduction of wages for the group of high salaries (top 5% and top 1%). The reason for this bump is the payment of bonuses in February, which implies that the difference of wages between April and the reference month is a sudden reduction.[18] This is the reason why analyzing the evolution of the distribution of wages only in 2020

[17] The same figures are available for March and May but the trends are essentially the same.

[18] As we explained before and it is the standard in official statistics, our definition considers all the payments to employees including extraordinary payments and bonus.

would not provide the right answer.

Figure 3 shows the difference in wages between April and February after accounting for the public support schemes. In this case the movement to the no income category has been largely mitigated with respect to the pre-benefits situation (Figure 2). There is still the concentration around 50% to 75% reduction of salaries for the high wages segments.

[Insert figure 2 here]

[Insert figure 3 here]

Figure 4 provides the same information of the previous two graphs but comparing in the same figure all the interval of wage of reference. Panel (a) shows the pre-benefits situation. Most of the distribution is concentrated around the no change, or small changes, in wages. There are two interesting parts of the distribution. First of all a significant portion of the probability mass is concentrated around the no income extreme. More interestingly, the lower is the salary in the period of reference, the higher is probability of losing the job. This monotonic behavior is not only present in the groups of income included in Figure 4 Panel (a) but it is a feature common to all the levels of wages of the distribution of reference. The second noticeable feature of the figure is a large bump in the reduction of wages around a 50% that we already discussed above. In fact, for the highest level of wages in our sample (top 0.01%) this bump reaches over 30%. Not surprisingly for low level of wages there is no bump in the distribution at that salary change.

[Insert figure 4 here]

This means that wages for high income individuals are drastically reduced from February to April. However, this is a seasonal effect that is observed every year. To eliminate this seasonality we will discuss the results of the diff-in-diffs distribution.[19] Figure 4 Panel (b) includes the same distribution calculated after considering public benefits. In this case the concentration of mass around the no income situation is largely reduced while part of the probability mass is moved to increases in wages. In the upper limit we observe the same monotonic classification: the highest proportion of increase corresponds to the lowest wages and increases with the salary of reference.

[19]The Economist has recently cautioned against putting too much weight in the measurement of economic activity using new real-time indicators claiming that seasonality is an important problem that cannot be addressed most of the time when using real-time indicators. See "Real-time danger", 25th of July, 2020.

To correct for seasonality Figure 5 presents the difference in the proportion of changes in salaries between April and March of 2020 versus 2019, separating by different levels of wages. Logically, the effect of the February bonus in the high wages has been eliminated. It is interesting to notice that the transmission of the shock to the shape of the wave is very different for low salaries and high salaries in the pre-subsidy situation. In the case of the wages below 1,300 euros the reduction of the no salary change category has push part of the probability mass to the no income category. For wages above 1,700 euros the vibration of the wave sends part of the no salary change to a small reduction of wages.

To facilitate the comparison across wages groups, Figure 6 shows the changes for all wage categories in the same figure. The movements are similar, although less pronounced than in Figure 4. The pre-benefits figure, Panel (a), reproduces the same monotonicity of increase in the proportion of no income in function of the wages of the month of reference. Panel (b) shows the post-benefits case where, as before, the increase in the no income category has been largely reduced.

[Insert Figure 5 here]

[Insert figure 6 here]

4.2 The evolution of inequality

One important objective of this research is to evaluate the evolution of inequality after the impact of the COVID-19 and the mitigating effect of the measures approved by the government. Figure 7 Panel (a) shows the evolution of the Gini index from February to May of 2020 and the same months for 2019. The pre and post-benefits curves are basically parallel until April of 2020 were there is a large increase of the Gini index pre-benefits while the post-benefit curve shows a more moderate increase. In May of 2020 the pre-benefits Gini index still stays very high while the post-benefits has returned to the pre-COVID situation.

[Include Figure 7 here]

We can check the evolution of the Gini index with respect to alternative measures of inequality. For instance, Figure 7 Panel B shows the evolution of the Theil index, which is an inequality measure related with the concept of entropy and Shannon's index. The Generalized entropy family

$$GE(\alpha) = \frac{1}{N\alpha(\alpha-1)} \sum_{i=1}^{N} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right] \;\; for\, \alpha \neq 0, 1 \tag{1}$$

where $\bar{y}$ is the mean of y. If $\alpha = 1$ then we obtain the Theil index

$$GE(1) = Theil = \frac{1}{N} \sum \left(\frac{y_i}{\bar{y}}\right) ln \left(\frac{y_i}{\bar{y}}\right) \tag{2}$$

Figure 7 Panel B confirms the evolution of inequality during this period, showing that the Theil index evolves very similarly to the Gini index. There is a large increase in inequality in March in both, pre and post-benefits, that continues in April in the pre-benefits situation but is tamed in the post-benefits case.

Figure 8 depicts the Lorentz curve in February of 2020, our reference month, before and after considering benefits. Obviously the effect of the public action moves the curve upward, mostly for the low percentiles. Figure 9 shows the change between the pre and the post-benefits Lorenz curve for April of 2020. In this case, as already seen in the Gini index, the movement between the pre-benefits Lorenz curve and the post-benefits is much larger. Figure 10 shows the movement of the pre-benefits Lorenz curve every month from February 2020 until May 2020. It shows, as expected after analyzing the evolution of the Gini index, that the downward movement accelerates in April and it stabilizes in May. The post-benefits movements are smoother than the ones of the pre-benefits Lorenz curves (Figure 11)

[Include Figure 8 here]

[Include Figure 9 here]

[Include Figure 10 here]

[Include Figure 11 here]

5 The evolution of inequality by subgroups and geographies

In this section we analyse the evolution of inequality measured by the Gini index in different subgroups of the population, differentiating by gender, age, regions, and country of origin. This allows us to analyse how the crisis is affecting the different segments of the population and, ultimately, the impact it might be having on the most vulnerable groups. Our preliminary analysis of decomposition between and within groups[20]

[20]We have used the Theil index which, differently from the Gini index, provides an additively decomposable measure of inequality.

shows that the between groups contribution is very small in all the cases, and inequality is dominated by the within component. Additionally, the change in the importance of the between component with respect to the situation in the reference month is also small. For this reason in the rest of the section we consider the evolution of the within group inequality[21]

5.1 Gender and inequality

Figure 12 shows the evolution of the Gini index by gender, before and after the outbreak of the crisis, and before and after public subsidies. As it can be seen, the evolution of the Gini index was relatively stable during 2019. In February 2020 its level was similar to the one observed in the previous year. We do not observe major differences between males and females before the shock. The magnitude of the increase in the Gini index after the declaration of the state of alarm is similar across gender types before public transfers, but slightly higher for females in the post-benefits case.

[Include Figure 12 here]

5.2 Inequality within age groups

Figure 13 shows the evolution of the Gini index by different age groups, before and after the shock, and pre and post-benefits. In this case, the differences across groups are remarkable. In the pre-transfers case, we observe a large increase in the Gini index among the youngest cohorts (between 16 and 29 years old). For the other age groups the impact is also notable, but substantially smaller, specially for the oldest cohorts (between 50 and 64 years old). The disbursement of the public transfers partially mitigates the large increase in the Gini index among the young. Nonetheless, the increase in the Gini index post-benefits is still remarkable both in absolute and in relative terms. The increase in the Gini index post-benefits for the older cohorts is relatively minor.

[Include Figure 13 here]

[21] Recent research (Liao (2016)) argues that the within component may contain part of the between group inequality if there is a part of the within distribution that is not shared by the different groups. This could be particularly relevant for the case of gender inequality. We plan to address these issues in future research.

5.3 Inequality and regions

The change in the Gini index across the different Spanish regions is also very different. The jump of the index pre-benefits is specially pronounced in the Balearic Islands and Canarias, where the weight of tourism is specially high, and the economic impact of the restrictions to international mobility is likely being specially intense. The increase in the Gini index post-benefits is significantly smaller, and the regional variation is also significantly reduced once public benefits are taken into account. Note that the variation in Gini index post-benefits is not only related to the variation in the Gini index pre-benefits but also to the coverage and processing speed of support programs. We leave this for further research.

[Include Figure 14 here]

5.4 Nationality and inequality

The internal data of CaixaBank also allows us to analyse the impact of the crisis depending on the country of origin of the population. The Gini index pre-benefits increases significantly more for foreign-born individuals than for native-born. The increase in the Gini index for foreign-born is smaller in the post-benefits case, but still significantly higher than for natives. If we further classify foreign-born individuals by the GDP per capita level of their country of origin, we observe significant differences between different income groups. For example, the Gini index post-benefits is not reduced over time for foreign-born from low income countries while for foreign-born from high-income countries inequality post-benefits is reduced (as it is the case for native-born individuals). These higher increases in inequality post-benefits for migrants show their vulnerability in times of crisis as their social welfare net is thinner. We can also aggregate all the foreign born population. Figure 16 shows the large increase in inequality of the foreign born sample and the improvement, although incomplete, after accounting for public benefits.

[Include Figure 15 here]

6 Concluding remarks

In this paper we study the evolution of inequality since the beginning of the COVID-19 pandemic, and its effect on different groups of the population. First, we show that the payroll data managed by banks are an extremely useful source of information to detect, timely and accurately, changes in the distribution of wages. Our data replicate very closely the distribution of wages from the official wage surveys. Second, we show that, in absence of the large public scheme activated soon after the beginning of the COVID-19 crisis, inequality would have increased dramatically. The impact of the crisis on inequality is explained mostly by its effect on low-wage workers. Pre-benefits wage inequality has increased significantly among foreign-born individuals and regions that have a heavy economic dependence on touristic activities. Finally, we show that the public benefits activated soon after the beginning of the pandemic have substantially mitigated the impact of the COVID-19 crisis on inequality.

References

Adams-Prassi, A., Boneva, T., Golin, M., & Rauh, C. (2020). *Inequality and the impact of the coronavirus schock: evidence from real time surveys* (Tech. Rep.). mimeo.

Alvargonzález, P., Pidkuyko, M., & Villanueva, E. (2020). La situación financiera de los trabajadores más afectados por la pandemia: un análisis a partir de la encuesta financiera de las familias.

Baker, S. (2018). Debt and the response of household income to shocks: validation and application of linked financial account data. *Journal of Political Economy*, *126*(4), 1504-1556.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). *Income, liquidity, and the consumption response to the 2020 economic stimulus payments* (Tech. Rep.). National Bureau of Economic Research.

Bick, A., & Blandin, A. (2020). Real time labor market estimates during the 2020 coronavirus outbreak. *Unpublished Manuscript, Arizona State University*.

Cajner, T., Crane, L. D., Decker, R. A., Grigsby, J., Hamins-Puertolas, A., Hurst, E., … Yildirmaz, A. (2020). *The us labor market during the beginning of the pandemic recession* (Tech. Rep.). National Bureau of Economic Research.

Carvalho, V. M., Hansen, S., Ortiz, Á., Garcia, J. R., Rodrigo, T., Rodriguez Mora, S., & Ruiz de Aguirre, P. (2020). Tracking the covid-19 crisis with high-resolution transaction data.

Chen, S., Igan, D., Pierri, N., & Presbitero, A. F. (2020). *Tracking the economic impact of covid-19 and mitigation policies in europe and the united states* (Tech. Rep.). Working Paper.

Chetty, R., Friedman, J. N., Hendren, N., & Stepner, M. (2020). *Real-time economics: A new platform to track the impacts of covid-19 on people, businesses, and communities using private sector data* (Tech. Rep.). Mimeo.

Cicala, S. (2020). *Early economic impacts of covid-19 in europe: A view from the grid.* unpublished manuscript.

Gelman, M., Kariv, S., Shapiro, M. D., Silverman, D., & Tadelis, S. (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science*, *345*(6193), 212–215.

Gonzalez, J., Urtasum, A., & Perez, M. (2020). Evolucion del consumo en españa durante la vigencia del estado de alarma: un análisis a paritr del gasto con tarjetas de pago. *Articulos Analiticos, 3/2020*.

Gourinchas, P.-O. (2020). Flattening the pandemic and recession curves. *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever*, 31.

Kueng, L. (2018). Excess sensitivity of high-income consumers. *Quarterly Journal of Economics*, *133*(4), 1693-1751.

Liao, T. (2016). Evaluating distributional differences in income inequality. *Socius*, *2*, 1-14.

Olafsson, A., & Pagel, M. (2018). The liquid hand-to-mouth: evidence from personal finance management software. *Review of Financial Economics*, *31*(11), 4398-4446.

Scheidel, W. (2017). *The great leverer.* Princeton University Press.

Wade, L. (2020). An unequal blow. *Science*, *368*(6492), 700-703.

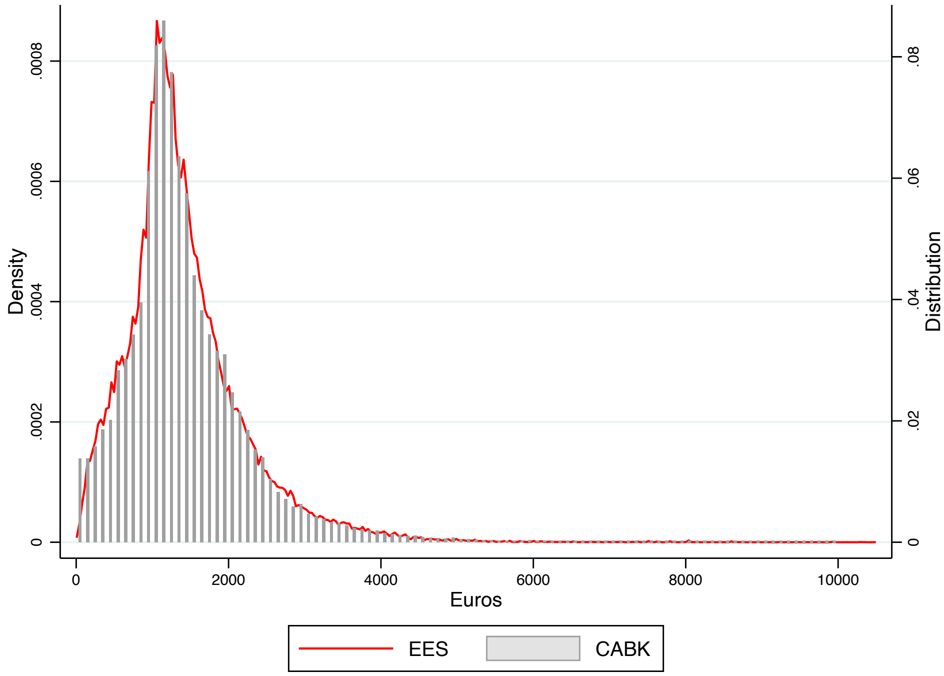

Figure 1: Distribution of monthly net wages: our sample (CABK) versus the sample of the Wage Survey (EES)

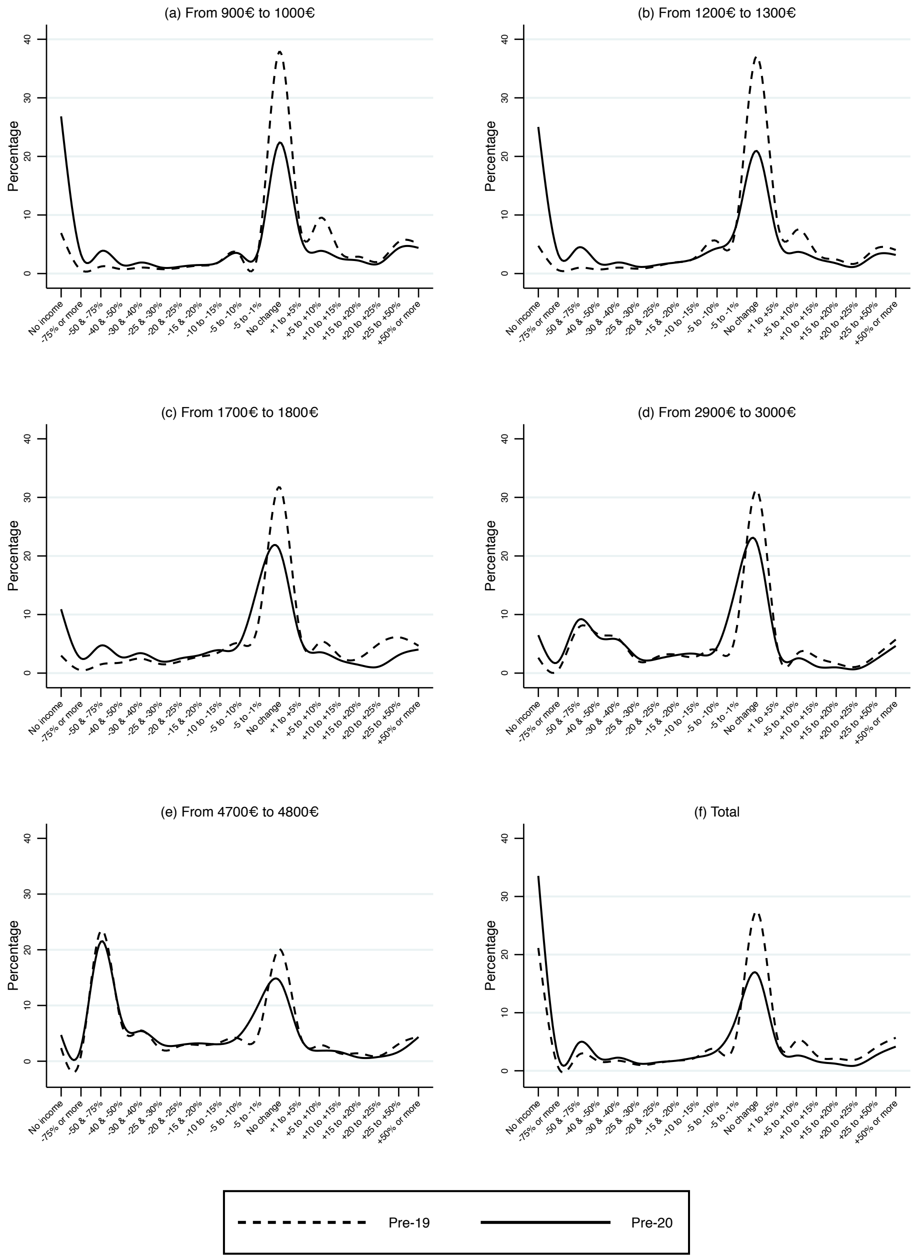

Figure 2: Changes in payments between April and February by level of wages in the reference period. Pre-benefits. Comparing 2020 and 2019

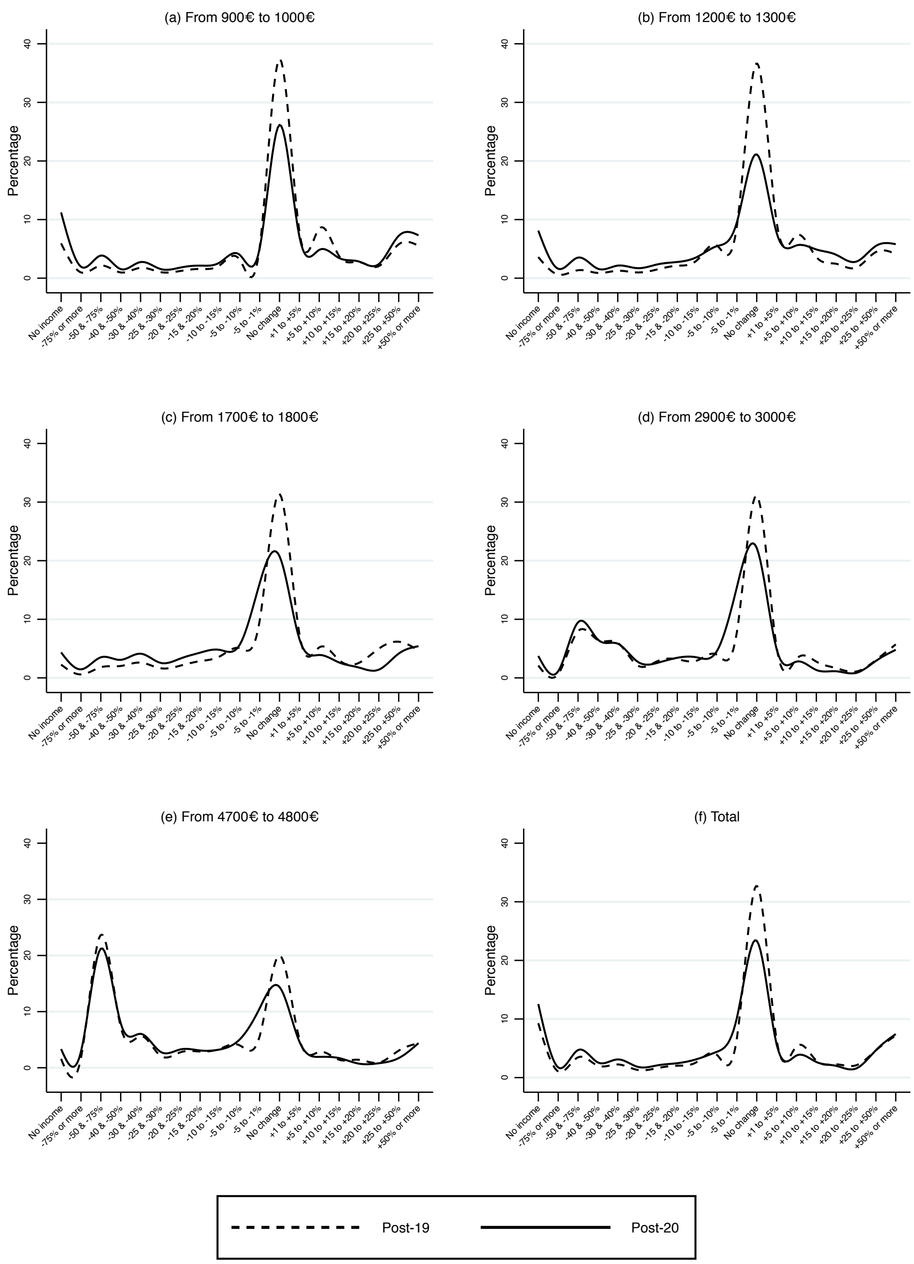

Figure 3: Changes in payments between April and February by level of wages in the reference period. Post-benefits. Comparing 2020 and 2019

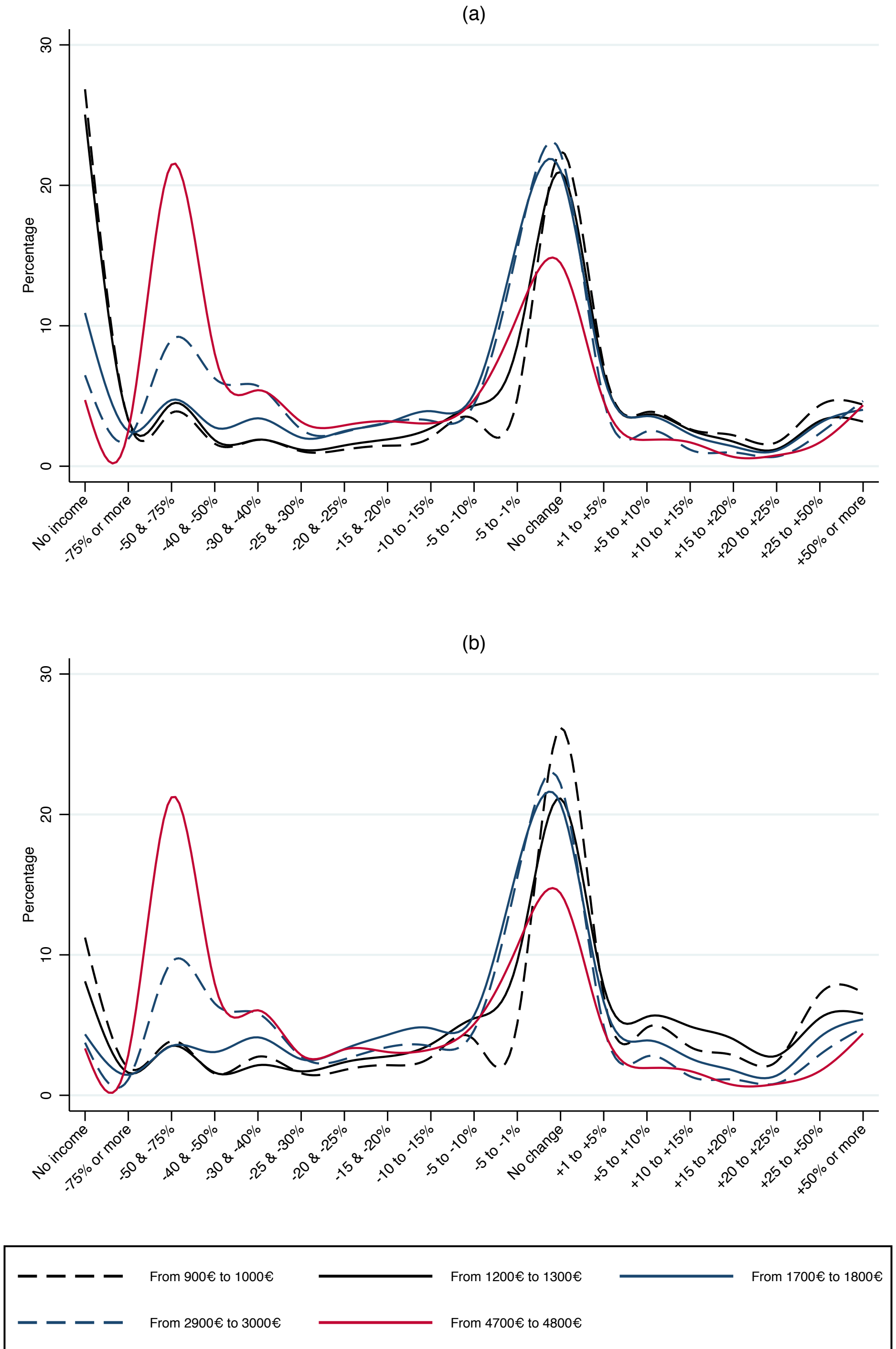

Figure 4: Change in payments received by workers between April 2020 and February 2020. (a) Pre-transfers. (b) Post-transfers

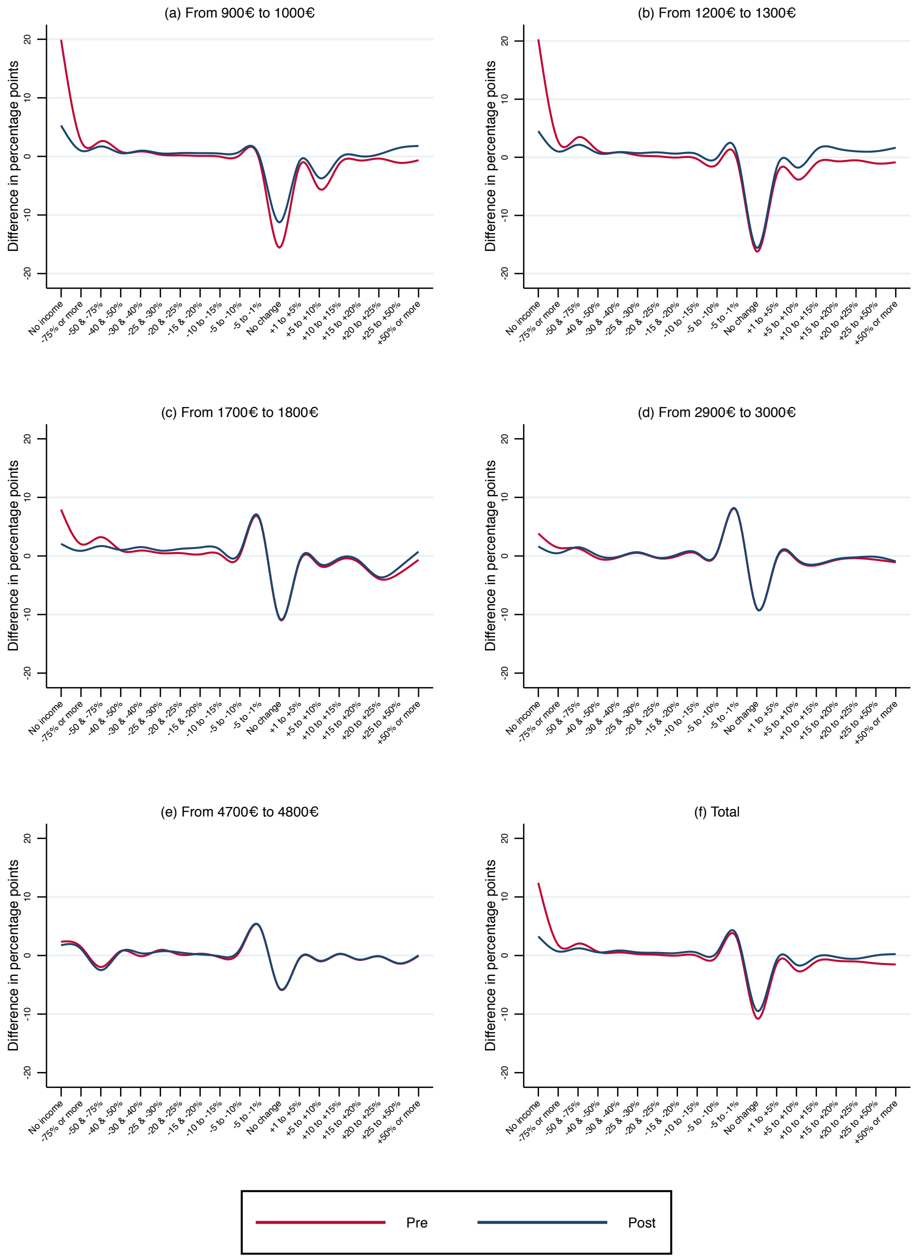

Figure 5: Diff-in-diffs in payments for each level of salaries in the reference month. April vs February - 2020 vs 2019

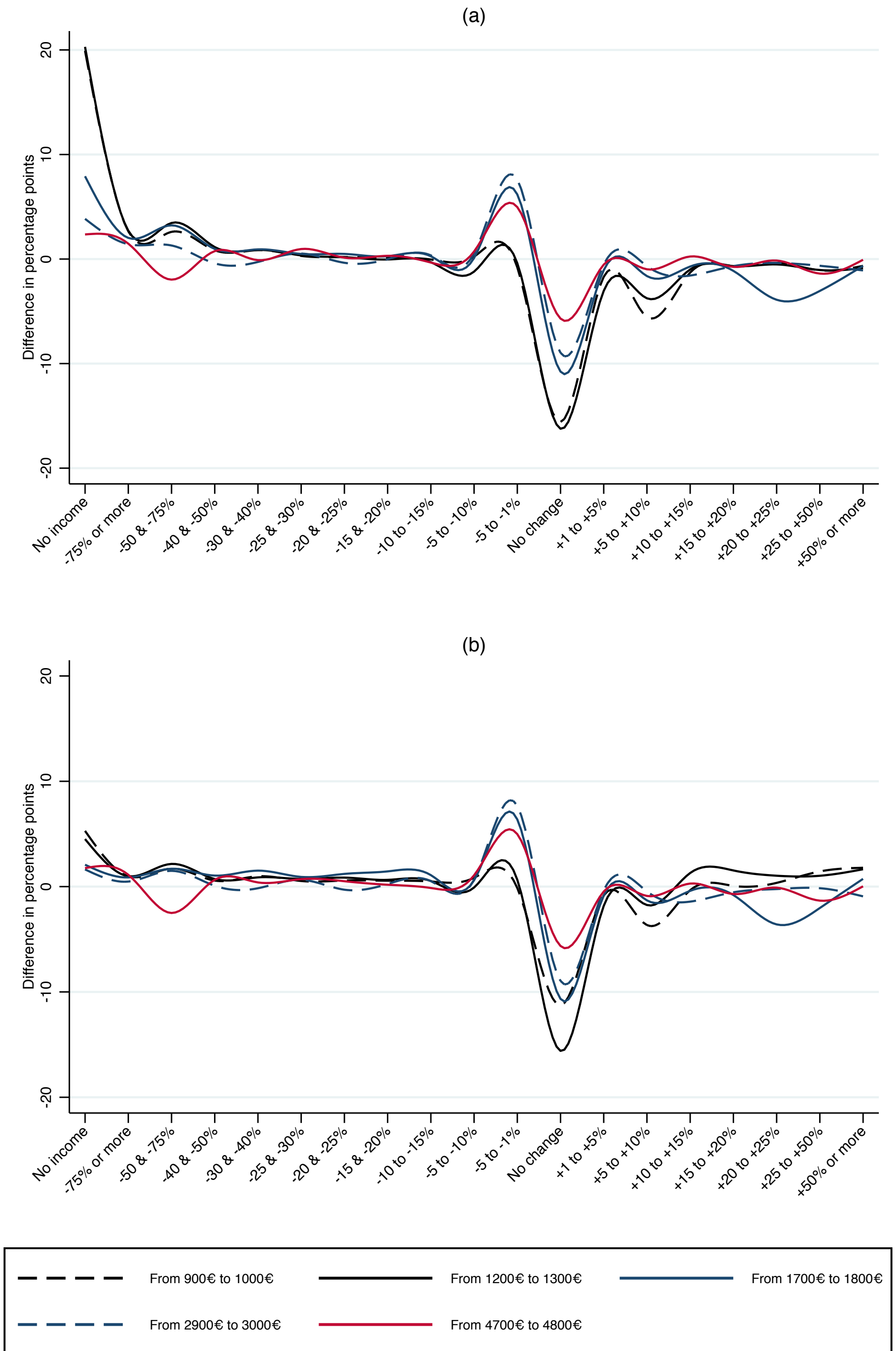

Figure 6: Diff-in diffs in payments received by workers: April vs February comparing 2020 vs 2019 (a) Pre-transfers. (b) Post-transfers

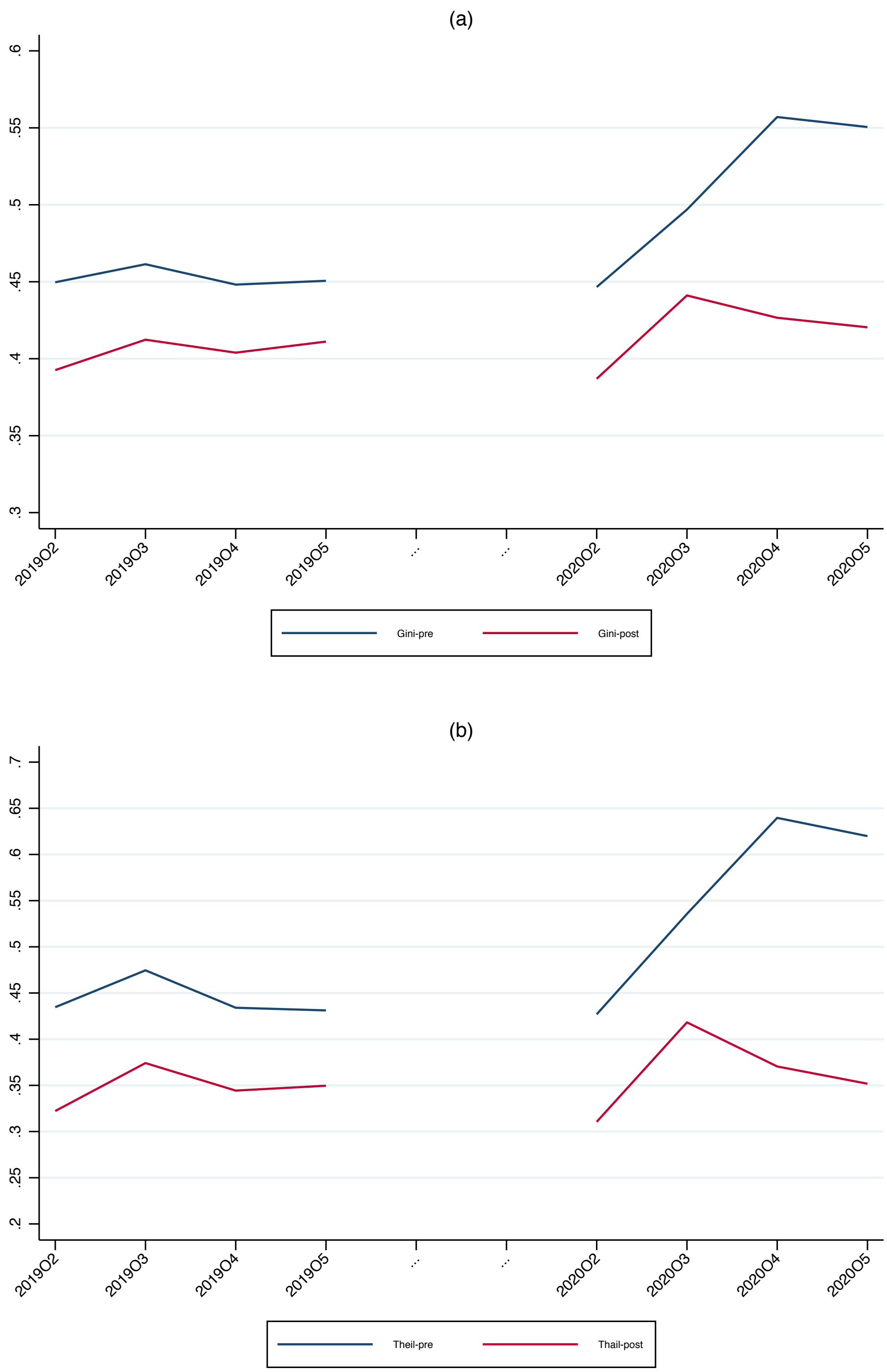

Figure 7: Inequality measures
(a) Gini index (b) Theil index ($\alpha = 1$)

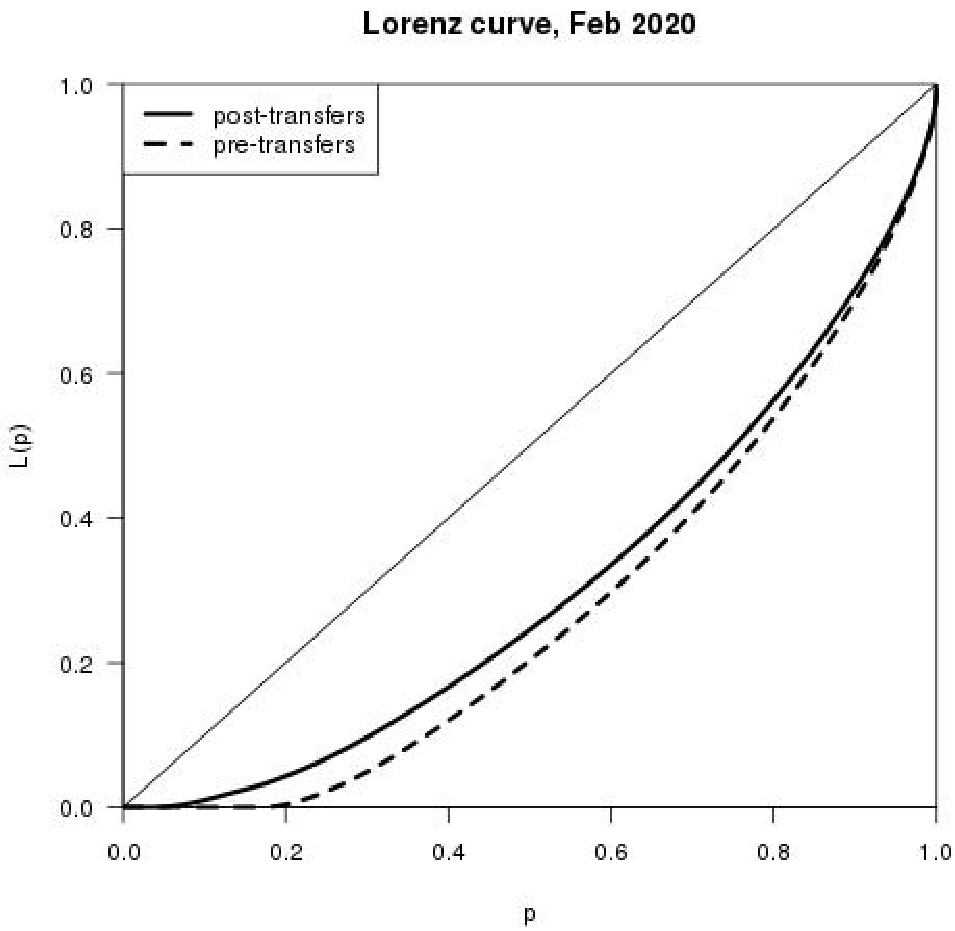

Figure 8: Lorenz curve (February 2020)

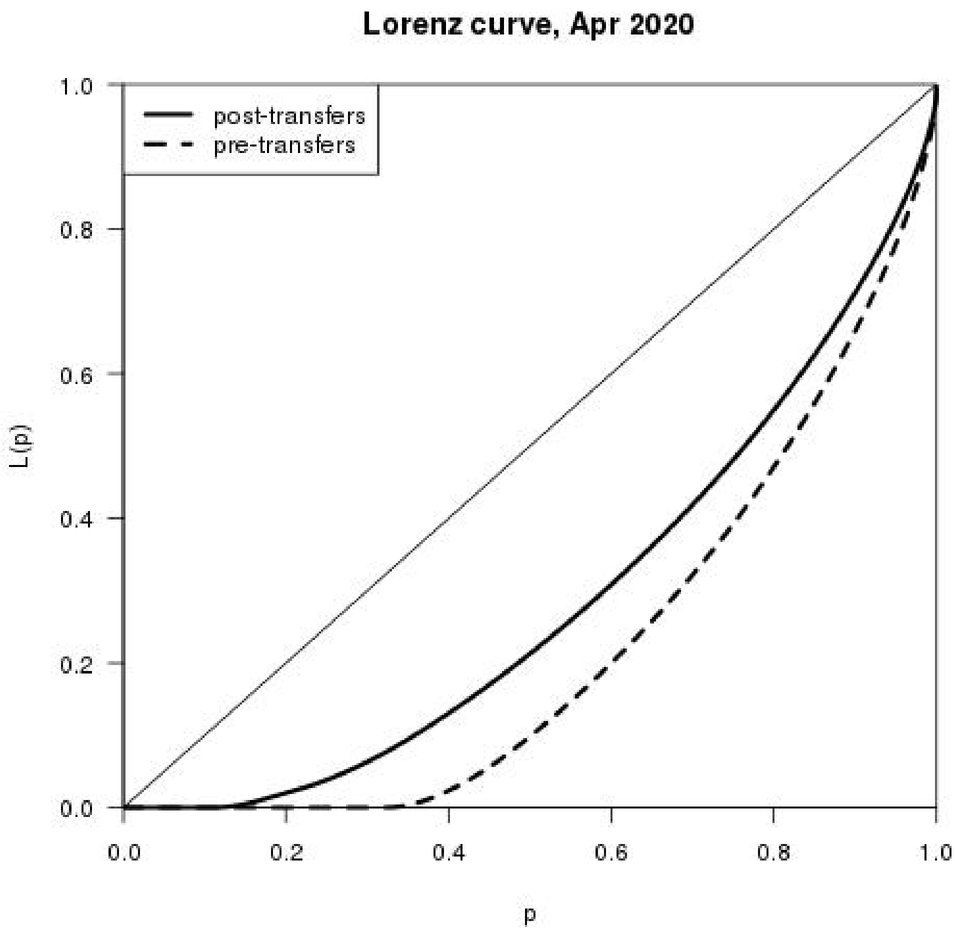

Figure 9: Lorenz curve (April 2020)

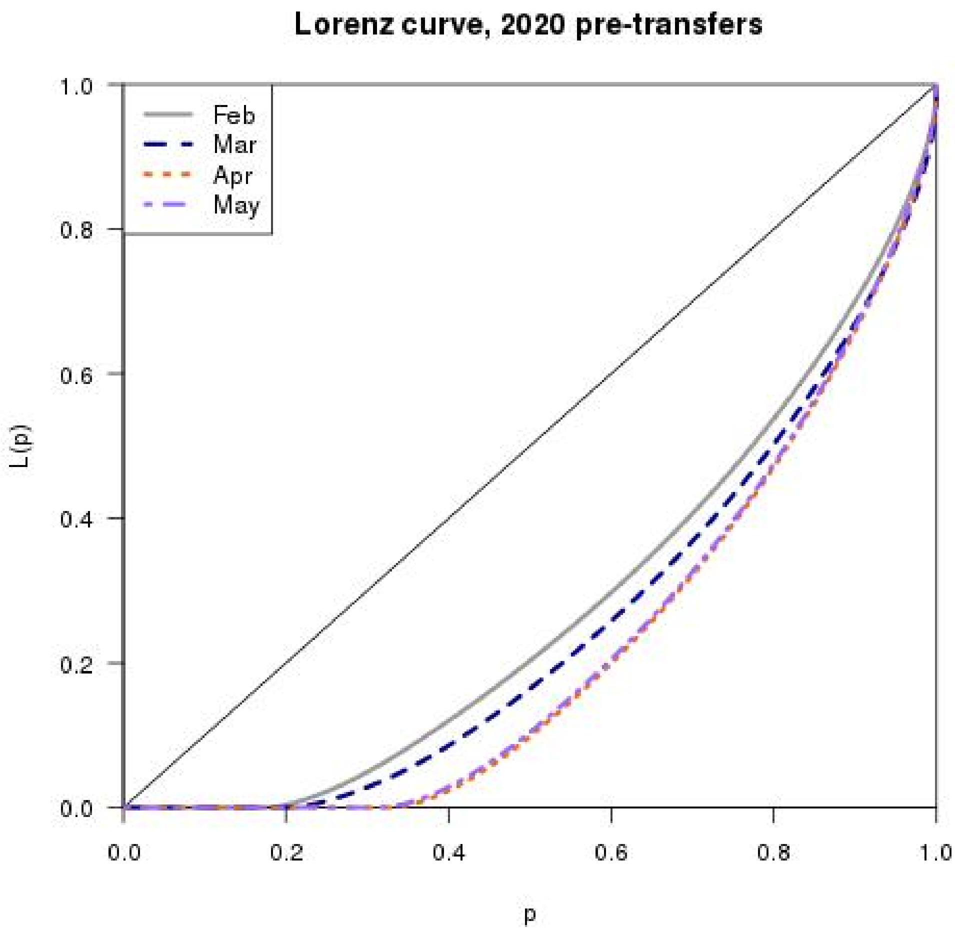

Figure 10: Movements of the Lorenz curve. Pre-benefits, 2020

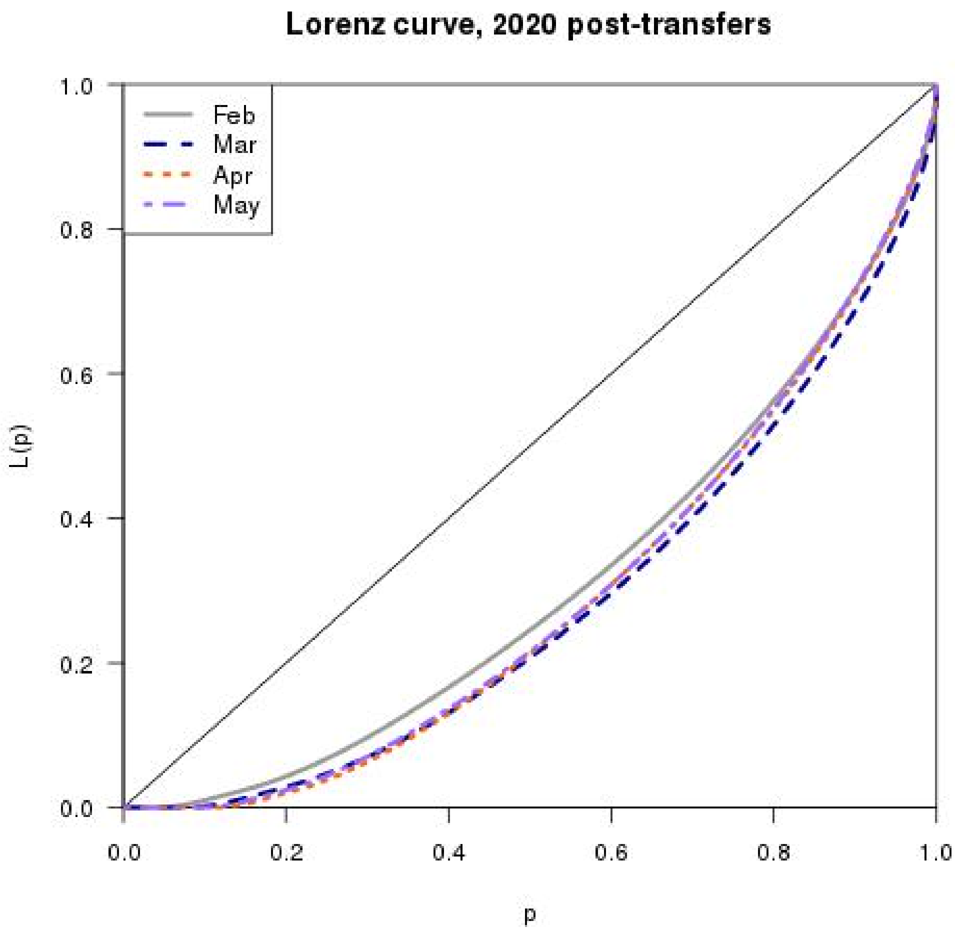

Figure 11: Movements of the Lorenz curve. Post-benefits, 2020

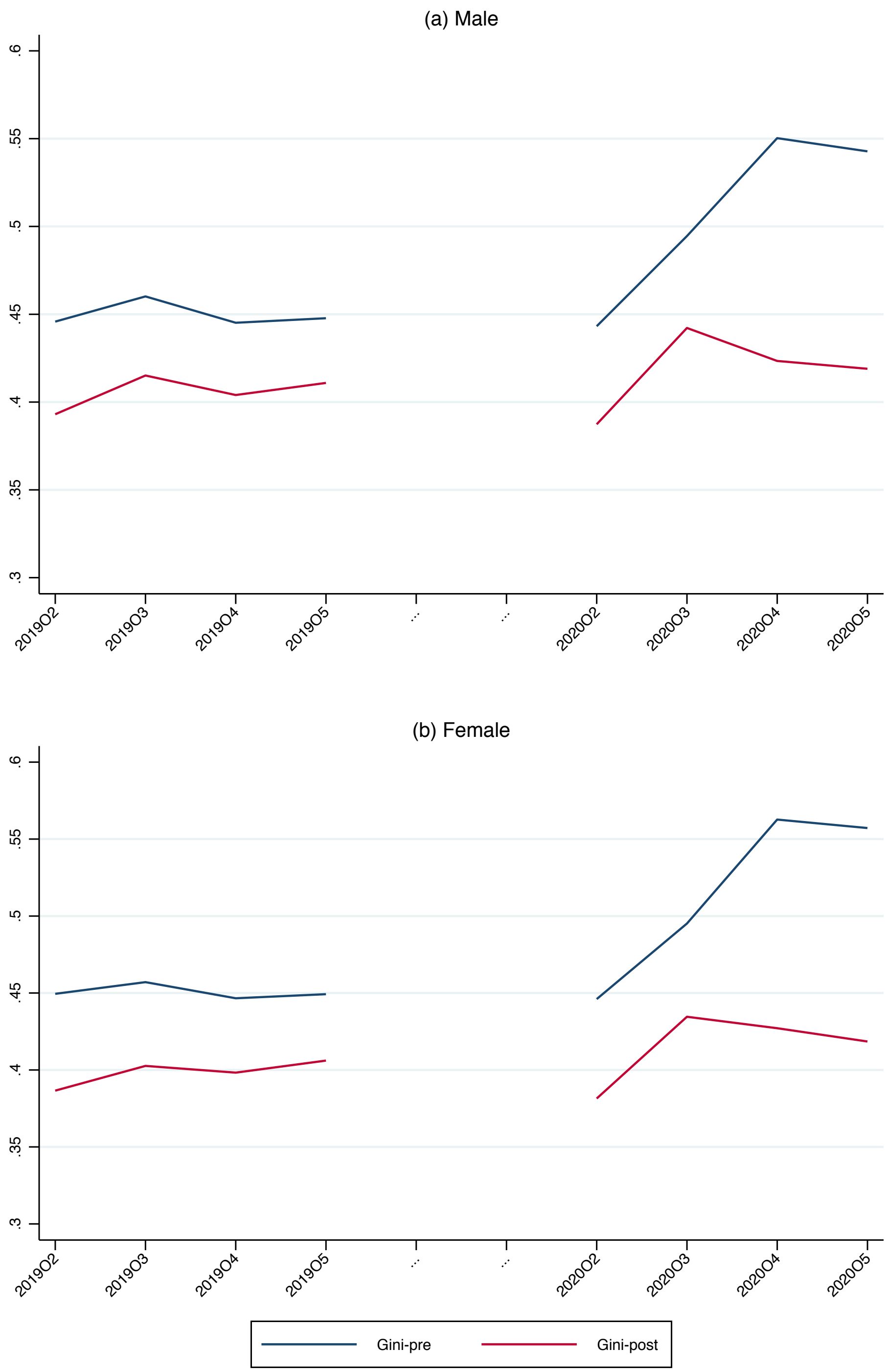

Figure 12: Evolution of the Gini coefficient by gender
(a) Males (b) Females

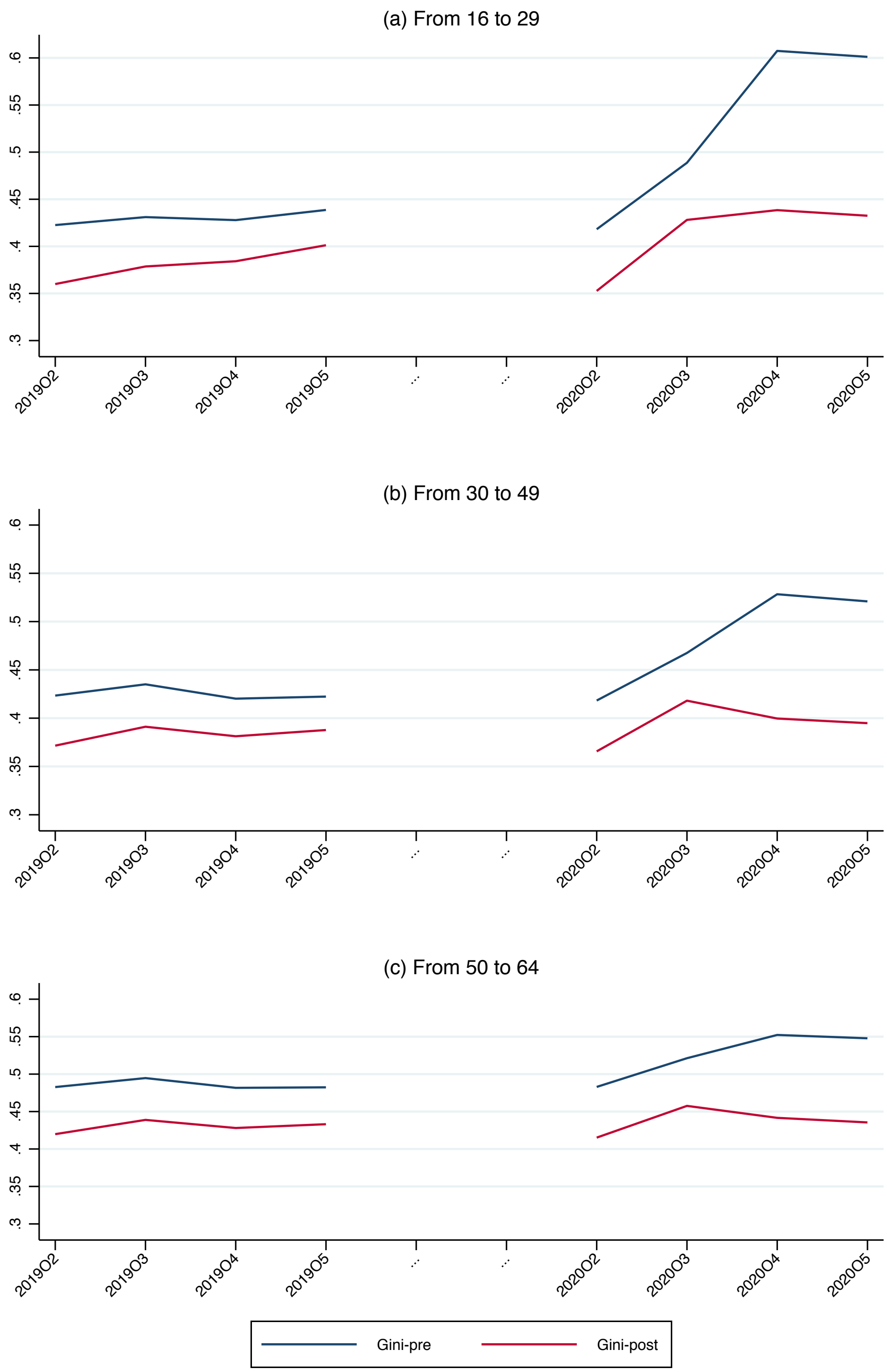

Figure 13: Evolution of the Gini coefficient by age
(a) From 16 to 29 years old (b) From 30 to 49 years old (c) From 50 to 64 years old

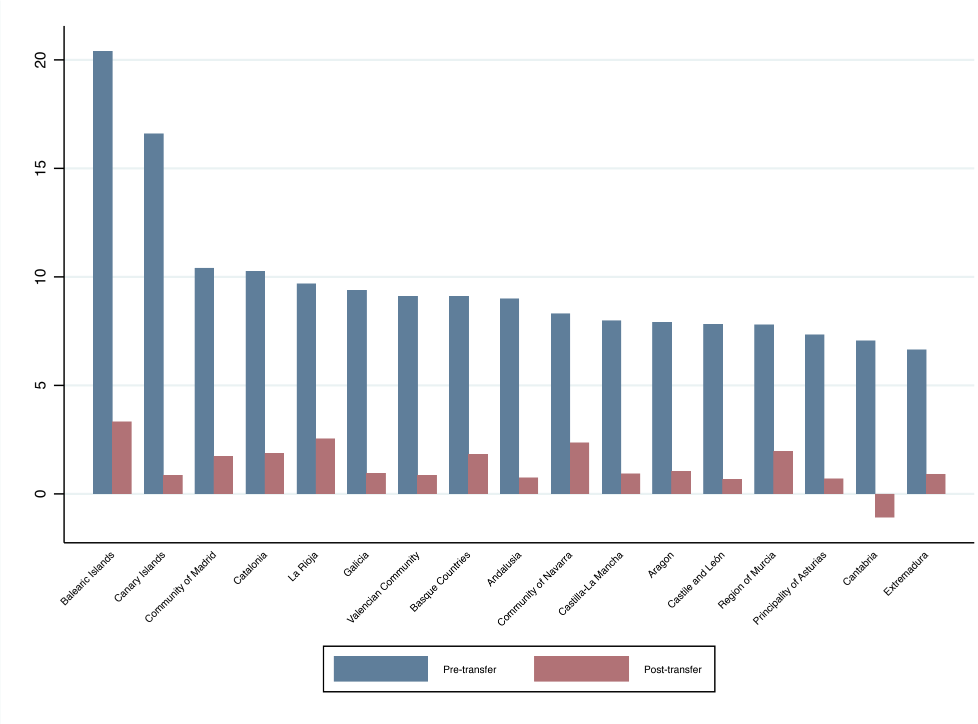

Figure 14: Changes in the Gini coefficient by regions

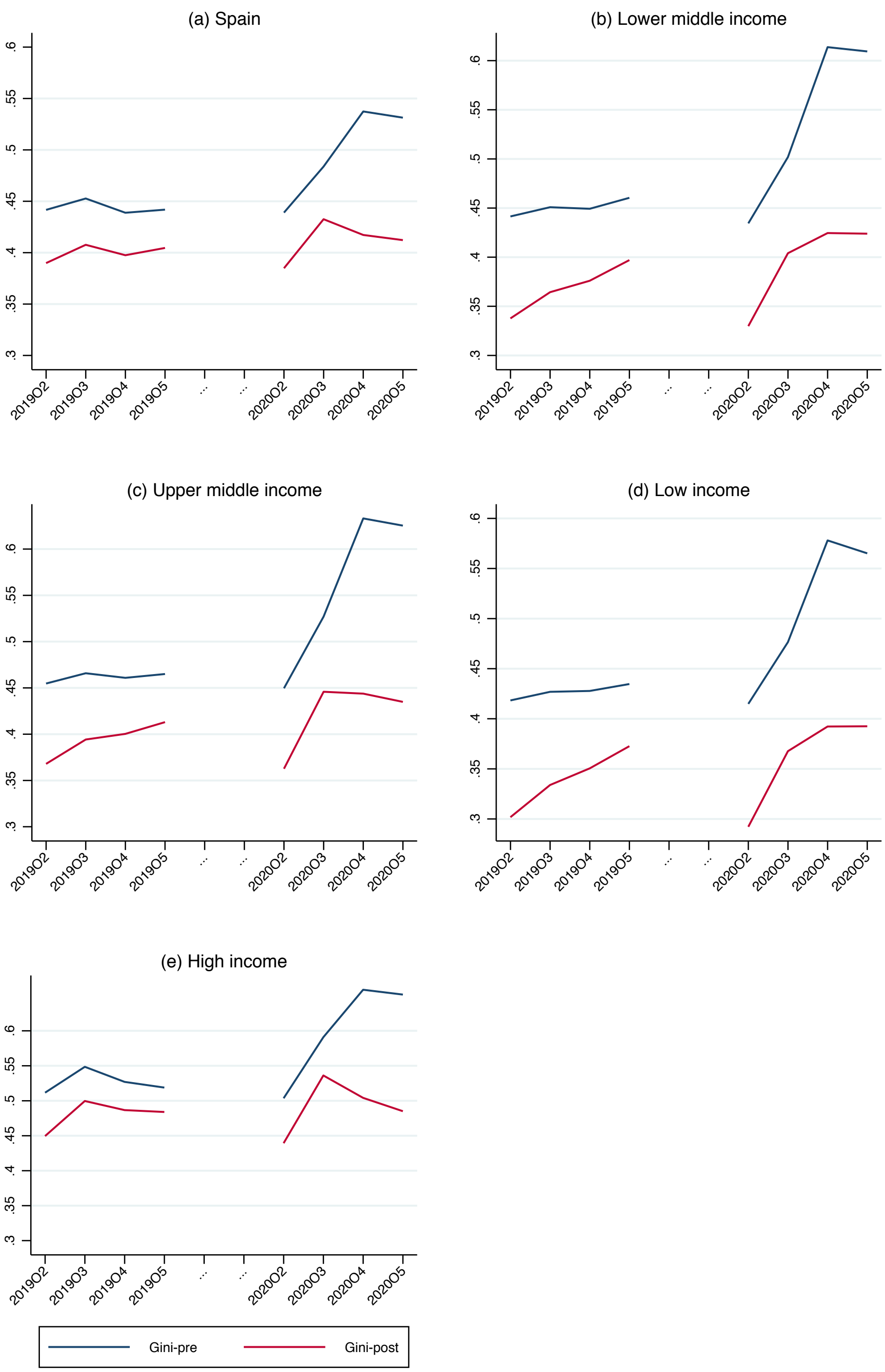

Figure 15: Evolution of the Gini coefficient by country of origin
(a) Spain (b) Lower middle income (c) Upper middle income (d) Low income (e) High income

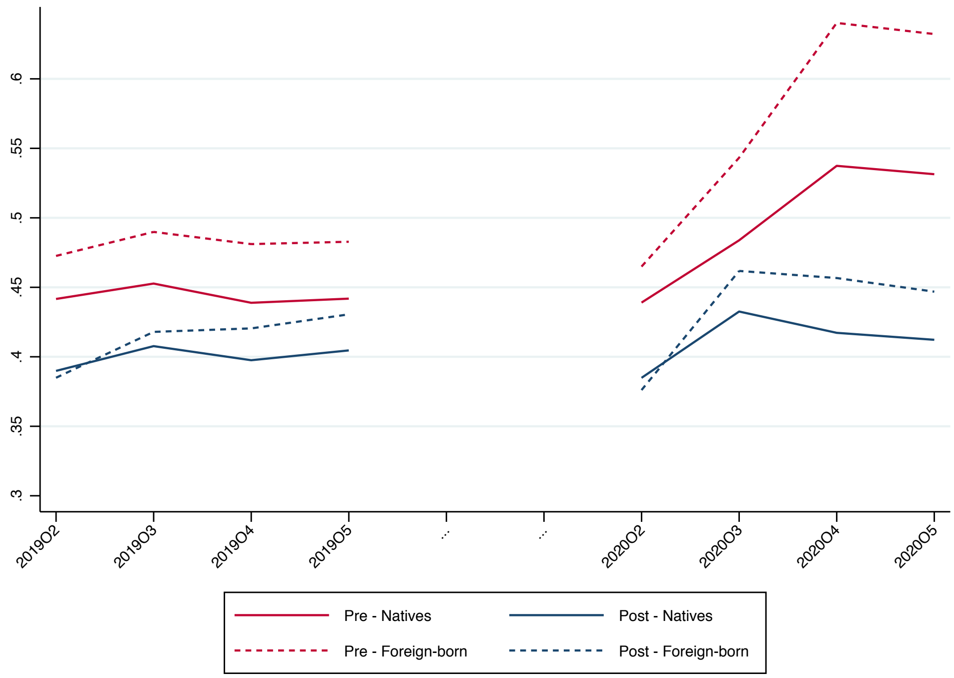

Figure 16: Evolution of the Gini coefficient: natives versus foreign born
(a) Natives (b) Foreign born