



Universidad de Valladolid



PROGRAMA DE DOCTORADO EN ECONOMÍA

TESIS DOCTORAL:

**DESCOMPOSICIÓN SIMULTÁNEA DE LA
DESIGUALDAD. ALGUNAS APLICACIONES**

Presentada por JAVIER MARTÍN ROMÁN para optar al
grado de
Doctor por la Universidad de Valladolid

Dirigida por:

LUIS AYALA CAÑÓN
JUAN VICENTE PERDIZ

Inicialmente, estaba previsto presentar esta Tesis Doctoral por compendio de publicaciones. Aunque aún dispongo de plazo dentro del Programa de Doctorado en Economía de la Universidad de Valladolid, razones contractuales me aconsejan optar por la modalidad ordinaria. En todo caso, para acreditar la calidad de esta Tesis Doctoral presento la siguiente aportación científica relacionada con la misma, que ya está publicada y que cumple con los criterios de la ANECA según exige la normativa de la Universidad de Valladolid. Ayala, L., Martín-Román, J., Vicente, J.: “The contribution of the spatial dimension to inequality: A counterfactual analysis for OECD countries”. *Papers in Regional Science*, 99(3), pp. 447-477, 2020. DOI: <https://doi.org/10.1111/pirs.12503>. Impact factor: 2.22. 2019 Journal Citation Reports (Clarivate Analytics): 96/373 (Economics).

Agradecimientos

Esta tesis doctoral es la culminación de un proyecto de investigación que comenzó hace cuatro años en la Universidad de Valladolid, donde tuve la oportunidad de iniciar mi carrera académica como Personal Docente e Investigador, y que finaliza ahora en la Universidad Nacional de Educación a Distancia, donde tengo el enorme privilegio de trabajar en la actualidad acompañado de excelentes profesores y amigos.

En primer lugar, quisiera expresar mi profundo agradecimiento a todas aquellas personas que aceptaron el reto de tutorizar mi trabajo desde que decidí matricularme en el Programa de Doctorado en Economía por primera vez. En concreto, me gustaría tener un recuerdo especial para Carlos Pérez, mi primer mentor, por depositar toda su confianza en mí de manera altruista y desinteresada. Por aquel entonces, yo aún vivía atenazado por las dudas y los temores que me generaban las diversas opciones de futuro en el terreno profesional. Carlos supo inculcarme mejor que nadie el valor de una vida dedicada al estudio y la formación continua. Y lo hizo con el mismo entusiasmo y pasión que demostraba en su trabajo diario. No podía concebir esta profesión de otro modo. Tampoco me puedo olvidar de Juan Gabriel Rodríguez, por animarme a realizar una primera incursión en el campo de la igualdad de oportunidades. Sus indicaciones, que a menudo aparecían vinculadas a irreverentes axiomas, me facilitaron sobremanera la obtención de la suficiencia investigadora. Por último, he de dedicar unas palabras de sincero reconocimiento a mis directores de tesis, Juan Vicente y Luis Ayala, quienes, a pesar de las dificultades sobrevenidas en diferentes etapas del proceso, mostraron siempre una encomiable calidad humana para guiarme de manera irreprochable. Gracias por los sabios y elegantes consejos, por vuestra generosidad, y por el apoyo incondicional. Me siento muy afortunado.

Desde una perspectiva institucional, me siento en deuda con mis buenos amigos y compañeros del Instituto de Estudios Fiscales (Adrián, Álvaro, Carmen, Isabela, Jaime, Marta, Mila...). En los años que disfruté allí realicé numerosos cursos formativos, reparé en la trascendencia de presentar trabajos y comunicaciones en congresos, y aprendí a preservar mi bienestar emocional para afrontar con éxito los momentos más críticos de este viaje. Esto último, fundamentalmente, gracias a mis apreciados bedeles y ordenanzas del edificio A, que destinaron un gran esfuerzo al fomento y práctica de actividades deportivas y me aleccionaron en temas capitales. Igualmente, me gustaría manifestar mi gratitud hacia todos mis compañeros del Departamento de Fundamentos del Análisis Económico e Historia e Instituciones Económicas de la Universidad de Valladolid. Fueron ellos quienes me brindaron la posibilidad de alcanzar un gran sueño: convertirme en profesor universitario. Soy plenamente consciente de que, en un periodo tan convulso como el actual, son muchos los que lo ambicionan y muy pocos quienes pueden lograrlo. Gracias a todos ellos por el afecto, el respeto y la consideración durante los meses que pudimos compartir juntos. Asimismo, debo expresar mi agradecimiento hacia todos y cada uno de los profesores que forman parte del Departamento de Economía Aplicada y Gestión Pública de la Universidad Nacional de Educación a Distancia, que apostaron firmemente por mí desde el principio y decidieron que prolongara mi carrera académica junto a ellos. Desde que me incorporé en el año 2016 no han hecho más que apoyarme para que pudiese concluir felizmente este proyecto.

En el espacio más personal he de mencionar, primeramente, a mis buenos amigos Chema, David, Iván, Javier, Juan, Raúl y Roberto. Aunque en los últimos meses se han sentido un tanto desatendidos, han terminado empatizando y reconociendo que la motivación que escondía este aislamiento puntual estaba más que justificada. Gracias por vuestra comprensión y, especialmente, por poner a buen recaudo los valores que mejor

os definen y que sois capaces de transmitir como nadie: la sencillez y la humildad. Por lo que se refiere a mis amigos “cuantitativos”, el reconocimiento principal debe ir dirigido hacia Alfonso Moral e Ignacio Moral, que pese a no ser hermanos se han comportado como tales conmigo. No tengo palabras para describir vuestra desprendida ayuda durante todos estos años. También quisiera dedicar unas líneas a Nacho González, que me invitó gentilmente a realizar una estancia de investigación en American University a finales de 2018. Fue una experiencia única de la que solo guardo extraordinarios recuerdos. Allí tuve el placer de participar en diversos seminarios organizados por su departamento y la fortuna de presentar los avances de mi segundo capítulo de tesis. Gracias por tus reflexiones sobre potenciales aplicaciones de la metodología de controles sintéticos, muy valiosas y pertinentes, y por los buenos ratos en Stead Park.

Dentro del ámbito familiar, las referencias a mi entorno más cercano son obligadas. Sin las constantes palabras de aliento y motivación de mis hermanos, sin su amor fraternal, este trabajo jamás se hubiese podido llevar a cabo. Gracias por vuestro inestimable respaldo y protección. Me ha resultado imprescindible. Mención aparte merecen mis padres. Ellos no tuvieron la opción de cursar estudios universitarios, pero se volcaron para que sus hijos sí gozaran de esta preciosa oportunidad. La honestidad con que desempeñaron su trabajo y la ejemplaridad de su esfuerzo son únicamente dos de sus grandes enseñanzas, pero ambas me han bastado para perseverar y no caer en el desaliento en las épocas más complicadas. Gracias, en definitiva, por renunciar a todo para que a mí no me faltara de nada. Estas palabras de gratitud no están a la altura de vuestro inmenso sacrificio, pero espero que las recibáis con el mismo cariño e ilusión que me produce a mí el mero hecho de escribirlas.

No me olvido de ti, pequeña Lucía. Gracias por tu contagiosa alegría y por tu optimismo vital. Nos haces mejores a todos cada día.

Y gracias a mi querida tía Concha, por cuidarme de la manera en que lo hizo y por proporcionarme tanta felicidad. Este trabajo también es por y para ti.

CONTENTS

Introducción.....	8
Introduction	17
Chapter 1. The Contribution of the Spatial Dimension to Inequality: A Counterfactual Analysis for OECD Countries.....	25
1. Introduction.....	26
2. Literature review	28
3. Data.....	31
4. Empirical strategy	35
4.1. Inequality decomposition by population subgroups.....	36
4.2. Dynamic decomposition.....	37
4.3. Counterfactual analysis	39
5. Main results.....	42
5.1. Between and within regional inequality	42
5.2. Decomposition of the <i>trend</i> of inequality	48
5.3. How much inequality is explained by territory?	51
6. Conclusions.....	54
References.....	57
Appendix.....	61
Chapter 2. Unemployment Shocks and Material Deprivation in the European Union: A Synthetic Control Approach.....	71
1. Introduction.....	72
2. Unemployment shocks and material deprivation in the EU-28.....	75
2.1. Unemployment shocks	75
2.2. Material deprivation in EU countries	77
3. Methodology	79
3.1. Model formalization.....	81
3.2. Inference.....	83
4. Data.....	84

5. Results.....	86
5.1. Main results	90
5.2. Inference.....	95
5.3. Sensitivity analysis.....	96
6. Conclusions.....	99
References.....	102
Appendix.....	105
Chapter 3. Key Drivers of Income Inequality in Large Cities: A Multi-Country Analysis Using Static and Dynamic Decompositions	108
1. Introduction.....	109
2. Income inequality and city size: review of literature.....	111
3. Methodology	116
3.1. Empirical strategy	116
3.2. Model formalization.....	119
4. Data.....	122
5. Main results.....	124
5.1. Descriptive analysis.....	124
5.2. Reweighted RIF-OLS decomposition: Dynamic approach.....	127
5.3. Reweighted RIF-OLS decomposition: Static approach	135
6. Conclusions.....	143
References.....	146
Appendix.....	149
Conclusiones	164
Conclusions	172

Introducción

El reciente aumento de la desigualdad de ingresos y de la privación material en un número considerable de países ha convertido su estudio en una de las prioridades del análisis económico. Los objetivos de desarrollo sostenible de las Naciones Unidas para 2030, así como la estrategia UE 2020, que exigen a los países mejoras sustanciales en ambos ámbitos, permiten confirmarlo.

Esta tesis doctoral se plantea tres grandes objetivos para abordar esta problemática en profundidad. En primer lugar, tratamos de determinar la importancia del territorio como factor explicativo de las tendencias recientes de la desigualdad de ingresos en los principales países de la OCDE. En otras palabras, intentamos identificar si esta dimensión continúa siendo determinante en la actualidad y examinamos cómo ha evolucionado su contribución a la desigualdad desde principios del siglo XXI. Aunque algunos estudios recientes otorgan al territorio un papel relevante como factor explicativo de la desigualdad de ingresos, solo unos pocos autores, sin embargo, han profundizado en el análisis de la influencia de los cambios territoriales en el proceso distributivo.

En segundo lugar, centramos la atención en las consecuencias derivadas de los shocks de desempleo sobre la privación material y tratamos de cuantificar este impacto en los países de la Unión Europea. El conocimiento sobre cómo cambian los índices de privación a medida que varía el ciclo económico es todavía muy limitado. Para cuantificar este efecto implementamos la metodología de los controles sintéticos, una novedosa técnica de evaluación de impacto que apenas ha sido utilizada en los análisis empíricos sobre pobreza y desigualdad de ingresos.

En tercer lugar, pretendemos determinar y calcular, para un conjunto de países de la OCDE, la contribución individual de los principales factores determinantes de las diferencias de desigualdad, tanto dentro de las *grandes ciudades* como entre las *grandes ciudades* y el resto de áreas o territorios. Para ello, aplicamos dos enfoques: un enfoque

dinámico, para identificar los factores que explican el aumento de la desigualdad de ingresos en las *grandes ciudades* durante las últimas dos décadas, y un enfoque estático, que permite analizar las diferencias en la desigualdad de ingresos en las *grandes ciudades* y en el resto de territorios.

En los tres capítulos que componen esta tesis doctoral se desarrollan diversas técnicas y ejercicios de descomposición con el propósito de dar respuesta a los tres principales objetivos mencionados, aportando evidencia empírica. La implementación de una amplia variedad de ejercicios contrafactuales en cada uno de los capítulos constituye el principal vínculo entre los mismos y representa el eje vertebrador de esta investigación.

El Capítulo 1 se centra en la dimensión espacial de la desigualdad y en el impacto de los cambios en la distribución territorial de la población sobre la desigualdad de ingresos, midiendo su relevancia. La contribución de dinámicas alternativas a los cambios en el mercado de trabajo y en el sistema de impuestos y prestaciones, que resultó ser importante en períodos precedentes para explicar la evolución de la desigualdad, ha sido descuidada por la literatura reciente. En la práctica, sin embargo, la movilidad de la población y la dimensión espacial de la desigualdad cobran un significado especial, principalmente cuando se alinean con las tensiones políticas (Kanbur y Venables, 2005).

Por otro lado, los recientes procesos de convergencia entre países y de divergencia entre regiones parecen encajar bien con los nuevos modelos de geografía económica que introducen rendimientos crecientes y otros fallos de mercado (Fujita y Krugman, 2004). Estos modelos suscitan diferentes respuestas de política para prevenir la aglomeración en el núcleo o compensar a la periferia (Commendatore et al., 2018), siendo la movilidad factorial un elemento esencial del análisis. Más específicamente, de la concentración geográfica de la actividad económica pueden resultar cambios en la desigualdad interregional dependiendo, en buena medida, de la movilidad de la población (Puga,

2002). Por tanto, un tema central es la relación entre las fluctuaciones en la distribución territorial de la población y los cambios en la desigualdad de ingresos. Apenas existen estudios recientes que consideren esta dimensión como un determinante de la desigualdad (Martino y Perugini, 2008), y solo algunos trabajos se centran en cambios en la demografía territorial (Dickey, 2014; Carrillo y Rothbaum, 2016).

Este capítulo realiza tres contribuciones. Primeramente, llevamos a cabo diferentes tipos de análisis de descomposición para identificar el efecto sobre la desigualdad tanto de las diferencias de ingresos entre regiones como dentro de las regiones. En segundo lugar, proporcionamos nueva evidencia sobre un grupo más amplio de países y para un periodo más reciente que en estudios anteriores. En tercer lugar, acompañando a los planteamientos tradicionales de análisis incorporamos un ejercicio de simulación. En concreto, examinamos el impacto sobre la desigualdad que podrían haber supuesto los cambios en la distribución interterritorial de la población. Utilizando la base de datos LIS, realizamos un análisis comparativo utilizando una muestra que representa más de dos tercios de la población de la OCDE. Nuestros resultados muestran un aumento significativo y generalizado de la desigualdad de ingresos en la mayoría de los países estudiados, así como una contribución reveladora, aunque modesta, a esta tendencia de las variables territoriales. A pesar de no haber encontrado un patrón firme entre los países seleccionados, nuestras especificaciones empíricas parecen capturar gran parte del cambio observado en la desigualdad de ingresos durante el período explorado. En particular, la variable regional ejerce un efecto al que merece la pena prestar atención.

En el Capítulo 2, tratamos de responder a las siguientes preguntas: ¿Deberíamos esperar un gran aumento de las privaciones materiales y un empeoramiento de las condiciones de vida inmediatamente después de un shock de desempleo? ¿Son las

medidas de privación material tan sensibles a cambios drásticos en las condiciones macroeconómicas como las medidas de pobreza monetaria?

Uno de los mayores avances en las investigaciones sobre pobreza ha sido el desarrollo de nuevos métodos para medir la privación material. Como han demostrado diferentes autores, la posibilidad de combinar diferentes indicadores parciales en un índice que mide sintéticamente el nivel de privación puede ser más eficaz que una amplia gama de indicadores para captar la atención pública y política. De hecho, algunas instituciones han incorporado el concepto de privación material en sus indicadores de pobreza y exclusión. La Unión Europea, por ejemplo, utilizó la tasa AROPE, la proporción de la población total en riesgo de pobreza o exclusión social, como su principal indicador para monitorear el objetivo de pobreza en la Estrategia UE 2020. La medida corresponde a la suma de las personas que se encuentran en riesgo de pobreza, en situación de privación material o que viven en un hogar con muy baja intensidad laboral.

Si bien los avances en la caracterización de este fenómeno han sido considerables, la evidencia sobre sus determinantes es considerablemente menos robusta. La extensa literatura empírica sobre los efectos de los cambios en las condiciones macroeconómicas en la distribución del ingreso [Blank y Blinder (1986), Cutler y Katz (1991), Jäntti (1994), Smeeding et al. (2011), Meyer y Sullivan (2011), Ayala et al. (2017)] ha tenido mucho menos desarrollo en el caso de la privación material. Una de las razones de esta asimetría radica en la naturaleza, a priori más estática, de las medidas de privación material en relación con las de desigualdad de ingresos o pobreza monetaria. Tal y como ha reconocido gran parte de la literatura, mientras que las últimas podrían considerarse variables flujo, las primeras se pueden asemejar a variables stock. Sin embargo, este razonamiento no parece corresponderse con lo ocurrido en varios países durante la

denominada Gran Recesión. En muchos países ricos, y especialmente en Europa, los indicadores de privación crecieron notablemente.

Las razones por las que este capítulo se centra en países de la UE son variadas. Primero, aunque la mayoría de los países europeos estuvieron expuestos a cambios significativos en el desempleo, en algunos su crecimiento fue mucho más rápido y las tasas de desempleo alcanzaron sus niveles máximos. En segundo lugar, la Unión Monetaria Europea se diseñó asignando el papel de estabilización fiscal a los presupuestos nacionales con muy pocas contrapartidas comunitarias. Una política monetaria común no fue suficiente para acomodar las necesidades de todos los Estados frente a los shocks asimétricos. El hecho de que no existiera un mecanismo estabilizador común en forma de seguro de desempleo europeo hizo que las respuestas de las condiciones sociales a los shocks de desempleo fueran muy diferentes en cada país (Ábráham et al., 2018).

Para abordar esta cuestión, aplicamos la metodología de los controles sintéticos. Utilizamos este enfoque para identificar el impacto de los shocks de desempleo en el nivel de privación material y realizamos diferentes análisis de sensibilidad para testar los resultados. Como hallazgo más importante, encontramos que los shocks de desempleo tienen un efecto rápido y significativo sobre el nivel de privación material en aquellos países donde tienen lugar (Grecia y España). Esta conclusión se mantiene incluso cuando se amplía el período de análisis, se cambia el indicador de privación material o se modifica la definición de shock de desempleo.

En el Capítulo 3 la base del análisis es el crecimiento reciente de las *grandes ciudades* en los países de renta alta y su relación con la desigualdad de ingresos. Entre las diversas preguntas que plantea la creciente concentración de población en las *grandes ciudades*, los efectos potenciales sobre la desigualdad van a constituir, sin duda, un núcleo

importante de investigación en los próximos años. La evidencia sobre la relación entre la desigualdad de ingresos y el tamaño de la ciudad no es tan obvia como la de cada uno de sus determinantes. Son nociones complejas y multivariantes, y cuando se intenta vincular las medidas de desigualdad y el tamaño de la ciudad la relación no siempre resulta bien definida. Las predicciones razonables únicamente son viables cuando es posible identificar un conjunto de variables significativas que simultáneamente son causa de la desigualdad de ingresos y consecuencia del proceso de urbanización.

Según la hipótesis estándar de Kuznets, ése fue el caso de la industrialización en los países avanzados, que en un principio provocó marcados aumentos de la desigualdad de ingresos, o también puede ser el caso de la inteligencia artificial en la actualidad (Frank et al., 2019). Por otro lado, hoy en día han surgido nuevos escenarios vinculados a los procesos de desarrollo de *grandes ciudades* y muchos analistas políticos y académicos han especulado sobre sus efectos sobre la desigualdad. Sin embargo, la falta de modelos empíricos adecuados, la insuficiente comprensión de ciertas interacciones y las limitaciones de datos representan importantes barreras que limitan una medición y conocimiento apropiados del nexo entre la desigualdad y el tamaño de la ciudad.

Así, los efectos potenciales de la creciente relevancia de las grandes ciudades sobre la desigualdad plantean numerosas preguntas interesantes que consideramos oportuno abordar: ¿Es la desigualdad mayor en las grandes ciudades que en otros territorios o hábitats? ¿Han aumentado las desigualdades en las grandes ciudades con el tiempo? ¿Son los factores impulsores de la desigualdad en las grandes ciudades diferentes de los de otras áreas? Estas preguntas motivan este capítulo. Nuestra preocupación es contribuir a una mejor comprensión de la relación entre desigualdad y tamaño de la ciudad. El objetivo principal es determinar y cuantificar la contribución individual de

diferentes factores explicativos a las diferencias de desigualdad tanto dentro de las grandes ciudades como entre las grandes ciudades y el resto de áreas.

Metodológicamente, en este tercer capítulo desarrollamos un doble ejercicio. Por un lado, aplicamos un enfoque dinámico, que se centra en el aumento de la desigualdad de ingresos en las grandes ciudades durante las últimas dos décadas. Por otro lado, a través de una aproximación estática analizamos las diferencias en la desigualdad entre las *grandes ciudades* y otros territorios. A través del primer enfoque podemos responder preguntas como las siguientes: ¿Existe algún patrón común de cambios de desigualdad en las grandes ciudades entre los países analizados? ¿Qué factores explican los cambios en la distribución del ingreso en las grandes ciudades? ¿Cuál es la magnitud del efecto de cada uno de estos *drivers* potenciales? A su vez, mediante el enfoque estático, nos centramos en los datos más recientes para realizar una comparación de la desigualdad de ingresos en las *grandes ciudades* y otras áreas.

En virtud de lo que conocemos, éste es el primer intento de poner en práctica la metodología propuesta por Firpo et al. (2009, 2018) desde una doble perspectiva. Asimismo, y por lo que sabemos, este es el primer estudio que la aplica a más de dos países al mismo tiempo, siendo ésta una de las principales contribuciones del capítulo. Los dos enfoques ofrecen una visión complementaria e integral de la influencia de algunos de los principales determinantes que explican la desigualdad de ingresos en las grandes ciudades en los países elegidos.

Nuestros resultados apoyan la tesis de que la desigualdad de ingresos es mayor en las grandes ciudades que en otras áreas. Esta conclusión se mantiene cuando las relaciones se prueban con diferentes medidas de desigualdad. Además, encontramos que la desigualdad aumentó en las grandes ciudades durante las dos primeras décadas del siglo XXI. Estos cambios se explican, básicamente, por lo que llamamos “efecto estructura”.

Entre los determinantes potenciales de estas tendencias, uno de los más importantes es el nivel educativo. La edad y el tamaño del hogar también son relevantes para explicar las diferencias de desigualdad dentro de las grandes ciudades y las observadas con respecto a otros territorios.

Introduction

The recent growth in income inequality and material deprivation in a considerable number of countries has made the study of these topics one of the main issues to be examined in economic analysis. The 2030 United Nations sustainable development goals as well as the EU 2020 strategy, requiring countries to improve substantially in the reduction of both problems, confirm it.

This dissertation has three main aims regarding how to tackle this problem in depth. First, we try to determine the importance of the territory in explaining recent trends in income inequality in OECD countries. Ultimately, we try to find out whether this dimension is still important and how its contribution has evolved from the beginning of the 21st century to the present. Some recent studies assign a relevant role to the territory as a driver of income inequality. However, only few authors have delved into the influence of these territorial changes. Second, we focus on the consequences derived from unemployment shocks (in the EU) on a composite measure of material deprivation and try to quantify their effects. Despite of the fact that an intensive literature has explored the effects of the business cycle on inequality or certain forms of poverty, we still know very little about how deprivation indicators change as the economic cycle varies. To measure this, we implement the synthetic control methodology, a novel impact evaluation technique that has hardly been used in the field of poverty and income inequality. Third, we intend to determine and estimate for a selected sample of OECD countries the individual contribution of different drivers to inequality differences both within *big cities* and between *big cities* and other areas. To do this, we perform both a *dynamic approach* – focusing on the increase in income inequality in big cities over the last two decades – and a *static approach* – focusing on inequality differences between big cities and other territories.

In the three chapters that make up this dissertation various decomposition techniques and exercises are conducted. The intention is none other than responding to the three main objectives mentioned above providing empirical evidence. The application of a variety of counterfactual analyses in each of the chapters is the common thread of this research project.

In Chapter 1 we focus on the spatial dimension of inequality and the impact of the changes in the territorial distribution of population on income inequality testing its relevance. The contribution of additional dynamics to changes in the labor market and in the tax and benefit systems that were important in earlier periods in explaining the evolution of inequality has been somewhat neglected by the recent literature. Nevertheless, in practice population mobility and the spatial dimension of inequality have particular significance, especially when aligned with political tensions (Kanbur and Venables, 2005).

On the other hand, recent processes of convergence between countries and divergence between regions seem to fit well with the new models of economic geography that introduce increasing returns and other market failures (Fujita and Krugman, 2004). These models elicit different policy responses in order to prevent agglomeration in the core or to compensate the periphery (Commendatore et al., 2018), being factor mobility an essential element of the analysis. More specifically, the geographical concentration of economic activity will be followed or not by shifts in interregional inequality depending critically on population mobility (Puga, 2002). Therefore, a central issue is the relationship between fluctuations in the territorial distribution of population and changes in income inequality. There are few recent studies that consider this dimension as a driver of inequality (Martino and Perugini, 2008), and only some papers focus on changes in territorial demographics (Dickey, 2014; Carrillo and Rothbaum, 2016).

This chapter makes three contributions. First, we carry out different kinds of decomposition analyses to identify the effect on inequality of both income differences between regions and within regions. Second, we provide new evidence over a more varied group of countries and a more recent period than in previous studies. Third, we add to previous studies the simulation of the impact on inequality that the changes in the inter-territorial distribution of the population may have had. Using the Luxembourg Income Study (LIS) database, we carry out a comparative analysis using a sample that represents more than two thirds of the OECD population. Our results show a significant and generalized increase in income inequality in most of the countries studied, as well as a revealing (but minor) contribution to this trend of the territorial variables. Despite not having found a strong pattern among the selected countries, our empirical specifications do seem to capture much of the change observed in income inequality during the period explored. In particular, the regional variable exerts an effect that is worth paying attention to.

In Chapter 2, we try to answer the following questions: Should we expect a large increase in material deprivation and a worsening of living conditions right after an unemployment shock? Are material deprivation measures as sensitive to drastic changes in macroeconomic conditions as monetary poverty measures? One of the greatest advances in the research on poverty has been the development of new methods for measuring material deprivation. As different authors have shown, the possibility of combining different partial indicators into an index that synthetically measures the level of deprivation can be more effective than a wide range of indicators to capture public and political attention. Some institutions have, in fact, incorporated the concept of material deprivation into their indicators of poverty and exclusion. The European Union, for instance, used the AROPE rate – the share of the total population at risk of poverty or

social exclusion – as its main indicator for monitoring the EU 2020 Strategy poverty target. The measure corresponds to the sum of persons who are at risk of poverty, severely materially deprived or living in a household with very low work intensity.

While advances in the characterization of this phenomenon have been considerable, the evidence on its determining factors is considerably less robust. The extensive empirical literature on the effects of changes in macroeconomic conditions on income distribution [Blank and Blinder (1986), Cutler and Katz (1991), Jäntti (1994), Smeeding et al. (2011), Meyer and Sullivan (2011), Ayala et al. (2017)] has had much less development in the case of material deprivation. One of the reasons for this asymmetry lies on the a priori more static nature of material deprivation measures relative to those of income inequality or monetary poverty. As the extensive literature on capabilities has recognized, while the latter could be considered flow variables, the former are more similar to stock variables. However, this reasoning does not seem to correspond well with what happened in several countries during the so-called Great Recession. In many rich countries and especially in Europe, deprivation indicators grew remarkably.

The reasons for focusing on EU countries are varied. First, while most European countries were exposed to significant unemployment changes, in some its growth was much faster and unemployment rates reached their highs. Second, the European Monetary Union was designed by assigning the role of fiscal stabilization to national budgets with very few community counterparts. A common monetary policy was not enough to accommodate the needs of all states against asymmetric shocks. The fact that there was no common stabilizing mechanism in the form of a European unemployment insurance made the responses of social conditions to unemployment shocks very different in each country (Ábráham et al., 2018).

To address this question, we apply the synthetic control methodology. We use this approach to identify the impact of unemployment shocks on material deprivation and conduct different sensitivity analyses to test the results. As our most important factual finding, we find that unemployment shocks have a rapid and significant effect on material deprivation in countries where they take place (Greece and Spain). This conclusion holds even when extending the period of analysis, changing the indicator of material deprivation or modifying the definition of unemployment shock.

In Chapter 3, we focus on the recent growth of *big cities* all over the world and study its connection with income inequality in a set of OECD countries. Among the various questions raised by this growing concentration of population in big cities, the potential effects on inequality will without doubt be a major focus of policy research for years to come. The evidence on the relationship between income inequality and city size it is not as obvious as that of each one of its drivers. They are complex and multivariate notions and when one tries to link inequality measures and city size, the relationship is not always well defined. Reasonable predictions are only feasible when it is possible to identify a set of significant variables which simultaneously are a cause of income inequality and a consequence of urbanization.

According to the standard Kuznets hypothesis, that was the case of industrialization in advanced countries, initially causing marked increases of income inequality, or it may be also the case of artificial intelligence nowadays (Frank et al., 2019). New scenarios linked to the development processes of big cities have emerged and many in the policy and research communities have speculated about their effects on inequality. However, the lack of empirical informed models, the insufficient understanding of certain interactions, and data constraints are important barriers limiting an adequate measurement and comprehension of the nexus between inequality and city size.

The potential effects of the growing relevance of big cities on inequality raise numerous interesting questions that must be addressed: Is inequality greater in big cities than in other territories or habitats? Have inequalities in big cities increased over time? Do inequality drivers in big cities differ from those in other areas? These questions motivate this chapter. Our concern is to contribute to the understanding of the relationship between inequality and city size. With this aim, we focus on big cities in a selected sample of OECD countries. The main goal is determining and quantifying the individual contribution of different explanatory factors to inequality differences both within big cities and between big cities and other areas.

We apply both a dynamic approach – focusing on the rise in income inequality in big cities over the last two decades – and a static approach – emphasizing inequality differences between big cities and other territories. By means of the first approach, we can answer questions such as: Is there any common pattern in inequality changes in big cities among the countries analyzed? Which factors account for the changes in the distribution of income in big cities? What is the magnitude of the effect of each one of these potential drivers? By means of the static approach, we focus on the most recent data to carry out a comparison of income inequality in big cities and other areas.

From what we know, this is the first attempt to put into practice this methodology from this double perspective. Also, and to the best of our knowledge, this is the first study that applies it to more than two countries at the same time, being this one of the contributions of the chapter. The two approaches bring about a complementary and comprehensive view of the influence of some of the main drivers explaining income inequality in big cities in the chosen countries.

Our findings lend support to the thesis that income inequality is higher in big cities than in other areas. This conclusion holds when the relationships are tested with different

inequality measures. We also find that inequality increased in big cities during the first two decades of the 21st century. These changes are explained, basically, by what we call as ‘structure effect’. Among the potential drivers of these trends, one of the most important is educational attainment. Age and household size are also relevant in explaining inequality differences within big cities, and those observed with respect to other territories.

Chapter 1. The Contribution of the Spatial Dimension to Inequality: A Counterfactual Analysis for OECD Countries

1. Introduction

“Reduce inequality within and among countries” is one of the 2030 United Nations sustainable development goals. Too high levels of inequality are a menace to equity and can hamper social cohesion. While inequality can influence growth positively by providing incentives for innovation and entrepreneurship and by raising saving and investment, it can also undermine progress in health and education, cause investment-reducing political and economic instability, and undercut the social consensus required to adjust in the face of shocks (Ostry et al., 2014).

The recent increase in income inequality in a large number of countries has made the study of inequality one of the main issues to be explored in economic analysis. Technical progress, changes in the labour market or the limited capacity of some tax-benefit systems to reduce inequality in market income are considered to be common drivers of the current trends in inequality. The contribution of additional dynamics that were important in earlier periods in explaining the evolution of inequality has been somewhat neglected by the recent literature. It seems that the spatial dimension of inequality and the impact of the changes in the territorial distribution of population on income inequality have received less attention than other dimensions. However, population mobility and the spatial dimension of inequality have special significance when aligned with political tensions (Kanbur and Venables, 2005).

At the global level, the between-countries component is still by far the main component of income inequality. By contrast, at the national level inequality between regions is a minor component of national inequality (Jesuit, 2003; Novotný, 2007 and Piaccentini, 2014). Neoclassical models predicted this greater convergence across regions than across countries, because frictions and mobility between-regions are lower than in the between-countries space. At the global level, however, the between-countries component has lost

some weight due to the China and India catching-up. At the national level, after a convergence period in many OECD countries (Rodríguez-Pose and Ezcurra, 2009), spatial inequalities have increased again (Lessmann, 2014).

Such recent processes of convergence between countries and divergence between regions seem to fit well with the new models of economic geography that introduce increasing returns and other market failures (Fujita and Krugman, 2004). These models elicit different policy responses in order to prevent agglomeration in the core or to compensate the periphery (Commendatore et al., 2018), being factor mobility an essential element of the analysis. More specifically, the geographical concentration of economic activity will be followed or not by shifts in interregional inequality depending critically on population mobility (Puga, 2002).

Therefore, a central issue is the relationship between fluctuations in the territorial distribution of population and changes in income inequality. There are few recent studies that consider this dimension as a driver of inequality (Martino and Perugini, 2008), and only some papers focus on changes in territorial demographics (Dickey, 2014; Carrillo and Rothbaum, 2016). This chapter aims at analysing and measuring the effect of territory on income inequality to determine whether this dimension is still important and how its contribution has evolved from the beginning of the 21st century to the present.

The chapter makes three contributions. First, we carry out different kinds of decomposition analyses to identify the effect on inequality of both income differences between regions and within regions. Second, we provide new evidence over a more varied group of countries and a more recent period than in previous studies. Third, we add to previous studies the simulation of the impact on inequality that the changes in the inter-territorial distribution of the population may have had.

Using the Luxembourg Income Study (LIS) database, we carry out a comparative analysis using a sample that represents more than two thirds of the OECD population. Our results show a significant and generalized increase in income inequality in most of the countries studied, as well as a revealing (but minor) contribution to this trend of the territorial variables. Despite not having found a strong pattern among the selected countries, our empirical specifications do seem to capture much of the change observed in income inequality during the period explored. In particular, the regional variable exerts an effect that is worth paying attention to.

The rest of the chapter is structured as follows. Section 2 provides an overview of the previous relevant literature on the issue under study. In section 3, the data used are described. In section 4, we present our methodological approach. In section 5, we present our main results, and section 6 summarizes the study and presents our conclusions.

2. Literature review

A number of studies have investigated the main sources of inequality to facilitate policy-making at the national level. For developed countries, globalization and technical progress are considered to be common drivers of the current trends in inequality, while the impact of other factors such as regulation, redistributive policies, as well as demographic changes depends more on national idiosyncrasies. In developing countries, different processes of transition also influence these tendencies.

One of the most relevant dimensions in the analysis of changes in inequality is the role that territorial differences have played in amplifying income differences. Perhaps one of the most important attempts to relate space, income and inequality are the different developments of the so called new economic geography (NEG). Since the early 1990s, NEG has provided different strands of theory in understanding the relationships between

these dimensions. The overall aim of NEG is to explain regional economic disparities on the basis of spatial agglomeration effects. The most recent NEG models have stressed path dependence and lock-ins as mechanisms to explain the persistence of spatial disparities through time (Hassing and Gong, 2019).

Under this framework, a key issue is labour mobility. NEG models relate labour migrations across regions to the geography of production through real wage differentials. The basic intuition of NEG models highlights the influence of access to markets on location choices of both firms and workers. The cumulative process of agglomeration rests on the complementarity of these two relations: agglomeration may occur only if migrants, like firms, are attracted by high market potential regions (Crozet, 2004). The concentration of manufacturing workers creates a large market, so making the location profitable for firms. And the entry of firms bids up wages, so making the location attractive for workers (Venables, 2016). However, as stressed by Garretsen and Martin (2010), some authors have questioned whether the formal economic models that are the focus of attention within NEG are can adequately capture the full range of factors and forces that help shape the economic landscape, particularly since some of these factors are social, institutional and cultural in nature.

The drivers of inequality have barely been examined at regional level and mostly for EU countries. Martino and Perugini (2008) use LIS and Eurostat data to estimate the influence of different economic, demographic and institutional variables on within-regional inequality finding a non-significant role of the only demographic variable included – the share of the population aged 65 years or over. Castells-Quintana et al. (2015) use data from ECHP and EU-SILC to estimate the relationship between within-regional inequality and regional per capita income and find a significant influence of some control variables, including population density. Mussini (2017) uses Eurostat data to decompose the

changes in between-regional inequality and finds a non-negligible role of the change of regional population weights. Finally, Dickey (2014) uses British-HPS data to estimate the impact of migration on the within-regional wage distributions finding a significant impact but with opposite signs across regions.

A number of works have also addressed some methodological issues relevant for the goals of this chapter, such as the level of aggregation, the choice of inequality measures or the spatial variation of prices. Regarding the level of spatial aggregation, we are aware that it affects the relative weights of the between and within components for a given distribution (Shorrocks and Wan, 2005). To solve this problem either the between-component can be reinterpreted to discount the effects of the number and size of regions (Elbers et al., 2008) or the classification can be redefined to homogenize the number of regions across countries (Novotný, 2007).

Regarding the choice of the inequality measure, each one implies not only a different concept of inequality but also a different rule of decomposition (Cowell, 2011). The decomposition of the Theil measures is more user friendly than that of the Gini due to their simpler structure, but such simplicity is achieved at the cost of losing useful information. The extra term added to the inter and intra components in the standard decomposition of the Gini index, accounts for the amount of overlapping between subgroups which can be interpreted in terms of stratification and other relevant concepts (Yitzhaki and Schechtman, 2013). Alternative decompositions of the Gini index also allow to analyze other concepts such as the spatial autocorrelation (Rey and Smith, 2013) that is both a nuisance for the analysis and a relevant feature of the topic at hand.

Lastly, regarding the spatial variation of prices, the correlation between price levels and living standards produces an overestimation of the between- component (Shorrocks and Wan, 2005). This is a recurring issue of concern to practically all researchers, but the lack

of data makes it difficult to measure properly. In countries that have tried to implement an alternative methodology based on relative region-specific poverty lines, such as France (Insee, 1997) or Spain (Ayala et al., 2014), it has been shown that these alternatives also have problems, such that they end up mixing the disparities in the cost of living with those related to the level of economic development (Brandolini, 2007).

From all the different possible approaches, we chose to focus on the role that changes in the regional population distribution play regarding income inequality. On the one hand, in two-thirds of OECD countries the share of population in predominantly urban regions has increased in the past 15 years (OECD, 2018). On the other hand, as anticipated by NEG theories, regional migration does not affect all regions of a country equally. Distance to labour market and services seems to explain migration within OECD countries. These flows may lead to persistent regional economic disparities also causing effects on the personal income distribution.

This chapter contributes to the existing literature by considering a sample of OECD countries that have very different administrative divisions of their national territory. The chapter also contributes to the integration of existing methods by using alternative ways to decompose inequality, which is a promising issue in the research agenda of the dynamics of regional disparities (Rey and Janikas, 2005).

3. Data

We have chosen the LIS database for three basic reasons: (1) it has a wider spatial scope than the EU-SILC, (2) it allows access to deeper content than the OECD-IDD, and (3) it allows us to identify the relative weight of the spatial dimension of income inequality and its changes over time. As far as we know, this is the first attempt to conduct such a study implementing a counterfactual analysis.

The LIS is the largest available income database of harmonized microdata and includes approximately 50 years of information on more than fifty countries all over the world. It currently gathers data from very different countries and continues to expand. LIS datasets contain household-level and person-level data on labour income, capital income, social security and private transfers, taxes and contributions, demography, employment, and expenditures. Its use and influence have been steadily increasing, although only a few papers report estimates of income inequality at the subnational level.

The period considered in this exercise is covered by the datasets available for the 21st century. We focus on the OECD area, including some of the most populous ones in 2016.¹ The selected sample represents more than two thirds of the OECD population (see Table A.1 in the appendix).

Another noteworthy issue addressed by this chapter concerns the definition and division of the territorial units used in each country. Bearing in mind that this is a somewhat arbitrary issue, we decided to proceed to use an administrative division as the main classification criterion. This proposal coincides with the sorting provided by the LIS database for the regional variable, and respects the Eurostat recommendations of using the classification closer to the framework adopted by the countries for their regional policy. It also clarifies the subsequent interpretation and justification of the results, being more comprehensible and straightforward

Our administrative criterium gives priority to institutional boundaries. In this manner, the different units used in the study are the following: 7 regions in Australia (the 6 federated states and Canberra); in Canada, the 10 administrative divisions that are responsible for

¹ According to 2016 OECD data, the eleven most populous OECD member countries were the United States, Japan, Mexico, Germany, Turkey, France, United Kingdom, Italy, South Korea, Spain and Poland. We had to exclude Japan (only had data for one year), Turkey (no data) and South Korea (no regional data). Instead, we were compelled to add the OECD countries ranked 12th, 13th, and 14th in population: Canada, Australia and Chile.

sub-national governance (the provinces); in Chile, we have data for 13 regions (all of the country's administrative units); in France the data available provide information on 8 regions; in Germany, the 16 *Länder*; in Italy, the 20 administrative regions; in Mexico, the 32 federal entities of the United Mexican States; in Poland, the 16 *voivodeships* (the largest unit of the Polish administrative political system); in Spain, the 18² *Comunidades Autónomas* (the Spanish territorial administrative entities established by the country's Constitution that are endowed with a certain legislative autonomy); in the United Kingdom, we have considered 12 divisions (the 9 regions of England, also known as *Government Office Regions*, Scotland, Wales and Northern Ireland); in the United States 51 territories were analysed, i.e., the states that share sovereignty with the federal government.

Additionally, it is important to mention that another interesting alternative is the criteria proposed by Novotný (2004, 2007), Hoffmeister (2009), and others. Their fundamental premise is to make comparisons among entities with a similar number of divisions, making relevant groupings according to a specific convention when needed. This procedure presents a clear advantage: it allows us to eliminate the discrepancies and biases that would otherwise occur if we calculate the *between* and *within* components in an inequality decomposition by population subgroups and if the units to be examined were not of a similar size. However, there is a major disadvantage in considering divisions of the same size: to assume that the productive structures of different regions are the same regardless of their dimension is a very restrictive assumption. Nonetheless, to test the sensitivity of our results to the criterion for aggregation, we also re-estimated all the calculations following the recommended guidelines of these authors.³ The new estimates

² The 17 Spanish *Comunidades Autónomas* and Ceuta and Melilla considered as a single unit.

³ Since the LIS database does not allow a lower level of disaggregation, we can only test the possible effects of a higher aggregation level. More precisely, in the case of the two EU countries with more regions we move from NUTS2 to NUTS 1 level (in Italy the change is from 20 to 5 regions, and in Spain from 18 to 7

can be tested in Table A.2 of the Appendix. In general terms, regional aggregation slightly reduces the contribution to national inequality of the between-regions component, but it hardly affects the trends of that contribution (see Table A.2 in the appendix).

Regarding the variables considered in the study, the key one is the *real equivalent household disposable income*. The *disposable income* includes both primary (labour and capital/market) and secondary (tax and transfers/non-market) incomes. Following the standard criteria found in the LIS database, we choose the *equivalent household disposable income*, obtained by dividing the disposable income by the square root of the household size. It is also important to clarify two ideas regarding the handling of the data. First, negative and zero income values have been replaced with 1/100 of the mean. In this manner, we can prevent relevant observations from being dropped by default. Second, observations with missing values for the regional variable have been removed to ensure consistency with all the results presented here.

The application of the aggregation criteria and the methodological options previously described to the LIS data allow us to calculate the extent of inequality both in each country of the sample and in each of the territorial units that we have defined. Table A.3 provides a thumbnail sketch of the general picture of inequality within each region. The Table gives general support to the notion of a very wide range in the inequality indicators within each region in all countries, with large differences between the highest and the lowest values of the mean logarithmic deviation (see Table A.3 in the appendix)..

regions). In the two non-EU countries with more regions, we take into consideration the regional division used by the United States Census Bureau (grouping the states into 9 divisions), and we group the Mexican states into 8 regions or conglomerations.

4. Empirical strategy

The methodology used follows the proposal by Cowell and Fiorio (2011) to reconcile the conventional theoretical schemes and the most recent regression techniques. This integration of existing methods has been also outlined as a promising issue in the research agenda of the dynamics of regional inequalities (Rey and Janikas, 2005). First, we carry out a standard subgroup decomposition to identify the corresponding weights of the within and between regional components. Second, we develop a dynamic decomposition to identify the weight of population changes. Finally, we apply counterfactual analyses to estimate the effect of these population changes on income inequality.

Among the many ways to quantify inequality in the distribution of income, the two classical measures are the coefficient of variation (C) and the Gini index (G). They can be expressed in terms of the ratios (λ_i) between income (q_i) and population (p_i) shares of the $i=1 \dots n$ receivers:

$$0 \leq C^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{q_i}{p_i} - 1 \right)^2 = \sum_{i=1}^n \frac{1}{n} (\lambda_i - 1)^2 \leq n - 1 \quad (1)$$

$$0 \leq G = \frac{1}{n} \sum_{i=1}^n \frac{i}{n} \left(\frac{q_i}{p_i} - 1 \right) = \sum_{i=1}^n \frac{2i}{n^2} (\lambda_i - 1) \leq \frac{n-1}{n}, \quad q_i \leq q_{i+1} \quad (2)$$

A third commonly used measure is the Theil index (T), which can be derived from the mean logarithmic deviation (L) by exchanging the population and income shares:

$$0 \leq T = \sum_{i=1}^n q_i \log(\lambda_i) \leq \log n \quad (3)$$

$$0 \leq L = \sum_{i=1}^n p_i \log(1/\lambda_i) \leq \infty \quad (4)$$

Cowell reported on other measures of inequality and extensively discussed a measure initially called the ‘generalized information measure’, which was later renamed

generalized entropy measure (E_θ) following modifications to allow the fulfilment of additional properties (Cowell, 2011):

$$0 \leq E_\theta = (\sum_{i=1}^n p_i \lambda_i^\theta - 1) / (\theta^2 - \theta) \leq \infty \quad (5)$$

The members of the entropy family with a parameter $\theta < 2$ concentrate their transfer sensitivity more at the lower end the further away parameter θ is from 2, while those with a $\theta > 2$ exhibit a transfer sensitivity more towards the top of the distribution the greater the parameter θ value. When $\theta = 2$, the transfer sensitivity increases symmetrically at two tails towards both sides.

In our analysis, and in order to determine the explanatory power of the territorial variable in the recent evolution of inequality, we have chosen the mean log deviation as the index to be examined. The main reasons that support this decision are presented in the following section.

4.1. Inequality decomposition by population subgroups

The analysis of regional inequality is mainly related to the decomposition by groups. The standard procedure to implement such a decomposition consists of defining the inequality between-groups (B) as that which remains after removing the within-groups inequality (W) by replacing individual incomes with their group mean. The W component is computed later by subtracting the B component from total inequality.

For the generalized entropy family, the weights are a first-order homogeneous function of the population and income shares of the groups: $w_g = p_g \lambda_g^\theta = p_g^{1-\theta} q_g^\theta$.

$$E_\theta = B E_\theta + W E_\theta = (\theta^2 - \theta)^{-1} [\sum_{g=1}^k w_g - 1] + \sum_g w_g E_{\theta g} \quad (6)$$

The weights w_g sum to unity when $\theta=0$ (the income weighted T) or $\theta=1$ (the population weighted L):

$$T = BT + WT = \sum_g q_g \log(\lambda_g) + \sum_g q_g T_g \quad (7)$$

$$L = BL + WL = \sum_g p_g \log(1/\lambda_g) + \sum_g p_g L_g \quad (8)$$

In the case of the squared coefficient of variation, the weights sum to unity if all subgroup distributions have the same mean:

$$C^2 = BC^2 + WC^2 = \sum_g p_g (\lambda_g - 1)^2 + \sum_g p_g \lambda_g^2 C_g^2 \quad (9)$$

The mean logarithmic deviation (L) is the only member of the entropy family that produces the same results with both approaches and generates assignments similar to the Shapley decomposition (Shorrocks, 2013). This “path independent” property may explain the preference for this measure in recent empirical work in this area and is also one the motives for choosing this index in the decompositions. Additionally, due to its simplicity, the decomposition of the mean log deviation (L) has been the most widely used in the literature. The results for other indices are not as easily interpreted (Shorrocks and Wan, 2005).

4.2. Dynamic decomposition

This decomposition allows observing the changes produced for a given period instead of for each specific year. This helps us know the importance of each component of the decomposition in explaining the general evolution of the index. The dynamic decomposition was initially proposed by Mookherjee and Shorrocks (1982) and, given its greater simplicity, this decomposition has been developed primarily for L . It can be described as follows:

$$\begin{aligned} \Delta L &= \Delta \left[\sum_g p_g L_g - \sum_g p_g \lambda_g \right] = \sum_g \bar{p}_g \Delta L_g + \sum_g \bar{L}_g \Delta p_g - \sum_g \bar{\lambda}_g \Delta p_g - \sum_g \bar{p}_g \Delta \lambda_g \\ &\simeq \sum_g \bar{p}_g \Delta L_g + \sum_g \bar{L}_g \Delta p_g + \sum_g \left[\overline{\left(\frac{\mu_g}{\mu} \right)} - \overline{\log \left(\frac{\mu_g}{\mu} \right)} \right] \Delta p_g + \sum_g (\bar{q}_g - \bar{p}_g) \Delta \log \mu_g \quad (10) \end{aligned}$$

where Δ shows the variation in the variables of interest from the initial year (t_0) to the final year (t_1). Following those authors, we can express the four terms of expression (10) as follows:

$$\sum_g \bar{p}_g \Delta L_g \quad (10a)$$

$$\sum_g \bar{L}_g \Delta p_g \quad (10b)$$

$$\sum_g \left[\overline{\left(\frac{\mu_g}{\mu} \right)} - \overline{\log \left(\frac{\mu_g}{\mu} \right)} \right] \Delta p_g \quad (10c)$$

$$\sum_g (\bar{q}_g - \bar{p}_g) \Delta \log \mu_g \quad (10d)$$

The first term (10a) denotes the changes in within-subgroup⁴ inequality; the second (10b) reflects the variations in the population shares of the “within group” component; expression (10c) reveals the same as the previous one, but for the case of “between-group” inequality; and the last expression (10d) displays the effect of changes attributable to differences in relative incomes for the groups of interest. This decomposition allows us to recognize the influence of each one of these four elements on the *trend* in aggregate inequality⁵ as well as link this section with the following one, where the contribution of the territorial dimension is presented through the methodology of counterfactual analysis.

⁴ In our case, the different groups are the territorial units for each country selected from the LIS database.

⁵ See Table 2.

Once again, the exercise can be understood in terms of the aforementioned counterfactual analysis. In the context of this chapter, by taking regions as groups, the sum of the second and the third terms can be interpreted as the change in inequality that would have occurred if the relative incomes *within* and *between* regions had not changed.

4.3. Counterfactual analysis

A growing number of scholars are investigating the drivers of inequality using alternative decompositions to those previously reviewed. Among the methods that go beyond the mean, the *reweighting* approach has been one of the most applied in practice and was first introduced in the decomposition literature by DiNardo, Fortin and Lemieux (DFL) (1996). DFL decomposition is a semi-parametric approach that enables us to extend the results to the whole distribution of income, since it works with the entire population; not only with the mean. Although this is not the first time that the DFL methodology is being applied to the analysis of income inequality focusing on the spatial dimension (Dickey, 2014; Carrillo and Rothbaum, 2016), as far as we know, an analysis similar to the one we describe below has not yet been conducted. We focus on the particular contribution of territory to income inequality not only from a semi-parametric perspective, but linking it to the traditional and more theoretical proposals.

The non-parametric estimation of density functions by means of kernel methods is a subtle and complementary approach to use when the model followed by our data is unknown. It consists of building a function based on the sample values. If we have several samples, one for each different population, and we do not know the function they describe, we could create a density function for each sample by classifying the new individuals through a simple assignment to the population that has more values. In particular, kernel density functions allow us to estimate the counterfactual density

distributions that we intend to study by putting into practice the DFL methodology. They can be defined as follows:

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{t - X_i}{h_n}\right) \quad (11)$$

where n refers to the number of observations, K is a predetermined density named kernel, and h_n is a chain of smoothing parameters (*bandwidths*) that must slowly tend to zero.

DFL approach

We propose to analyse a selected sample of OECD countries to determine what would have happened to the distribution of the equivalent disposable household income in the final period ($t = 1$) if the territorial distribution of the population had remained constant and equal to that of the analysis starting point ($t = 0$). The individual observations of the income y , a vector of individual attributes x , and a date t belong to a joint distribution $F(y, x, t)$ that, at a given date, becomes the conditional distribution $F(y, x|t)$. Thus, at date $t = 1$, the actual density of incomes can be written as:

$$f(y; t_y = 1, t_x = 1) = \int_x f(y|x, t_y = 1) dF(x|t_x = 1) \quad (12)$$

We handled eight representative attributes of the households of interest in addition to the regional variable: one housing variable (owned/rented); one variable related to household composition and living arrangements (number of household members); five socio-demographic characteristics (age, marital status, immigration, health, and education); and one variable reporting on labour market activity (employment). The main reasons for choosing these variables and not others are the large number of countries selected and the different LIS waves covered by the analysis⁶ (see Table A.4 in the appendix).

⁶ Data are lacking for some periods in certain countries.

Assuming independence between the income structures and the distribution of attributes, the following represents the hypothetical or counterfactual density of incomes that would have prevailed if the distribution of attributes had remained the same as on the initial date:

$$\begin{aligned}
 f(y|t_y = 1, t_x = 0) &= \int f(y|x, t_y = 1)dF(x|t_x = 0) \\
 &= \int f(y|x, t_y = 1)\Psi(x)dF(x|t_x = 1)
 \end{aligned}
 \tag{13}$$

The counterfactual distribution (13) is similar to the actual one (12), except that it introduces a "reweighting" function:

$$\Psi(x) = dF(x|t_x = 0)/dF(x|t_x = 1)
 \tag{14}$$

Estimating Ψ is not straightforward, but DFL solves the implementation problem with the application of Bayes' rule in order to obtain:

$$\widehat{\Psi}(x) = \frac{Pr(t_x=0|x) Pr(t_x=1)}{Pr(t_x=1|x) Pr(t_x=0)}
 \tag{15}$$

Unlike (12), equation (13) can be readily estimated by first pooling the individual observations from the two dates and then estimating a probit model for the likelihood that an observation is from date t given x . The estimates allow us to determine $\widehat{\Psi}(x)$ for each observation, which can be used to obtain the counterfactual density through weighted kernel methods.

In expression (15), $Pr(t_x = 0|x)$ represents the probability that a randomly chosen individual with characteristics x (variables we have considered relevant in our analysis) belongs to the starting year when all the individuals in the sample are pooled together. $Pr(t_x = 1|x)$ would reflect the same idea, but for the second period.

Once the pool of data is created, we have to estimate two probit models. The first model is an estimation considering all the attributes of interest, while the second would include

all the explanatory variables of the previous estimation, except for the territory variable. The contribution of the territorial variable to income inequality is determined by the difference between the two counterfactual distributions generated.

5. Main results

The first analysis carried out is descriptive and allows us to illustrate the heterogeneity existing in the territorial units under examination. Table A.3 includes, for all the territories studied, the initial and final values of the mean logarithmic deviation (L), the population shares, and the number of observations used. As a complement, Figure A.1 provides an overview of data distribution through some of the most relevant measures of position and dispersion. These figures make it easy to check the symmetry in each particular case and to identify the presence of outliers (see Figure A.1 in the appendix).

5.1. Between and within regional inequality

In this subsection, we present a set of results corresponding to the standard decomposition by population subgroups. This analysis (see Table 1) provides us a first picture of income inequality in two specific moments in time (2000 and 2016 or nearest year available), and the figures are presented in both absolute and relative terms according to the original regional groupings provided by the LIS database. The general trend reveals a significant increase in total inequality in most countries, as well as a relevant growth for the within-region component in almost all of them. Equally striking from a general perspective is the process of convergence that took place during the first years of the 21st century.

Table 1. Spatial decomposition of the Mean Log Deviation (*L*) in the OECD: 1999-2016

Country	Units	Year	Total inequality (<i>L</i>)	Between (<i>B</i>)	Within (<i>W</i>)		
<i>Australia</i> *	7	2001	0.20674	0.00255	1.2%	0.20419	98.8%
	7	2014	0.20977	0.00318	1.5%	0.20660	98.5%
<i>Canada</i> ***	10	2000	0.17509	0.00497	2.8%	0.17012	97.2%
	10	2013	0.19453	0.00587	3.0%	0.18866	97.0%
<i>Chile</i> ⁷ ***	13	2000	0.51730	0.03214	6.2%	0.48516	93.8%
	13	2015	0.41412	0.02398	5.8%	0.39015	94.2%
<i>France</i> ⁸ **	8	2000	0.13408	0.00821	6.1%	0.12587	93.9%
	8	2010	0.15161	0.00381	2.5%	0.14780	97.5%
<i>Germany</i> ***	16	2000	0.11965	0.00316	2.6%	0.11649	97.4%
	16	2015	0.15813	0.00402	2.5%	0.15411	97.5%
<i>Italy</i> **	20	2000	0.22779	0.03144	13.8%	0.19635	86.2%
	20	2014	0.26626	0.02342	8.8%	0.24284	91.2%
<i>Mexico</i>	32	2000	0.45041	0.06910	15.3%	0.38131	84.7%
	32	2012	0.42132	0.04450	10.6%	0.37682	89.4%
<i>Poland</i>	16	1999	0.16887	0.00336	2.0%	0.16551	98.0%
	16	2016	0.17442	0.00343	2.0%	0.17099	98.0%
<i>Spain</i> ⁹ ***	18	2004	0.21216	0.01007	4.7%	0.20209	95.3%
	18	2013	0.24624	0.01713	7.0%	0.22911	93.0%
<i>United Kingdom</i> **	12	2004	0.23035	0.00910	4.0%	0.22124	96.0%
	12	2013	0.20528	0.00666	3.2%	0.19863	96.8%
<i>United States</i> ***	51	2000	0.27129	0.00458	1.7%	0.26672	98.3%
	51	2016	0.29662	0.00398	1.3%	0.29263	98.7%

Note: Asterisks indicate that initial and final values are significantly different at the 90 (*), 95 (**), and 99 (***) % confidence levels.

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

⁷ Arica y Parinacota and Tarapaca have been considered as a single region (the data appear disaggregated in 2015, but not in 2000). For the same reason, Los Lagos and Los Ríos have been analysed as a single territory.

⁸ The data corresponding to the Départements d'Outre-Mer (Guadeloupe, Martinique, French Guiana, Réunion and Mayotte) are only available for 2010 and have been deleted. In this way, we can perform a homogeneous analysis between the 8 territories that correspond to the NUTS-1 level.

⁹ The Spanish constitutional legal system divides the country into 17 regions and 2 *Autonomous Cities* (Ceuta and Melilla). For the purposes of this exercise, and because of the disaggregation of the region variable in the LIS database, Ceuta and Melilla have been considered as a single entity.

This fact is supported by the main figures of the population subgroups decompositions and by reading-through the actual density functions¹⁰ corresponding to the first and last years of interest.

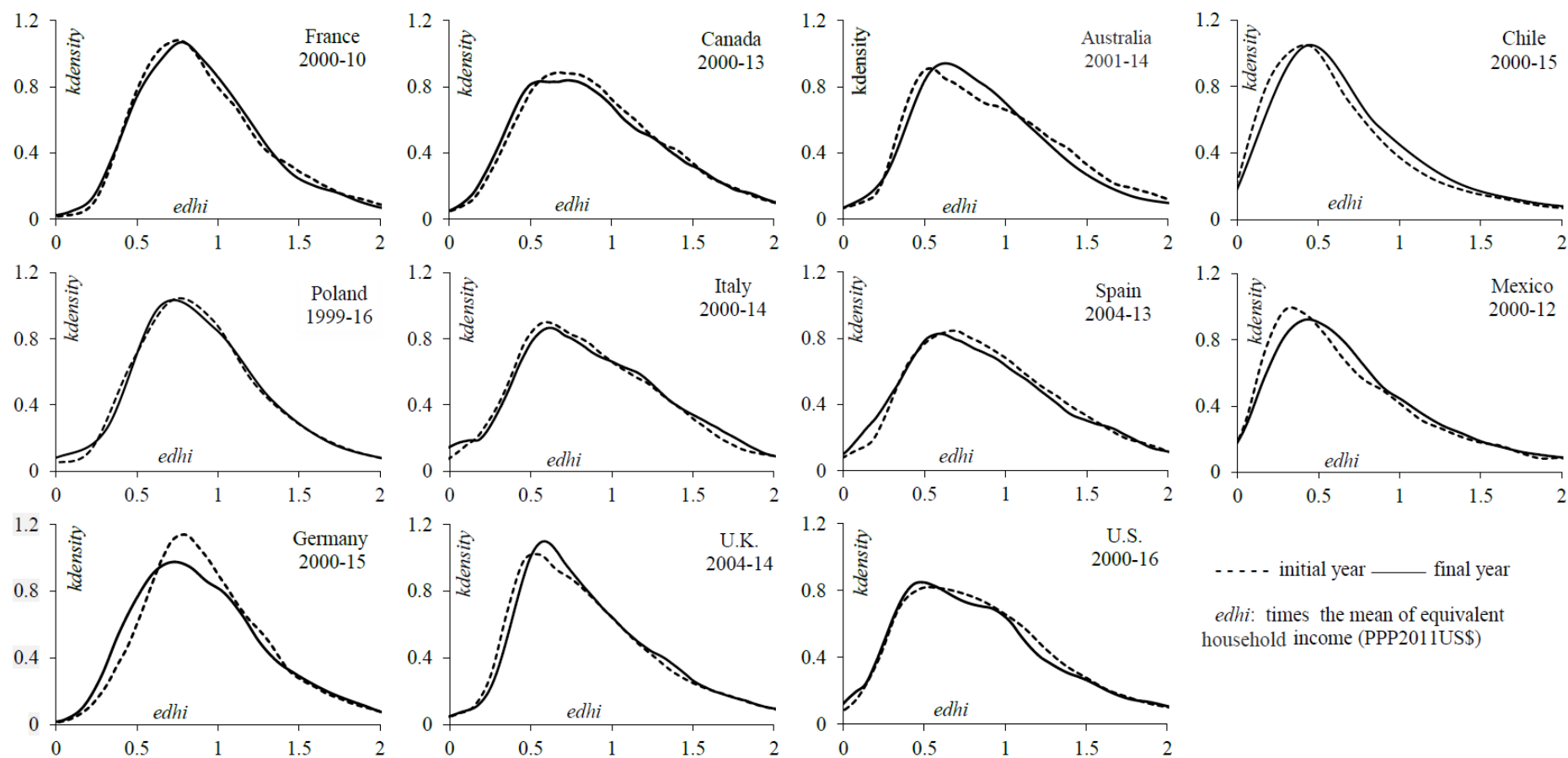
Figure 1 allows us to clearly see that the 2015 distribution in Germany is flatter and slightly to the left of the 2000 distribution. Conversely, the comparison for Chile between the two actual distributions provides very clear evidence of the decrease in income inequality from 2000-2015.

Regarding the specific countries examined, it is necessary to provide some clarifications. Chile experienced the greatest reduction in income inequality. This fall in income inequality since the beginning of the 21st century is explained, on one hand, by the implementation of three inclusive policies that benefited the most precarious households: *Chile solidario*, the gradual health reform called AUGE, and a noteworthy Social Security reform (Contreras and Ffrench-Davis, 2014). On the other hand, Parro and Reyes (2017) attributed it to the link between the factors that determine economic growth, and the focus on education that prompted individuals to invest in higher education.

Mexico has also experienced a remarkable decrease in inequality. The main drivers for this reduction during the last decade seem to be the labour market forces (a remarkable decrease in the wage ratio between skilled and unskilled workers), and institutional factors (minimum wage and unionization rate agreements) (Esquivel et al., 2010). Another noteworthy fact pertaining to Mexico is that it had the greatest weight of the *between* component among the eleven countries analysed.

¹⁰ In these graphs, the outcome variable has been relativized to the median to obtain a straightforward interpretation of the results.

Figure 1. Initial and final actual density functions in the most populated OECD countries: 1999-2016



Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

In the European context, the United Kingdom deserves special mention, as it is the main exception among European countries regarding the evolution of the mean log deviation from 2004 to 2013. In Germany, the results are in accordance with the ones obtained by Biewen and Juhasz (2012) for an earlier period (2000-2006), who attributed the increase in income inequality to changes in the tax system, restructuring of household organization and to other variations in some highly important socioeconomic characteristics (such as age or education). The results observed in Italy could find justification in the works of Torrisi et al. (2015), who mainly attributed the reduction of regional disparities between the northern regions and the southern ones to changes in population dynamics, and Usseglio (2016), whose main findings detected that human capital has not played such an important role as employment opportunities. In a similar way, Tirado et al. (2016) confirmed the existence of great disparities between the most prosperous Spanish regions, those of the North-east, and the poorest territories, located in the South.¹¹ As regards France, Combes et al. (2011) underscored the recent convergence across their territories in labor productivity and drew special attention to the influence of the agglomeration economies. To conclude with the European scanning, Czyż and Hauke (2011) observed that the development-activating elements have not managed to reduce inter-regional differences in Poland and could explain the increase we see in our results.

Concerning Canada, some factors driving the recent evolution of family inequality *within* Canadian provinces could be the roles played by human capital and the life cycle (inter-temporal dynamics) (Gray et al., 2004).

In Australia, the discrepancies noticed around the between-region and within-region component do seem to lay on the effects of the ‘mining boom decade’ of 2001–2011

¹¹ They highlight that “as NEG literature stresses, the presence of economic activities characterized by the emergence of economies of scale makes accessibility to the nodes act as a catalyst in boosting regional development”.

(Fleming and Measham, 2015).¹² Miranti et al. (2013) revealed, on the other hand, that income inequality varied noticeably when focusing on small spatial areas and justified the relevance of exploring territorial characteristics for a better understanding of both *between* and *within* regional complexity.

Lastly, in the United States, spatial-specific income dynamics characterized by segmented income classes of neighbours, among others, could be the reason for the increase in income inequality from 2000 to 2016. On the opposite side, there is some evidence indicating that innovation is not one of the main drivers in explaining increasing income inequality in the United States (Lee and Rodríguez-Pose, 2013).¹³

In general, the results obtained are consistent with the evidence reported in much of the literature, and they corroborate that inequality of disposable income not only differs between countries but also between regions within countries (Jesuit, 2003). In the same manner, the results lead us to confirm some stylized facts, including the very high weight of the within-region component (Piacentini, 2014).

These results are also confirmed when a spatial perspective is added to the distributive analysis taking stratification into account as well. In Table A.5 of the Appendix we present an analysis of the Gini index (ANOI) that jointly accounts for the contribution of inter- and intra- regional inequality to total inequality and the contribution of the overlapping of both components' regional distributions (Frick et al., 2006). Perfect stratification (the inverse of overlapping) occurs when the incomes of each region belong to a specific range and the ranges of the regions do not overlap (see Yitzhaky and

¹² Athanasopoulos and Vahid (2003) had previously confirmed the predominance of the within-region component in accounting for total income inequality. They also pointed out that this country's increase in income inequality during the 1990s had been more pronounced around major metropolitan areas.

¹³ The relationship between innovation and income inequality is another interesting variant in the study of spatial inequality. These authors found a strong link in the case of European regions. However, their findings only provided inconclusive results for the territories of the United States. Differences in labor market flexibility as well as dissimilarities in the levels of migration look to be behind these discrepancies.

Schetchetman, 2013). The measure employed here is $I=Gb/Gbp$, where G_b and G_{pb} are the between group component of the Gini decomposition in Yitzhaki (1994) and Pyatt (1967), respectively¹⁴ (see Table A.5 in the appendix).

The last column presents the stratification index, which increases as inter-regional inequality rises and also when intra-regional inequality falls. The two more stratified countries in our sample are Italy and Mexico, followed by Chile and Spain. During the analysis period, France registered a significant improvement due to the fall in interregional inequality, enhanced by an increase in intraregional inequalities. In Spain stratification worsened due to the rise of interregional disparities barely compensated by the increase in inequality between regions.

Explaining the varied country specific causes for this change is beyond the scope of this chapter, as it would require bearing in mind not only the common drivers of the evolution of regional disparities, but also a variety of idiosyncratic factors. Among the most relevant ones, the dominant role played by some big cities like Paris (Dormard, 2004), the isolation of some of the poorest regions of East-Germany (Frick and Goebel, 2008) or the mining boom of 2000`s in Australia (Fleming and Measham, 2015) stand out. The evidence shown in this section not only makes updated results on the inter- and intra-regional components available, but also provides information on regional cohesion measured through the overlapping degree.

5.2. Decomposition of the *trend* of inequality

The static decomposition offers us a “picture” of inequality at one specific moment. This second decomposition will provide us with an idea of the evolution of the index. More

¹⁴ It is identified in Milanovic and Yitzhaki (2002) as the loss of between group inequality due to overlapping and latter reinterpreted by Monti and Santoro (2011) for the two groups case, and by Allanson (2014) for the general case, as an index of non-overlapping or stratification.

importantly, it allows us to know and quantify the contribution of territory to income inequality through two of its four components.

This method is based on the mean log deviation (L) because of the difficulties involved in implementing it with other inequality indices. It breaks down income inequality into two blocks that also correspond to the four identities listed in equation (10). The extensive and generalizable increase in aggregate inequality observed can be disaggregated into these two blocks, as detailed below (see Table 2).

One effect would be due to inequality within the regions and corresponds entirely to component [a]. It gathers the effect of the changes in the intra-group component (*pure within effect*) and is the most important component in our decomposition. However, even more significant is the fact that this first component is greater than the observed change in L for many countries, being especially notable in Italy. Therefore, variations in relative means, or group compositions, can hardly account for the rise in inequality during the period of reference.

The second effect refers to inequality between the regions and can be divided into the *income effect* and the *allocation effect*. The *income effect* (component [d]) denotes the effect on the variation of the relative income of the regions considered. Also noteworthy is its negative contribution (in most of the cases), which represents the changes attributable to shifting relative income between groups. This contribution is positive in two of the three countries where inequality grew the most (Germany and Spain) and in Australia and Canada.

The *allocation effect* (components [b] and [c]) incorporates modifications attributable to changing numbers in the different regions (variations in their populations). Components [b] and [c] acquire special relevance concerning the question of population shares.

Table 2. Decomposition of the trend in aggregate inequality ($L \times 10^2$): 1999-2016

Countries and time	Change in aggregate inequality $\Delta L = L_f - L_i$ = [11a] + [11b] + [11c] + [11d]	Contribution to ΔL attributable to variations in ¹⁵							
		Within group inequality		Population shares				Mean group incomes	
		[11a]	[11b]	[11c]		[11d]			
	$\Delta L = L_f - L_i$	$\sum_g \bar{p}_g \Delta L_g$	$\sum_g \bar{L}_g \Delta p_g$	$\sum_g \left[\left(\frac{\mu_g}{\mu} \right) - \log \left(\frac{\mu_g}{\mu} \right) \right] \Delta p_g$		$\sum_g (\bar{q}_g - \bar{p}_g) \Delta \log \mu_g$			
<i>Australia</i> 2001-2014	0.303	0.224 (71.7%)	0.017 (5.3%)	0.000 (0.0%)	0.072 (23.0%)				
<i>Canada</i> 2000-2013	1.945	1.809 (92.1%)	0.046 (2.3%)	0.010 (0.5%)	0.099 (5.0%)				
<i>Chile</i> 2000-2015	- 10.317	- 9.479 (91.8%)	- 0.022 (0.2%)	- 0.017 (0.2%)	- 0.810 (7.8%)				
<i>France</i> 2000-2010	1.753	2.141 (122.2%)	0.051 (2.9%)	- 0.007 (- 0.4%)	- 0.433 (- 24.7%)				
<i>Germany</i> 2000-2015	3.848	3.690 (96.0%)	0.072 (1.9%)	- 0.019 (- 0.5%)	0.102 (2.6%)				
<i>Italy</i> 2000-2014	3.847	5.567 (144.7%)	- 0.919 (- 23.9%)	- 0.123 (- 3.2%)	- 0.678 (- 17.6%)				
<i>Mexico</i> 2000-2012	- 2.909	- 0.329 (11.3%)	- 0.120 (4.1%)	- 0.136 (4.7%)	- 2.316 (79.8%)				
<i>Poland</i> 1999-2016	0.555	0.324 (58.4%)	0.225 (40.5%)	0.019 (3.3%)	- 0.012 (- 2.2%)				
<i>Spain</i> 2004-2013	3.408	2.670 (78.4%)	0.032 (0.9%)	0.006 (0.2%)	0.699 (20.5%)				
<i>United Kingdom</i> 2004-2013	- 2.507	- 2.378 (94.9%)	0.116 (- 4.6%)	0.010 (- 0.4%)	- 0.255 (10.2%)				
<i>United States</i> 2000-2016	2.532	2.514 (99.3%)	0.078 (3.1%)	- 0.010 (- 0.4%)	- 0.050 (- 2.0%)				

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

¹⁵ See equation (11).

They show a very revealing fact: territory exerts some influence on income inequality, but the magnitude of change as well as the sign of variation differ markedly depending on the country. In this sense, the evolution of country disparities is driven fundamentally by the role of population movements within the regions themselves [b], rather than by the effect of movements between territories [c], which is the decomposition component that contributes the least.

Of the three countries where a reduction in aggregate inequality is observed, Chile and Mexico prove how population shares contributed to reduce inequality. In the United Kingdom, the opposite happens. Likewise, it is important to clarify that in Italy, the country (along with Germany) where aggregate inequality increases the most, territory has the greatest reducing effect, mainly through component [b].

The opposite behaviour of two of the countries analysed also draws our attention: Spain and the United Kingdom. In Spain, where inequality increased meaningfully, the four elements of the decomposition contribute with a positive sign. The opposite is true for the United Kingdom, where there was a remarkable decrease in income inequality, with all components contributing to this reduction.

All these dissimilar effects lead us to ask the following questions. What would have happened in the levels of income inequality if the population weights of the different regions had remained constant during the period investigated? This is precisely what we try to elucidate and measure in the simulation exercise implemented through the counterfactual analysis.

5.3. How much inequality is explained by territory?

We carry out a decomposition of the mean log deviation index into three components (see Table 3).

Table 3. Estimation results of the DFL decomposition ($L \times 10^2$): 1999-2016

Countries chosen and time period	<i>L</i> change: observed	<i>L</i> change: estimated (contribution of the decomposition components)					
	= [1] + [2] + [3] $\Delta L = L_f - L_i$	<i>Unexplained</i> inequality		<i>Explained</i> inequality			
		[1] Other characteristics		[2] Territory		[3] Control variables	
		$L_f(c2) - L_i$		$L_f(c1) - L_f(c2)$		$L_f - L_f(c1)$	
<i>Australia</i> 2001-2014	0.303	1.632	(538.5%)	- 0.006	(- 2.0%)	- 1.323	(- 436.5%)
<i>Canada</i> 2000-2013	1.945	3.106	(159.7%)	0.088	(4.5%)	- 1.250	(- 64.3%)
<i>Chile</i> 2000-2015	- 10.317	- 10.762	(104.3%)	- 0.166	(1.6%)	0.610	(- 5.9%)
<i>France</i> 2000-2010	1.753	0.888	(50.6%)	- 0.063	(- 3.6%)	0.928	(52.9%)
<i>Germany</i> 2000-2015	3.848	3.453	(89.7%)	- 0.028	(- 0.7%)	0.423	(11.0%)
<i>Italy</i> 2000-2014	3.847	3.266	(84.9%)	- 0.479	(- 12.5%)	1.060	(27.5%)
<i>Mexico</i> 2000-2012	- 2.909	- 4.331	(148.9%)	- 0.192	(6.6%)	1.614	(- 55.5%)
<i>Poland</i> 1999-2016	0.555	0.406	(73.1%)	0.093	(16.8%)	0.056	(10.1%)
<i>Spain</i> 2004-2013	3.408	0.117	(3.4%)	0.057	(1.7%)	3.235	(94.9%)
<i>United Kingdom</i> 2004-2013	- 2.507	- 4.621	(184.3%)	- 0.200	(8.0%)	2.314	(- 92.3%)
<i>United States</i> 2000-2016	2.532	0.785	(31.0%)	0.059	(2.3%)	1.689	(66.7%)

Notes: (1) L_f and L_i are the observed inequality in the last and first year, respectively; (2) $L_f(c1)$ is the inequality estimated for the first counterfactual income distribution (the one that would have prevailed in the final year if the eight selected attributes and the regional variable had stayed at the same values of the first year); (3) $L_f(c2)$ is the inequality estimated for the second counterfactual income distribution (remaining constant the eight selected attributes, with the figures of the first year, but not the regional variable).

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

On the one hand, we can identify an *unexplained inequality* that would be included in the variation attributable to the "other characteristics" column. It would show the contribution to inequality of those characteristics or explanatory variables not taken into account in the analysis. On the other hand, we find two other elements (last two columns) that we group together, and call *explained inequality*. The first element represents the contribution of the region variable to inequality (variation attributable to "territory"), our main objective in this chapter. The second element of *explained inequality* is the contribution to inequality of the other attributes relevant to the analysis – variation attributable to “control variables”. The latter correspond to the variables described in the data section and are further detailed in Table A.4.

First, territory has a “reducing effect” on inequality in most of the countries examined (7 of 11). That is, if the demographic weights of the regions examined had not changed in recent years, income inequality would have been higher. Although the magnitude of this effect varies significantly among the countries analyzed, it is necessary to highlight this as the most striking result. This simulation enhances the results presented in the dynamic decomposition and confirms the importance of this variable in countries such as Italy and Mexico. On the other hand, it is equally remarkable to note that this “reducing effect” is not only found in countries where aggregate inequality has diminished but also in those where it has increased considerably.

A second and more specific comment relates to the relevance of the attributes chosen for the analysis. As expected, the weight assigned to "other characteristics" is very high in this decomposition. However, the results obtained also show that the control variables play an important role in explaining income inequality, and the influence of territory is not at all negligible.

To test the sensitivity of the results to the measure chosen, we also performed the three decompositions with other inequality indices. The results change very slightly, and the general conclusions do not vary significantly with other measures. In general terms, our results confirm and update those of previous works, as well as some of the theoretical premises reviewed in the previous sections. First, income differences between regions have a small explanatory capacity of inequality in the selected countries. As mentioned earlier, considering only a single explanatory dimension significantly reduces the weight of this component, as previous studies showed for other periods and countries.

Our results are in keeping with empirical evidence on the determinants of inequalities in market income in several of the countries included in the sample. As stressed by different authors, wages and unemployment are by far the most important channels of adjustment to macro-economic shocks in these countries being also the main drivers of inequality, while labour migration has played a secondary role (Brandsma et al., 2013). Second, we confirm that the changes in the territorial distribution of population affect inequality. Among other possible causes, these flows are related to spatial agglomeration effects -as OECD (2018) data seems to confirm-, which, in turn, are at the root of regional economic disparities that can affect inequality in income distribution.

6. Conclusions

Our study makes three contributions. First, we carry out different kinds of decomposition analyses to identify the effect on inequality of both income differences between regions and within regions. Second, we provide new evidence over a more varied group of countries and a more recent period than in previous studies. Third, we add to previous studies the simulation of the impact on inequality that the changes in the inter-territorial distribution of the population may have had.

We have implemented several complementary techniques in using the LIS database and by relying on the potential of the decompositions of inequality, useful tools for the correct design of redistributive policies. The idea was not only to provide a particular perspective on the same problem (the relevance of territory in the description of income inequality) but to offer a global and joint perspective by aggregating all of them. The question we tried to answer was concise: To what extent does territory drive the main changes observed in the recent evolution of income inequality? Despite not having found a strong pattern among the countries selected, the empirical specification applied for the DFL decomposition does seem to capture much of the change observed in income inequality during the years analysed. The region variable, in particular, exerts a noteworthy effect. In particular, the evidence found reveals a generalized and interesting "reducing effect" in income inequality that is directly linked to territory. All this takes on great importance because high levels of spatial inequality are undesirable for the development and economic growth of any society. They can constrain progress, are often associated with crime problems, reveal institutional weaknesses, and even hinder social cohesion among their regions (Atkinson, 2015).

In closing, we have shed some light on the transcendence of the territorial dimension in inequality by implementing an application not developed thus far: a counterfactual analysis for population weights in the context of international comparative analysis. In general terms, our results confirm and update those of previous works, as well as some of the theoretical premises of the related literature. Income differences between regions have a small explanatory capacity of inequality in the selected countries while the changes in the territorial distribution of population affect inequality. Even with the challenges of providing conclusive evidence -the influence of social and institutional factors is very diverse in each country- we have obtained some interesting results that underscore the

relevance of this determining factor and that could encourage the pursuit of future research on spatial inequality.

These results are important both to inform the theory and to contribute to the design of public policies related to social cohesion. Regarding the first issue, we have confirmed that changes in the territorial distribution of population affect inequality. As stated by NEG models, these flows are related to spatial agglomeration effects, which, in turn, are at the root of regional economic disparities that can affect inequality in the income distribution.

A political implication of our results is that if population mobility is conceived as a possible tool to influence inter- and intra-inequality, it is necessary to coordinate this type of specific policies with other redistributive measures in order to optimize their joint impact. In any case, we need to remain cautious when drawing other policy implications, given the difficulty of going from an accounting exercise to the design of specific policies. As stated by Fujita and Krugman (2005), “because geography is such a crucial factor in development, and there are undoubtedly strong policy implications of some sort, it is an important subject for further research”. In this regard, some of the limitations of our work may be interpreted as further promising extensions.

One of these caveats is the very high weight of “other characteristics” suggesting that some relevant variables might have been omitted. Since we focus exclusively on the information included in LIS data, relevant issues such as technology, trade, or decentralization are not considered here. Second, a larger and more varied sample might enrich the analysis. Third, most of our analyses are accounting exercises of how much the changes in the territorial distribution of population may influence recent inequality trends. A more detailed analysis of population changes at a higher level of disaggregation should be a promising extension of this work.

References

- Allanson, P. (2014): "Income stratification and between-group inequality". *Economics letters*, 124(2), 227-230.
- Athanasopoulos, G. and Vahid, F. (2003): "Statistical Inference and Changes in Income Inequality in Australia", *The Economic Record*, vol.79, no. 247: 412-424.
- Atkinson, A. B. (2015): *Inequality*. Harvard University Press.
- Ayala, L., Jurado, A. and J. Pérez-Mayo (2014): "Drawing the poverty line: do regional thresholds make a difference?". *Applied Economic Perspectives & Policy*, 36: 309-322.
- Biewen, M. and Juhasz, A. (2012): "Understanding rising income inequality in Germany, 1999/2000–2005/2006". *Review of Income and Wealth*, 58(4): 622-647.
- Brandolini, A. (2007): "Measurement of Income Distribution in Supranational Entities: The Case of the European Union", In P. Jenkins and J. Micklewright (eds), *Inequality and Poverty Reexamined*, Oxford, 2007, Oxford University Press.
- Brandsma, A., Kancs, D.A. and Persyn, D. (2013): "Modelling migration and regional labour markets: an application of the new economic geography model rhomolo". European Commission Joint Research Centre, Technical Report.
- Carrillo, P. E. and Rothbaum, J. L. (2016): "Counterfactual spatial distributions". *Journal of Regional Science*, 56(5): 868-894.
- Castells-Quintana, D., Ramos, R. y Royuela, V. (2015): "Income inequality in European Regions. Recent trends and determinants", *Review of Regional Research*, 35: 123-146.
- Combes, P. P., Lafourcade, M., Thisse, J. F. and Toutain, J. C. (2011): "The rise and fall of spatial inequalities in France: A long-run perspective". *Explorations in Economic History*, 48(2), 243-271.
- Commendatore, P., Kubin, I. and Mossay, P. (2018): "On the new economic geography of a multicore world". *Review of International Economics*, 26(3), 539-554.
- Contreras, D. and Ffrench-Davis, R. (2014): "Policy regimes, inequality, poverty and growth: The Chilean experience, 1973-2010". In *Falling inequality in Latin America: policy changes and lessons* (pp.94-118). OUP Oxford.
- Cowell, F. A. (2011): *Measuring Inequality* (First ed.), Oxford: Philip Allan, (Third ed.), Oxford University Press.
- Cowell, F. A. and Fiorio, C. (2011): "Inequality Decompositions. A Reconciliation". *Journal of Economic Inequality*, 9: 509-528.
- Crozet, M. (2004): "Do migrants follow market potentials? An estimation of a new economic geography model". *Journal of Economic Geography* 4, 439-458.
- Czyż, T. and Hauke, J. (2011): "Evolution of regional disparities in Poland". *Quaestiones Geographicae*, 30(2), 35-48.
- Dickey, H. (2014): "The Impact of Migration on Regional Wage Inequality: a Semiparametric Approach". *Journal of Regional Science*, 54(5), 893-915.

- DiNardo, J, Fortin, N. and Lemieux, T. (1996): “Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”. *Econometrica*, 64(5): 1001-1044.
- Dormard, S. (2004): “Economic development and regional disparities in France”. In *Employment and regional development policy: Market efficiency versus policy intervention* (pp. 50-67). Hannover: Verlag der ARL-Akademie für Raumforschung und Landesplanung.
- Elbers, C., Lanjouw, P., Mistiaen, J. A. and Özler, B. (2008): “Reinterpreting between-group inequality”. *The Journal of Economic Inequality*, 6(3), 231-245.
- Esquivel, G., Lustig, N. and Scott, J. (2010): “A decade of falling inequality in Mexico: market forces or state action?”. *Declining inequality in Latin America: A decade of progress*, 175-217.
- Ezcurra, R. and A. Rodríguez-Pose (2009): “Measuring the regional divide”. In Capello, and Nijkamp (Eds.): *Handbook of regional growth and development theories*, (pp. 329–353). Cheltenham UK: Edward Elgar.
- Fleming, D. A. and Measham, T. G. (2015): “Income inequality across Australian regions during the mining boom: 2001–11”. *Australian Geographer*, 46(2), 203-216.
- Frick, J. R. and Goebel, J. (2008): “Regional income stratification in unified Germany using a Gini decomposition approach”. *Regional Studies*, 42(4), 555-577.
- Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G. and Yitzhaki, S. (2006): “Using analysis of Gini (ANOI) for detecting whether two subsamples represent the same universe: The German Socio-Economic Panel Study (SOEP) experience”. *Sociological Methods & Research*, 34(4), 427-468.
- Fujita, M. and Krugman, P. (2005): “The new economic geography: Past, present and the future”. *Papers in Regional Science* 83, 139–164.
- Garretsen, H. and Martin, R. (2010): “Rethinking (New) Economic Geography Models: Taking Geography and History More Seriously”, *Spatial Economic Analysis* 5, 127-160.
- Gray, D., Mills, J. A. and Zandvakili, S. (2004): “An analysis of differential provincial income inequality trends in Canada”. In *Studies on Economic Well-Being: Essays in the Honor of John P. Formby* (pp. 443-461). Emerald Group Publishing Limited.
- Hassing, R. and Gong, H. (2019): “New Economic Geography”. In Orum, A.M. (2019): *The Wiley Blackwell Encyclopedia of Urban and Regional Studies*. Wiley Online Library.
- Hoffmeister, O. (2009): “The spatial structure of income inequality in the enlarged EU”. *Review of Income and Wealth*, 55(1): 101-127.
- Insee (1997): *Revenus et patrimoine des ménages*. Édition 1997. Synthèses, 11. Paris: Insee.
- Jesuit (2003): “The regional Dynamics of European Electoral Politics”, *European Journal of Politics*, 4(2): 139-164.
- Kanbur, R. and Venables, A. J. (Eds.) (2005): *Spatial inequality and development*. Oxford University Press.
- Lee, N. and Rodríguez-Pose, A. (2013): “Innovation and spatial inequality in Europe and USA”. *Journal of economic geography*, 13(1), 1-22.

- Lessmann, C. (2014): “Spatial inequality and development—Is there an inverted-U relationship?”. *Journal of Development Economics*, 106, 35-51.
- Martino, G. and Perugini, C. (2008): “Income inequality within European regions: determinants and effects on growth”, *Review of Income and Wealth*, 564(3): 337-406.
- Milanovic, B. and Yitzhaki, S. (2002) “Decomposing World Income Distribution: Does the World have a Middle Class?”. *Review of Income and Wealth*, 48, 155–78, 2002.
- Miranti, R., Cassells, R., Vidyattama, Y. and McNamara, J. (2013): “Measuring Inequality for Small Areas Using Spatial Microsimulation: An Australian Case Study”. In *Proceedings of the The 42nd Australian Conference of Economists, Murdoch University, Perth, Western Australia* (pp. 7-10).
- Monti, M. and Santoro, A. (2011): Stratification and between-group inequality: A new interpretation. *Review of Income and Wealth*, 57(3):412-427.
- Mookherjee, D. and Shorrocks, A. (1982): “A Decomposition Analysis of the Trend in UK Income Inequality”. *The Economic Journal*, 92(368): 886–902.
- Mussini, M. (2017): “Decomposing changes in inequality and welfare between the EU regions”, *Social Indicators Research*, 130:455-478.
- Novotný, J. (2004): “Decomposition of global and European socio-economic inequalities with attention to their regional dimensions”. *Regional and Sectoral Economic Studies*, 4(2): 59-78.
- Novotný, J. (2007): “On the measurement of regional inequality: does spatial dimension of income inequality matter?”. *The Annals of Regional Science*, 41(3): 563-580.
- OECD (2018): Job creation and local economic development 2018: Preparing for the future of work. Paris: OECD Publishing.
- Ostry, M.J.D., Berg, M.A. and Tsangarides, M.C.G. (2014): “Redistribution, Inequality, and Growth”. IMF Staff Discussion Note SDN/14/02.
- Parro, F. and Reyes, L. (2017): “The rise and fall of income inequality in Chile”. *Latin American Economic Review*, 26(1): 3.
- Puga, D. (2002): “European regional policies in light of recent location theories”. *Journal of economic geography*, 2(4), 373-406.
- Piacentini, M. (2014): “Measuring Income Inequality and Poverty at the Regional Level in OECD Countries”, *OECD Statistics Working Papers*, 2014/03.
- Perugini, C and Martino, G. (2008): “Income inequality within European regions: determinants and effects on growth”, *Review of Income and Wealth*, 564(3): 373-406.
- Pyatt, G. (1976): “On the Interpretation and Disaggregation of Gini Coefficient,” *Economic Journal*, 86, 243–55.
- Rey, S. J. and Janikas, M.V. (2005): “Regional convergence, inequality and space”. *Journal of Economic Geography*, 5(2): 155-176.
- Rey, S. J. and Smith, R. J. (2013): “A spatial decomposition of the Gini coefficient”. *Letters in Spatial and Resource Sciences*, 6(2), 55-70.

- Rodríguez-Pose, A. and Ezcurra, R. (2009): “Does decentralization matter for regional disparities? A cross-country analysis”. *Journal of Economic Geography*, 10(5), 619-644.
- Shorrocks, A. F. (2013): “Decomposition procedures for distributional analysis: a unified framework based on the Shapley value”. *Journal of Economic Inequality*, 1-28.
- Shorrocks, A. F. and Wan, G. (2005): “Spatial Decomposition of Inequality”. *Journal of Economic Geography*, 5(1): 59-81.
- Tirado, D. A., Díez-Minguela, A. and Martínez-Galarraga, J. (2016): “Regional inequality and economic development in Spain, 1860–2010”. *Journal of Historical Geography*, 54, 87-98.
- Torrisi, G., Pike, A., Tomaney, J. and Tselios, V. (2015): “(Re-) exploring the link between decentralization and regional disparities in Italy”. *Regional Studies, Regional Science*, 2(1), 123-140.
- Usseglio, A. N. (2016): “The North-South Divide: Regional Economic Inequality in Contemporary Italy”. *UVM Honors College Senior Theses*. 122.
- Venables, A.J. (2016): “New Economic Geography”. In *The New Palgrave Dictionary of Economics*. Palgrave Macmillan.
- Yitzhaki, S. (1994): “Economic Distance and Overlapping of Distributions,” *Journal of Econometrics*, 61, 147–59.
- Yitzhaki, S. and Schechtman, E. (2013): “More than a dozen alternative ways of spelling Gini”. In *The Gini Methodology* (pp. 11-31). Springer, New York, NY.

Appendix

Table A.1. Basic descriptive statistics for the most populated OECD countries: 1999-2016

Wave	Years	Countries and time										
		AU 2001-2014	CA 2000-2013	CL 2000-2015	FR 2000-2010	DE 2000-2015	IT 2000-2014	MX 2000-2012	PL 1999-2016	ES 2004-2013	UK 2004-2013	US 2000-2016
(a) Sample size (number of household respondents)												
V	99/01	6,786	28,881	65,036	10,305	11,796	8,000	10,108	31,428	-	-	78,054
VI	03/05	11,361	27,665	68,153	10,240	11,294	8,012	22,595	32,214	12,996	27,753	76,447
VII	06/08	9,345	26,560	73,720	-	10,921	7,977	29,468	37,366	13,014	24,977	75,872
VIII	09/11	18,008	24,826	71,460	10,342	16,703	7,941	27,655	37,412	13,109	25,350	75,188
IX	12/14	14,115	23,014	66,725	-	15,946	8,151	9,002	37,181	11,965	20,135	51,498
X	15/16	-	-	83,887	-	14,426	-	-	36,886	-	-	69,957
(b) Population size (inhabitants: Ix10 ³)												
V	99/01	18,747	29,798	15,039	59,329	83,150	56,635	98,163	38,666	-	-	275,662
VI	03/05	19,521	30,851	15,571	59,500	83,086	57,208	102,989	37,784	42,874	57,945	286,674
VII	06/08	20,508	31,889	16,115	-	82,739	58,360	111,612	37,708	45,109	59,829	292,009
VIII	09/11	21,343	32,945	16,583	61,781	80,579	58,854	114,560	37,726	45,900	61,041	296,992
IX	12/14	22,389	34,017	17,256	-	81,351	60,439	117,284	38,101	45,977	62,853	305,234
X	15/16	-	-	17,530	-	83,157	-	-	38,004	-	-	310,964
(c) Equivalent household disposable income (PPP\$US2011)												
V	99/01	23,692	27,602	9,417	23,683	27,464	21,334	6,664	9,538	-	-	37,803
VI	03/05	26,504	29,335	9,217	24,408	27,760	28,844	6,950	9,647	21,109	26,612	38,226
VII	06/08	33,481	31,904	10,057	-	28,225	22,247	8,072	11,708	23,086	28,747	39,040
VIII	09/11	33,524	32,690	10,681	27,982	28,295	21,527	7,038	13,637	21,069	27,617	37,557
IX	12/14	39,977	35,044	13,121	-	28,229	19,270	7,520	13,684	21,161	27,128	37,882
X	15/16	-	-	13,080	-	28,870	-	-	15,925	-	-	41,209

Notes: (1) AU=Australia, CA=Canada, CL=Chile, FR=France, DE=Germany, IT=Italy, MX=Mexico, PL=Poland, ES=Spain, UK=United Kingdom, US=United States; (2) To ensure the consistency of all the tables, data from 1999 in the United Kingdom and from 2000 in Spain, relative to Wave V, have been excluded. There are no data for Northern Ireland in 1999 and there is no regional information for Spain at a NUTS-2 level of disaggregation in 2000; (3) Overseas regions are not taken into account in France.

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.2. Checking aggregation effects

Table A.2.1. Spatial decomposition of the Mean Log Deviation (L) in the OECD: 1999-2016

Country	Units	Year	Total inequality (L)	Between (B)		Within (W)	
<i>Italy</i> **	5	2000	0.22779	0.02580	11.3%	0.20199	88.7%
	5	2014	0.26626	0.01713	6.4%	0.24914	93.6%
<i>Mexico</i> ***	8	2000	0.45041	0.04727	10.5%	0.40314	89.5%
	8	2012	0.42132	0.02756	6.5%	0.39376	93.5%
<i>Spain</i> ***	7	2004	0.21216	0.00794	3.7%	0.20422	96.3%
	7	2013	0.24624	0.01308	5.3%	0.23316	94.7%
<i>United States</i> **	9	2000	0.27129	0.00171	0.6%	0.26958	99.4%
	9	2016	0.29662	0.00174	0.6%	0.29488	99.4%

Note: Asterisks indicate that initial and final values are significantly different at the 90 (*), 95 (**), and 99 (***) % confidence levels.

Table A.2.2. Decomposition of the *trend* in aggregate inequality ($L \times 10^2$): 1999-2016

Countries and time	Change in aggregate inequality $\Delta L = L_f - L_i$	Contribution to ΔL attributable to variations in ¹⁶							
		Within group inequality		Population shares				Mean group incomes	
		[11a]		[11b]		[11c]		[11d]	
		$\sum_g \bar{p}_g \Delta L_g$		$\sum_g \bar{L}_g \Delta p_g$		$\sum_g \left[\left(\frac{\mu_g}{\mu} \right) - \log \left(\frac{\mu_g}{\mu} \right) \right] \Delta p_g$		$\sum_g (\bar{q}_g - \bar{p}_g) \Delta \log \mu_g$	
<i>Italy</i> (5) 2000-2014	3.847	5.076	(132.1%)	-0.362	-(9.4%)	-0.054	-(1.4%)	-0.816	-(21.2%)
<i>Mexico</i> (8) 2000-2012	-2.909	0.835	(28.8%)	-0.103	(3.6%)	-0.054	(1.9%)	-1.903	(65.8%)
<i>Spain</i> (7) 2004-2013	3.408	2.868	(84.1%)	0.026	(0.8%)	0.007	(0.2%)	0.508	(14.9%)
<i>United States</i> (9) 2000-2016	2.532	2.441	(96.4%)	0.088	(3.5%)	-0.003	-(0.1%)	0.006	(0.2%)

Table A.2.3. Estimation results of the DFL decomposition ($L \times 10^2$): 1999-2016

Countries chosen and time period	L change: observed	L change: estimated (contribution of the decomposition components)					
	= [1] + [2] + [3] $\Delta L = L_f - L_i$	<i>Unexplained inequality</i>		<i>Explained inequality</i>			
		[1] Other characteristics		[2] Territory		[3] Control variables	
		$L_f(c2) - L_i$		$L_f(c1) - L_f(c2)$		$L_f - L_f(c1)$	
<i>Italy</i> (5) 2000-2014	3.847	2.542	(66.1%)	0.245	(6.4%)	1.060	(27.5%)
<i>Mexico</i> (8) 2000-2012	-2.909	-4.470	(153.7%)	-0.054	(1.8%)	1.614	-(55.5%)
<i>Spain</i> (7) 2004-2013	3.408	0.106	(3.1%)	0.066	(1.9%)	3.235	(94.9%)
<i>United States</i> (9) 2000-2016	2.532	0.744	(29.4%)	0.100	(3.9%)	1.689	(66.7%)

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

¹⁶ See equation (11).

Table A.3. Descriptive statistics in the OECD countries by territories: 1999-2016

Territories of Australia: 2001-2014 <i>6 Federated States and Canberra</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Canberra region and Northern T.	0.18079	2.32	381	0.16821	2.42	1,395
New South Wales	0.20682	32.75	1,530	0.23913	31.82	2,532
Queensland	0.19940	19.08	1,218	0.22140	20.17	2,272
South Australia	0.16315	8.33	889	0.19465	7.66	2,131
Tasmania	0.13893	2.60	482	0.15753	2.39	1,704
Victoria	0.20055	25.30	1,400	0.21083	25.10	2,287
Western Australia	0.16391	9.62	886	0.28533	10.42	1,841

Territories of Canada: 2000-2013 <i>10 Provinces</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Alberta	0.17309	9.85	2,342	0.19393	11.49	2,213
British Columbia	0.19719	13.15	2,514	0.20585	12.94	2,434
Manitoba	0.16173	3.61	2,172	0.17011	3.46	2,049
New Brunswick	0.15330	2.45	1,728	0.14225	2.14	1,278
Newfoundland	0.14688	1.72	1,177	0.16608	1.52	884
Nova Scotia	0.15204	3.04	1,954	0.17474	2.66	1,388
Ontario	0.17213	38.47	8,384	0.20684	38.93	5,968
Prince Edward Island	0.14949	0.45	821	0.14001	0.42	622
Quebec	0.15434	24.11	5,755	0.15791	23.44	4,510
Saskatchewan	0.16554	3.15	2,034	0.19219	3.01	1,718

Territories of Chile: 2000-2015 <i>13 Regions</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Antofagasta	0.43746	3.05	1,804	0.34863	3.06	2,025
Arica y Parinacota and Tarapacá	0.40927	2.59	2,278	0.40753	2.73	3,419
Atacama	0.36747	1.59	1,930	0.30354	1.49	3,989
Aysén	0.42563	0.62	878	0.34652	0.64	1,152
Biobío	0.49256	12.38	11,432	0.34151	12.05	11,490
Coquimbo	0.40592	3.89	3,146	0.32915	4.17	3,745
La Araucanía	0.57609	5.56	6,434	0.36124	5.71	7,040
Libertador Bernardo O'Higgins	0.34134	5.20	4,746	0.31847	5.29	7,165
Los Lagos and Los Ríos	0.44695	7.12	5,880	0.34265	7.24	9,562
Magallanes and La Antártica Chile	0.57442	1.03	845	0.38106	0.94	1,892
Maule	0.51325	6.04	6,370	0.32077	6.13	5,687
Región Metropolitana Santiago	0.51783	40.20	13,100	0.43930	39.74	17,723
Valparaíso	0.36132	10.69	6,193	0.34529	10.82	8,998

Territories of France: 2000-2010 <i>NUTS-1 (8 ZEAT¹⁷)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Bassin parisien	0.13464	17.39	1,842	0.12641	17.28	1,922
Centre-est	0.12448	11.83	1,221	0.13911	12.04	1,138
Est	0.10225	9.19	1,081	0.12760	8.63	936
Méditerranée	0.16017	12.19	1,224	0.15202	13.00	1,193
Nord	0.13920	6.47	675	0.13915	6.14	761
Ouest	0.12266	13.49	1,535	0.14600	13.35	1,568
Région parisienne	0.14956	18.06	1,609	0.19289	18.18	1,625
Sud-ouest	0.13501	11.38	1,118	0.15531	11.37	1,199

¹⁷ Zones d'études et d'aménagement du territoire (Research and National Development Zones).

Descomposición simultánea de la desigualdad. Algunas aplicaciones

Territories of Germany: 2000-2015 <i>NUTS-1 (16 Länder)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Baden-Wuerttemberg	0.12017	12.46	1,381	0.15616	12.81	1,645
Bavaria	0.13369	14.54	1,634	0.16694	15.48	2,322
Berlin	0.15679	4.81	487	0.15221	4.81	614
Brandenburg	0.08048	3.08	503	0.15835	2.98	569
Bremen	0.23559	1.07	100	0.18832	0.87	105
Hamburg	0.22795	2.21	182	0.14440	2.40	265
Hesse	0.15632	7.35	789	0.21360	7.39	989
Lower Saxony	0.13515	9.32	967	0.15217	9.54	1,357
Mecklenburg-Western Pomerania	0.10637	2.18	289	0.15525	2.03	330
North Rhine-Westphalia	0.13366	21.77	2,512	0.21261	21.36	2,930
Rhineland-Palatinate	0.11541	4.46	551	0.12869	4.70	713
Saarland	0.08532	1.56	171	0.13989	1.21	140
Saxony	0.08323	5.52	857	0.11861	5.42	878
Saxony-Anhalt	0.09106	3.19	508	0.13457	2.85	516
Schleswig-Holstein	0.15073	3.63	367	0.13038	3.43	505
Thuringia	0.08210	2.87	498	0.16015	2.72	548

Territories of Italy: 2000-2014 <i>NUTS-2 (20 Regioni)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Abruzzo	0.23960	2.02	228	0.14706	2.04	204
Basilicata	0.11683	1.99	95	0.15293	3.12	128
Calabria	0.15522	2.61	210	0.16736	3.62	217
Campania	0.21439	8.82	815	0.22106	7.63	716
Emilia Romagna	0.14742	7.66	751	0.16686	7.78	677
Friuli	0.15542	2.17	255	0.14613	3.49	214
Lazio	0.12401	9.74	425	0.34521	9.48	452
Liguria	0.14654	3.25	316	0.10185	3.66	347
Lombardia	0.18214	16.62	860	0.18865	12.17	944
Marche	0.18889	2.36	328	0.15291	3.16	345
Molise	0.22230	0.68	83	0.27921	1.78	111
Piemonte	0.13844	8.02	732	0.21700	9.02	725
Puglia	0.20749	6.17	471	0.24561	5.72	453
Sardegna	0.18040	2.69	308	0.21700	2.45	343
Sicilia	0.28245	8.14	630	0.26316	5.50	618
Toscana	0.13941	6.12	598	0.15180	6.01	605
Trentino	0.12557	1.52	161	0.13015	4.90	238
Umbria	0.10067	1.42	271	0.24910	1.55	277
Valle d'Aosta	0.06396	0.17	25	0.10929	0.71	43
Veneto	0.20700	7.84	439	0.14882	6.22	499

Territories of Poland: 1999-2016 <i>NUTS-2 (16 Voivodeships)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Dolnoslaskie	0.13657	7.93	2,489	0.13035	8.10	2,899
Kujawsko-Pomorskie	0.12881	5.46	1,728	0.14503	5.24	1,968
Lodzkie	0.13947	8.47	2,665	0.13559	7.02	2,515
Lubelskie	0.14980	5.71	1,791	0.16936	5.33	2,062
Lubuskie	0.10497	2.81	889	0.12231	2.60	999
Malopolskie	0.13171	7.36	2,307	0.12876	8.26	3,117
Mazowieckie	0.17919	13.03	4,059	0.19126	14.99	5,535
Opolskie	0.15615	2.69	838	0.11968	2.49	973
Podkarpackie	0.12130	5.06	1,592	0.10656	4.85	1,878
Podlaskie	0.13883	2.80	883	0.15070	3.12	1,152
Pomorskie	0.13809	5.36	1,683	0.20087	5.92	2,230
Slaskie	0.10355	13.40	4,228	0.11558	12.42	4,297
Swietokrzyskie	0.14022	3.07	960	0.12758	3.13	1,193
Warminsko-Mazurskie	0.12267	4.23	1,332	0.14271	3.65	1,478
Wielkopolskie	0.16812	8.44	2,661	0.12619	8.41	3,066
Zachodnio-Pomorskie	0.14879	4.19	1,323	0.10365	4.49	1,524

Descomposición simultánea de la desigualdad. Algunas aplicaciones

Territories of Mexico: 2000-2012 <i>32 Federal Entities</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Aguascalientes	0.21190	0.99	226	0.31265	1.06	266
Baja California Norte	0.24744	2.59	264	0.31505	2.85	267
Baja California Sur	0.18527	0.44	209	0.29350	0.60	264
Campeche	0.37318	0.72	218	0.42726	0.74	287
Chiapas	0.55252	4.15	325	0.44411	4.31	296
Chihuahua	0.21598	2.91	247	0.28289	3.08	266
Ciudad de México	0.40714	8.44	456	0.36397	7.54	365
Coahuila de Zaragoza	0.24660	2.23	289	0.21898	2.44	266
Colima	0.35353	0.56	203	0.29144	0.59	275
Durango	0.29121	1.49	294	0.44399	1.46	272
Guanajuato	0.58894	4.79	288	0.36621	4.85	266
Guerrero	0.49396	3.16	311	0.39153	2.99	254
Hidalgo	0.35198	2.31	254	0.34716	2.36	280
Jalisco	0.23771	6.53	345	0.45714	6.53	334
México	0.39081	13.58	402	0.43785	13.80	372
Michoacán de Ocampo	0.35990	4.12	353	0.28050	3.82	288
Morelos	0.45057	1.61	197	0.35123	1.58	265
Nayarit	0.31086	0.94	213	0.49357	0.99	281
Nuevo León	0.23969	3.94	367	0.30357	4.17	243
Oaxaca	0.52580	3.55	238	0.52159	3.35	280
Puebla	0.40518	5.28	308	0.34095	5.13	275
Querétaro	0.61300	1.45	214	0.46545	1.63	280
Quintana Roo	0.41285	0.92	194	0.29955	1.24	277
San Luis Potosí	0.28936	2.36	275	0.37650	2.29	263
Sinaloa	0.32482	2.61	252	0.39448	2.48	275
Sonora	0.44931	2.29	251	0.30667	2.41	261
Tabasco	0.47245	1.95	224	0.42890	1.97	263
Tamaulipas	0.25368	2.85	279	0.32535	2.93	243
Tlaxcala	0.28028	1.01	232	0.22711	1.05	273
Veracruz-Llave	0.37690	7.13	1,735	0.42836	6.72	360
Yucatán	0.53801	1.71	234	0.26250	1.74	267
Zacatecas	0.33491	1.38	211	0.42849	1.31	278

Territories of Spain: 2004-2013 <i>NUTS-2 (17 Regiones + CA¹⁸)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Andalucía	0.17243	16.80	1,610	0.24659	17.25	1,478
Aragón	0.21117	3.06	582	0.18065	2.95	541
Canarias	0.20308	4.11	645	0.23558	4.39	501
Cantabria	0.20838	1.24	344	0.14411	1.32	295
Castilla La Mancha	0.23397	4.23	680	0.19186	4.31	564
Castilla y León	0.22080	5.96	913	0.21670	5.68	845
Cataluña	0.17737	16.49	1,376	0.19581	16.18	1,264
Ceuta and Melilla	0.34812	0.29	265	0.31958	0.27	258
Comunidad de Madrid	0.18062	13.26	801	0.22300	13.85	1,134
Comunidad Foral de Navarra	0.19075	1.38	429	0.15204	1.39	426
Comunidad Valenciana	0.17656	10.90	1,089	0.18500	10.87	891
Extremadura	0.20687	2.48	554	0.19476	2.35	506
Galicia	0.18238	6.31	911	0.16736	5.86	811
Illes Balears	0.19178	2.31	508	0.23154	2.38	373
La Rioja	0.17093	0.69	411	0.17045	0.70	385
País Vasco	0.17090	5.11	731	0.15756	4.86	696
Principado de Asturias	0.20210	2.59	593	0.18328	2.50	498
Región de Murcia	0.18230	2.79	557	0.20979	2.91	499

Territories of UK: 2004-2013 <i>NUTS-1 (12 Regions)</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
East Midlands	0.19447	7.00	1,849	0.17747	7.17	1,294
East of England	0.21671	8.96	2,216	0.19318	9.19	1,674

¹⁸ CA = *Ciudades Autónomas* (Ceuta and Melilla).

Descomposición simultánea de la desigualdad. Algunas aplicaciones

London	0.31422	12.70	2,548	0.26678	12.44	1,748
North East (including Cumbria)	0.17741	4.32	1,103	0.16176	4.26	743
North West	0.18978	11.52	2,996	0.17316	11.29	1,967
Northern Ireland	0.18873	2.60	1,927	0.16218	2.74	1,965
Scotland	0.18281	8.87	4,523	0.18119	8.90	3,000
South East (excluding London)	0.26375	13.28	3,294	0.20641	13.43	2,421
South West	0.20921	8.69	2,163	0.18056	8.60	1,448
Wales	0.17993	4.86	1,234	0.16364	4.92	875
West Midlands	0.18559	8.73	2,172	0.16094	8.65	1,519
Yorkshire and Humber	0.16903	8.47	2,016	0.19046	8.41	1,483

Territories of US: 2000-2016 <i>50 States and Distric of Columbia</i>	Initial year			Final year		
	L_i	% Pop.	N	L_f	% Pop.	N
Alabama	0.25001	1.60	1,255	0.36027	1.54	1,487
Alaska	0.24050	0.21	1,081	0.27291	0.21	769
Arizona	0.22902	1.78	1,047	0.32178	2.15	1,123
Arkansas	0.24447	1.00	903	0.34505	0.96	1,294
California	0.28731	11.03	5,611	0.35227	11.04	6,226
Colorado	0.25764	1.57	1,522	0.25882	1.83	895
Connecticut	0.26327	1.25	1,339	0.31166	1.13	685
Delaware	0.24366	0.28	943	0.25084	0.31	746
District of Columbia	0.33622	0.23	1,015	0.42728	0.25	1,468
Florida	0.28213	6.03	3,455	0.30774	6.70	3,367
Georgia	0.23652	2.86	1,094	0.33338	3.14	1,566
Hawaii	0.19624	0.38	1,031	0.27041	0.38	1,094
Idaho	0.24410	0.46	996	0.27659	0.50	1,031
Illinois	0.27631	4.35	2,721	0.30223	4.01	1,982
Indiana	0.27836	2.24	1,389	0.27601	2.08	1,064
Iowa	0.19144	1.08	1,293	0.25753	1.05	758
Kansas	0.22011	1.01	1,276	0.30665	0.92	850
Kentucky	0.24158	1.48	1,062	0.26942	1.45	864
Louisiana	0.26233	1.56	870	0.32899	1.47	1,680
Maine	0.20792	0.50	1,268	0.26116	0.47	545
Maryland	0.27886	1.92	1,274	0.27643	1.84	885
Massachusetts	0.28009	2.37	1,433	0.29150	2.19	1,374
Michigan	0.25455	3.56	2,172	0.24912	3.21	1,496
Minnesota	0.25672	1.81	1,406	0.23488	1.76	853
Mississippi	0.27511	1.02	800	0.31658	0.91	1,347
Missouri	0.27557	2.05	1,180	0.27897	1.93	950
Montana	0.21289	0.33	866	0.25419	0.35	1,397
Nebraska	0.21559	0.62	1,217	0.22765	0.60	805
Nevada	0.24943	0.70	1,417	0.25034	0.90	916
New Hampshire	0.23964	0.46	1,281	0.23014	0.43	833
New Jersey	0.26053	2.98	1,963	0.30917	2.70	1,373
New Mexico	0.28593	0.62	1,017	0.37996	0.64	1,482
New York	0.31062	6.70	4,202	0.32928	6.13	2,912
North Carolina	0.24438	2.93	1,652	0.32193	3.19	1,619
North Dakota	0.20654	0.25	1,101	0.26671	0.25	937
Ohio	0.22308	4.19	2,475	0.28857	3.76	1,793
Oklahoma	0.26576	1.26	1,091	0.31001	1.24	1,069
Oregon	0.25090	1.27	1,198	0.25803	1.36	997
Pennsylvania	0.25206	4.48	2,659	0.28285	4.08	1,793
Rhode Island	0.25060	0.40	1,326	0.27223	0.36	613
South Carolina	0.31398	1.45	1,014	0.32509	1.61	1,156
South Dakota	0.24155	0.28	1,210	0.30113	0.28	690
Tennessee	0.31564	2.10	973	0.31969	2.18	1,298
Texas	0.32035	6.95	3,504	0.31745	8.00	4,207
Utah	0.18434	0.68	932	0.26970	0.83	932
Vermont	0.34441	0.24	1,174	0.26549	0.22	821
Virginia	0.27686	2.53	1,331	0.34010	2.56	1,286
Washington	0.31934	2.15	1,350	0.31141	2.26	1,229
West Virginia	0.23042	0.69	1,132	0.30432	0.60	1,465
Wisconsin	0.27140	1.94	1,488	0.26541	1.89	925
Wyoming	0.22652	0.18	1,045	0.27274	0.19	1,010

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.4. Control variables used in the probit regressions

	Variables	Definition
<i>Geography and housing</i>	Owned/Rented housing (h-file) LIS variable: own	Indicator of housing tenure. We have defined it as a dummy variable (1=owned; 0=rented/others).
<i>Household composition and living arrangements</i>	Household members (h-file) LIS variable: <i>nhhmem</i>	Number of household members. We have defined it as a dummy variable (1=one member; 0=more than one member).
<i>Socio-demographic characteristics</i>	Age (p-file) LIS variable: <i>age</i>	Age in years (min/35=1, 36/55=2, 36/max=3)
	Marital status (p-file) LIS variable: <i>marital</i>	Classification of persons according to their marital status, as provided in relation to the marriage laws or customs of the country. We have defined it as a dummy variable (1=married/in consensual union; 0=another status).
	Immigration (p-file) LIS variable: <i>immigr</i>	All persons who have that country as their country of usual residence and (in order of priority): whom the data provider identified as immigrants; who self-identify as immigrants; who are a citizen/national of another country; who were born in another country. It is defined as a dummy variable (1=immigrant; 0=not immigrant).
	Health (p-file) LIS variable: <i>disabled</i>	Disabled persons who have a permanent disability condition, defined as a (physical or mental) health condition that permanently limits an individual in his/her basic activity functioning (such as walking or hearing), even if the limitation is ameliorated by the use of assistive devices or a supportive environment. It is defined as a dummy variable (1=disabled; 0=not disabled).
	Education (p-file) LIS variable: <i>educ</i>	Recoding of highest level of education completed into three categories: low: less than secondary education completed (never attended, no completed education or education completed at the ISCED levels 0, 1 or 2); medium: secondary education completed (completed ISCED levels 3 or 4); high: tertiary education completed (completed ISCED levels 5 or 6).
<i>Labour-market information</i>	Employment (p-file) LIS variable: <i>emp</i>	Indicator that employment is the status of the main current activity as self-assessed by the respondent. It is defined as a dummy variable (0=not employed; 1=employed).

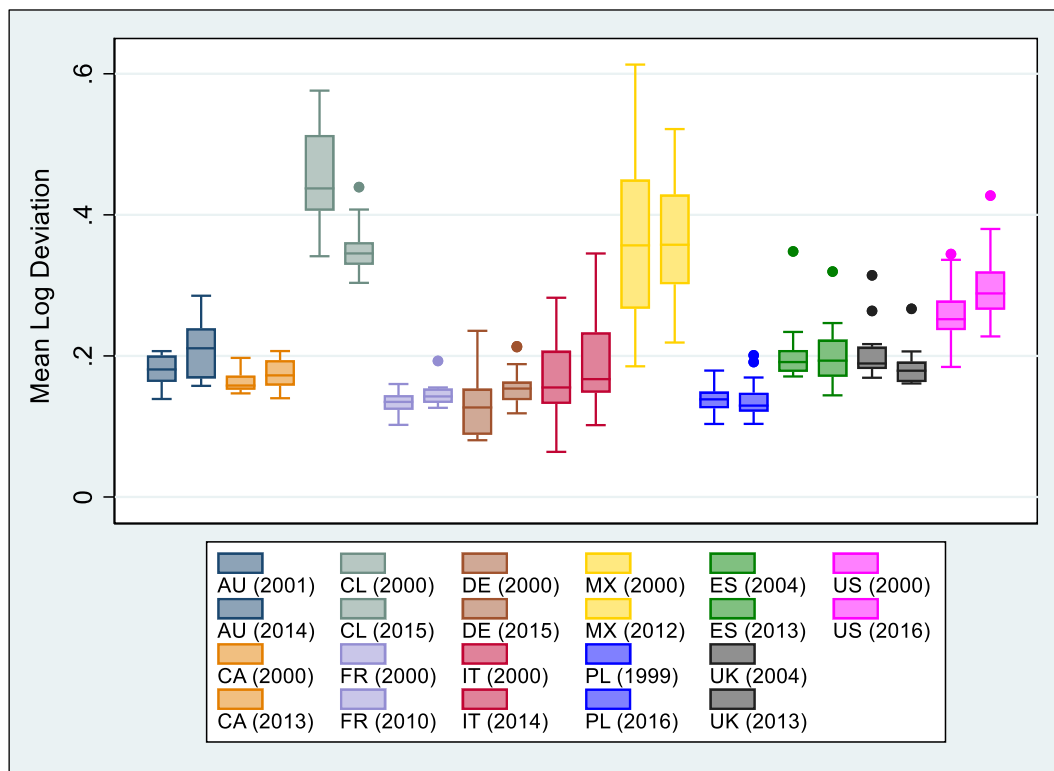
Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

Table A.5. ANOGI decomposition and regional stratification in OECD countries: 1999-2016

Countries	Units	Year	Overall Gini	G_{wo}	G_b	IG	IGO	BGp	BGO	Stratification Index % G_b/G_{bp}
				$\sum q_i G_i O_i$		$\sum q_i G_i$	$\sum q_i G_i (O_i - 1)$	G_{bp}	$G_b - G_{bp}$	
			1=2+3	2=4+5	3=6+7	4	5	6	7	100*(3/6)
<i>Australia</i>	7	2001	0.3199	0.3153	0.0046	0.3173	-0.0020	0.0342	-0.0296	3.16
(2001-2014)	7	2014	0.3315	0.3265	0.0050	0.3291	-0.0026	0.0404	-0.0354	3.81
<i>Canada</i>	10	2000	0.3029	0.2928	0.0101	0.2977	-0.0048	0.0542	-0.0441	4.89
(2000-2013)	10	2013	0.3212	0.3109	0.0103	0.3159	-0.0050	0.0568	-0.0465	5.16
<i>Chile</i>	13	2000	0.5270	0.4933	0.0337	0.5144	-0.0211	0.1347	-0.1010	12.21
(2000-2015)	13	2015	0.4674	0.4403	0.0271	0.4566	-0.0163	0.1164	-0.0892	10.55
<i>France</i>	8	2000	0.2781	0.2605	0.0176	0.2689	-0.0084	0.0624	-0.0449	5.32
(2000-2010)	8	2010	0.2870	0.2807	0.0063	0.2834	-0.0027	0.0458	-0.0395	4.22
<i>Germany</i>	16	2000	0.2591	0.2521	0.0070	0.2550	-0.0029	0.0394	-0.0324	3.53
(2000-2015)	16	2015	0.2966	0.2889	0.0077	0.2926	-0.0037	0.0473	-0.0396	4.32
<i>Italy</i>	20	2000	0.3347	0.2746	0.0601	0.3049	-0.0304	0.1340	-0.0739	10.43
(2000-2014)	20	2014	0.3322	0.2866	0.0456	0.3073	-0.0207	0.1133	-0.0677	8.84
<i>Mexico</i>	32	2000	0.4998	0.4106	0.0892	0.4587	-0.0481	0.2102	-0.1210	16.91
(2000-2012)	32	2012	0.4720	0.4111	0.0609	0.4463	-0.0352	0.1660	-0.1052	14.03
<i>Poland</i>	16	1999	0.2917	0.2857	0.0059	0.2887	-0.0029	0.0450	-0.0390	4.20
(1999-2016)	16	2016	0.2899	0.2835	0.0064	0.2863	-0.0028	0.0461	-0.0397	4.25
<i>Spain</i>	18	2004	0.3213	0.3015	0.0199	0.3112	-0.0097	0.0799	-0.0601	6.98
(2004-2013)	18	2013	0.3454	0.3138	0.0316	0.3287	-0.0149	0.1043	-0.0728	8.76
<i>United Kingdom</i>	11	2004	0.3543	0.3427	0.0116	0.3486	-0.0060	0.0752	-0.0636	6.95
(2004-2013)	11	2013	0.3336	0.3233	0.0103	0.3285	-0.0052	0.0645	-0.0542	5.93
<i>United States</i>	51	2000	0.3721	0.3648	0.0073	0.3684	-0.0036	0.0517	-0.0444	4.80
(2000-2016)	51	2016	0.3860	0.3796	0.0064	0.3824	-0.0028	0.0498	-0.0435	4.62

Note: The intragroup (2) and between group (3) components of overall inequality (1) are decomposed to account for the effects of overlapping on within (5) and between (7) components. IG is an average of regions' Ginis, weighted by income shares (q_i), G_{bp} is the between-groups component (based on Pyatt 1967), O_i is the overlapping index of the group i with the entire population. The last column presents the stratification index, which increases as interregional inequality rises and also when intraregional inequality falls. Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

Figure A.1. Box-and-whisker plot for the mean log deviation (L) in OECD countries: 1999-2016



Notes: (1) AU=Australia, CA=Canada, CL=Chile, FR=France, DE=Germany, IT=Italy, MX=Mexico, PL=Poland, ES=Spain, UK=United Kingdom, US=United States; (2) For each country, the *box-and-whisker plot* on the left corresponds to the initial year, while the one on the right is relative to the final year. Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2016). Luxembourg: LIS.

Chapter 2. Unemployment Shocks and Material Deprivation in the European Union: A Synthetic Control Approach

1. Introduction

Should we expect a large increase in material deprivation and a worsening of living conditions right after an unemployment shock? Are material deprivation measures as sensitive to drastic changes in macroeconomic conditions as monetary poverty measures? In this chapter, we try to determine the effects of an unemployment shock on a composite measure of material deprivation.

One of the greatest advances in the research on poverty has been the development of new methods for measuring material deprivation. As different authors have shown, the possibility of combining different partial indicators into an index that synthetically measures the level of deprivation can be more effective than a wide range of indicators to capture public and political attention. Some institutions have, in fact, incorporated the concept of material deprivation into their indicators of poverty and exclusion. The European Union, for instance, used the AROPE rate – the share of the total population at risk of poverty or social exclusion – as its main indicator for monitoring the EU 2020 Strategy poverty target. The measure corresponds to the sum of persons who are at risk of poverty, severely materially deprived or living in a household with very low work intensity.

While advances in the characterization of this phenomenon have been considerable, the evidence on its determining factors is considerably less robust. For instance, while numerous studies have explored inequality or certain forms of poverty, we still know very little about how these indicators change as the economic cycle changes. The extensive empirical literature on the effects of changes in macroeconomic conditions on income distribution [Blank and Blinder (1986), Cutler and Katz (1991), Jäntti (1994), Smeeding

et al. (2011), Meyer and Sullivan (2011), Ayala et al. (2017)] has had much less development in the case of material deprivation.

One of the reasons for this asymmetry lies in the a priori more static nature of material deprivation measures relative to those of income inequality or monetary poverty. As the extensive literature on capabilities has recognized, while the latter could be considered flow variables, the former are more similar to stock variables. However, this reasoning does not seem to correspond well with what happened in several countries during the so-called Great Recession. In many rich countries and especially in Europe, deprivation indicators grew remarkably (Duiella and Turrini, 2014).

Such a difference in the extent of this strand of the literature does not mean that the relationship between unemployment and material deprivation has not been addressed. Figari (2012) analyzed the drivers of deprivation in eleven European countries and found strong impacts of unemployment in most of them. Some studies have also used multilevel techniques to test the possible effects on unemployment on differences in multiple deprivation in EU countries (Whelan and Maître, 2012, 2013). Visser et al. (2014) found that the stronger the rise in the unemployment rate, the more economic deprivation individuals experience. Bárcena-Martín et al. (2014) found that long-term unemployment rates have a significant effect on deprivation when only macro-level variables are considered but that this effect vanishes when micro-level variables are introduced. More recently, Verbunt and Guio (2019) also used single- and multilevel methods to confront the respective within and between-country explanatory power of both types of models in the measuring of severe multiple deprivation. The authors also employed the Shapley decomposition method to compare the relative contributions of independent variables at the household and country levels to find that macroeconomic and institutional variables explain a large share of between-country differences in the risk of material deprivation.

None of these studies specifically analyze what happens when a significant change in the unemployment rate occurs over a very short time period, such as those changes that took place in the so-called Great Recession or in the more recent downturn resulting from COVID-19. During the Great Recession, unemployment rates in some European countries more than tripled and in some cases exceeded the 20% level. This chapter analyzes how material deprivation responds to drastic changes in unemployment levels taking as reference the unemployment shocks registered in some European Union countries during the Great Recession.

The reasons for focusing on EU countries are varied. First, while most European countries were exposed to significant unemployment changes, in some its growth was much faster and unemployment rates reached their highs. Second, the European Monetary Union was designed by assigning the role of fiscal stabilization to national budgets with very few community counterparts. A common monetary policy was not enough to accommodate the needs of all states against asymmetric shocks. The fact that there was no common stabilizing mechanism in the form of a European unemployment insurance made the responses of social conditions to unemployment shocks very different in each country (Ábráham et al., 2018).

To address this question, we apply the synthetic control methodology, which has not yet been widely used in the field of distributive studies. We use this approach to identify the impact of unemployment shocks on material deprivation and conduct different sensitivity analyses to test the results. As our most important factual finding, we find that unemployment shocks have a rapid and significant effect on material deprivation in countries where they take place (Greece and Spain). This conclusion holds even when extending the period of analysis, changing the indicator of material deprivation or modifying the definition of unemployment shock.

This chapter is structured as follows. In the following section, we introduce our definitions of unemployment shocks and material deprivation. In the third section, we present our empirical strategy. Section 4 presents our main results. The article ends with a brief list of conclusions.

2. Unemployment shocks and material deprivation in the EU-28

2.1. Unemployment shocks

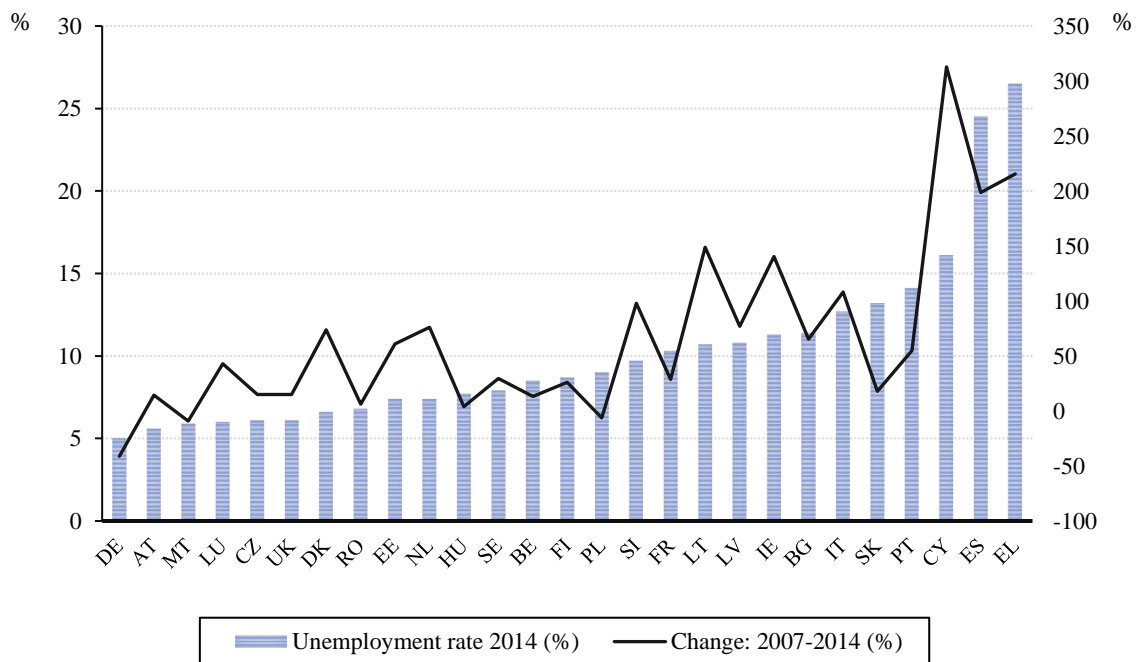
As the main goal of this chapter is to evaluate the effects of unemployment shocks on material deprivation rates within the EU-28, a necessary first step is to define this event. In practice, there is not a sufficient consensus on an empirically testable definition for unemployment shocks. It is worth mentioning, as an example, Burda and Hamermesh's (2009) tentative definition as the difference between the current year's unemployment rate and the unemployment rate averaged over the previous five years. The authors interpret this as the cyclical shock to the labor market in the corresponding area or country. In a similar vein, Dibooglu and Enders (2001) use one standard deviation of the unemployment rate to test whether real wages asymmetrically respond to unemployment shocks.

Other studies that explicitly try to estimate the effects of unemployment shocks on dimensions of well-being do not use such specific definitions. Aaberge et al. (2000) take as a reference general changes in unemployment in Nordic countries from the early 1980s to the mid-1990s. Christelis et al. (2015) define an individual unemployment shock as a significant change in consumption with the transition to unemployment. Alt et al. (2017) define unemployment shocks by comparing expectations of unemployment for a calendar year – asking respondents to provide their best estimate of the probability that they will

experience unemployment in a given year – to actual unemployment with a larger share of the year involving unemployment denoting a negative unemployment shock.

In the absence of a standard definition, we formulate a new proposal focused on the economic and financial crisis that started in 2007/2008 and our sample of countries (EU-28). As shown by Figure 1, between 2007 and 2014, unemployment grew in practically all EU countries. However, differences in growth rates were considerable. While in Lithuania, Latvia, Cyprus, Spain and Greece, the rate more than doubled, in ten countries it grew by less than 20%. There is also broad variability in the resulting unemployment rates. While in Spain and Greece the unemployment rate increased to above 20%, in sixteen countries it remained at below 10%.

Figure 1. Unemployment rates in EU countries



Notes: (1) The axis on the left denotes the unemployment rate in 2014; the one on the right denotes the change from 2007-2014; (2) BE: Belgium; BG: Bulgaria; CZ: Czech Republic; DK: Denmark; DE: Germany; EE: Estonia; IE: Ireland; EL: Greece; ES: Spain; FR: France; IT: Italy; CY: Cyprus; LV: Latvia; LT: Lithuania; LU: Luxembourg; HU: Hungary; MT: Malta; NL: Netherlands; AT: Austria; PL: Poland; PT: Portugal; RO: Romania; SI: Slovenia; SK: Slovakia; FI: Finland; SE: Sweden; UK: United Kingdom. Source: Own elaboration from the Eurostat Database.

We define a country as suffering an unemployment shock – starting in approximately 2007 – when the two following circumstances occur: (a) over 200% growth in the unemployment rate from 2007-2014 and (b) an unemployment rate exceeding 20% in 2014. When applying these criteria, two EU-27 countries are identified as being affected by an unemployment shock: Spain and Greece.¹⁹ These countries are thus considered as the countries affected by the event studied. The remaining EU-27 countries will form the control group (*donor pool*) in our evaluation of the effects of unemployment shocks on material deprivation.

2.2. Material deprivation in EU countries

Relative to standardized relative measurement procedures for monetary poverty, the range of composite indices of material deprivation available is broad. In recent years, several works have aimed at more precisely identifying the extension and characteristics of multidimensional deprivation [Chakravarty and D'Ambrosio (2003), Atkinson (2003), Bourguignon and Chakravarty (2003), Dutta et al. (2003), Deutsch and Silber (2005), Duclos et al. (2006). Guio and Marlier (2013), Boarini and Mira D'Ercole (2013), Alkire et al. (2015)]. These approaches have been developed in an attempt to answer to the two main questions that the measurement of this phenomenon focuses on. The two standard ways of measuring material deprivation include the selection of partial deprivation indicators (*items*) and the calculation of a synthetic index that combines these partial indicators into a single value.

The policy-oriented nature of our research forces these elections to reflect as closely as possible the official items proposed by EU institutions and the indicators recommended by these institutions for monitoring the problem. We use the definition for standard

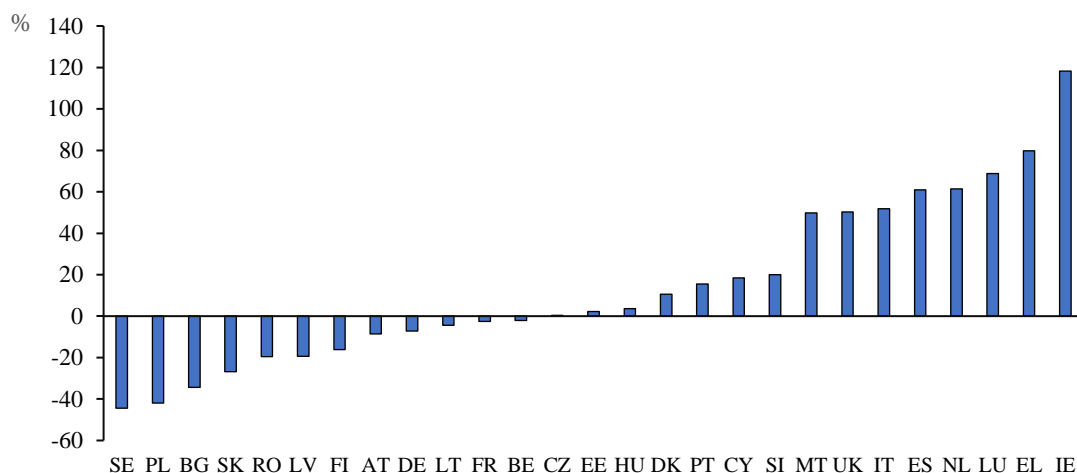
¹⁹ We have not used the EU-28 as the focus of our analysis due to a lack of data for Croatia.

material deprivation defined by the European Commission and the index currently employed under the Europe 2020 strategy (together with low income and very low work intensity). This definition – and our analysis – takes as a starting point a subset of material deprivation indicators available in *European Statistics of Income and Living Conditions* (EUSILC) and the deprivation index included in Eurostat statistics. This is defined as the percentage of the population that cannot afford at least three of the following nine items: (1) to pay their rent, mortgage or utility bills; (2) to keep their home adequately heated; (3) to pay for unexpected expenses; (4) to eat meat or protein regularly; (5) to go on holiday; and (6) to have a television set, (7) washing machine, (8) car, or (9) telephone.

This standard index presents certain limitations that reduce its usefulness for the analysis of levels and changes in material deprivation in European countries. On one hand, as stressed by Martínez and Navarro (2016), four of the nine indicators are consumer durables whose possession is highly generalized in Western Europe to the point at which their enforced lack is typically rare. The index has also been criticized for its inclusion of durable goods, which may reduce the index's sensitivity to the economic cycle. In our case, this issue, more than posing a disadvantage, serves as an important argument to try to test for whether the effect of an unemployment shock can be so great that it can increase an indicator with limited expected variations.

Figure 2 shows how the rate of standard material deprivation changed over the period studied for the identification of unemployment shocks. The most important finding illustrated in the figure is the considerable heterogeneity of the indicator's behavior in EU countries in the period studied. It cannot be concluded that during the Great Recession deprivation increased in a generalized way nor that it was a problem of a fundamentally static nature.

Figure 2. Rate of growth in the standard material deprivation rate in EU countries (2007-2014)



Note: BE: Belgium; BG: Bulgaria; CZ: Czech Republic; DK: Denmark; DE: Germany; EE: Estonia; IE: Ireland; EL: Greece; ES: Spain; FR: France; IT: Italy; CY: Cyprus; LV: Latvia; LT: Lithuania; LU: Luxembourg; HU: Hungary; MT: Malta; NL: Netherlands; AT: Austria; PL: Poland; PT: Portugal; RO: Romania; SI: Slovenia; SK: Slovakia; FI: Finland; SE: Sweden; UK: United Kingdom.

Source: Own elaboration from the Eurostat Database.

In a third of the countries the change was relatively minor and in almost the same number there was a significant reduction (greater than 15%) with a marked drop observed in Sweden and Poland. On the other hand, in a meaningful proportion of countries, deprivation increased by more than 50%; in particular, the rate of material deprivation more than doubled in Ireland.

3. Methodology

To assess the effects of unemployment shocks on material deprivation in EU countries, we apply the Synthetic Control Methodology (SCM). This technique was originally proposed by Abadie and Gardeazabal (2003) as a means to analyze the effects of terrorism on GDP per capita, and with Abadie et al. (2010) the generalized application of the methodology was established. Since this work, the method has been widely used to examine effects caused by a broad variety of specific events (see Craig (2015) for a review).

The SCM has been applied in numerous studies ranging from an evaluation of the economic impact of natural disasters (Cavallo et al., 2013) to an assessment of the effect of institutional interventions on a population's consumption and welfare (Abadie et al., 2010), among others. Within the framework of public policy evaluation, the SCM has been consolidated as one of the most powerful methodologies for conducting impact evaluations in the last decade. Nevertheless, and as far as we are concerned, practically no studies have implemented this method to study poverty and inequality (one exception is Grier and Maynard, 2016).

The most important advantages associated with the SCM are the following. (1) A number of public policy interventions affect aggregate units. The management of and access to macro-level data are more common and simple than the treatment of micro-level data, and there are many series available at that level of aggregation. (2) Regressions applied to samples of countries have been frequently questioned (Harberger, 1987). Such regressions involve carrying out comparisons of entities with potentially different characteristics. In applying the SCM methodology, we resort to data-driven procedures that reduce the discretion in the choice of comparison control units and that allow us to create appropriate comparison groups. (3) The SCM does not involve making strict hypotheses to make precise estimations as with other quantitative techniques such as those of the difference-in-differences approach.²⁰ (4) Finally, the standard results inform us of the individual contributions of each *donor* units that form the synthetic control group.

Among restrictions applied, it is important to point out the following. (1) Some units in the *donor pool* should present both higher and lower values in predictor variables in comparison to that affected by the intervention. Otherwise, it would be impossible to

²⁰ See Abadie et al. (2010) for a more detailed explanation.

appropriately recreate the unit of treatment. (2) In the preintervention period, units of control should have predictor values comparable to those of the treated unit.²¹ In addition, these variables should have an approximately linear effect on the result. (3) It has been recommended that using all preintervention outcomes together with covariates as predictors be avoided (Kaul et al., 2018). Otherwise, one would restrain the predictive power of the remaining covariates. (4) Finally, the statistical inference procedure is much less formal than those implemented by other quantitative methods and more traditional techniques.

3.1. Model formalization

Initially, let us assume that there are $J + 1$ countries where $j = 1$ denotes the country treated and $j = 2, \dots, J + 1$ denote untreated or control countries (the EU-27 members not conditioned by the unemployment shock). It is thus assumed that a single country is affected by the event considered and that J units are available to contribute to the synthetic control (*donor pool*).

We also assume that $1, 2, \dots, T_0$ are the periods preceding the unemployment shock already referenced, and $T_{0+1}, T_{0+2}, \dots, T$ are the post-treatment periods. Two results in relation to the outcome of interest are possible: (1) Y_{1t}^N is the result observed for country i at time t if country i was not affected by the unemployment shock and (2) Y_{1t}^I is the observed result for country i at time t if country i was exposed to the event. The magnitude of the effect, impact (α), is simply the difference between the two potential results for the periods $T_{0+1}, T_{0+2}, \dots, T$:

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N \quad (1)$$

²¹ We proceed this way to avoid interpolation bias and overfitting (Abadie et al., 2015; Grier and Maynard, 2016).

For the affected country, Y_{1t}^N cannot be observed in the post-treatment periods. Data are available for the actual path of the outcome (Y_{1t}^I), but it is unknown what would have happened with that trajectory if the treated country had not suffered effects of the unemployment shock. Therefore, we look for an estimate of Y_{1t}^N that, following Abadie et al. (2010), is given by a linear factor model. This is necessary to quantify the effect of the event by calculating the difference specified in (1).

To find optimal weights, Abadie and Gardeazabal (2003) defined a $(K \times 1)$ vector X_1 of the preunemployment shock values of K predictors of the outcome variable and a $(K \times J)$ matrix X_0 , which measures the values of the same variables for the *donor pool*. The vector of optimal weights referring to the control countries, W^* , is the one that minimizes the following problem:

$$\| X_1 - X_0 W \|_v = (X_1 - X_0 W)' V (X_1 - X_0 W) \quad (2)$$

where $W^* = (w_1^*, w_2^*, \dots, w_{j+1}^*)'$ is a $(J \times 1)$ vector of non-negative weights that sums to one, and V is a symmetric, diagonal matrix of non-negative components that represents the relative importance of the selected predictors.

Once we have obtained the matrix $W^*(V^*)$ formed by the estimated optimal weights that each country of the control group receives for the design of the synthetic control unit, it is enough to apply these weights in (1) to obtain the estimate of the effect of the unemployment shock:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \quad (3)$$

3.2. Inference

With the SCM methodology, neither confidence intervals nor statistical significance parameters are calculated, which are typical procedures in an inference analysis. Alternatively, the SCM offers complementary options also known as *falsification* tests. With “in-space” placebos, each country integrating the original *donor pool* is separately conceived as a treated entity and the SCM is applied as if countries were affected by the unemployment shock (Abadie et al., 2010; Abadie et al., 2015).

By applying this iterative mechanism, we obtain a distribution of estimated placebo treatment effects for all countries in which no event occurred. Considering that none of these control countries has been influenced by the unemployment shock studied, we should only observe great disparities between these *placebo* countries and their corresponding synthetic control randomly and in sporadic cases. A more accurate mechanism for identifying the significance of the results is based on the Root Mean Squared Prediction Error (RMSPE), which is the index typically used to assess the goodness of fit when applying the SCM. It measures for a given unit of analysis the fit – or lack thereof – between the actual outcome variable and its synthetic counterpart. In other words, it represents the distance or discrepancy between the path drawn by each variable. Formally, it is defined as follows:

$$RMSPE = \sqrt{\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2} \quad (4)$$

Ultimately, we must calculate the ratio between the postintervention RMSPE (the average for 2009-2014) and preintervention RMSPE (the average for 2004-2008) and determine how many control countries present an effect as large as that observed in the treated country (Spain or Greece). Within this ratio, the numerator quantifies the magnitude of

the impact (the higher the RMSPE, the greater the impact) and the denominator quantifies the goodness of fit (the lower the RMSPE, the better the fit).

4. Data

We use annual country-level data from Eurostat for 2004-2014 for EU-27 countries. As EU-SILC begins in 2004 (corresponding to 2003 income data), we include the five years preceding the event analyzed. The endpoint is set to 2014 based on one of the two conditions we consider in defining the unemployment shock.

The two countries considered to be affected by the event – unemployment shock – are Greece and Spain. First, we use Spain as our unit of treatment. Next, the same analysis is conducted for Greece. The remaining EU-27 countries form the control group (*donor pool*) in our evaluation of effects on material deprivation. The defined event – the unemployment shock – captures the effects of the economic cycle in all EU-27 countries, but we can quantify the intensity of impacts in the countries where there is a differential increase in the evolution of the two parameters chosen as a reference to define the unemployment shock.

As stressed above, the successful use of the SCM requires an important assumption to be fulfilled: it is essential to dispense with all units suffering the effects of a similar event in some years of the preintervention period – in our case, 2004 to 2008. If these were included, they could interfere with and condition the true effects of the intervention examined. Of the considered countries, Cyprus is excluded from the group of potential controls to satisfy one of the two proposed requirements for defining an unemployment shock.

According to the above definition, the unemployment shock took place in 2008, so we have a pretreatment period of five years (we observe effects from 2009 onwards) and a

post-treatment period of six years. We study effects on the standard material deprivation rate and, as a measure of sensitivity, we also consider the severe material deprivation rate and use a *counting* approach.²² For the predictors considered, we use the Gini index, work intensity, GDP per capita, social protection benefits as a percentage of GDP, and the lagged outcome variable for several periods preceding the unemployment shock (see Table A.1 in the appendix).

Some authors have stressed that the SCM might be an adequate methodology with a fairly short preintervention time period inasmuch as the duration of the post-treatment period is reasonably long and the fit between the synthetic and treated units is adequate (Carling and Li, 2016), as is the case in our empirical exercise. Barreix and Corrales (2019), for instance, use a period of four years for their preintervention period, and Heim and Lurie (2014) also use a relatively short pretreatment period (eight years) to analyze the effects of a Massachusetts health reform on self-employment.

With respect to the number of predictors used, it should be underscored that increasing their number does not necessarily improve the fit, and similarly eliminating some of them inevitably worsens the fit (McClelland and Gault, 2017). Additionally, regarding the predictors considered, one of the most common practices in the application of this methodology involves the use of the lagged outcome variable (Abadie et al., 2010). By including several lags of the outcome variable, we measure the effect of other predictors. This strategy somehow mitigates the effects of not incorporating relevant predictors into the analysis. However, there is no consensus on what a suitable number of lags is.

Some authors have drawn attention to the desirability of encompassing all outcome lags available as predictors. Furthermore, they believe that including other covariates has

²² This option simply involves counting the number of items a household is deprived of while assigning the same weight to each item (Mayer y Jencks, 1989; Atkinson, 2003; Bárcena-Martín et al., 2013).

hardly any influence on the final estimates (Athey and Imbens, 2006). On the other hand, other scholars claim that only using the lags of the outcome variable is not the best solution (Kaul et al., 2016). Without any additional predictor, the estimated model cannot be supported by economic theory and does not have any justification. Ferman et al. (2016) recommend working with different specifications, using several combinations of lags and generating all possible results. This latter option is the one we use in this investigation.

We initially determined which model provides a better fit (the one that presents the lowest RMSPE) when selecting a maximum of three lags of the outcome variable from the set of predictors.²³ For Spain and Greece, the best model is the one that selects the lags of standard material deprivation rates corresponding to 2008, 2007 and 2005 (see Tables 1a and 1b). This initial specification helps us then choose the best model when we use two lags and when we only use one²⁴ (insert Table 2).

5. Results

We are interested in determining how the standard material deprivation rates of Spain and Greece would have evolved in absence of the unemployment shock that, according to the definition set out in the above section, took place in 2008. For this purpose, we use a combination of different European countries to construct a synthetic control unit for each of these countries that resembles as much as possible the actual evolution of the material deprivation rate before the outset of the shock. The subsequent track of this counterfactual Spain (and Greece) without effects of the treatment is then compared to the actual path.²⁵

²³ We rule out using four or five lags for the reasons stated above.

²⁴ Table A.2 in the appendix shows the country weights in the synthetic units.

²⁵ In cases where multiple units are affected by the event of interest, as is the case that concerns us, the SCM can be applied to each affected unit separately or to an aggregate of all units involved (Abadie et al., 2015). As it would not make much sense to consider Spain and Greece as a single unit of treatment, we developed two exercises in parallel.

Table 1a. Best fit for the standard material deprivation rate as the outcome of interest using three lags (SPAIN)

	Predictor variables	All possible combinations of choosing 3 lags from the 5 years of the preunemployment shock period									
		[ES_1]	[ES_2]	[ES_3]	[ES_4]	[ES_5]	[ES_6]	[ES_7]	[ES_8]	[ES_9]	[ES_10]
Standard material deprivation	Gini index	X	X	X	X	X	X	X	X	X	X
	Work intensity (%)	X	X	X	X	X	X	X	X	X	X
	Ln (GDP per capita)	X	X	X	X	X	X	X	X	X	X
	Social protection benefits (% GDP)	X	X	X	X	X	X	X	X	X	X
	Standard material deprivation rate 2008	X	X	X	X	X	X	—	—	—	—
	Standard material deprivation rate 2007	X	X	X	—	—	—	X	X	X	—
	Standard material deprivation rate 2006	X	—	—	X	X	—	X	X	—	X
	Standard material deprivation rate 2005	—	X	—	X	—	X	X	—	X	X
	Standard material deprivation rate 2004	—	—	X	—	X	X	—	X	X	X
	RMSPE	0.170	0.078	0.104	0.162	0.163	0.187	0.210	0.146	0.226	0.211

Note: ES = Spain.

Source: Own elaboration.

Table 1b. Best fit for the standard material deprivation rate as the outcome of interest using three lags (GREECE)

	Predictor variables	All possible combinations of choosing 3 lags from the 5 years of the preunemployment shock period									
		[EL_1]	[EL_2]	[EL_3]	[EL_4]	[EL_5]	[EL_6]	[EL_7]	[EL_8]	[EL_9]	[EL_10]
Standard material deprivation	Gini index	X	X	X	X	X	X	X	X	X	X
	Work intensity (%)	X	X	X	X	X	X	X	X	X	X
	Ln (GDP per capita)	X	X	X	X	X	X	X	X	X	X
	Social protection benefits (% GDP)	X	X	X	X	X	X	X	X	X	X
	Standard material deprivation rate 2008	X	X	X	X	X	X	—	—	—	—
	Standard material deprivation rate 2007	X	X	X	—	—	—	X	X	X	—
	Standard material deprivation rate 2006	X	—	—	X	X	—	X	X	—	X
	Standard material deprivation rate 2005	—	X	—	X	—	X	X	—	X	X
	Standard material deprivation rate 2004	—	—	X	—	X	X	—	X	X	X
	RMSPE	0.454	0.314	0.461	0.387	0.564	0.564	0.571	0.639	0.592	0.766

Note: EL = Greece.

Source: Own elaboration.

Table 2. Choice of model: sensitivity test to different specifications

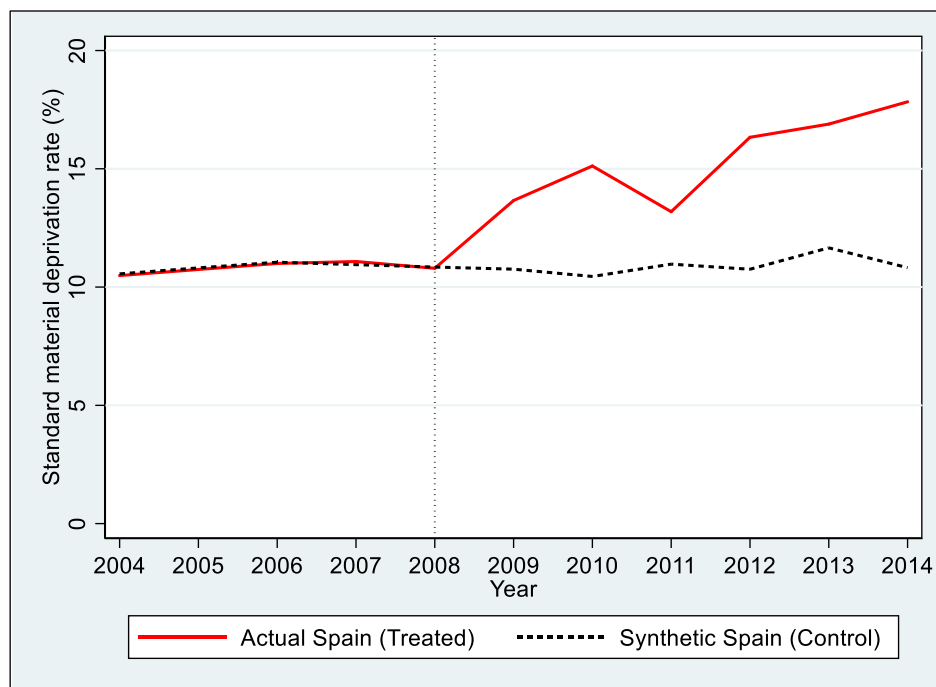
Predictor variables		3 lags	2 lags			1 lag		
		[1]	[2]	[3]	[4]	[5]	[6]	[7]
Standard material deprivation (SPAIN)	Gini index	X	X	X	X	X	X	X
	Work intensity (%)	X	X	X	X	X	X	X
	Ln (GDP per capita)	X	X	X	X	X	X	X
	Social protection benefits (% GDP)	X	X	X	X	X	X	X
	Standard material deprivation rate 2008	X	X	—	X	X	—	—
	Standard material deprivation rate 2007	X	X	X	—	—	X	—
	Standard material deprivation rate 2005	X	—	X	X	—	—	X
	RMSPE	0.078	2.391	0.211	0.209	3.237	2.979	0.211
Standard material deprivation (GREECE)	Gini index	X	X	X	X	X	X	X
	Work intensity (%)	X	X	X	X	X	X	X
	Ln (GDP per capita)	X	X	X	X	X	X	X
	Social protection benefits (% GDP)	X	X	X	X	X	X	X
	Standard material deprivation rate 2008	X	X	—	X	X	—	—
	Standard material deprivation rate 2007	X	X	X	—	—	X	—
	Standard material deprivation rate 2005	X	—	X	X	—	—	X
	RMSPE	0.313	1.559	0.876	0.752	1.180	1.866	0.876

Source: Own elaboration.

5.1. Main results

Regarding what constitutes a good fit or how to appraise similarities, the most direct and immediate option is to resort to the *eyeball test* by comparing the evolution of the material deprivation rate in the treatment country (Spain and Greece) to that of the control group. Starting with Spain, our first result is that the evolution of actual Spain and its synthetic counterpart practically overlap in the three cases analyzed²⁶ with the first requirement being met if we want to rely on estimates of the causal impact of the unemployment shock. From the moment that the unemployment shock occurs, the two curves separate (see Figure 3a).

Figure 3a. Trends of the material deprivation rate: SPAIN and synthetic SPAIN

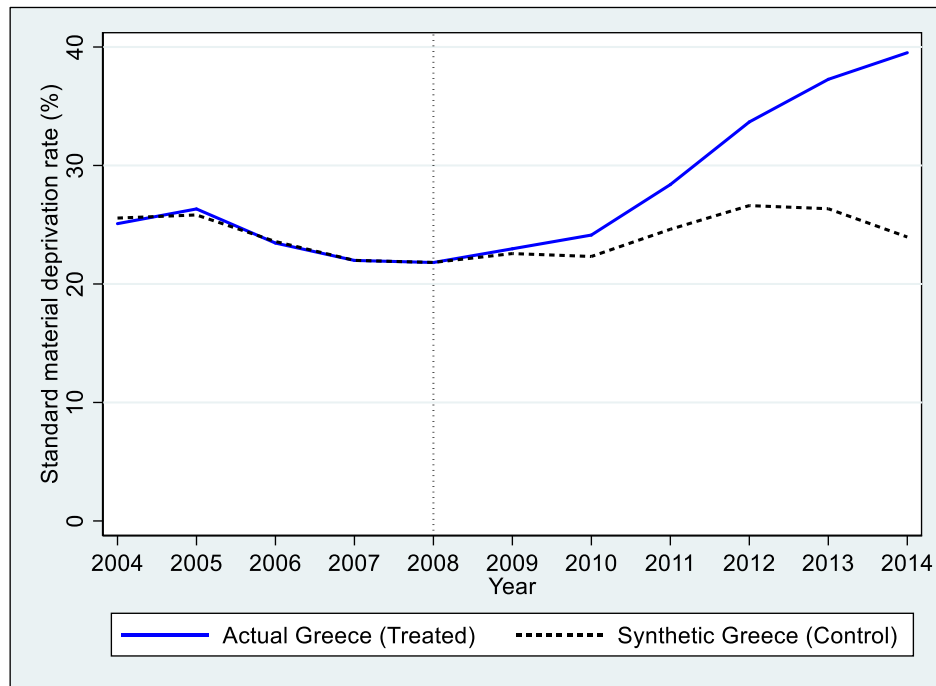


Source: Own elaboration from the Eurostat database.

²⁶ For both Spain and Greece, we only include the figure corresponding to specification [1], which presents the lowest RMSPE and which is the model we follow henceforth. The other figures are available upon request.

As observed for Spain, what first draws our attention when examining Greece is the accuracy of the pretreatment fit across the different specifications. The three figures reveal extraordinarily homogeneous behavior, providing an initial guarantee for subsequent estimates (see Figure 3b).

Figure 3b. Trends of the material deprivation rate: GREECE and synthetic GREECE



Source: Own elaboration from the Eurostat database.

Second, another precondition relates to the similarities of real predictor values for the treated country to those of the synthetic version. Table 3a shows these values for the three models under analysis – the specifications with the lowest RMSPE including one, two and three lags. While not all of them match exactly, the approximation can be accepted as reasonably good.

Table 3a. Predictor means: Results for SPAIN

Predictor variables	Actual Spain	Synthetic Spain		
		[1]	[4]	[7]
Gini index	31.88	28.50	27.16	27.27
Work intensity (%)	59.73	61.80	60.33	60.80
Ln (GDP per capita)	10.02	10.51	10.08	10.03
Social protection benefits (% GDP)	19.98	23.36	20.48	20.04
Standard material deprivation rate 2008	10.79	10.85	11.05	—
Standard material deprivation rate 2007	11.08	10.95	—	—
Standard material deprivation rate 2005	10.74	10.81	10.81	10.79

Note: Numbers enclosed in square brackets refer to the econometric specifications used.
Source: Own elaboration from the Eurostat database.

For Greece, it is important to also note that the predictor means are again very close to the actual values. On the other hand, we find that some models that in principle provide a better fit – a lower RMSPE – show a greater mismatch in their predictor values. This is due to the predictive power assigned to each of them, since it varies depending on the specification used and with the total number of variables involved in the estimate. Achieving the best possible fit regardless of these considerations is what truly matters (see Table 3b).

The indicators on the fit of the estimates therefore confirm the validity of our evaluation of the impact of the unemployment shock in both countries on the standard material deprivation rates. The gap between the actual rates and those of the synthetic units reports and quantifies the impact. The drastic increase in unemployment denotes a significant and rapid increase in material deprivation in both Spain and Greece.

Table 3b. Predictor means: Results for GREECE

Predictor variables	Actual Greece	Synthetic Greece		
		[1]	[4]	[7]
Gini index	33.64	32.44	33.67	31.93
Work intensity (%)	58.60	62.89	58.66	59.19
Ln (GDP per capita)	9.88	9.88	9.85	9.88
Social protection benefits (% GDP)	23.55	19.28	22.52	23.57
Standard material deprivation rate 2008	21.81	21.82	20.72	—
Standard material deprivation rate 2007	21.99	21.99	—	—
Standard material deprivation rate 2005	26.33	25.83	25.63	25.58

Note: Numbers enclosed in square brackets refer to the econometric specifications used.
Source: Own elaboration from the Eurostat database.

For Spain, the double-rip recession and its W-shaped recovery path seem to be the main explanatory factor behind the sharp fall in the actual material deprivation rate observed for 2011.²⁷ With the exception of this drop in 2011, in the remaining years the rate increased. In fact, between 2008 and 2014, there was a dramatic rise of 65%. Martínez and Navarro (2014) drew attention to this issue – the sudden increase in the material deprivation rate during the Great Recession – and highlighted the early impact of material deprivation on the main indicators. According to the authors, one of the first and most intense effects of the crisis involved a reduction in the capacity to face unexpected expenses. This item increased from 36% in 2008 to 42% in 2009 and then continued to grow until it reached 48% in 2013. Likewise, the authors find that the number of families declaring they could not go on holiday at least one week a year increased from 30% in 2008 to 36% in 2009 and then to 42% in 2013. These factors caused a notable increase in the material deprivation rate during the treatment period (2009-2014) (see Table 4).

²⁷ See Figure 3a.

Table 4. Impact results (estimated gap) (%)

Year	Treatment unit: SPAIN			Treatment unit: GREECE		
	[1]	[4]	[7]	[1]	[4]	[7]
2009	2.903	1.823	1.596	0.404	3.306	3.225
2010	4.671	3.498	3.144	1.808	2.310	2.881
2011	2.218	0.396	0.003	3.774	6.960	7.383
2012	5.579	2.459	1.940	7.058	9.905	10.743
2013	5.231	2.935	2.299	10.921	13.382	14.971
2014	7.011	3.775	2.966	15.553	18.379	19.944
Average	4.602	2.898	2.389	7.823	9.040	9.858

Note: Numbers enclosed in square brackets refer to the econometric specifications used.

Source: Own elaboration.

In the absence of the 2008 unemployment shock and according to the estimates made, the scenario could have been a very different one. The results of the best model show that on average, the standard material deprivation rate would have been 4.6% lower than that actually observed. In addition, and with the exception of 2011, the impact seems to follow a growing trend with 2014 being the year in which the impact reaches its maximum. This last finding is extensible to the three proposed models.

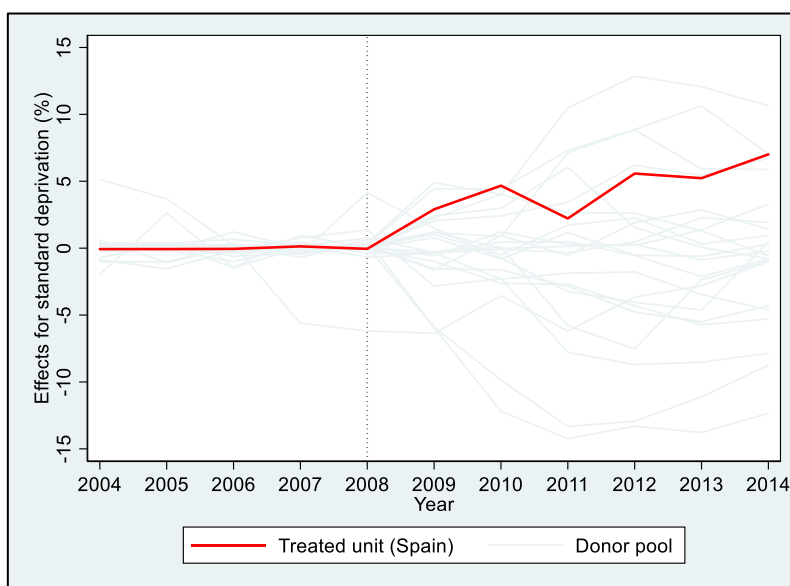
For Greece, Papanastasiou and Papatheodorou (2018), in the same way as Martínez and Navarro (2014) did for Spain, found that more than half of the population in 2015 experienced difficulties paying unexpected financial expenses and could not afford a week-long holiday. Both studies coincide in finding that these two items were the most sensitive to effects of the crisis and heavily conditioned the evolution of the actual material deprivation rate. Here, an exception is observed in 2009 when the effects of the Great Recession on the deprivation rate were barely noticeable. Nonetheless, the growth occurring from 2009 to 2014 rose to 72% (almost 17 percentage points). On average, the

impact is roughly 8% in the model with three lags, 9% in the model with two lags and close to 10% in the model including only one lag. All of them also share a remarkable feature: extraordinary growth from 2011 onwards reaching its greatest increase in 2014. For 2013 and 2014, the figures provided by all models exceed 10%.

5.2. Inference

As stated above, we are interested in measuring similarities between the actual trajectory of the material deprivation rate and the path described by the same variable for the comparison group or synthetic unit. The ratio between the postunemployment shock RMSPE and the preunemployment shock RMSPE allows us to evaluate the gap in Spain relative to those of the remaining countries of the *donor pool*. Only Poland, where the postevent RMSPE is roughly 70 times the RMSPE of the pre-event period, remains ahead of Spain, where the ratio is quite similar (68). This information confirms that the good fit shown by the *eyeball test* is not a product of chance. We can also check these figures by applying *placebo runs* (see Figure 4a).

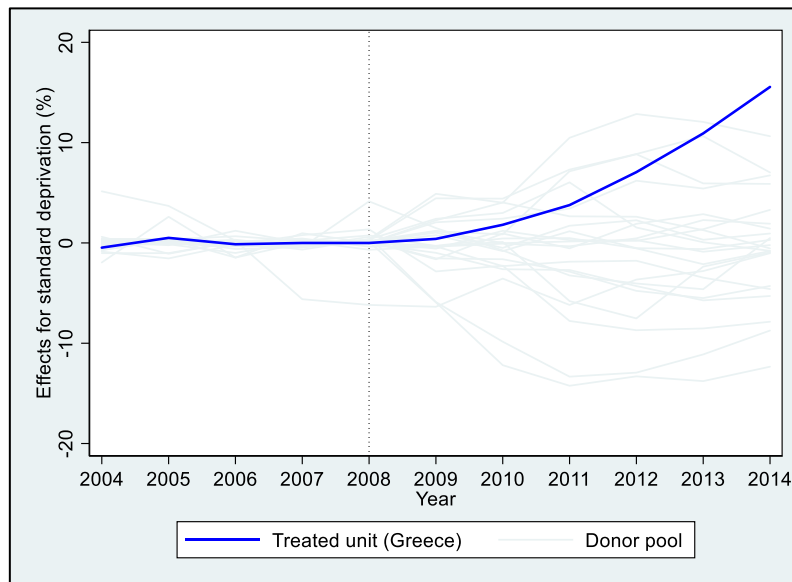
Figure 4a. “In-space” placebos: Gaps in the donor pool (treated unit: SPAIN)



Source: Own elaboration from the Eurostat database.

For the distribution of post-/preunemployment shock RMPSE using Greece as the unit of treatment, the calculations made place Greece in fourth position with a postevent RMPSE that is roughly 26 times that of the pre-event period. This ratio is higher than those observed in 20 countries of all 24 members of the *donor pool*. Therefore, these results also reveal that the probability of the effects being entirely attributable to chance is extremely low (see Figure 4b).

Figure 4b. “In-space” placebos: Gaps in the *donor pool* (treated unit: GREECE)



Source: Own elaboration from the Eurostat database.

5.3. Sensitivity analysis

To test the validity of our finding of a large impact of unemployment shocks on material deprivation, we propose different alternative scenarios that evaluate their sensitivity to changes in the length of the pretreatment period and in the number of control countries used (*donor pool*) and to a new definition (a stricter one) for unemployment shock.

Extension of the preunemployment shock time period: 1996-2004

Our first sensitivity exercise involves extending the number of years included in the pretreatment period. We start our analysis in 2004 because this is the year for which data for all EU-27 countries are available. Obtaining information on previous years implies restricting the number of countries in the donor pool. This is what we do here. We exploit microdata from the European Community Household Panel (ECHP).²⁸ Using information for a new sample of 12 countries²⁹, we reconstruct the series for 1996-2001. For 2002 and 2003, years in which there is “a survey gap,” we link the series by applying, for the different variables used, the rate of variation observed from 2000-2001.

Figure 5. Extension of the preunemployment shock time period: 1996-2008

Figure 5a. SPAIN and synthetic SPAIN

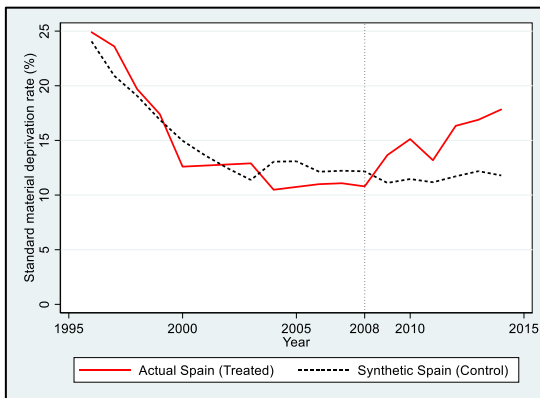
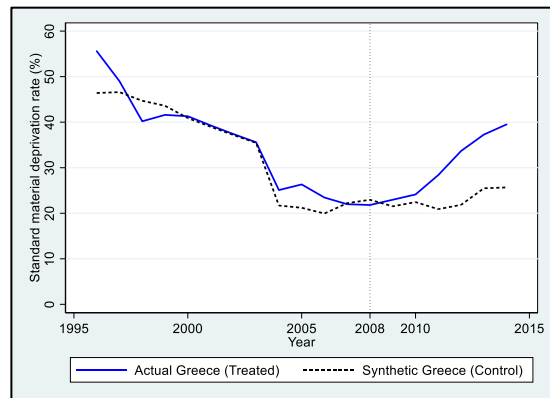


Figure 5b. GREECE and synthetic GREECE



Source: Own elaboration from the Eurostat database.

Alternative definition of unemployment shock

We also reformulate our definition of unemployment shock. As specified above, while unemployment grew in practically all of the countries studied, the magnitude of this growth and the resulting rates were very different. One way to isolate the treatment more

²⁸ For the United Kingdom, data were drawn from the British Household Panel Survey (BHPS).

²⁹ The new sample includes the following countries: Belgium, Denmark, Ireland, France, Italy, the Netherlands, Austria, Portugal, Finland, the United Kingdom, Spain and Greece.

³⁰ We have not used data for 1994 and 1995 due to a large number of missing values.

precisely involves draw a more radical divide between countries exposed to the shock and those not exposed. To do so, we discard as potential controls countries registering an unemployment rate of 10% to 20% in 2014 or a 50% to 200% increase in the unemployment rate from 2007-2014. In applying these more rigorous new criteria, the list of countries excluded from the donor pool is extended to the following: Slovenia, France, Lithuania, Latvia, Ireland, Bulgaria, Italy, Slovakia and Portugal. The similarities between the new figures and the original ones are remarkable.

Figure 6. A stricter criterion for the unemployment shock definition

Figure 6a. SPAIN and synthetic SPAIN

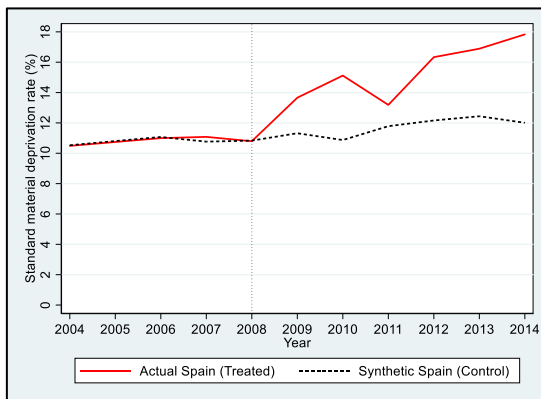
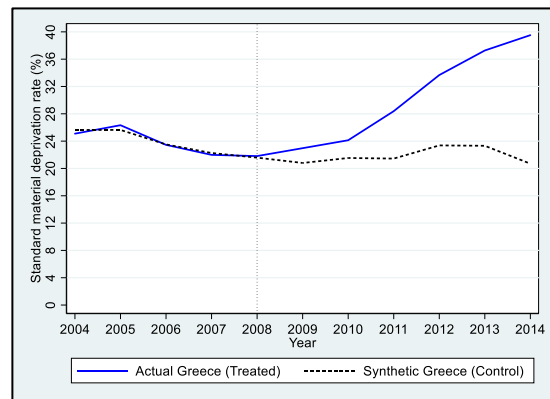


Figure 6b. GREECE and synthetic GREECE



Source: Own elaboration from the Eurostat database.

In brief, the new evidence exposed in this section is broad and strong enough to show that the unemployment shock tested in the chapter did indeed have a special and particular impact on material deprivation in the countries considered.

Different outcome variables

The final sensitivity test conducted involved replacing the standard material deprivation rate with two alternative measures. First, we replicate the above estimates using the severe material deprivation rate. This measure was the first official measure of deprivation used in the EU and is more restrictive than the original one – the percentage of the population that cannot afford at least four rather than three items. We also use the *counting* approach

proposed by Atkinson (2003). The fits obtained are quite good and the effects, despite being slightly smaller for Spain, do not present major changes from what was previously found (see Figure 7).

Figure 7. Alternative outcome variables

Figure 7a.1. Severe material deprivation rate: SPAIN and synthetic SPAIN

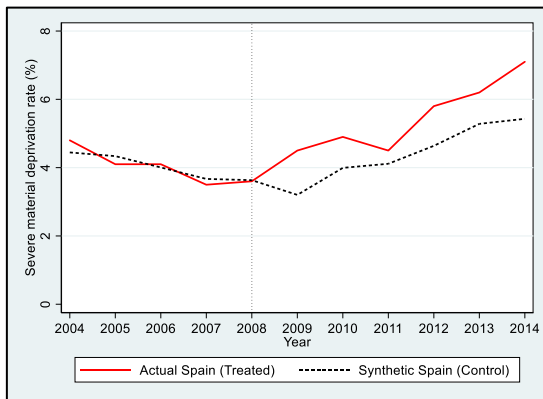


Figure 7a.2. Counting approach: SPAIN and synthetic SPAIN

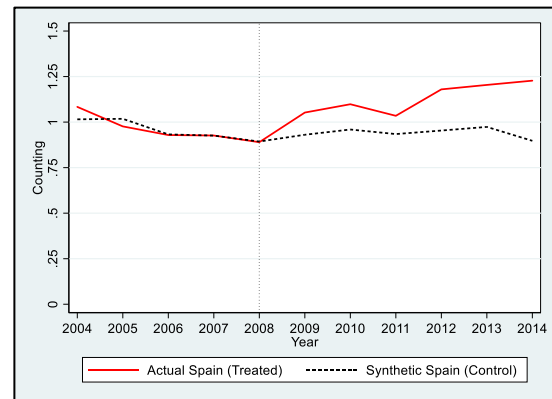


Figure 7b.1. Severe material deprivation rate: GREECE and synthetic GREECE

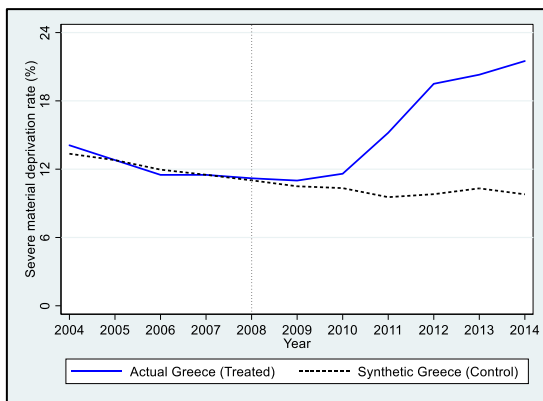
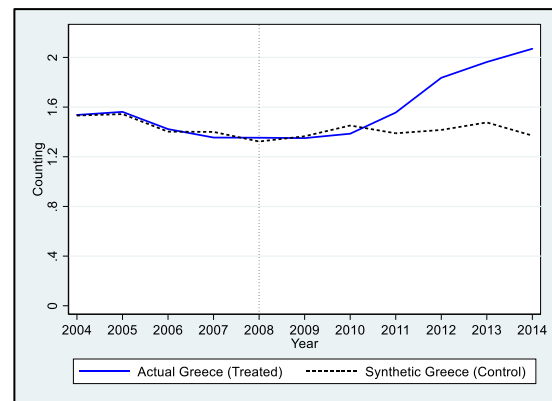


Figure 7b.2. Counting approach: GREECE and synthetic GREECE



Source: Own elaboration from the Eurostat database.

6. Conclusions

Unlike the extensive literature on the relationship between income distribution and macroeconomic conditions, the evidence on the sensitivity of material deprivation indicators to unemployment changes is much more limited. The less dynamic nature of deprivation measures compared to monetary indicators has meant that interest in

relationships to the economic cycle has traditionally been less widespread. The remarkable increase in material deprivation observed during the Great Recession puts this assumption at risk.

In this chapter, we have tried to establish causality relationships between changes in material deprivation and unemployment shocks. In focusing on the recent EU experience, we use a combination of European countries to construct a synthetic control unit for each country that as much as possible resembles the actual evolution of outcome variables before the outset of the shock.

An important and novel element of our approach relates to our proposed definition of an unemployment shock. A lack of consensus in the literature has led us to propose a specific definition that could be used in other studies. The use of the double criterion of the growth rate of the unemployment rate and its level has allowed us to differentiate two countries in which such shocks took place (Spain and Greece). However, this is a relative criterion in which the demarcation of countries affected by an event depends on the severity of the problem involved. Fortunately, through our sensitivity analyses we have been able to use more stringent criteria in defining these shocks, which has served to more clearly delimit the countries affected by these shocks and those that were not.

Our results show that in the countries for which the proposed criteria confirm the existence of an unemployment shock, a significant increase in material deprivation occurred. Based on the natural limits for establishing causal relationships, these results refute the traditional assumption of the low sensitivity of material deprivation measures to changes in the economic cycle.

This conclusion holds when other methods are used to identify the observed effect. To cover a broader pretreatment period, we extended the series by combining it with ECHP data. Even at the cost of reducing the number of countries analyzed, the effect of the

unemployment shock on material deprivation remains. The same occurs when other material deprivation measures are considered and above all when countries relatively similar to Spain and Greece based on any of the criteria used to define the unemployment shock are removed from the analysis.

Our results, in short, allow us to anticipate how drastic changes in the unemployment rate can lead to rapid well-being losses among households, which are not limited to increased monetary poverty and insufficient income but extend to material well-being and living conditions. Such results, derived from this study of what happened in the so-called Great Recession in a high-income area such as the European Union, could be even more severe in the face of even greater and rapid increases in unemployment such as those registered in these same countries due to the COVID-19 crisis.

References

- Aaberge, R., Wennemo, T., Bjorklund, A., Jantti, M., Pedersen, P. J. and Smith, N. (2000): "Unemployment shocks and income distribution: how did the Nordic countries fare during their crises?" *Scandinavian Journal of Economics* 102: 77-99.
- Abadie, A. and Gardeazabal, J. (2003): "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* 93: 113-132.
- Abadie, A., Diamond, A. and Hainmueller, J. (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105: 493-505.
- Abadie, A., Diamond, A. and Hainmueller, J. (2015): "Comparative politics and the synthetic control method." *American Journal of Political Science* 59: 495-510.
- Ábrahám, Á., Brogueira de Sousa, J., Marimón, R. and Mayr, L. (2018): "On the Design of a European Unemployment Insurance System". ADEMU WP2018/105.
- Alt, J. E., Barfort, S. and Lassen, D. D. (2017): "The effects of income and unemployment shocks on political preferences." NBER Political Economy Conference. Cambridge, MA, USA.
- Athey, S. and Imbens, G. W. (2006): "Identification and inference in nonlinear difference-in-differences models." *Econometrica* 74: 431-497.
- Atkinson, A. B. (2003): "Multidimensional deprivation: contrasting social welfare and counting approaches." *Journal of Economic Inequality* 1: 51-65.
- Ayala, L., Cantó, O. y Rodríguez, J.G. (2017): "Poverty and the business cycle: The role of intrahousehold distribution of unemployment," *Journal of Economic Inequality* 15: 47-73.
- Bárcena-Martín, E., Arias, B. L., Moro-Egido, A. I. and Moreno, S. P. (2013): "Privación material y desigualdad de género: una comparativa entre países europeos." *Papeles de trabajo del Instituto de Estudios Fiscales. Serie economía*, (6): 3-20.
- Bárcena-Martín, E., Lacomba, B., Moro-Egido, A.I., and Pérez-Moreno, S. (2014): "Country Differences in Material Deprivation in Europe." *Review of Income and Wealth* 60: 802-820.
- Barreix, A. D. and Corrales, L. F. (eds) (2019): *Reglas fiscales resilientes en América Latina* (Vol. 767). Inter-American Development Bank.
- Blank, R.M. and Blinder, A.S. (1986): "Macroeconomics, Income Distribution, and Poverty," in S.Danziger (ed.): *Fighting Poverty: What Works and What Does Not*. Cambridge: Harvard University Press.
- Carling, K. and Li, Y. (2016): "The power of the synthetic control method." Working papers in transport, tourism, information technology and microdata analysis. No.2016:10, Dalarna University.
- Cavallo, E., Galiani, S., Noy, I. and Pantano, J. (2013): "Catastrophic natural disasters and economic growth." *Review of Economics and Statistics* 95: 1549-1561.
- Christelis, D., Georgarakos, D. and Jappelli, T. (2015): "Wealth shocks, unemployment shocks and consumption in the wake of the Great Recession." *Journal of Monetary Economics* 72: 21-41.

- Craig, P. (2015): “Synthetic Controls: A New Approach to Evaluating Interventions.” Working Paper. What Works Scotland.
- Cutler, D.M. and Katz, L.F. (1991): “Macroeconomic Performance and the Disadvantaged,” *Brookings Papers on Economic Activity* 2: 1-74.
- Dibooglu, S. and Enders, W. (2001): “Do real wages respond asymmetrically to unemployment shocks? Evidence from the US and Canada.” *Journal of Macroeconomics* 23: 495-515.
- Duiella, M. and Turrini, A. (2014): “Poverty developments in the EU after the crisis: a look at main drivers.” ECFIN Economic Brief n°31.
- Ferman, B., Pinto, C. and Possebom, V. (2016): “Cherry Picking with Synthetic Controls.” FGV Working Paper 420. São Paulo, Brazil: Sao Paulo School of Economics.
- Figari, F. (2012): “Cross-National Differences in Determinants of Multiple Deprivation in Europe.” *Journal of Economic Inequality* 10: 397–418.
- Grier, K. and Maynard, N. (2016): “The economic consequences of Hugo Chavez: A synthetic control analysis”. *Journal of Economic Behavior & Organization* 125: 1-21.
- Harberger, A. C. (1987): [The Macroeconomics of Successful Development: What Are the Lessons?]: Comment. *NBER macroeconomics annual* 2: 255-258.
- Heckman, J. and Hotz, V. J. (1989): “Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.” *Journal of the American Statistical Association* 84: 862-74.
- Heim, B. T. and Lurie, I. Z. (2014): “Does health reform affect self-employment?”. Evidence from Massachusetts. *Small Business Economics* 43: 917-930.
- Jäntti, M. (1994): “A More Efficient Estimate of the Effects of Macroeconomic Activity on the Distribution of Income,” *The Review of Economics and Statistics* 76: 372-378.
- Kaul, A., Klößner, S., Pfeifer, G. and Schieler, M. (2018): “Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictors.” Working Paper. Saarbrücken, Germany: Saarland University.
- Martínez, R. and Navarro, C. (2014): “Pobreza y privación: Tendencias y determinantes.” Madrid: Fundación FOESSA. Documento de Trabajo 2.2.
- Martínez, R. and Navarro, C. (2014): “Has the Great Recession Changed the Deprivation Profile of Low Income Groups? Evidence from Spain.” *Hacienda Pública Española / Review of Public Economics*, 218-(3/2016): 79-104
- Mayer, S. E. and Jencks, C. (1989): “Poverty and the distribution of material resources.” *Journal of Human Resources* 24: 88-113.
- Meyer, B.D., and Sullivan, J.X. (2011): “Consumption and Income Poverty Over the Business Cycle,” *Research in Labor Economics*, in: Herwig Immervoll & Andreas Peichl & Konstantinos Tatsiramos (ed.), Who Loses in the Downturn? Economic Crisis, Employment and Income Distribution, volume 32: 51-82.
- McClelland, R. and Gault, S. (2017): “The Synthetic Control Method as a Tool to Understand State Policy.” *Washington, DC: Urban-Brookings Tax Policy Center*.

- Papanastasiou, S. and Papatheodorou, C. (2018): “The Greek Depression: Poverty Outcomes and Welfare Responses.” *Journal of Economics and Business* 21: 197–214
- Siwach, G. (2018): “Unemployment shocks for individuals on the margin: Exploring recidivism effects.” *Labour Economics* 52: 231-244.
- Smeeding, T., Thompson, J.P., Levanon, A. and Burak, E. (2011): “Income, Inequality, and Poverty over the Early Stages of the Great Recession,” in Grusky, D. and Western, B. and C. Wimer (eds.): *The Great Recession*, New York: Russell Sage Foundation.
- Verbunt, P. and Guio, A.C. (2019): “Explaining Differences Within and Between Countries in the Risk of Income Poverty and Severe Material Deprivation: Comparing Single and Multilevel Analyses.” *Social Indicators Research* 144: 827-868.
- Visser, M., Gesthuizen, M., & Scheepers, P. (2014): “The impact of macro-economic circumstances and social protection expenditure on economic deprivation in 25 European Countries, 2007–2011.” *Social Indicators Research* 115: 1179–1203.
- Whelan, C.T. and Maître, B. (2012): “Understanding Material Deprivation in Europe: A Multilevel Analysis.” *Research in Social Stratification and Mobility* 30: 489–503.
- Whelan, C.T. and Maître, B. (2013): “Material Deprivation, Economic Stress and Referencing Groups in Europe: An Analysis of EU-SILC2009.” *European Sociological Review* 29: 1162-1174

Appendix

Table A.1. Description of the variables

	Variables	Definition
Outcome/Dependent variables	Standard material deprivation rate (%)	Measures the percentage of the population that cannot afford at least three of the following nine items: (1) to pay their rent, mortgage or utility bills; (2) to keep their home adequately heated; (3) to pay for unexpected expenses; (4) to eat meat or protein regularly; (5) to go on holiday; and (6) to have a television set, (7) washing machine, (8) car, (9) or telephone.
	Severe material deprivation rate (%)*	Measures the percentage of the population that cannot afford at least four of the following nine items: (1) to pay their rent, mortgage or utility bills; (2) to keep their home adequately heated; (3) to pay for unexpected expenses; (4) to eat meat or protein regularly; (5) to go on holiday; and (6) to have a television set, (7) washing machine, (8) car, or (9) telephone.
	Counting approach (%)*	Number of dimensions under which people suffer deprivation.
Predictor variables	Gini index	Indicator measuring the extent to which the distribution of income within a country deviates from a perfectly equal distribution.
	Work intensity (%)	The ratio of the total number of months in which all working-age household members worked in the income reference year and the total number of months in which the same household members theoretically could have worked in the same period.
	Temporary employment (%)*	Employees who cannot find a permanent or full-time job.
	Ln (GDP per capita)	Ratio of real GDP to the average population of a specific year in natural logarithm form.
	Social protection benefits (% GDP)	Transfers to households, in cash or in kind, intended to relieve them of the financial burden of several risks and needs as defined in ESSPROS ³¹ . These include disability, sickness/healthcare, old age, survivor, family/child, unemployment, housing and social exclusion provisions not covered elsewhere.

Source: Eurostat and own elaboration.

Notes: (1) The asterisk (*) is denoting variables used in sensitivity tests; (2) *Temporary employment* has been used instead of *Work intensity* when extending the preunemployment shock time period.

³¹ ESSPROS refers to the European system of integrated social protection statistics.

Table A.2. Country weights in the synthetic units: SPAIN and GREECE

EU-27 countries	Composition of the <i>donor pool</i>					
	Synthetic SPAIN			Synthetic GREECE		
	[1]	[4]	[7]	[1]	[4]	[7]
Austria	0	0	0	0	0	0
Belgium	0	0	0	0	0	0
Bulgaria	0.032	0	0	0	0.118	0.133
Cyprus*	—	—	—	—	—	—
Czech Republic	0	0	0	0	0	0
Denmark	0	0	0	0	0	0
Estonia	0	0	0	0	0	0
Finland	0	0.110	0	0	0	0
France	0	0	0	0	0.062	0.291
Germany	0.314	0.203	0.190	0	0	0
Greece**	—	—	—	—	—	—
Hungary	0	0	0	0.116	0	0
Ireland	0	0	0	0.400	0	0
Italy	0	0	0	0	0	0
Latvia	0	0	0	0	0	0
Lithuania	0	0	0	0	0	0
Luxembourg	0.396	0.195	0.192	0	0	0
Malta	0	0.492	0.555	0	0	0
Netherlands	0	0	0.063	0	0	0
Poland	0	0	0	0.255	0.133	0.131
Portugal	0.132	0	0	0.106	0.117	0
Romania	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0
Spain**	—	—	—	—	—	—
Sweden	0.126	0	0	0	0	0
United Kingdom	0	0	0.175	0.123	0.571	0.446

Note: (1) (*) Conflicting country excluded; (**) Countries of treatment; (2) Numbers enclosed in square brackets refer to the econometric specifications used.

Chapter 3. Key Drivers of Income Inequality in Large Cities: A Multi-Country Analysis Using Static and Dynamic Decompositions

1. Introduction

According to data from the United Nations, between 1950 and 2018, the world's urban population grew more than four times. In that period, the percentage increased from 30% to 55% (United Nations, 2018). Furthermore, the predictions estimate that the urbanization process will continue for decades, with an increasing proportion of the world's population concentrated in large metropolitan areas.

Among the various questions raised by this growing concentration of population in large cities, its potential effects on inequality will undoubtedly be a major focus of policy research for years to come. The evidence on the relationship between income inequality and city size is not as obvious as that of each of its drivers. Inequality and size are complex and multifaceted concepts, and when one tries to link inequality measures and city size, the relationship is not always well defined. Reasonable predictions are only feasible when it is possible to identify a set of significant variables that are simultaneously a cause of income inequality and a consequence of urbanization.

According to the standard Kuznets hypothesis, industrialization was such a variable in advanced countries, initially causing marked increases in income inequality, and artificial intelligence may be one today (Frank et al., 2019). New scenarios linked to the processes through which large cities developed have emerged, and many in the policy and research communities have speculated about their effects on inequality. However, the lack of empirically informed models, the insufficient understanding of certain interactions, and data constraints are important barriers preventing the adequate measurement and comprehension of the nexus between inequality and city size.

The potential effects of the growing relevance of large cities on inequality raise numerous interesting questions that must be addressed: Is inequality greater in large cities than in

other territories or areas? Has inequality in large cities increased over time? Do the drivers of inequality in large cities differ from those in other areas? These questions motivate this chapter. Our concern is to contribute to the understanding of the relationship between inequality and city size. With this aim, we focus on large cities in a selected sample of OECD countries. The main goal is to determine and quantify the individual contributions of different explanatory factors to differences in inequality both within large cities and between large cities and other areas.

We apply the methodology proposed by Firpo et al. (2009, 2018) (FFL henceforth), implementing it using both a *dynamic approach*—focusing on the increase in income inequality in large cities over the last two decades—and a *static approach*—focusing on inequality differences between large cities and other territories. By means of the first approach, we can answer questions such as the following: Is there any common pattern in inequality changes in large cities among the countries analyzed? Which factors account for the changes in the distribution of income in large cities? What is the magnitude of the effect of each of these potential drivers? By means of the static approach, we focus on the most recent data to compare income inequality in large cities and other areas.

To the best of our knowledge, this is the first attempt to put into practice this methodology from a double perspective, which is one of the contributions of this chapter. Furthermore, as far as we know, few studies apply the aforementioned methodology to more than one country. The two approaches provide a complementary and comprehensive view of the influence of some of the main drivers of income inequality in large cities in the chosen countries.

Our findings lend support to the thesis that income inequality is higher in large cities than in other areas. This conclusion holds when the relationships are tested with different inequality measures. We also find that inequality increased in large cities during the first

two decades of the 21st century. These changes are explained, essentially, by what we call the ‘structure effect’. Among the potential drivers of these trends, one of the most important is educational attainment. Age and household size are also relevant in explaining inequality differences within large cities and relative to other territories.

The chapter is structured as follows. In section two, we review the literature connecting income inequality and city size. In section three, we describe the methodology used. In section four, we introduce the data sources used and the variables chosen. In section five, we present and comment on the main results of the chapter. Section six concludes.

2. Income inequality and city size: review of literature

An extant literature has tried to quantify the relevance of income inequality in cities and its circumstantial and political drivers. The availability of data and analytical methods has guided the empirical research on inequality in these areas. New or updated data and novel testing tools explain the sequence by which the causes and consequences of inequality in different areas, identified by the theoretical literature, are tested in the empirical literature. This background helps to clarify why, until recently, most of the evidence on inequality and city size has referred to US metropolitan areas and why the focus was primarily on the labor market.

This review includes only contributions explicitly addressing the relationship between inequality and city size. In Garofalo and Fogarty (1979), we find a pioneering theoretical framework for analyzing the urban income distribution based on agglomeration economies and the amenity structure of cities. The former causes the demand curve of a large city to lie above that of a smaller city. Given a common amenity structure, if the supply curve of unskilled workers is perfectly elastic and that of skilled workers is upward sloping, then the wage of skilled workers would increase more in large cities while the

wage of unskilled workers would remain constant. However, if amenities vary with city size, inequality may follow different paths. Assuming that amenities increase until some urban size threshold is reached and then reverse, the authors derived a U-shaped relationship between inequality and city size. However, their empirical results for the US metropolitan areas in 1970 were not robust and were sensitive to the choice of inequality measure. Nord (1980a) provided empirical support for the U-shaped hypothesis after extending the sample used by previous studies to include smaller cities. He accounted separately for some of the factors outlined by the previous literature (1980b) finding that race (1982, 1984) reinforced the positive relationship between city size and inequality.

Later evidence for US metropolitan areas from the 1980s is not conclusive. Galster et al. (1988) found only weak support for the hypothesis that population has a direct effect on inequality after controlling for industry and occupational structure. Cloutier (1997) found a positive effect of population size and population growth on inequality after controlling for spatial, demographic and industrial structures, although it was not statistically significant. Under a general equilibrium framework for an open system of cities, Alperovich (1995) offered different propositions about the relationship between city size and income inequality. A relevant one was that inequality rises (declines) with city size if the relative preference for nontraded goods increases (decreases) with the level of income. Pooling data on US metropolitan areas from 1970, 1980 and 1990, Wheler (2004a, 2004b) found a negative association between the changes in population density and the 90/10 percentile ratio of wages after controlling for college education levels, the share of manufacturing, the immigration rate, the unemployment rate and the level of unionization.

With the turn of the century, the hypothesis of a positive relationship between population size and inequality in cities received increasing empirical support. Using longitudinal US

data from the 1979 to 1998 Surveys of Labor Market Experience, Gould (2007) found evidence of a city wage premium, but only for white-collar workers and not for blue-collar workers, which implicitly meant that inequality was higher in cities than in rural areas. Glaeser et al. (2009), using data on US metropolitan areas from the Censuses and the American Community Survey (ACS) for the period 1980-2006, found that the causes of household income inequality in cities are a higher skill wage premium in fields such as finance or computing, immigration and, above all, the skill composition within the population. After controlling for these factors, the authors found a positive link between population size and inequality, which has become stronger since 1980.

Baum-Snow and Pavan (2013), using data from the same datasets and over a similar period but restricting the analysis to working white men aged 25-54, confirmed this increase in inequality with city size. They found that city-size specific factors explained at least a quarter of the overall increase in the variance in wages between 1979 and 2007. City differences in the skill wage premium were more relevant than differences in skill composition for explaining the city size effect on inequality. The authors also found a special case of city size affecting inequality in the 1990s, especially in the top half of the income distribution. Baum-Snow et al. (2018), using manufacturing data from core-based statistical areas (CBSAs) from 1980 to 2007, theoretically and empirically examined some potential causes of the more rapid increase in wage inequality in larger cities over time. The high estimated elasticity of substitution between unskilled labor and capital explains why unskilled wages are much less variable across locations than skilled wages. They also suggest that the increasing complementarity in production between human capital and market scale indicates the growing role of knowledge spillovers in generating agglomeration economies.

Finally, Davis and Dingel (2019) developed a spatial equilibrium model to look inside the black box of knowledge spillovers. Their model replicates a range of empirical facts, including the fact that skill premia are higher in larger cities. Using SMA data from the 1990 and 2000 US censuses and the 2005-2007 ACS on full-time/full-year workers aged 25-55 with a high school diploma or bachelor's degree, the authors found a positive relationship between skill premia and city size.

The evidence for countries other than the US is quite fragmented, and there are very few comparative analyses. In the case of high-income countries, the positive relationship between inequality and city size is generally confirmed, although not all the US results are transferable to other countries. Using 1971 census data, Soroka (1984) found that city size did not have a direct effect on overall urban income distributions in Canada. Using bivariate regressions and different controls, Lee et al. (2016) found that larger cities were more unequal in Great Britain. This relationship did not hold when the mean wage was included, something that the authors interpreted as the consequence of the abundance of more highly skilled and better paid workers in large cities. Henkel (2017) used the Sample of Integrated Labor Market Biographies from (West) Germany on working males aged 21-60 to obtain results similar to those of Baum-Snow and Pavan (2013) for the US. The variance in wages by location size increased in (West) Germany between 1985 and 2009, and location size accounted for a quarter of the rise in wage inequality during this period, though the relationship between location size and inequality was stronger and more positive when only the upper part of the income distribution was considered. Hortas-Rico and Rios (2019) used microdata from Spanish personal income tax records for the period 2000-2006 and methods that account for spatial interactions and weights; they found that population size was a moderate determinant of inequality in Spanish cities. Finally, Mastronardi and Cavallo (2020), also using data on gross income from tax returns, found

a significant positive association between population density and income inequality at the municipal level in Italy.

In the case of developing and emerging countries, the evidence from the two most populous countries is apparently contradictory. Chen et al. (2018), using data from China's 2005 population survey for employed individuals aged 18-53 living in 252 cities with 190,000 or more residents, found a positive relationship between income inequality and city size. This association held after introducing different control variables and regardless of the inequality measure used. Dubey and Mahadevia (2001), using data from an Indian consumer expenditure survey for the period 1987-1994 for metropolitan cities (those with over one million residents in 1991), found that while the Gini coefficient seems to be unrelated to city size, the incidence of poverty decreases with city size. A possible explanation is that the caste-based segregation in India diminishes with city size (Haque et al., 2019). Another explanation is that the effects of city size on inequality and on poverty have opposite signs when poverty lines are defined at the national level.

There are few comparative studies covering different countries. Royuela et al. (2014) used the OECD (2012) metropolitan database and the concept of functional urban areas (FUAs) to find that regional inequality is positively correlated with urbanization (the share of people living in FUAs). This correlation increases when the definition of 'urban' is restricted to people living in large metropolitan FUAs.

Boulant et al. (2016) provided the first estimation of the distribution of household disposable income for 153 metropolitan areas in 11 OECD countries using data from tax records (Austria, Belgium, Canada, France, Italy, Norway), household surveys (Chile, and the United States), and other registers (Denmark and Sweden) or estimations (Mexico). Using the same data and sample, Castells-Quintana et al. (2020) provided a deeper analysis of the relationship between inequality and city size considering the United

States, Canada, Latin America (Chile and Mexico) and some European countries (Austria, Belgium, Denmark, France, Italy, Norway, and Sweden). After controlling for income, education, demographic factors and industrial composition, their estimations suggested that as cities double in size, the Gini index grows by approximately one percentage point.

3. Methodology

3.1. Empirical strategy

The empirical strategy we propose in this chapter consists of implementing an extension of the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973)—OB hereafter—to study recent changes in income inequality trends in a sample of OECD countries. We use recentered influence function (RIF) regressions and analyze four different distributional measures: the Gini coefficient, the P90-P10 ratio (the difference between the 90th and 10th percentiles of equivalent disposable income), and the P90-P50 and P50-P10 ratios. By carrying out a twofold procedure, an ‘aggregate decomposition’ and a ‘detailed decomposition’, we try to identify and quantify the differences between two groups: what we have called large cities—territories with over 500,000 inhabitants—and other areas—those with fewer than 500,000 citizens. In particular, we focus on the contributions of a set of covariates to explanations of the disparities in the inequality measures chosen.

This exercise is implemented in each country from a double perspective. On the one hand, we examine the existing differences in inequality between the two groups mentioned above with the most recent data (static approach). On the other hand, we pay attention to the evolution of inequality within large cities during approximately the last two decades (dynamic approach). To the best of our knowledge, this is the first attempt to apply this methodology with two approaches and using a comparative perspective for more than two countries.

Following Fortin et al. (2011), who presented a review of the main decomposition procedures used to evaluate changes in wage distributions, we chose the FFL proposal (Firpo et al., 2009) with the improvements suggested by Firpo et al. (2018). In this way, we can measure the specific contribution of each covariate included in the model, and the numerous drivers explaining income inequality can be identified.

This technique consists of a two-stage process that can be illustrated as follows. First, assuming the premises established in OB as the starting point, we conduct an ‘aggregate decomposition’. This task allows us to identify the so-called *composition effect*—variation attributable to changes in characteristics—and the *structure effect*—changes associated with the returns to these characteristics. Following DiNardo et al. (1996), the first stage is performed by means of a reweighting approach. In the second step, to complete the “detailed decomposition”, we need to make use of the regression strategy set down in FFL. This method is based on the estimation of a regression where the dependent variable—real equivalent disposable income in our case—is replaced by its transformation through the so-called recentered influence function. This function calculates the effect of small changes in the corresponding distribution on distributional statistics.

Once the previous regressions have been estimated, a standard OB can be developed based on their results. Under the assumption of linearity, this method allows us to decompose the income inequality gap between two groups in a straightforward way. On the one hand, we try to identify the variables correlated with the changes in inequality, in what we have called large cities, over time. On the other hand, we use the latest available data to capture the differences in inequality between the latter compared to the other areas. The RIF regressions implemented in the analysis are quite easy to estimate since they can

be performed via ordinary least squares (OLS), just as OB can be. As a result, we estimate a RIF-OLS model.

Regarding other methodologies developed in the economics literature that focus on decomposition exercises to estimate the differences between distributions, the FFL proposal has a remarkable advantage. While the semiparametric approach of DiNardo et al. (1996) based on the reweighting of samples, the parametric approximation of Juhn et al. (1993) involving the distributions of the residuals and the conditional quantile regressions (CQR)³² of Machado and Mata (2005) or Melly (2006) only allow for the calculation of the aggregate effects of characteristics and their returns, the FFL scheme provides a detailed decomposition. In this sense, it identifies the individual contributions of each explanatory factor considered in the model through the profile of characteristics and their corresponding returns. Another advantage is that in contrast to the classical OB approach, FFL allows us to take into account the entire distribution, not only the mean. In other words, we can not only distinguish the importance of individual contributions to income inequality in average, but can also delve into which factors have been the most relevant sources of change in the different segments of the distribution by decomposing the variation at different percentiles.

There are also drawbacks to this approach, as stressed by Rothe (2015). The approximations of certain nonlinear functions obtained with RIF regressions may not be as rigorous and precise as might be expected in some particular cases.

³² These proposals, additionally, are *path dependent*, as the result of the decomposition is affected by the order in which the mentioned decomposition is implemented. The FFL methodology applied here and derived from RIF regressions is *path independent*.

3.2. Model formalization

Following the main guidelines of Firpo et al. (2018), the methodology can be described in the following terms:

a) *Aggregate decomposition*

Let us suppose there is a joint distribution function defining all relationships between the following three variables: the real equivalent disposable income (Y), the regressors or exogenous characteristics (X), and a categorical variable (T) indicating the group to which each individual belongs: $f_{Y,X,T}(y_i, x_i, T_i)$.

Bearing in mind that we have only two groups,³³ the joint probability distribution function and the cumulative distribution of real equivalent disposable income conditional on T can be described as follows:

$$f_{Y,X}^k(y, x) = f_{Y|X}^k(Y|X) \cdot f_X^k(X) \quad [1]$$

$$F_Y^k(y) = \int F_{Y|X}^k(Y|X) \cdot dF_X^k(X) \quad [2]$$

The superscript k indicates that the density is conditional on $T = k$, with $k \in [0,1]$. The way to compute the gap between the two groups, given a distributional statistic v , such as the median, would be:

$$\Delta_0^v = v_1 - v_0 = v(F_Y^1) - v(F_Y^0) \quad [3]$$

$$\Delta_0^v = v \left(\int F_{Y|X}^1(Y|X) \cdot dF_X^1(X) \right) - v \left(\int F_{Y|X}^0(Y|X) \cdot dF_X^0(X) \right) \quad [4]$$

To assess the relevance of the differences in characteristics ($dF_X^1(X) \neq dF_X^0(X)$) and their returns ($F_{Y|X}^1(Y|X) \neq F_{Y|X}^0(Y|X)$) when determining the overall differential between both

³³ In the static approach, $T = 0$ represents households living in territories with less than 500,000 inhabitants and $T = 1$ denotes households residing in territories with more than 500,000 inhabitants. In the dynamic approach, $T = 0$ stands for the initial year and $T = 1$ refers to the last year.

groups, we need to design a hypothetical scenario.³⁴ The counterfactual statistic can be denoted in the following terms:

$$v_C = v(F_Y^c) = v\left(\int F_{Y|X}^0(Y|X) \cdot dF_X^1(X)\right) \quad [5]$$

Finally, the ‘aggregate decomposition’ can be expressed as the difference between the two groups (*v-overall income gap*):

$$\Delta_0^v = (v_1 - v_C) + (v_C - v_0) = \Delta_S^v + \Delta_X^v \quad [6]$$

The *structure effect* (Δ_S^v) is given by $v_1 - v_C$, while $v_C - v_0$ captures the *composition effect* (Δ_X^v).

b) *Detailed decomposition*

Formally, the influence function for the τ -th quantile can be defined as follows:

$$IF(y; q_\tau, F) = \frac{\tau - l(y \leq q_\tau)}{f_Y(q_\tau)} \quad [7]$$

where $l(y \leq q_\tau)$ is an indicator function showing whether the value of real equivalent disposable income is below q_τ , and $f_Y(q_\tau)$ is the marginal density of the same outcome of interest at q_τ , which is determined by kernel estimation.

For operational reasons, it seems appropriate to center the influence function on the statistic of interest (the Gini coefficient, for instance). All we have to do is add this statistic to the influence function. The RIF formula becomes:

$$RIF(y; q_\tau, F) = q_\tau + IF(y; q_\tau, F) \quad [8]$$

³⁴ To recreate the counterfactual scenario, a situation that cannot be checked in the available data, we apply a reweighting approach similar to those described in DiNardo et al. (1996) or Barsky et al. (2002). The alternative proposed by these authors is to multiply the distribution of characteristics $dF_X^0(X)$ with a reweighting factor $\psi(X)$ so that it provides a distribution similar to $dF_X^1(X)$.

Once the RIF function has been calculated, we obtain the value of the transformed variable for each observation in the sample. As stated by Firpo et al. (2011), the key point is to assume that the conditional expectation of the RIF function can be modeled as a linear function of the explanatory variables. This assumption translates into the fact that RIF regressions can be estimated by simply running OLS.

The RIF regressions (*unconditional quantile regressions*, UQR) provide estimations of the marginal impact of the explanatory variables on the statistic chosen. In other words, the $\hat{\gamma}$ estimated coefficients can be interpreted as the (average partial) effect of an increase in the average value of an explanatory variable on the corresponding statistic—Gini coefficient, variance, percentile, etc.

Finally, it should be highlighted that the detailed decomposition³⁵ embodies a RIF-OLS decomposition combined with a semiparametric reweighting estimator³⁶, again applying DiNardo et al. (1996). This decomposition can be disaggregated into four terms as follows:

$$\hat{\Delta}_O^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_C^v) + (\bar{X}_1 - \bar{X}_0^C)' \cdot \hat{\gamma}_C^v + (\bar{X}_0^C - \bar{X}_0)' \cdot \hat{\gamma}_0^v + \bar{X}_0^{C'} \cdot (\hat{\gamma}_C^v - \hat{\gamma}_0^v) \quad [9]$$

Where

$$\hat{\Delta}_{S,p}^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_C^v) \quad [10]$$

$$\hat{\Delta}_{S,e}^v = (\bar{X}_1 - \bar{X}_0^C)' \cdot \hat{\gamma}_C^v \quad [11]$$

$$\hat{\Delta}_{X,p}^v = (\bar{X}_0^C - \bar{X}_0)' \cdot \hat{\gamma}_0^v \quad [12]$$

$$\hat{\Delta}_{X,e}^v = \bar{X}_0^{C'} \cdot (\hat{\gamma}_C^v - \hat{\gamma}_0^v) \quad [13]$$

³⁵ A very common problem here is the choice of a specific reference for the dummy variables used in the analysis, since this decision can have an impact on the results. This is shown in Oaxaca and Ransom (1999). For this reason, we have applied a normalization strategy following Yun (2005) that let us overcome this identification problem and develop an adequate estimation of the real contribution of each covariate.

³⁶ The OB-type decomposition, without reweighting, would be: $\hat{\Delta}_{OB}^v = \bar{X}_1' \cdot (\hat{\gamma}_1^v - \hat{\gamma}_0^v) + (\bar{X}_1 - \bar{X}_0)' \cdot \hat{\gamma}_0^v$.

This model, an improved version of the original FFL, amends some misspecification and reweighting drawbacks existing in the model without reweighting. It is achieved by including two error terms in the decomposition (Firpo et al., 2018): the *specification error* ($\hat{\Delta}_{X,e}^v$), due to misspecifications in the model (i.e., nonlinearities), and the *reweighting error* ($\hat{\Delta}_{S,e}^v$), used to evaluate the quality of the reweighting strategy. The first term equals zero if the model is truly linear; the last term tends toward zero when the samples are large. For the two other terms, $\hat{\Delta}_{X,p}^v$ reflects the *pure composition effect*, and $\hat{\Delta}_{S,p}^v$ represents the *pure structure effect*. The sum of $\hat{\Delta}_{X,p}^v + \hat{\Delta}_{X,e}^v$, on the other hand, shows the aggregate *composition effect* of the detailed decomposition, whereas $\hat{\Delta}_{S,p}^v + \hat{\Delta}_{S,e}^v$ reproduces its counterpart for the *structure effect*.

4. Data

In this chapter, we focus on some of the most populous countries in the OECD area. In particular, the selected sample represents almost 50% of the total OECD population.³⁷ We use the LIS database for the following set of countries: Canada, Germany, Italy, Poland, the United States and Spain.³⁸

With regard to income inequality and poverty analysis, and in comparison with other data sources, such as the World Income Inequality Database (WIID) of the World Institute for Development Economics Research (WIDER) or the World Bank's PovcalNet, the LIS database includes a characteristic feature that is its greatest value-added. It provides access to a set of harmonized microdata files generated from survey data at the country

³⁷ The latest information available can be checked here: <https://data.oecd.org/pop/population.htm>

³⁸ The criteria followed for the selection of countries was as follows: we identified the 20 most populous countries in the OECD based on the most recent data available and chose those in which the LIS variable *size of the locality of residence* was defined. This is the reason why important countries such as France and the United Kingdom, among others, are not analyzed in this study. In Spain, the population strata provided by the LIS did not allow for working with the two areas we propose in this chapter. To ensure homogeneous information, data for Spain are taken from the Spanish Family Budget Survey (EPF).

level, and the information is provided by the national statistical agencies. These datasets permit us to handle standardized income distributions in several countries, allowing us to implement international comparisons with homogenous data. On the other hand, it should be emphasized that one of the original weaknesses of this dataset was its limited coverage. Nonetheless, several new countries have recently been included in the database. Therefore, the range of and possibilities for analysis have expanded not only to high-income countries but also to middle-income countries.

Regarding the variables included, the fundamental variable is *size of the locality of residence*. We have used this variable to delimit and distinguish what we have called large cities (those with over 500,000 inhabitants) from the other territorial areas (those with fewer than 500,000 inhabitants). Following Dijkstra and Poelman (2012), we consider large cities to be those urban centers with the size labels ‘XL’ (500,000–1,000,000 inhabitants), ‘XXL’ (1,000,000–5,000,000 inhabitants) and ‘Global city’ (more than 5,000,000 inhabitants). The other group is made up of the rest of the categories: ‘S’ (50,000–100,000 inhabitants), ‘M’ (100,000–250,000 inhabitants) and ‘L’ (250,000–500,000 inhabitants). The list of metropolitan areas defined as large cities according to this classification can be found in the appendix.

We have opted to use real equivalent disposable income for the income variable. This variable is derived by dividing household disposable income by the square root of household size.³⁹ As controls, we consider a set of variables related to the following fields: geography and housing (tenure), household composition and living arrangements (household composition and the number of household members), sociodemographic characteristics (age, sex, marital status, immigration status, disability status, health status,

³⁹ Negative and zero incomes have been replaced by 1/100 of the mean to prevent relevant observations from being dropped by default.

and education) and labor market information (employment status and part-time employment status). A comprehensive definition of each variable can be found in the appendix.

Finally, we use two waves of LIS data for the countries mentioned above: one corresponding to the early years of the 21st century and the other corresponding to the most recently available data.⁴⁰ These two years are used for the implementation of the *dynamic approach* set out in the preceding section. For the *static approach*, we consider the most recent year available, making it possible to identify, at a more recent moment and for each country, the contributions of the different determinants. Whereas the *dynamic approach* focuses exclusively on large cities, the static approach is estimated for both large cities and the other areas. In this way, the two methods applied provide a complementary and comprehensive view of the influence of the main drivers explaining income inequality in the chosen countries.

5. Main results

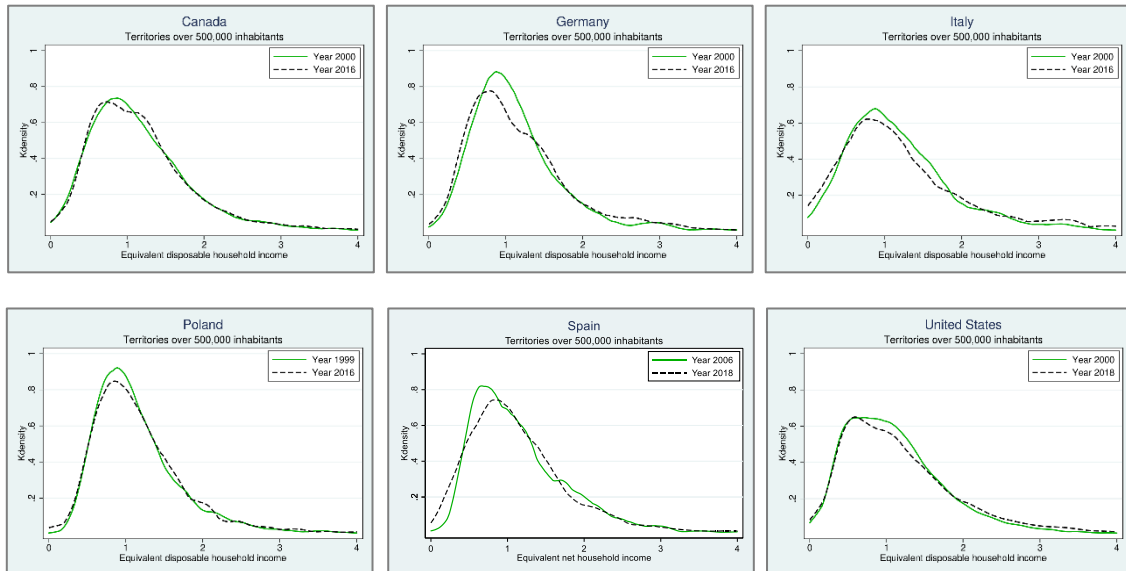
5.1. Descriptive analysis

Figure 1 and Figure 2 present general and descriptive evidence on income inequality disparities in the set of countries examined. The first figure presents the evolution of equivalent disposable income within large cities in the selected countries. The time period analyzed covers approximately 20 years, with one curve reflecting the disposable income distributions at the beginning of the 21st century and the other curve reflecting the distribution corresponding to the most recent year available. Notably, the distributions in

⁴⁰ In the case of Spain, we use the EPF waves of 2006 and 2018. In 2006, a number of methodological improvements were incorporated, such as the change of periodicity (from quarterly to annual) as well as a notable increase in the sample size (up to 24,000 households).

which a greater probability mass is concentrated near the median are those of the initial wave.

Figure 1. Density of equivalent disposable household income (PPP 2017 USD)
Dynamic approach

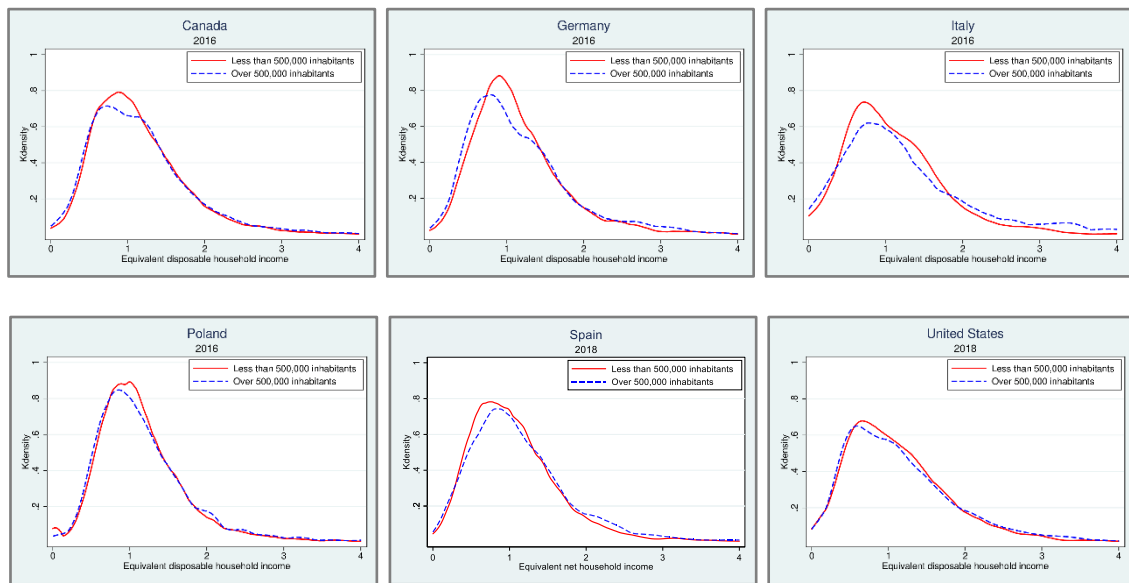


Notes: (1) Income values are expressed in relation to the median. (2) Income values in Spain are expressed in constant euros (reference year 2016 = 100).

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

Figure 2 shows the divergence between the density of the same income variable in large cities compared to the other areas. In general, all the curves represented are characterized by slight positive skewness. More revealing, however, is the information provided by the kurtosis measure (tailedness) in the probability density functions. The probability mass appears to be more concentrated around the median in territories with fewer than 500,000 inhabitants than in large cities, where the curves are flatter. The shape of the upper tail in the different distributions also seems to point in the same direction.

Figure 2. Density of equivalent disposable household income (PPP 2017 USD)
Static approach



Notes: (1) Income values are expressed in relation to the median. (2) Income values in Spain are expressed in constant euros (reference year 2016 = 100).

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

Other additional descriptive results that yield interesting information are, on the one hand, those related to the sample means of the socioeconomic characteristics examined and, on the other hand, several inequality indicators measuring the distributions of equivalent disposable income (see Tables A3 to A8 in the Appendix). First, it is worth noting that in four of the six countries under study, the percentage of territories with a population over 500,000 inhabitants has increased since the early 2000s. The United States, which accounts for the greatest number of large cities, registered a figure very close to 60% in 2018. In Italy, in contrast, only 8.4% of individuals resided in a large city in 2016. Second, the United States exhibits the largest sample size not only globally but also with regard to large cities (40,835 observations). For this same group, Italy and Germany, with 625 and 2,736 observations, respectively, are at the opposite extreme. This is a central point to bear in mind, since the low significance of the results in these two countries (see the next section) is very likely to be affected by this circumstance. Third and last, the reported

mean differences in observable characteristics are significant for the vast majority of variables taken into consideration in the different countries. The same can be said regarding the differences in the income distribution by percentile.⁴¹ A last comment refers to the Gini coefficient. In general, Gini values are higher in large cities than in territories with fewer than 500,000 inhabitants. Within the former, the values are also higher in the most recent year than at the beginning of the time period analyzed. We present the explanatory factors driving these differences in subsections 5.2 and 5.3, which mainly focus on the estimates of the individual contribution of each of the regressors to the observed gap in equivalent disposable income as well as to the differences observed in inequality.

5.2. Reweighted RIF-OLS decomposition: Dynamic approach

Income inequality as measured by the Gini coefficient increased in large cities in all the countries analyzed except Poland. Nevertheless, the difference in Gini values in Poland between 1999 and 2016 is not significant. The differences in Gini values in the two years considered are determined, basically, by the *structure effect*. In the countries with significant results in the aggregate decomposition when using the Gini index, Germany and the United States, this effect has a relative importance of 92.2% and 71.3%, respectively.

Similar to the changes in the Gini coefficient, the changes in the three percentile ratios during the period analyzed are also mainly explained by the *structure effect*.⁴² The magnitude of this effect is noticeably higher in the bottom of the distribution in countries

⁴¹ The only exception is Italy in the *dynamic approach*.

⁴² Regarding the *composition effect*, and only in the case of Spain, specification errors turn out to be significant in the different income gaps. The strong assumption of nonlinear functional forms required to implement the RIF regressions may be behind this finding.

such as Germany and Poland. In others, such as the United States and Canada, the relative weight of the changes assigned to returns to characteristics is greater in the upper tail of the distribution.

**Table 1. RIF-OLS decomposition results with reweighting
CANADA**

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2016	53,038.79***	33,793.29***	19,245.50***	29.718***
Year 2000	41,469.61***	25,318.18***	16,151.43***	29.199***
Difference / Total change	11,569.18***	8,475.11***	3,094.07***	0.519
[1] Composition effect	1,417.65*** (12.25%)	659.83 (7.79%)	757.82*** (24.49%)	0.571*** (110.02%)
[2] Structure effect	10,151.53*** (87.75%)	7,815.28*** (92.21%)	2,336.25*** (75.51%)	-0.052 (-10.02%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	1,953.84***	999.50***	954.35***	0.730***
Tenure	20.13	-4.89	25.03	-0.031
Household composition	15.57	-33.15	48.71	0.052
Household members	118.31	37.80	80.51*	-0.020
Age	665.28***	432.88**	232.39**	0.171
Sex	11.26	43.22	-31.97	0.063
Marital status	-98.37	-56.43	-41.94	-0.020
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	853.59***	449.63***	403.96***	0.158*
Employed	-	-	-	-
Part-time employment	368.07***	130.42**	237.65***	0.358***
$\hat{\Delta}_{X,e}^v$ = Specification error	-536.19*	-339.66	-196.52	-0.159***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	10,117.77***	7,808.37***	2,309.40***	-0.041
Tenure	2,616.15	264.15	2,351.99***	-0.874
Household composition	-1,488.86*	-363.15	-1,125.71***	-0.371
Household members	-10,568.61***	-7,046.02*	-3,522.04***	-2.568
Age	4,053.91	-3,799.41	7,853.33	10.072
Sex	-9,630.90***	-7,686.69**	-1,944.21	-3.172*
Marital status	3,094.94**	2,953.20*	141.74	0.757
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2,870.80**	2,535.59**	336.21	0.318
Employed	-	-	-	-
Part-time employment	-367.61	-210.88	-156.73	-0.033
Constant	19,536.96	21,161.59	-1,624.63	-4.172
$\hat{\Delta}_{S,e}^v$ = Reweighting error	33.75	6.91	26.84	-0.011

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 2. RIF-OLS decomposition results with reweighting
GERMANY**

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2016	49,221.64***	32,506.90***	16,714.74***	32.241***
Year 2000	38,770.78***	23,420.82***	15,349.96***	26.495***
Difference / Total change	10,450.86***	9,086.08***	1,364.78***	5.746***
[1] Composition effect	2,376.09 (22.74%)	2,324.87 (25.59%)	51.22 (3.75%)	0.448 (7.80%)
[2] Structure effect	8,074.78*** (77.26%)	6,761.21** (74.41%)	1,313.56*** (96.25%)	5.298*** (92.20%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	693.23	919.20	-225.97	0.258
Tenure	323.58	183.37	140.21	0.034
Household composition	618.43	577.15	41.28	-0.135
Household members	-1,801.45**	-1,190.70	-610.75*	-0.054
Age	130.75	169.16	-38.41	-0.235
Sex	-360.66	-303.87	-56.78	-0.188
Marital status	95.15	89.78	5.37	-0.013
Immigrant	787.90	1,228.86*	-440.95	0.273
Disabled	31.20	35.00	-3.80	0.023
Health status	-498.78	-380.21	-118.57	-0.038
Education	648.46	270.29	378.17*	0.031
Employed	95.54	-7.37	102.91	0.073
Part-time employment	623.11*	247.74	375.37*	0.486**
$\hat{\Delta}_{X,e}^v$ = Specification error	1,682.86	1,405.67	277.19	0.190
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	7,399.95**	6,251.51**	1,148.44	5.220***
Tenure	2,139.80	1,361.48	778.31	1.042
Household composition	850.94	1,122.00	-271.06	1.345
Household members	2,866.67	5,599.46	-2,732.79	3.507
Age	-33,216.85	-45,325.33	12,108.49	-10.923
Sex	3,204.68	2,275.28	929.40	1.328
Marital status	2,340.14	160.37	2,179.77	1.151
Immigrant	-6,567.80***	-5,357.84**	-1,209.96	-1.919**
Disabled	372.61	297.38	75.23	0.124
Health status	-182.83	603.67	-786.50	2.896
Education	-3,938.43*	-2,274.55	-1,163.89	-2.016*
Employed	-4,462.70	4,284.55	-8,747.25	-9.380
Part-time employment	-2,415.16	-379.18	-2,035.98*	-0.722**
Constant	46,408.88	44,384.22	2,024.66	18.785
$\hat{\Delta}_{S,e}^v$ = Reweighting error	674.82	509.70	165.12	0.077

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 3. RIF-OLS decomposition results with reweighting
ITALY**

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2016	45,265.33***	31,587.85***	13,677.48***	36.234***
Year 2000	40,516.49***	27,391.27***	13,125.22***	31.791***
Difference / Total change	4,748.84	4,196.57	552.26	4.443*
[1] Composition effect	10,545.85 (222.07%)	10,598.16 (252.54%)	-52.31 (-9.47%)	1.759 (39.59%)
[2] Structure effect	-5,797.01 (-122.07%)	-6,401.58 (-152.54%)	604.57 (109.47%)	2.684 (60.41%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	3,509.39	3,125.31	384.08	-0.018
Tenure	516.19	569.29	-53.10	-0.397
Household composition	-2,951.46	-2,288.95	-662.54	-0.067
Household members	2,980.92	1,833.38	1,147.53	-1.575
Age	2,536.38	1,992.39	543.98	2.784
Sex	-2,002.36	-884.87	-1,117.48*	-0.959
Marital status	-152.94	11.63	-164.56	0.341
Immigrant	-30.89	75.55	-106.44	0.029
Disabled	-	-	-	-
Health status	-	-	-	-
Education	3,156.53	2,115.39	1,041.13	0.087
Employed	-	-	-	-
Part-time employment	-542.98	-298.52	-244.46	-0.259
$\hat{\Delta}_{X,e}^v$ = Specification error	7,036.46	7,472.85	-436.39	1.777*
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	-13,243.65	-14,472.03	1,228.38	0.501
Tenure	8,230.72	3,661.64	4,569.07	2.669
Household composition	-26,538.89	-29,700.45*	3,161.57	-8.051
Household members	-19,811.01	-26,851.01	7,040.00	-5.869
Age	193,930.70	146,042.30	47,888.40	100.654
Sex	26,163.07	14,189.26	11,973.81*	-16.016*
Marital status	-2,080.97	-1,026.83	-1,054.14	-7.962
Immigrant	2,589.95	1,916.04	673.91	0.360
Disabled	-	-	-	-
Health status	-	-	-	-
Education	-532.63	-332.31	-200.31	-0.251
Employed	-	-	-	-
Part-time employment	-456.79	-684.46	227.67	1.029
Constant	-194.737.80	-121,686.20	-73,051.59**	-66.059
$\hat{\Delta}_{S,e}^v$ = Reweighting error	7,446.64	8,070.45	-623.81	2.182*

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 4. RIF-OLS decomposition results with reweighting
POLAND**

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2016	28,659.98***	18,536.37***	10,123.62***	29.362***
Year 1999	17,867.00***	12,051.07***	5,815.93***	30.843***
Difference / Total change	10,792.98***	6,485.29***	4,307.69***	-1.480
[1] Composition effect	4,000.26*** (37.06%)	2,999.07*** (46.24%)	1,001.19*** (23.24%)	0.127 (-8.58%)
[2] Structure effect	6,792.73*** (62.94%)	3,486.23** (53.76%)	3,306.50*** (76.76%)	-1.607 (108.58%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	4,202.32***	2,823.80***	1,378.52***	1.376
Tenure	670.90***	443.71**	227.19***	0.704
Household composition	196.19***	271.87	-75.68	1.029
Household members	740.77	501.54*	211.20***	-0.873
Age	-191.44	-287.58	-201.62	-1.066***
Sex	-60.11	-42.09	161.09	-0.147
Marital status	-858.48***	-700.06***	1,100.86*	-0.235
Immigrant	-	-	-	-
Disabled	-3.99	-11.02	7.03	-0.012
Health status	-	-	-	-
Education	3,705.78***	2,653.17***	1,052.60***	1.952*
Employed	-	-	-	-
Part-time employment	2.71	-5.74	8.45	0.026
$\hat{\Delta}_{X,e}^v$ = Specification error	-202.06	175.27	-377.33**	-1.249***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	6,835.48***	3,502.29**	3,333.19***	-1.898
Tenure	429.64	-158.68	588.32	-1.379
Household composition	-2,200.98**	-1,547.12	-653.86**	-1.484*
Household members	287.60	446.67	-159.06	-2.895
Age	8,879.83	5,714.22	3,165.61	6.094
Sex	1,579.49	1,845.81	-266.32	6.838**
Marital status	-3,691.37*	-2,400.66	-1,290.71**	-1.757
Immigrant	-	-	-	-
Disabled	-220.87	-191.76	-29.10	-0.183
Health status	-	-	-	-
Education	-1,732.79	-1,606.39	-126.39	-1.213
Employed	-	-	-	-
Part-time employment	258.18	158.90	99.28	0.050
Constant	3,246.74	1,241.31	2,005.43	-5.966
$\hat{\Delta}_{S,e}^v$ = Reweighting error	-42.75	-16.06	-26.69	0.291*

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Table 5. RIF-OLS decomposition results with reweighting SPAIN

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2018	2,028.98***	1,301.16***	727.82***	31.585***
Year 2006	1,926.10***	1,292.09***	634.01***	30.407***
Difference / Total change	102.88**	9.06	93.81***	1.177***
[1] Composition effect	34.04 (33.09%)	34.03 (375.61%)	0.01 (0.02%)	0.342 (29.06%)
[2] Structure effect	68.84 (66.91%)	-24.97 (-275.61%)	93.80*** (99.98%)	0.835 (70.94%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v =$ Pure explained	94.15***	62.16***	31.99***	1.115***
Tenure	-19.43***	-4.89	-14.53***	0.299***
Household composition	0.29	0.46	-0.17	0.255***
Household members	-18.83***	-18.79***	-0.03	-0.438***
Age	20.92***	22.10***	-1.17	0.480***
Sex	3.68	1.39	2.29	0.502***
Marital status	0.30	-1.51	1.81	-0.073*
Immigrant	-8.27	-1.71	-6.55**	0.038
Disabled	-	-	-	-
Health status	-	-	-	-
Education	115.48***	65.13***	50.36***	0.052
Employed	-	-	-	-
Part-time employment	-	-	-	-
$\hat{\Delta}_{X,e}^v =$ Specification error	-60.11***	-28.13*	-31.98***	-0.773***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v =$ Pure unexplained	64.76	-23.31	88.07***	0.888*
Tenure	-296.46***	-6.72	-289.74***	-2.565***
Household composition	-14.21	12.89	-27.11***	-0.108
Household members	-50.59	-76.31	25.72	-2.276
Age	100.76	-694.05	794.82***	5.582
Sex	59.23	8.15	51.08*	0.574
Marital status	41.65	6.18	35.47**	-0.326
Immigrant	41.49**	32.60*	8.89	0.679***
Disabled	-	-	-	-
Health status	-	-	-	-
Education	13.39	26.11	-12.72	-0.416
Employed	-	-	-	-
Part-time employment	-	-	-	-
Constant	169.49	667.84	-498.35*	-2.276
$\hat{\Delta}_{S,e}^v =$ Reweighting error	4.08	-1.65	5.73**	0.053

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly for rounding reasons. (5) Empty cells are due to a lack of data.

Source: Own elaboration from EPF.

**Table 6. RIF-OLS decomposition results with reweighting
UNITED STATES**

<i>Dynamic approach</i>				
Territories over 500,000 inhabitants				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Year 2018	82,118.39***	55,769.74***	26,348.65***	37.074***
Year 2000	61,807.52***	38,573.80***	23,233.72***	34.491***
Difference / Total change	20,310.87***	17,195.94***	3,114.94***	2.582***
[1] Composition effect	3,904.41*** (19.22%)	3,008.79*** (17.50%)	895.62*** (28.75%)	0.739*** (28.62%)
[2] Structure effect	16,406.46*** (80.78%)	14,187.15*** (82.50%)	2,219.31*** (71.25%)	1.842*** (71.34%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	4,361.57***	3,320.48***	1,041.09***	0.974***
Tenure	-257.27***	-70.75*	-186.53***	0.121***
Household composition	-292.65***	-216.94***	-75.71***	-0.068**
Household members	797.16***	485.03***	312.13***	0.015
Age	1,732.28***	1,509.66***	222.63***	0.316**
Sex	-156.06***	-95.52***	-60.54***	-0.087***
Marital status	-129.78***	-19.85	-109.93***	0.174***
Immigrant	65.13	131.62**	-66.48**	0.095**
Disabled	-29.63*	-25.79*	-3.84	0.001
Health status	-318.29***	-193.04***	-125.25***	-0.013
Education	2,865.66***	1,772.46***	1,093.21***	0.269***
Employed	22.88	6.14	16.74	0.028
Part-time employment	62.15*	37.47	24.69	0.121***
$\hat{\Delta}_{X,e}^v$ = Specification error	-457.16	-311.69	-145.47*	-0.234***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	16,807.93***	14,602.61***	2,205.32***	2.032***
Tenure	-608.24	-846.53	238.29	-0.129
Household composition	-609.40	-467.85	-141.55	0.443
Household members	-1,537.84	-1,439.84	-98.00	-2.729**
Age	26,316.52*	22,115.59	4,200.93	5.679
Sex	-5,651.04*	-5,215.51*	-435.52	3.490**
Marital status	4,182.29***	2,471.24*	1,711.04***	1.208**
Immigrant	-337.72	-113.56	-224.15	0.453*
Disabled	225.29	221.96*	3.33	0.014
Health status	4,109.54**	2,702.85*	1,406.69**	1.199
Education	3,292.56***	2,182.75***	1,109.81***	0.049
Employed	-917.14	608.18	-1,525.32	-0.482
Part-time employment	232.61	159.11	73.50	0.216
Constant	-11,889.50	-7,775.79	-4,113.71	-7.380
$\hat{\Delta}_{S,e}^v$ = Reweighting error	-401.47	-415.56*	13.99	-0.189***

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

The *unexplained* part of the model, which is attributable to differences in prices associated with characteristics, shows that the most important driver of inequality dynamics is the education premium. Changes in the returns to education (the skill premium) account for a remarkable amount of the rise in inequality in Canada, especially in the top half of the income distribution, and in the United States. In Germany, the trend observed is just the opposite. It should also be emphasized that changes in the returns to the other variables analyzed make it difficult to find a clear pattern among countries. Heterogeneity can be appreciated not only by determining the relative importance of the diverse controls but also through other perspectives, such as identifying the impact over the entire distribution or the sign of the direction of change. However, in three countries (Canada, Poland and Spain), changes in the returns linked to one-person households exhibit a common feature: a reduction in income differences in the lower tail of the distribution. Additionally, it is worth highlighting that the figures reflecting variations in the returns to marital status are associated with a significant spread of income differences in Spain and the United States, especially at the half bottom of the distribution. As expected, the effect of the reweighting errors is practically imperceptible, which is a desirable result.

For the detailed decomposition of endowments and as far as the *pure explained* component is involved, changes in educational attainment seem to have a great impact on increases in inequality differences. This result is found in Canada, Poland, Spain and the United States throughout the entire distribution and in Germany at the bottom half of the distribution. In each instance, the magnitude of this impact—the extent to which inequality differences are increased—is the largest of all the control variables. In relative terms, the impact of changes in educational attainment is greater when considering the P50-P10 ratio in Spain and the United States, while in Canada, it is slightly greater at the half top of the distribution. At the same time, reductions in the number of household

members have a smaller effect on inequality in large cities than they did two decades ago. To this extent, compositional changes in household size contribute to an inequality-augmenting trend in Canada in the lower tail of the distribution, in Poland when looking at the P90-P50 and P50-P10 ratios, and in the United States across the whole distribution. For the P90-P10 and P50-P10 ratios in Germany and the P90-P10 and P90-P50 ratios in Spain, changes in this control variable partially offset the positive effect of education.

5.3. Reweighted RIF-OLS decomposition: Static approach

In a second approach, we analyze the differences between the drivers of income inequality in large cities and the other territories—those with fewer than 500,000 inhabitants. As mentioned above, we use the most recent wave of data as the basis for this analysis. Similar to the results of the dynamic approach, the results here also confirm widespread growth in Gini coefficient differences in all countries.⁴³ As in the previous analysis, divergences among the drivers of inequality in both areas are mainly explained by the *structure effect*. Conversely, the two components of aggregate decomposition are now significant in practically all countries. The exceptions are the *composition effect* in Italy and the *structure effect* in Poland. Germany (59.9% vs. 40.1%) and Spain (66.4% vs. 33.6%) are, in this order, the two countries with the most balanced percentages. On the other hand, the *structure effect* has a higher relative importance in the bottom half of the distribution in Canada, Germany, Italy and the United States (see Tables from 7 to 12).

⁴³ Again, the differential in Poland is not significant.

**Table 7. RIF-OLS decomposition results with reweighting
CANADA**

<i>Static approach</i>				
Year 2016				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	53,038.79***	33,793.29***	19,245.50***	29.718***
Less than 500,000 inhabitants	48,525.05***	30,840.31***	17,684.73***	28.022***
Difference / Total change	4,513.74***	2,952.98***	1,560.76***	1.696***
[1] Composition effect	-1,004.67*** (-22.26%)	-486.34* (-16.47%)	-518.32*** (-33.21%)	0.177* (10.44%)
[2] Structure effect	5,518.41*** (122.26%)	3,439.32*** (116.47%)	2,079.09*** (133.21%)	1.519*** (89.56%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	-930.95***	-552.15***	-378.79***	0.222**
Tenure	-134.91	130.43	-265.34***	0.454***
Household composition	-55.02	-31.65	-23.37	0.016
Household members	-736.38***	-379.93***	-356.46***	-0.119***
Age	-378.92***	-466.64***	87.72**	-0.054
Sex	-198.73***	-155.55***	-43.18**	6.90e-05
Marital status	-14.09	-5.06	-9.03	-4.56e-04
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	623.52***	361.62***	261.89***	-0.026
Employed	-	-	-	-
Part-time employment	-36.40	143.99	-31.02	-0.048
$\hat{\Delta}_{X,e}^v$ = Specification error	-73.72	65.81	-139.53	-0.046***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	5,438.63***	3,396.69***	2,041.94***	1.531***
Tenure	-3,516.69	-1,150.14	-2,366.55	1.845
Household composition	274.67	286.53	-11.85	-0.066
Household members	-2,856.18	-1,632.13	-1,224.04	-0.187
Age	-12,834.48	-9,923.35	-2,911.13	-2.218
Sex	-1,141.62	-462.90	-678.72	-2.304*
Marital status	3,004.44**	3,306.38**	-301.95	1.385**
Immigrant	-	-	-	-
Disabled	-	-	-	-
Health status	-	-	-	-
Education	1,231.37	902.20	-329.17	0.656
Employed	-	-	-	-
Part-time employment	-240.09	-254.15	14.07	0.092
Constant	21,517.19	12,324.25	9,192.95	2.326
$\hat{\Delta}_{S,e}^v$ = Reweighting error	79.78***	42.63***	37.15***	-0.012

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 8. RIF-OLS decomposition results with reweighting
GERMANY**

<i>Static approach</i>				
Year 2016				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	49,221.64***	32,506.90***	16,714.74***	32.241***
Less than 500,000 inhabitants	39,309.42***	25,247.03***	14,062.39***	28.289***
Difference / Total change	9,912.22***	7,259.87***	2,652.35***	3.952***
[1] Composition effect	691.50 (6.98%)	568.62 (7.83%)	122.88 (4.63%)	1.586*** (40.13%)
[2] Structure effect	9,220.72*** (93.02%)	6,691.25*** (92.17%)	2,529.47*** (95.37%)	2.366* (59.87%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	1,501.26***	1,375.66***	125.60	2.003***
Tenure	-1,806.68***	-790.79***	-1,015.89***	0.185
Household composition	-700.41***	-354.92*	-345.49***	0.067
Household members	1,402.20***	713.40***	688.80***	0.388***
Age	-654.95***	-699.54***	44.59	-0.190*
Sex	11.87	8.69	3.17	0.009
Marital status	-36.33	-7.64	-28.70	0.391***
Immigrant	-208.12*	82.19	-290.32***	0.105
Disabled	65.99	57.96	8.03	0.060*
Health status	309.35***	266.11**	43.24	0.025
Education	3,037.69***	2,083.01***	954.69***	1.012***
Employed	101.60	33.73	67.88	0.091
Part-time employment	-20.95	-16.55	-4.40	-0.141**
$\hat{\Delta}_{X,e}^v$ = Specification error	-809.76**	-807.04***	-2.72	-0.417***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	9,339.22***	6,795.53***	2,543.69***	2.405*
Tenure	2,443.89*	2,252.14	191.76	1.692**
Household composition	-1,201.38	-1,308.75	107.37	0.935
Household members	-3,855.40	-2,749.64	-1,105.77	7.931
Age	-51,254.37**	-45,941.39**	-5,312.98	-31.813*
Sex	2,844.87	-511.55	3,396.42	5.225
Marital status	5,133.14**	3,540.31	1,592.83*	1.785
Immigrant	-2,634.06**	-1,966.11	-667.95	-0.503
Disabled	-112.51	-89.38	-23.13	0.152
Health status	-748.40	-817.10	68.70	-0.013
Education	-467.08	-670.12	203.04	-1.236*
Employed	2,341.17	3,472.65	-1,131.47	2.224
Part-time employment	-39.19	723.73	-762.91	0.226
Constant	56,848.54**	50,860.76**	5,987.78	15.799***
$\hat{\Delta}_{S,e}^v$ = Reweighting error	-118.50*	-104.28**	-14.22	-0.039

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 9. RIF-OLS decomposition results with reweighting
ITALY**

<i>Static approach</i>				
Year 2016				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	45,265.33***	31,587.85***	13,677.48***	36.234***
Less than 500,000 inhabitants	30,310.16***	18,317.70***	11,992.46***	29.296***
Difference / Total change	14,955.17***	13,270.15***	1,685.01**	6.938***
[1] Composition effect	3,911.50*** (26.15%)	3,870.15*** (29.16%)	41.36 (2.45%)	0.812 (11.70%)
[2] Structure effect	11,043.67** (73.85%)	9,400.00** (70.84%)	1,643.66** (97.55%)	6.126** (88.30%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	2,798.55***	2,326.53***	472.03	0.837
Tenure	7.55	-2.57	10.12	-0.019
Household composition	-456.34	-374.99	-81.34	-0.177
Household members	142.42	210.87	-68.44	-0.539*
Age	766.20	707.62	58.58	0.672
Sex	140.24	133.16	7.07	0.161
Marital status	-154.50	-155.95	1.44	0.013
Immigrant	5.06	-15.20	20.26	-0.021
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2,398.49***	1,781.14***	617.35***	1.100**
Employed	-	-	-	-
Part-time employment	-50.55	42.46	-93.01	-0.353***
$\hat{\Delta}_{X,e}^v$ = Specification error	1,112.95	1,543.62*	-430.67	-0.025
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	10,955.83**	9,330.58**	1,625.25**	6.146**
Tenure	-557.44	-2,779.79	2,222.35	2.908
Household composition	-1,301.26	-1,903.00	871.75	-2.107
Household members	14,320.88	10,691.86	3,629.03	5.506
Age	-5,532.37	19,006.97	-24,559.34	42.669
Sex	-10,833.86	-12,480.40	1,646.54	-11.772
Marital status	-9,140.04	-8,587.29	-552.75	-9.816**
Immigrant	-522.54	73.78	-596.32	0.014
Disabled	-	-	-	-
Health status	-	-	-	-
Education	-185.63	-70.47	-115.15	-0.302
Employed	-	-	-	-
Part-time employment	381.79	230.25	151.54	0.862
Constant	24,076.29	5,148.68	18,927.60	-21.816
$\hat{\Delta}_{S,e}^v$ = Reweighting error	87.83	69.42	18.41	-0.020

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 10. RIF-OLS decomposition results with reweighting
POLAND**

<i>Static approach</i>				
Year 2016				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	28,659.98***	18,536.37***	10,123.62***	29.362***
Less than 500,000 inhabitants	20,571.21***	13,154.53***	7,416.68***	29.368***
Difference / Total change	8,088.77***	5,381.84***	2,706.93***	-0.006
[1] Composition effect	1,715.54*** (21.21%)	1,054.82*** (19.60%)	660.72*** (24.41%)	-0.917** (15.283%)
[2] Structure effect	6,373.23*** (78.79%)	4,327.02*** (80.40%)	2,046.21*** (75.59%)	0.911 (-15.183%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	2,203.18***	1,620.82***	582.36***	-0.637
Tenure	-372.56***	-112.81	-259.75***	-0.247***
Household composition	7.74	111.28	-103.54**	0.405*
Household members	357.21***	180.53*	176.68***	-0.317
Age	-169.18**	-23.91	-145.27***	-0.156**
Sex	-391.42***	-253.58***	-137.84***	0.021
Marital status	-339.74***	-243.32***	-96.42***	-0.243*
Immigrant	118.73*	141.75**	-23.02	0.229*
Disabled	8.19*	4.57	3.61*	0.004
Health status	-	-	-	-
Education	2,957.24***	1,848.37***	1,108.87***	-0.443
Employed	-	-	-	-
Part-time employment	26.97	-32.06**	59.03***	0.111***
$\hat{\Delta}_{X,e}^v$ = Specification error	-487.64*	-566.00**	78.36***	-0.280
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	6,371.96***	4,301.73***	2,070.23***	0.864
Tenure	-325.25	34.20	-359.46	-1.596
Household composition	-577.37	-339.49	-237.88	-0.716
Household members	-2,519.08	-2,921.45	402.37	-1.448
Age	20,072.80*	24,701.92**	-4,629.11	19.091*
Sex	281.55	-767.36	1,048.91	2.848
Marital status	-943.43	-823.36	-120.06	-1.532
Immigrant	-88.59	-187.78	99.19**	-0.035
Disabled	-63.57	-91.94	28.36	-0.012
Health status	-	-	-	-
Education	-73.99	237.36	-311.35	-0.149
Employed	-	-	-	-
Part-time employment	-30.66	90.57	-121.23	-0.162
Constant	-9,360.43	-15,630.93	6,270.50*	-15.424*
$\hat{\Delta}_{S,e}^v$ = Reweighting error	1.27	25.29	-24.02*	0.047

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

**Table 11. RIF-OLS decomposition results with reweighting
SPAIN**

<i>Static approach</i>				
Year 2018				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	2,028.98***	1,301.16***	727.82***	31.585***
Less than 500,000 inhabitants	1,625.78***	995.22***	630.56***	30.045***
Difference / Total change	403.20***	305.94***	97.26***	1.540***
[1] Composition effect	93.75*** (23.25%)	61.45*** (20.09%)	32.30*** (24.41%)	0.518*** (33.63%)
[2] Structure effect	309.45*** (76.75%)	244.49*** (79.91%)	64.96*** (75.59%)	1.022*** (66.37%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	110.21***	73.43***	36.77***	0.853***
Tenure	13.67***	10.64***	3.02***	0.621***
Household composition	-1.85***	-0.37	-1.47***	0.075***
Household members	0.32	0.36	-0.04	-0.101***
Age	-11.45***	-6.35***	-5.10***	-0.099***
Sex	8.01***	5.93***	2.14***	0.111***
Marital status	-1.55**	-1.45**	-0.10	-0.008
Immigrant	5.33***	5.61***	-0.28	0.341***
Disabled	-	-	-	-
Health status	-	-	-	-
Education	97.67***	59.06***	38.61***	-0.087**
Employed	-	-	-	-
Part-time employment	-	-	-	-
$\hat{\Delta}_{X,e}^v$ = Specification error	-16.45***	-11.97*	-4.47**	-0.335***
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	310.27***	245.33***	64.94***	1.044***
Tenure	31.93	114.15*	-82.22***	1.444**
Household composition	204.42	13.09	-7.74	0.015
Household members	-514.18*	194.61**	9.81	-2.544
Age	520.99	251.67	269.31	1.004
Sex	36.72	22.22	14.50	0.982**
Marital status	12.69	7.31	5.38	0.070
Immigrant	9.76	5.49	4.27	-0.138
Disabled	-	-	-	-
Health status	-	-	-	-
Education	2.57	28.07	-25.50**	-0.366
Employed	-	-	-	-
Part-time employment	-	-	-	-
Constant	-514.18	-391.30	-122.87	-2.544
$\hat{\Delta}_{S,e}^v$ = Reweighting error	-0.82***	-0.83***	0.02	0.022***

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding. (5) Empty cells are due to lack of data.

Source: Own elaboration from EPF.

**Table 12. RIF-OLS decomposition results with reweighting
UNITED STATES**

<i>Static approach</i>				
Year 2018				
Inequality measures	P90-P10	P90-P50	P50-P10	Gini (x100)
<i>Overall</i>				
Over 500,000 inhabitants	82,118.39***	55,769.74***	26,348.65***	37.074***
Less than 500,000 inhabitants	62,049.97***	40,248.00***	21,801.96***	34.129***
Difference / Total change	20,068.43***	15,521.74***	4,546.69***	2.944***
[1] Composition effect	1,548.17*** (7.71%)	1,334.22*** (8.60%)	213.95 (4.71%)	0.497*** (16.88%)
[2] Structure effect	18,520.25*** (92.29%)	14,187.52*** (91.40%)	4,332.74*** (95.29%)	2.447*** (83.12%)
<i>[1] Composition effect (Explained)</i>				
$\hat{\Delta}_{X,p}^v$ = Pure explained	1,659.75***	1,572.53***	87.22	0.853***
Tenure	-296.03***	16.47	-312.51***	0.228***
Household composition	-134.97**	-108.76**	-26.21*	0.007
Household members	-119.53	-76.09	-43.43	-0.021
Age	-213.45**	-243.32***	29.87	-0.048*
Sex	16.13	13.97	2.16	0.002
Marital status	-160.54***	-67.64**	-92.89***	0.043**
Immigrant	249.77	546.89*	-297.12**	0.339**
Disabled	46.15*	44.53*	1.62	0.029***
Health status	203.16***	88.44	114.71***	-0.029
Education	2,171.08***	1,408.68***	762.40***	0.220***
Employed	-59.01**	-17.14	-41.87**	-0.068**
Part-time employment	-42.98	-33.50	-9.48	-0.086***
$\hat{\Delta}_{X,e}^v$ = Specification error	-111.57	-238.30	126.73*	-0.119**
<i>[2] Structure effect (Unexplained)</i>				
$\hat{\Delta}_{S,p}^v$ = Pure unexplained	18,411.02***	14,123.58***	4,287.43***	2.478***
Tenure	597.99	261.71	336.28	-0.282
Household composition	-775.44	-589.33	-186.10	-0.106
Household members	-1,723.99	-1,219.55	-504.44	0.538
Age	1,270.94	-1,001.50	2,272.45	2.084
Sex	-3,807.38	-2,626.51	-1,180.87	0.819
Marital status	2,346.82*	1,920.21	426.61	0.109
Immigrant	-360.09	-69.62	-290.47	0.308
Disabled	104.92	97.62	7.29	0.123**
Health status	1,620.72	1,460.77	159.94	0.677
Education	1,718.04*	1,772.76**	-54.71	-0.238
Employed	1,345.07	2,017.66	-672.59	0.012
Part-time employment	-256.40	-156.53	-99.86	0.056
Constant	16,329.81	12,255.90	4,073.91	-1.623
$\hat{\Delta}_{S,e}^v$ = Reweighting error	109.23	63.93	45.29	-0.031

Notes: (1) Statistical significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$. (2) Bootstrapped standard errors were used to compute the p-values (500 replications). (3) Age: aggregates age original variable and age squared. Education: includes all three education group categories: low, medium and high. Health status: gathers all five health status groupings: very good, good, satisfactory, poor and bad. (4) Some sums may not match exactly due to rounding.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Beginning with the *structure effect*, the first relevant result is the change in the returns to education (skill premium), which accounts for a great amount of the increase in inequality in Canada—markedly, in the upper tail of the distribution—and in the United States. In Germany, the trend is the opposite. The partial contributions of the rest of the drivers, in contrast to what is observed for the *composition effect*, barely allow us to establish a minimum correspondence or similarity among countries. In contrast, they are characterized by a remarkable lack of uniformity as well as by differences that are less significant. The influence of the reweighting errors in this second approach is, to an extent, more visible and important to their significance, though not in terms of their percentage weight in the total *structure effect*.

Regarding the *composition effect*, a similar conclusion to the one expressed above holds when the analysis focuses on the differences in inequality between large cities and other areas. The detailed decomposition, which is related to the *pure explained* component, shows that compositional changes in education have a notable effect on income gaps. The contribution of the maximum level of education achieved is again the key driver of the *composition effect*. This is so not only because of the magnitude of the effect, which is the greatest in regard to increasing the differentials in disposable income, but also because of its significance throughout the whole distribution for each of the six countries. Likewise, the impact is much more pronounced in the lower tail of the distribution (the P50-P10 ratio) for all countries. This result is in accordance with the literature reviewed (Bacolod et al., 2009; Baum-Snow and Pavan, 2013), being the consequence of the growth of more highly skilled and better paid workers in large cities—the ‘paradox of progress’ (Bourguignon et al., 2005). The wages and the number of skilled workers have increased more in large cities, while the wage of unskilled workers have remained constant or decreased in those areas.

Another variable playing a prominent role in income differences with regard to endowments is age. In large cities from countries such as Canada, Germany, Spain and the United States, aging has a strong negative effect on inequality. This result is consistent with the findings of Alimi et al. (2018), who found that metropolitan areas experienced rapid growth in inequality but slower rates of aging in New Zealand, which is mainly attributable to net inward migration rather than greater fertility, while nonmetropolitan areas had slower growth in inequality and faster rates of aging.

6. Conclusions

The study of the relationships between income inequality and city size has traditionally been constrained by a number of limitations. The lack of empirically well-versed models, the inadequate understanding of variable interactions and data restrictions have been severe barriers to the construction of suitable measurements and an understanding of their interactions.

In an attempt to partially overcome these problems, this chapter contributes a better understanding of this relationship to the literature. To do so, we have conducted empirical research focused on inequality differences in large cities and other areas within a selected sample of OECD countries. The goal has been to identify and quantify the individual contributions of different drivers to explanations of these differences. For this purpose, we have used a methodological approach (the FFL decomposition) that allows us to look at these differences both from a dynamic and static perspective.

One of the main findings of this chapter is the remarkable increase in income inequality in large cities in the selected countries. This is found with both approaches and with different inequality measures. These differences are mainly determined by a *structure effect* rather than changes in the distribution of endowments. The evidence obtained sheds

light on the relevance of education as one of the most important drivers of income inequality in these areas.

We have examined income inequality trends in large cities for a period covering close to two decades. Changes in the returns to education (the skill premium) account for a large amount of the increase in inequality in Canada—noticeably, in the top half of the income distribution—and in the United States. In Germany, the trend is the opposite. Changes in the returns to the rest of the variables analyzed present a different picture. The returns to the geography and housing control variables, as well as those related to labor market status, seem to describe more idiosyncratic behavior in each country. Regarding endowments, changes in educational attainment seem to also play a central role in shaping income inequality in large cities over time. We find that most countries share this result throughout the whole distribution. In all countries, the magnitude of this impact is the largest of all the control variables. Household size is also observed to increase inequality in several countries.

We have also used the most recent data available to compare large cities with other areas. In keeping with previous studies, we find that the returns to educational skills are higher in larger agglomerations of the United States. However, the opposite is observed in other countries, with larger cities exhibiting smaller returns for high-skilled workers. In the case of composition effects, the same conclusion stressed using the *dynamic approach* applies, but here, the effects are even stronger. Therefore, changes in educational attainment again play a predominant role in explaining the growth in income inequality. This ‘paradox of progress’ reveals that “a rightward movement in the distribution of years of schooling shifts population density to steeper segments of the earnings-education profile, leading to wider earnings gaps” (Ferreira et al., 2017). On the other hand, the evidence shown reveals a negative age-composition effect in five of the six countries examined.

We contribute here to this growing body of empirical evidence with a more accurate measurement of the effect of this and other potential drivers by considering a larger number of countries and a dual static and dynamic perspective.

In a socioeconomic context such as the one in which we are living today, full of uncertainty, it is vital to shed light on the particular contributions of the main drivers of the changes in the income distribution in different areas. It is especially relevant to clarify the reasons behind the recent rise in income inequality in areas where an increasing percentage of the population resides. COVID-19 has suddenly emerged as a dangerous threat to large cities. To prevent its effects on the most vulnerable groups from becoming more marked, policymakers should refer to the most complete information on the influence of these determinants. An adequate response at the present time is crucial not only to minimize the devastating effects of this crisis but also to anticipate responses to future shocks.

References

- Alimi, O. B., Maré, D. C. and Poot, J. (2018): “More pensioners, less income inequality? The impact of changing age composition on inequality in big cities and elsewhere.” In *Modelling aging and migration effects on spatial labor markets* (pp. 133-159). Springer, Cham.
- Alperovich, G. (1995): “The Relationship between Income Inequality and City Size: A General Equilibrium Model of an Open System of Cities Approach.” *Urban Studies* 32(6): 853-862.
- Bairoch, P. (1988): *Cities and economic development. From the dawn of history to the present*. The University Chicago Press.
- Barsky, R., Bound, J., Charles, K. K. and Lupton, J. P. (2002): “Accounting for the black–white wealth gap: a nonparametric approach.” *Journal of the American statistical Association*, 97(459), 663-673.
- Baum-Snow, N. and Pavan, R. (2013): “Inequality and city size.” *Review of Economics and Statistics*, 95(5), 1535-1548.
- Baum-Snow, N., Freedman, M. and Pavan, R. (2018): “Why has urban inequality increased.” *American Economic Journal: Applied Economics* 2018, 10(4): 1–42.
- Bacolod, M., Blum, B. S. and Strange, W. C. (2009): “Skills in the city.” *Journal of Urban Economics*, 65(2), 136-153.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455.
- Boulant, J., Brezzi, M. and Veneri, P. (2016): *Income levels and Inequality in Metropolitan Areas: A comparative Approach in OECD Countries*. OECD Publishing, Paris.
- Bourguignon, F., Ferreira, F. and Lustig, N. (2005): *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*. Washington, DC, United States: Oxford University Press and World Bank.
- Castells-Quintana, D., Royuela, V. and Veneri, P. (2020): “Inequality and city size: An analysis for OECD functional urban areas.” *Papers in Regional Science*. Forthcoming.
- Chen, B., Liu, D. and Lu, M. (2018): “City size, migration and urban inequality in China.” *China Economic Review*, 51, 42-58.
- Cloutier, N.R. (1997): “Metropolitan income inequality during the 1980s: The impact of urban development, industrial mix an, and family structure.” *Journal of Regional Science*, 37(3): 459-478.
- Davis, D.R. and Dingel, J.I. (2019): “A spatial knowledge economy.” *American Economic Review*, 109(1): 153-170.
- Dijkstra, L. and Poelman, H. (2012): “Cities in Europe: the new OECD-EC definition.” *Regional focus*, 1(2012), 1-13.
- DiNardo, J. M. Fortin N. and Lemieux T. (1996): “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach.” *Econometrica*, 64(5), 1001.

- Dubey, A. and Mahadevia, D. (2001): "Poverty and inequality in Indian metropolises." *Indian Journal of Labour Economics*, 44(2): 223-242.
- Ferreira, F. H., Firpo, S. P. and Messina, J. (2017): *Ageing Poorly? Accounting for the decline in earnings inequality in Brazil, 1995–2012*. The World Bank.
- Firpo, S. P., Fortin, N. M. and Lemieux, T. (2009): "Unconditional quantile regressions." *Econometrica*, 77(3), 953-973.
- Firpo, S. P., Fortin, N. M. and Lemieux, T. (2018): "Decomposing wage distributions using recentered influence function regressions." *Econometrics*, 6(2), 28.
- Fortin, N., Lemieux, T. and Firpo, S. (2011): Decomposition methods in economics. In *Handbook of labor economics* (Vol. 4, pp. 1-102). Elsevier.
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh M., Lobo, J., Moro, E., Wang, D., Youn H. and Rahwan, I. (2019): "Toward understanding the impact of artificial intelligence on labor." *Proceedings of the National Academy of Sciences of the United States of America*, 116(14), 653–6539.
- Galster, G., McCorkhill, G. and Gopalan, S. (1988): "The Determinants of Income Inequality, In Metropolitan Areas." *Review of Business*, 10(2): 17-22.
- Garofalo, G. and Fogarty, M.S. (1979): "Urban Income Distribution and the Urban Hierarchy-Equality Hypothesis." *Review of Economics and Statistics*, 61(3): 381-388.
- Glaeser, E., Resseger, M. and Tobio, C. (2009): "Inequality in cities." *Journal of Regional Science*, 49(4): 617–646.
- Gould, E.D. (2007): "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium." *Review of Economic Studies*, 74: 477-506.
- Gradín, C. (2016): "Why is Income Inequality so High in Spain?", *Income Inequality Around the World (Research in Labor Economics, Vol. 44)*, Emerald Group Publishing Limited, pp. 109-177.
- Haque, I., Das, D. N. and Patel, P. P. (2018): "Spatial segregation in Indian cities- does the city size matter?" *Environment and Urbanization Asia*, 9(1): 52–68.
- Henkel (2017): "Decomposing the role of urbanization in wage inequality." in *Three essays on the causes and consequences of spatial inequality*, Doktors dissertation. Universität Duisburg-Essen.
- Hortas-Rico, M. and Rios, V. (2019): "The drivers of local income inequality: a spatial Bayesian model-averaging approach." *Regional Studies*, 53(8): 1207–1220.
- Juhn, C., Murphy, K. M. and Pierce, B. (1993): "Wage inequality and the rise in returns to skill." *Journal of political Economy*, 101(3), 410-442.
- Lee N, Sissons P, Jones K (2016): "The Geography of Wage Inequality in British Cities." *Regional Studies*, 50(10): 1714-1727.
- Machado, J. y Mata, J.A.F. (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics*, 20(4), pp. 445-465.

- Mastronardi, L. and Cavallo, A. (2020): “The spatial dimension of income inequality: an analysis at municipal level”, *Sustainability*, 12(4): 1622.
- Melly, B. (2006): “Estimation of counterfactual distributions using quantile regression.” Mimeo, Swiss Institute for International Economics and Applied Economic Research, University of St. Gallen.
- Nord, S. (1980a): “An Empirical Analysis of Income Inequality and City Size.” *Southern Economic Journal*, 46(3): 863-872.
- Nord, S. (1980b): “Income inequality and city size: An examination of alternative hypothesis for large and small cities.” *Review of Economics and Statistics* 62: 502-508.
- Oaxaca, R. L. (1973): “Male-female wage differentials in urban labor markets.” *International economic review*, 693-709.
- Oaxaca, R. L. and Ransom, M. R. (1999): “Identification in detailed wage decompositions.” *Review of Economics and Statistics*, 81(1), 154-157.
- OECD (2012): Redefining “Urban”: A New Way to Measure Metropolitan Areas, OECD Publishing.
- OECD (2020): Cities Policy Responses, TACKLING CORONAVIRUS (COVID-19): CONTRIBUTING TO A GLOBAL EFFORT, OECD Publishing.
- Rothe, C. (2015): “Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes.” *Journal of Business & Economic Statistics*, 33(3), 323-337.
- Royuela, V., Veneri, P. and Ramos, R. (2014): *Income Inequality, Urban Size and Economic Growth in OECD Regions*, OECD Regional Development Working Papers, 2014/10, OECD Publishing.
- Soroka, L.A. (1984): “City size and income distributions: The Canadian experience.” *Urban Studies*, 21: 359-365.
- United Nations (2018): *The World’s Cities in 2018*. (ST/ESA/ SER.A/417). Department of Economic and Social Affairs, Population Division. New York: United Nations.
- United Nations (2019): *World Urbanization Prospects: The 2018 Revision* (ST/ESA/SER.A/420). Department of Economic and Social Affairs, Population Division. New York: United Nations.
- Wheeler, C. (2004a): “On the distributional aspects of urban growth.” *Journal of Urban Economics*, 55: 371-397.
- Wheeler, C. (2004b): “Wage inequality and urban density.” *Journal of Economic Geography*, 4:421-437.
- Yun, M. (2005): “A Simple Solution to the Identification Problem in Detailed Wage Decompositions.” *Economic Inquiry*, Vol. 43, pp. 766-772.

Appendix

Table A.1. Large cities by country and year

CANADA (2016) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Toronto	–	–	7,120,321
Montreal	–	4,393,811	–
Vancouver	–	2,575,988	–
Calgary	–	1,570,523	–
Ottawa	–	1,486,404	–
Edmonton	–	1,398,763	–
Quebec	887,952	–	–
Winnipeg	873,361	–	–
Hamilton	642,171	–	–
London	576,626	–	–
Kitchener	579,198	–	–
Total	5	5	1

CANADA (2001) ⁴⁴ Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Toronto	–	–	5,467,380
Montreal	–	3,819,791	–
Vancouver	–	1,982,265	–
Calgary	–	1,037,040	–
Ottawa	974,495	–	–
Edmonton	968,780	–	–
Quebec	774,533	–	–
Winnipeg	736,910	–	–
Hamilton	567,570	–	–
Total	5	3	1

GERMANY (2016) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Berlin	–	–	5,141,777
Ruhr	–	–	5,105,158
Hamburg	–	3,232,955	–
Munich	–	2,848,720	–
Stuttgart	–	2,734,661	–
Frankfurt am Main	–	2,648,379	–
Cologne	–	1,971,296	–
Dusseldorf	–	1,548,250	–
Dresden	–	1,342,639	–
Nuremberg	–	1,323,484	–
Hanover	–	1,299,281	–
Bremen	–	1,254,966	–
Mannheim-Ludwigshafen	–	1,172,753	–
Leipzig	–	1,016,335	–
Braunschweig-Salzgitter Wolfsburg	995,693	–	–
Bonn	915,640	–	–
Saarbrücken	802,268	–	–
Karlsruhe	743,648	–	–
Heidelberg	698,956	–	–
Augsburg	662,855	–	–
Freiburg im Breisgau	645,798	–	–
Kiel	640,186	–	–
Aachen	551,919	–	–
Erfurt	527,401	–	–
Münster	527,130	–	–
Osnabrück	520,481	–	–
Würzburg	501,022	–	–

⁴⁴ No data available for the year 2000.

Descomposición simultánea de la desigualdad. Algunas aplicaciones

Magdeburg	500,757	–	–
Total	14	12	2

GERMANY (2000) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Ruhr	–	–	5,375,423
Berlin	–	4,920,475	–
Hamburg	–	3,046,521	–
Stuttgart	–	2,600,384	–
Frankfurt am Main	–	2,479,572	–
Munich	–	2,409,918	–
Cologne	–	1,849,456	–
Dusseldorf	–	1,521,492	–
Dresden	–	1,396,127	–
Hanover	–	1,278,373	–
Nuremberg	–	1,267,010	–
Bremen	–	1,224,966	–
Mannheim-Ludwigshafen	–	1,165,989	–
Leipzig	–	1,012,529	–
Braunschweig-Salzgitter Wolfsburg	–	1,008,968	–
Bonn	874,299	–	–
Saarbrücken	869,585	–	–
Karlsruhe	693,700	–	–
Heidelberg	662,453	–	–
Kiel	629,360	–	–
Augsburg	611,702	–	–
Freiburg im Breisgau	591,613	–	–
Erfurt	555,425	–	–
Aachen	547,450	–	–
Magdeburg	543,739	–	–
Osnabrück	517,576	–	–
Würzburg	507,055	–	–
Göttingen	502,907	–	–
Total	13	14	1

ITALY (2016) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Milan	–	4,944,254	–
Rome	–	4,421,905	–
Naples	–	3,436,482	–
Turin	–	1,761,884	–
Palermo	–	1,018,430	–
Florence	801,112	–	–
Bologna	773,245	–	–
Bari	732,568	–	–
Genoa	695,371	–	–
Catania	638,866	–	–
Venice	536,379	–	–
Padua	531,619	–	–
Verona	506,172	–	–
Total	8	5	0

ITALY (2000) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Milan	–	4,478,932	–
Rome	–	3,785,190	–
Naples	–	3,365,751	–
Turin	–	1,679,348	–
Palermo	995,385	–	–
Florence	739,755	–	–
Genoa	723,951	–	–

Descomposición simultánea de la desigualdad. Algunas aplicaciones

Bari	705,574	–	–
Bologna	697,587	–	–
Catania	604,584	–	–
Venice	505,183	–	–
Total	7	4	0

POLAND (2016) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Warsaw	–	3,132,671	–
Katowice	–	2,541,191	–
Cracow	–	1,407,564	–
Gdansk	–	1,154,488	–
Poznan	983,006	–	–
Lodz	915,942	–	–
Wroclaw	876,511	–	–
Lublin	668,070	–	–
Rzeszow	510,949	–	–
Total	5	4	0

POLAND (2002) ⁴⁵ Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Warsaw	–	2,886,637	–
Katowice	–	2,699,020	–
Cracow	–	1,327,292	–
Gdansk	–	1,071,402	–
Lodz	979,154	–	–
Poznan	910,433	–	–
Wroclaw	824,925	–	–
Lublin	673,275	–	–
Total	4	4	0

SPAIN (2018) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Madrid	–	–	6,791,667
Barcelona	–	–	4,985,549
Valencia	–	1,733,606	–
Seville	–	1,545,862	–
Bilbao	–	1,005,328	–
Malaga	862,518	–	–
Saragossa	759,060	–	–
Palma de Mallorca	685,786	–	–
Las Palmas	632,018	–	–
Murcia	629,612	–	–
Granada	563,090	–	–
Vigo	542,277	–	–
Santa Cruz de Tenerife	510,019	–	–
Total	8	3	2

SPAIN (2006) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
Madrid	–	–	6,146,639
Barcelona	–	–	4,727,204
Valencia	–	1,654,049	–
Seville	–	1,444,514	–

⁴⁵ No data available for the years 1999, 2000 and also for 2001.

Descomposición simultánea de la desigualdad. Algunas aplicaciones

Bilbao	–	1,000,146	–
Malaga	791,489	–	–
Saragossa	711,183	–	–
Las Palmas	622,411	–	–
Palma de Mallorca	613,705	–	–
Murcia	572,944	–	–
Vigo	531,260	–	–
Granada	514,672	–	–
Total	7	3	2

UNITED STATES (2018) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
New York	–	–	19,961,045
Los Angeles	–	–	17,913,847
Chicago	–	–	9,498,716
Washington	–	–	9,115,152
Dallas	–	–	7,712,624
Houston	–	–	7,097,351
San Francisco	–	–	6,728,591
Philadelphia	–	–	6,466,183
Miami	–	–	6,359,694
Atlanta	–	–	5,600,743
Phoenix	–	4,857,962	–
Boston	–	4,436,124	–
Detroit	–	4,388,853	–
Seattle	–	3,939,363	–
Minneapolis	–	3,575,736	–
San Diego	–	3,343,364	–
Denver	–	2,932,415	–
St. Louis	–	2,602,213	–
Orange	–	2,572,962	–
San Antonio	–	2,537,852	–
Portland	–	2,371,808	–
Sacramento	–	2,345,210	–
Las Vegas	–	2,282,194	–
Austin	–	2,180,018	–
Cincinnati	–	2,139,570	–
Columbus	–	2,070,508	–
Jackson (MO)	–	2,063,645	–
Cuyahoga	–	2,057,009	–
Charlotte	–	2,038,115	–
Indianapolis	–	2,033,469	–
New Haven	–	1,801,443	–
Salt Lake	–	1,661,471	–
Jacksonville	–	1,629,078	–
Milwaukee	–	1,576,113	–
Tampa-Pinellas	–	1,514,910	–
Davidson	–	1,442,050	–
Tampa-Hillsborough	–	1,436,888	–
Pittsburgh	–	1,425,798	–
Jefferson (KY)	–	1,402,813	–
Oklahoma	–	1,356,710	–
Memphis	–	1,311,917	–
Wake	–	1,294,980	–
New Orleans	–	1,249,362	–
Hartford	–	1,206,300	–
Virginia Beach	–	1,185,493	–
Richmond	–	1,170,325	–
Fresno	–	1,152,072	–
Erie (NY)	–	1,130,152	–
Jefferson (AL)	–	1,085,248	–
Pima	–	1,039,073	–
Tulsa	–	1,034,904	–
Albany	980,115	–	–
Providence	974,773	–	–

Douglas (NE)	949,191	–	–
Albuquerque	942,673	–	–
Kern	896,764	–	–
Hidalgo	865,939	–	–
Ventura	850,967	–	–
Rochester (NY)	846,313	–	–
El Paso (TX)	845,553	–	–
East Baton Rouge	831,310	–	–
Worcester	830,839	–	–
Sarasota	821,573	–	–
Richland	812,122	–	–
Charleston	787,643	–	–
Ada	757,685	–	–
Lee	754,610	–	–
San Joaquin	752,660	–	–
El Paso (CO)	738,939	–	–
Lehigh	737,134	–	–
Kent	717,754	–	–
Montgomery (OH)	700,326	–	–
Polk	664,832	–	–
Onondaga	650,502	–	–
Utah	633,768	–	–
Hampden	631,761	–	–
Sedgwick	603,368	–	–
Dane	599,722	–	–
Dauphin	599,363	–	–
Brevard	596,849	–	–
Pulaski	587,758	–	–
Hamilton (TN)	560,793	–	–
Lucas	560,595	–	–
Stanislaus	549,815	–	–
Volusia-Daytona Beach	547,538	–	–
Fayette	546,699	–	–
Lancaster (PA)	543,557	–	–
Summit	541,918	–	–
Spokane	538,973	–	–
Greene	535,788	–	–
Guilford	533,670	–	–
Washoe	525,572	–	–
Greenville	514,213	–	–
Forsyth	504,842	–	–
Total	43	41	10

UNITED STATES (2000) Metropolitan areas	City size (inhabitants)		
	XL 500,000 – 1,000,000	XXL 1,000,000 – 5,000,000	Global city > 5,000,000
New York	–	–	18,941,708
Los Angeles	–	–	15,669,726
Chicago	–	–	9,113,234
Washington	–	–	7,467,081
Philadelphia	–	–	6,043,433
San Francisco	–	–	5,874,603
Dallas	–	–	5,387,746
Miami	–	–	5,153,196
Houston	–	4,801,921	–
Detroit	–	4,513,596	–
Boston	–	4,010,014	–
Atlanta	–	4,007,660	–
Phoenix	–	3,273,477	–
Seattle	–	3,052,187	–
Minneapolis	–	2,996,224	–
San Diego	–	2,827,366	–
St. Louis	–	2,482,576	–
Denver	–	2,192,840	–
Cuyahoga	–	2,147,532	–
Cincinnati	–	1,949,645	–

Portland	–	1,849,594	–
Sacramento	–	1,807,949	–
Jackson (MO)	–	1,748,839	–
San Antonio	–	1,736,206	–
New Haven	–	1,709,275	–
Orange	–	1,656,890	–
Indianapolis	–	1,648,997	–
Columbus	–	1,647,969	–
Milwaukee	–	1,502,420	–
Pittsburgh	–	1,482,659	–
Las Vegas	–	1,431,002	–
New Orleans	–	1,318,079	–
Charlotte	–	1,314,561	–
Austin	–	1,273,445	–
Tampa-Pinellas	–	1,269,188	–
Salt Lake	–	1,235,509	–
Jefferson (KY)	–	1,219,405	–
Jacksonville	–	1,206,482	–
Memphis	–	1,181,538	–
Erie (NY)	–	1,169,060	–
Hartford	–	1,150,872	–
Davidson	–	1,075,603	–
Virginia Beach	–	1,073,818	–
Oklahoma	–	1,066,496	–
Tampa-Hillsborough	–	1,003,435	–
Jefferson (AL)	993,496	–	–
Providence	964,534	–	–
Richmond (Greater)	935,082	–	–
Albany	924,990	–	–
Fresno (Greater)	924,875	–	–
Tulsa	899,650	–	–
Pima	848,019	–	–
Rochester (NY)	847,862	–	–
Douglas (NE)	777,286	–	–
Albuquerque	757,853	–	–
Wake	756,546	–	–
Ventura	756,506	–	–
Worcester	752,512	–	–
East Baton Rouge	707,589	–	–
Montgomery (OH)	706,923	–	–
El Paso (TX)	685,086	–	–
Kern	663,803	–	–
Onondaga	649,961	–	–
Lehigh	638,915	–	–
Kent	637,662	–	–
Richland	630,376	–	–
Hampden	608,897	–	–
Sarasota	592,809	–	–
Lucas	576,103	–	–
Hidalgo	573,216	–	–
San Joaquin	567,885	–	–
Charleston	550,916	–	–
Summit	543,442	–	–
El Paso (CO)	540,593	–	–
Sedgwick	539,332	–	–
Dauphin	532,301	–	–
Total	31	37	8

Note: The OECD, in cooperation with the European Union (EU), has developed a harmonised definition of FUAs. Being composed of a city and its commuting zone, FUAs encompass the economic and functional extent of cities based on daily people's movements (OECD, 2012). The definition of FUA aims at providing a functional/economic definition of cities and their area of influence, by maximising international comparability and overcoming the limitation of using purely administrative approaches. At the same time, the concept of FUA, unlike other approaches, ensure a minimum link to the government level of the city or metropolitan area.

Source: Own elaboration from OECD.

Table A.2. Variables used in the estimations

Categorization	Name	Definition
OUTCOME VARIABLE		
<i>Major economic aggregates (Income aggregates)</i>	Equivalent disposable household income LIS variable: <i>dhi</i> EPF proxy: <i>impexac</i>	Sum of cash and non-cash income from labour, income from capital, income from pensions (including private and public pensions) and non-pension public social benefits stemming from insurance, universal or assistance schemes (including in-kind social assistance transfers), as well as cash and non-cash private transfers, less the amount of income taxes and social contributions paid.
CONTROL VARIABLES		
<i>Geography and housing</i>	Tenure (h-file) LIS variable: <i>own</i> EPF proxy: <i>regten</i>	Indicator of housing tenure. We have redefined it as a dummy variable: 1 = owned; 0 = rented/other.
<i>Household composition and living arrangements</i>	Household composition (h-file) LIS variable: <i>hhtype</i> EPF proxy: <i>tiphogar1</i>	The composition of the household with respect to the head. We have redefined it as a dummy variable: 1 = one-person household; 0 = other values.
	Household members (h-file) LIS variable: <i>nghmem</i> EPF proxy: <i>nmiemb</i>	Number of household members.
<i>Socio-demographic Characteristics</i>	Age ⁴⁶ (p-file) LIS variable: <i>age</i> EPF proxy: <i>edadsp</i>	Age in years.
	Sex (p-file) LIS variable: <i>sex</i> EPF proxy: <i>sexosp</i>	Classification of individuals according to their sex. It is defined as a dummy variable: 1 = female; 0 = male.
	Marital status (p-file) LIS variable: <i>marital</i> EPF proxy: <i>ecivil / ecivillegal1</i>	Classification of individuals according to their marital status, as provided in relation to the marriage laws or customs of the country. We have redefined it as a dummy variable: 1 = married/in union; 0 = another status.

⁴⁶ We have also included in the estimations the variable age squared.

<i>Socio-demographic characteristics</i>	Immigrant (p-file) LIS variable: <i>immigr</i> EPF proxy: <i>nacionasp</i>	Individuals who have that country as their country of usual residence and (in order of priority): whom the data provider identified as immigrants; who self-identify as immigrants; who are a citizen/national of another country; who were born in another country. It is defined as a dummy variable: 1 = immigrant; 0 = not immigrant.
	Disabled (p-file) LIS variable: <i>disabled</i> EPF proxy: -	Individuals who have a permanent disability condition, defined as a (physical or mental) health condition that permanently limits an individual in his/her basic activity functioning (such as walking or hearing), even if the limitation is ameliorated by the use of assistive devices or a supportive environment. It is defined as a dummy variable: 1 = disabled; 0 = not disabled.
	Health status (p-file) LIS variable: <i>health_c</i> EPF proxy: -	Subjective evaluation of one own's self-perceived health status, including any dimension as considered appropriate by the individual (physical, emotional, mental, etc.). It is reported in a scale of ratings.
	Education (p-file) LIS variable: <i>educ</i> EPF proxy: <i>estudredsp</i>	Recoding of highest level of education completed into three categories: low: less than upper secondary education completed (never attended, no completed education or education completed at the ISCED 2011 levels 0, 1 or 2); medium: upper secondary education completed or post-secondary non-tertiary education (completed ISCED 2011 levels 3 or 4); high: tertiary education completed (completed ISCED 2011 levels 5 to 8).
<i>Labour market information</i>	Employed (p-file) LIS variable: <i>emp</i> EPF proxy: -	Indicator that employment is the main current activity status as self-assessed by the respondent. It is defined as a dummy variable: 1 = employed; 0 = not employed.
	Part-time employment (p-file) LIS variable: <i>emp</i> EPF proxy: -	Time schedule in the first job, as self-reported by the individual or defined by the data provider. It is defined as a dummy variable: 1 = part-time; 0 = full-time.

Notes: (1) The outcome variable used has been obtained dividing the disposable income by the square root of the household size. Negative and zero income values have been replaced with 1/100 of the mean. In such wise, we can prevent relevant observations from being dropped by default. (2) In regard to the Spanish EPF, the following clarifications must be made. The proxy *impexac* is defined as the exact amount of total net monthly household income. The proxy *estudredsp* has been recoded as follows: 1. *Inferior a la primera etapa de Educación Secundaria* has been reordered to: 1 = *Low*; 2. *Primera etapa de Educación Secundaria* and 3. *Segunda etapa de Educación Secundaria* transformes into: 2 = *Medium*; and, finally, 4. *Educación superior* is redesigned as: 3 = *High*. The proxy *ecivil / ecivillegal1* is the same variable with a different name considering the EPF surveys of 2006 and 2018, respectively. The proxy variables *regten* and *tiphogar1* have been recoded in the same terms as their corresponding LIS variables: *own* and *hhtype*, respectively. There is no information about “health variables” in the EPF. There is no comparable data for the labour market variables in the two waves under study in Spain (2006 and 2018).

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; 1999-2018). Luxembourg: LIS and EPF.

Table A.3. Descriptive statistics for CANADA

Table A.3a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Geography and housing</i>						
Tenure	0.651	0.731	-0.080***	0.651	0.641	0.010
<i>Household composition and living arrangements</i>						
Household composition	0.285	0.287	-0.002	0.285	0.271	0.014*
Household members	2.460	2.282	0.178***	2.460	2.504	-0.044**
<i>Socio-demographic characteristics</i>						
Age	47.979	50.283	-2.304***	47.979	46.644	1.335***
Sex	0.431	0.408	0.023***	0.431	0.394	0.037***
Marital status	0.475	0.475	0.000	0.475	0.486	-0.011
Immigrant	-	-	-	-	-	-
Disabled	-	-	-	-	-	-
Health status	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.108	0.165	-0.057***	0.108	0.194	-0.086***
Medium	0.223	0.234	-0.011*	0.223	0.267	-0.044***
High	0.659	0.590	0.069***	0.659	0.537	0.122***
<i>Labour market information</i>						
Employed	-	-	-	-	-	-
Part-time employment	0.086	0.093	-0.007	0.086	0.057	0.029***
% of sample of the year	26.74	73.26		26.74	26.41	
Sample size (N)	7,133	19,542		7,133	7,650	

Table A.3b. Distribution of the equivalent disposable household income

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Mean</i>	38,461.95	35,520.90	2,941.05***	38,461.95	31,496.75	6,965.20***
<i>Income percentiles</i>						
D10	18,898.64	19,071.89	-173.25	18,898.64	15,329.65	3,568.99***
D20	24,120.94	24,265.89	-144.95	24,120.94	20,291.97	3,828.97***
D30	28,747.20	28,354.11	393.09	28,747.20	24,105.80	4,641.39***
D40	33,139.20	32,517.41	621.79	33,139.20	27,721.31	5,417.89***
D50	38,144.13	36,756.59	1,387.54***	38,144.13	31,481.07	6,663.06***
D60	42,989.31	41,676.02	1,313.29***	42,989.31	35,487.75	7,501.56***
D70	48,993.43	47,414.58	1,578.85***	48,993.43	40,351.19	8,642.24***
D80	57,481.91	55,023.84	2,458.06***	57,481.91	46,503.22	10,978.69***
D90	71,937.42	67,596.90	4,340.52***	71,937.42	56,799.25	15,138.17***
<i>Inequality measures</i>						
P90-P10	53,038.79	48,525.05	4,513.74***	53,038.79	41,469.61	11,569.18***
P90-P50	33,793.29	30,840.31	2,952.98***	33,793.29	25,318.18	8,475.11***
P50-P10	19,245.50	17,684.73	1,560.76***	19,245.50	16,151.50	3,094.07***
Gini index	0.297	0.280	0.017***	0.297	0.292	0.005

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Table A.4. Descriptive statistics for GERMANY

Table A.4a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i>			<i>Dynamic approach</i>		
	Year 2016			Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Geography and housing</i>						
Tenure	0.255	0.487	-0.232***	0.255	0.228	0.027
<i>Household composition and living arrangements</i>						
Household composition	0.310	0.237	0.073***	0.310	0.386	-0.076***
Household members	2.375	2.542	-0.167***	2.375	2.067	0.308***
<i>Socio-demographic characteristics</i>						
Age	49.316	51.626	-2.310***	49.316	49.233	0.083
Sex	0.502	0.491	0.109	0.502	0.435	0.067***
Marital status	0.453	0.546	-0.093***	0.453	0.422	0.031**
Immigrant	0.263	0.186	0.077***	0.263	0.146	0.117***
Disabled	0.119	0.126	-0.007	0.119	0.143	-0.024**
<i>Health status</i>						
Very Good	0.106	0.085	0.021***	0.106	0.109	-0.003
Good	0.397	0.377	0.020*	0.397	0.390	0.007
Satisfactory	0.305	0.340	-0.035***	0.305	0.313	-0.008
Poor	0.145	0.152	-0.007	0.145	0.136	0.009
Bad	0.038	0.037	0.001	0.038	0.044	-0.006
<i>Education</i>						
Low	0.093	0.114	-0.021***	0.093	0.115	-0.022**
Medium	0.449	0.561	-0.112***	0.449	0.516	-0.067***
High	0.434	0.306	0.128***	0.434	0.331	0.103***
<i>Labour market information</i>						
Employed	0.652	0.651	0.001	0.652	0.596	0.056***
Part-time employment	0.245	0.269	-0.024**	0.245	0.174	0.071***
Percentage of sample size (%)	17.30	82.70		17.30	13.71	
Sample size (N)	2,736	13,080		2,736	1,617	

Table A.4b. Distribution of the equivalent disposable household income

	<i>Static approach</i>			<i>Dynamic approach</i>		
	Year 2016			Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Mean</i>	34,477.69	31,829.29	2,648.40***	34,477.69	31,208.15	3,269.44***
<i>Income percentiles</i>						
D10	16,422.89	16,492.46	-519.57	16,422.89	16,161.52	261.37
D20	20,761.67	21,041.43	-279.76	20,761.67	21,832.96	-1,071.29
D30	24,595.33	24,411.09	184.24	24,595.33	25,094.05	-498.72
D40	28,535.24	27,736.67	798.56	28,535.24	28,250.69	284.55
D50	33,137.63	31,004.77	2,132.86***	33,137.63	31,512.45	1,625.19**
D60	37,774.47	35,025.21	2,749.25***	37,774.47	35,041.66	2,732.81***
D70	43,315.37	39,551.55	3,763.83***	43,315.37	39,671.25	3,644.12***
D80	50,186.77	45,610.59	4,576.18***	50,186.77	45,644.27	4,542.49***
D90	65,644.53	56,251.47	9,393.06***	65,644.53	54,933.04	10,711.49***
<i>Inequality measures</i>						
P90-P10	49,221.64	39,309.42	9,912.22***	49,221.64	38,770.78	10,450.86***
P90-P50	32,506.90	25,247.03	7,259.87***	32,506.90	23,420.82	9,086.08***
P50-P10	16,714.74	14,062.39	2,652.35***	16,714.74	15,349.96	1,364.78
Gini index	0.322	0.282	0.040***	0.322	0.265	0.057***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Table A.5. Descriptive statistics for ITALY

Table A.5a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Geography and housing</i>						
Tenure	0.702	0.720	-0.018	0.702	0.590	0.112***
<i>Household composition and living arrangements</i>						
Household composition	0.404	0.337	0.067***	0.404	0.214	0.190***
Household members	2.138	2.226	-0.088*	2.138	2.647	-0.509***
<i>Socio-demographic characteristics</i>						
Age	61.976	62.180	-0.204	61.976	55.143	6.833***
Sex	0.441	0.447	-0.036*	0.441	0.329	0.112***
Marital status	0.486	0.534	-0.048**	0.486	0.638	-0.152***
Immigrant	0.069	0.071	-0.002	0.069	0.025	0.044***
Disabled	-	-	-	-	-	-
Health status	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.500	0.556	-0.056***	0.500	0.552	-0.052*
Medium	0.296	0.336	-0.040**	0.296	0.297	-0.001
High	0.203	0.107	0.096***	0.203	0.151	0.052**
<i>Labour market information</i>						
Employed	-	-	-	-	-	-
Part-time employment	0.071	0.128	-0.057***	0.071	0.086	-0.015
Percentage of sample size	8.42	91.58		8.42	8.64	
Sample size (N)	625	6,795		625	691	

Table A.5b. Distribution of the equivalent disposable household income

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	2000	Difference
<i>Mean</i>	26,146.63	22,078.65	4,067.73***	26,146.63	25,153.25	993.38
<i>Income percentiles</i>						
D10	11,953.69	11,123.39	830.30	11,953.69	11,910.95	42.74
D20	15,998.34	14,399.22	1,599.11	15,998.34	14,826.22	1,172.11
D30	19,771.48	17,254.34	2,517.14	19,771.48	18,343.68	1,427.79
D40	22,812.93	20,304.76	2,508.18**	22,812.93	22,238.14	574.79
D50	25,631.17	23,115.70	2,515.47***	25,631.17	25,036.43	594.75
D60	28,031.66	25,708.84	2,322.82	28,031.66	28,941.64	-909.98
D70	34,050.51	28,867.72	5,182.79***	34,050.51	33,146.51	904.00
D80	42,802.20	33,223.67	9,578.53***	42,802.20	39,017.34	3,784.86
D90	57,219.02	41,431.76	15,787.26***	57,219.02	52,427.64	4,791.38
<i>Inequality measures</i>						
P90-P10	45,265.33	30,310.16	14,955.17***	45,265.33	40,516.49	4,748.84
P90-P50	31,587.85	18,317.70	13,270.15***	31,587.85	27,391.27	4,196.58
P50-P10	13,677.48	11,992.46	1,685.02**	13,677.48	13,125.22	552.26
Gini index	0.362	0.293	0.069***	0.362	0.318	0.044*

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database

Table A.6. Descriptive statistics for POLAND

Table A.6a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	1999	Difference
<i>Geography and housing</i>						
Tenure	0.702	0.837	-0.135***	0.702	0.490	0.212***
<i>Household composition and living arrangements</i>						
Household composition	0.318	0.189	0.129***	0.318	0.229	0.089***
Household members	2.231	2.758	-0.527***	2.231	2.584	-0.353***
<i>Socio-demographic characteristics</i>						
Age	50.619	53.254	-2.634***	50.619	50.234	0.385
Sex	0.469	0.375	0.094***	0.469	0.437	0.032***
Marital status	0.506	0.644	-0.138***	0.506	0.628	-0.122***
Immigrant	0.014	0.006	0.008***	0.014	-	-
Disabled	0.078	0.099	0.021***	0.078	0.127	-0.049***
Health status						
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.055	0.148	-0.093***	0.055	0.134	-0.079***
Medium	0.482	0.660	-0.178***	0.482	0.629	-0.147***
High	0.463	0.191	0.272***	0.463	0.237	0.226***
<i>Labour market information</i>						
Employed	-	-	-	-	-	-
Part-time employment	0.065	0.047	0.018***	0.065	0.074	-0.009***
Percentage of sample size	12.91	87.09		12.91	13.96	
Sample size (N)	4,762	32,124		4,762	4,388	

Table A.6b. Distribution of the equivalent disposable household income

	<i>Static approach</i> Year 2016			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2016	1999	Difference
<i>Mean</i>	21,424.85	15,903.10	5,521.75***	21,424.85	13,674.88	7,749.97***
<i>Income percentiles</i>						
D10	11,502.36	8,372.74	3,129.62***	11,502.36	7,305.44	4,196.92***
D20	14,433.93	10,625.04	3,808.89***	14,433.93	8,926.55	5,507.38***
D30	16,953.51	12,437.87	4,515.64***	16,953.51	10,370.43	6,583.08***
D40	19,379.97	14,092.52	5,287.45***	19,379.97	11,650.38	7,729.59***
D50	21,625.97	15,789.41	5,836.56***	21,625.97	13,121.15	8,504.82***
D60	24,449.82	17,687.58	6,762.24***	24,449.82	14,852.06	9,597.76***
D70	27,702.15	20,039.24	7,662.91***	27,702.15	16,670.74	10,941.41***
D80	31,879.78	23,140.58	8,739.20***	31,879.78	19,703.14	12,176.64***
D90	40,162.34	28,943.91	11,218.44***	40,162.34	25,172.12	14,990.22***
<i>Inequality measures</i>						
P90-P10	28,659.98	20,571.21	8,088.77***	28,659.98	17,687.00	10,792.98***
P90-P50	18,536.37	13,154.53	5,381.84***	18,536.37	12,051.07	6,485.30***
P50-P10	10,123.62	7,416.68	2,706.93***	10,123.62	5,815.93	4,307.69***
Gini index	0.293	0.293	0.000	0.293	0.308	-0.015

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD. (3) Empty cells are due to a lack of data.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Table A.7. Descriptive statistics for SPAIN

Table A.7a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i> Year 2018			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2006	Difference
<i>Geography and housing</i>						
Tenure	0.714	0.813	-0.098***	0.714	0.815	-0.101***
<i>Household composition and living arrangements</i>						
Household composition	0.106	0.091	0.015***	0.106	0.073	0.033***
Household members	3.110	3.192	-0.082***	3.110	3.325	-0.215***
<i>Socio-demographic characteristics</i>						
Age	53.646	54.066	-0.420***	53.646	51.394	2.252***
Sex	0.629	0.602	0.027***	0.629	0.587	0.042***
Marital status	0.435	0.470	-0.035***	0.435	0.475	-0.040***
Immigrant	0.139	0.087	0.052***	0.139	0.083	0.056***
Disabled	-	-	-	-	-	-
Health status	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Education						
Low	0.121	0.175	-0.054***	0.121	0.231	-0.109***
Medium	0.421	0.495	-0.074***	0.421	0.403	0.018**
High	0.456	0.328	0.127***	0.456	0.365	0.091***
<i>Labour market information</i>						
Employed	-	-	-	-	-	-
Part-time employment	-	-	-	-	-	-
Percentage of sample size	10.72	89.28		10.72	10.49	
Sample size (N)	5,974	49,751		5,974	5,847	

Table A.7b. Distribution of the equivalent net household income

	<i>Static approach</i> Year 2018			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2006	Difference
<i>Mean</i>	1,515.12	1,291.80	223.32***	1,515.12	1,455.89	59.23***
<i>Income percentiles</i>						
D10	572.60	539.17	33.43***	572.60	618.85	-46.25***
D20	810.59	708.81	101.78***	810.59	789.71	20.88*
D30	973.62	842.73	130.88***	973.62	941.59	32.03*
D40	1,161.73	998.38	163.34***	1,161.73	1,088.22	73.51***
D50	1,300.43	1,169.75	130.67***	1,300.43	1,252.86	47.57***
D60	1,520.55	1,303.59	216.95***	1,520.55	1,441.93	78.62***
D70	1,776.50	1,512.88	263.61***	1,776.50	1,656.45	120.05***
D80	2,046.80	1,775.24	271.56***	2,046.80	2,008.06	38.74
D90	2,601.58	2,164.98	436.60***	2,601.58	2,544.95	56.63
<i>Inequality measures</i>						
P90-P10	2,028.98	1,625.78	403.20***	2,028.98	1,926.10	102.88***
P90-P50	1,301.15	995.21	305.94***	1,301.15	1,292.09	9.06
P50-P10	727.82	630.56	97.26***	727.82	634.01	93.81***
Gini index	0.315	0.300	0.015***	0.315	0.304	0.011**

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in constant euros (reference year 2016 = 100). (4) Empty cells are due to a lack of data.

Source: Own elaboration from EPF.

Table A.8. Descriptive statistics for UNITED STATES

Table A.8a. Socioeconomic characteristics of households. Sample means

	<i>Static approach</i> Year 2018			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2000	Difference
<i>Geography and housing</i>						
Tenure	0.622	0.700	-0.078***	0.622	0.651	-0.029***
<i>Household composition and living arrangements</i>						
Household composition	0.254	0.255	-0.001	0.254	0.226	0.028***
Household members	2.660	2.581	0.078***	2.660	2.809	-0.149***
<i>Socio-demographic characteristics</i>						
Age	50.492	51.832	-1.339***	50.492	46.864	3.628***
Sex	0.500	0.498	0.002	0.500	0.470	0.030***
Marital status	0.520	0.527	-0.007*	0.520	0.551	-0.031***
Immigrant	0.239	0.083	0.156***	0.239	0.191	0.048***
Disabled	0.097	0.128	-0.031***	0.097	0.102	-0.005***
Health status						
Excellent	0.251	0.212	0.039***	0.251	0.294	-0.043***
Very good	0.328	0.312	0.016***	0.328	0.320	0.008**
Good	0.280	0.291	-0.011***	0.280	0.243	0.037***
Fair	0.105	0.132	-0.027***	0.105	0.099	0.006***
Poor	0.034	0.051	-0.017***	0.034	0.042	-0.008***
Education						
Low	0.099	0.101	-0.002	0.099	0.155	-0.056***
Medium	0.403	0.500	-0.097***	0.403	0.463	-0.060***
High	0.497	0.397	0.100***	0.497	0.381	0.116***
<i>Labour market information</i>						
Employed	0.657	0.600	0.057***	0.657	0.706	-0.049***
Part-time employment	0.133	0.159	-0.026***	0.133	0.116	0.017***
Percentage of sample size	59.75	40.25		59.75	57.56	
Sample size (N)	40,835	27,510		40,835	44,927	

Table A.8b. Distribution of the equivalent disposable household income

	<i>Static approach</i> Year 2018			<i>Dynamic approach</i> Territories over 500,000 inhabitants		
	> 500,000 inhabitants	< 500,000 inhabitants	Difference	2018	2000	Difference
<i>Mean</i>	49,655.28	40,592.91	9,062.38***	49,655.28	43,312.54	6,342.74***
<i>Income percentiles</i>						
D10	19,599.46	17,721.03	1,878.43***	19,599.46	17,870.59	1,728.87***
D20	26,144.83	23,562.47	2,582.35***	26,144.83	24,277.80	1,867.03***
D30	32,882.96	28,710.61	4,172.35***	32,882.96	30,191.69	2,691.27***
D40	39,412.65	33,973.13	5,439.52***	39,412.65	35,609.69	3,802.96***
D50	45,948.12	39,522.99	6,425.13***	45,948.12	41,104.30	4,843.82***
D60	53,750.22	45,349.90	8,400.32***	53,750.22	46,817.26	6,932.96***
D70	62,955.50	52,306.12	10,649.38***	62,955.50	53,485.38	9,470.12***
D80	76,876.32	61,735.61	15,140.72***	76,876.32	62,924.31	13,952.02***
D90	101,717.90	79,770.98	21,946.88***	101,717.90	79,678.09	22,039.77***
<i>Inequality measures</i>						
P90-P10	82,118.39	62,049.97	20,068.43***	82,118.39	61,807.52	20,310.87***
P90-P50	55,769.74	40,248.00	15,521.74***	55,769.74	38,573.80	17,195.94***
P50-P10	26,348.65	21,801.96	4,546.69***	26,348.65	23,233.72	3,114.93***
Gini index	0.370	0.341	0.029***	0.370	0.344	0.026***

Notes: (1) Statistical significance levels: ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10. (2) Calculations made using sample weights. (3) Income values expressed in PPP 2017 USD.

Source: Own elaboration from Luxembourg Income Study (LIS) Database.

Conclusiones

Esta tesis doctoral, tal y como se indicó en la introducción, se marcaba tres objetivos fundamentales. En primer lugar, queríamos saber cuál era la relevancia del territorio a la hora de explicar las tendencias recientes de la desigualdad de ingresos en los principales países de la OCDE. En otras palabras, queríamos saber si esta dimensión seguía siendo importante y cómo había evolucionado su contribución desde principios del siglo XXI hasta la actualidad. Para dar respuesta a esta cuestión, en el Capítulo 1 nos apoyamos en el potencial de las descomposiciones de la desigualdad, herramientas útiles para el correcto diseño de políticas redistributivas, y desarrollamos varios ejercicios complementarios a partir de los datos de LIS. La idea era mostrar no solo una perspectiva particular sobre el mismo problema (la relevancia del territorio en la descripción de la desigualdad de ingresos), sino ofrecer una perspectiva global y conjunta agregando todos ellos.

La pregunta a la que intentábamos responder en este capítulo era directa y sencilla: ¿En qué medida el territorio recoge los principales cambios observados en la evolución reciente de la desigualdad de ingresos? A pesar de no haber encontrado un patrón firme entre los países seleccionados, la especificación empírica aplicada para la descomposición DFL llevada a cabo parece captar gran parte del cambio observado en la desigualdad de ingresos durante los años analizados. La variable región, en particular, ejerce un efecto notable. En particular, la evidencia encontrada revela un “efecto reductor” generalizado de la desigualdad de ingresos que está directamente vinculado al territorio. Todo esto cobra gran importancia porque los altos niveles de desigualdad espacial son indeseables para el desarrollo y crecimiento económico de cualquier sociedad. Pueden limitar el progreso, a menudo se asocian con problemas de delincuencia, revelan debilidades institucionales e incluso obstaculizan la cohesión social entre sus regiones (Atkinson, 2015).

A través de los resultados presentados en el Capítulo 1, tal y como acabamos de señalar, arrojamamos algo de luz sobre la trascendencia de la dimensión territorial en la explicación de la desigualdad. Y lo logramos implementando una aplicación metodológica no desarrollada hasta ahora: un análisis contrafactual para pesos de población en el contexto del análisis comparativo internacional. En términos generales, nuestros resultados confirman y actualizan los de trabajos anteriores, así como algunas de las premisas teóricas de la literatura relacionada. Las diferencias de ingresos entre regiones tienen una pequeña capacidad explicativa de la desigualdad en los países seleccionados, mientras que los cambios en la distribución intraterritorial de la población predominan de manera destacada sobre los anteriores.

Aun con los desafíos de aportar evidencia concluyente – la influencia de factores sociales e institucionales es muy diversa en cada país – hemos obtenido algunos resultados interesantes que subrayan la relevancia de este determinante y que podrían incentivar el desarrollo de futuras investigaciones sobre la desigualdad espacial. Estos resultados son importantes tanto para informar la teoría como para contribuir al diseño de políticas públicas relacionadas con la cohesión social. En cuanto a la primera observación, hemos confirmado que los cambios en la distribución territorial de la población afectan la desigualdad. Como señalan los modelos NEG, estos flujos están relacionados con efectos de aglomeración espacial que, a su vez, están en la raíz de las disparidades económicas regionales que pueden afectar la desigualdad en la distribución del ingreso.

Una implicación política de nuestros resultados es que si la movilidad poblacional se concibe como una posible herramienta para incidir en la desigualdad inter- e intra-territorial, es necesario coordinar este tipo de políticas específicas con otras medidas redistributivas para optimizar su impacto conjunto. En cualquier caso, debemos ser cautelosos al extraer otras implicaciones de política, dada la dificultad de pasar de un

ejercicio empírico al diseño de políticas específicas. En este sentido, algunas de las limitaciones de nuestro trabajo pueden interpretarse como nuevas extensiones prometedoras. En primer lugar, se ha constatado un elevado peso de “otras características”, que sugiere que algunas variables relevantes podrían haberse omitido. Dado que nos centramos exclusivamente en la información incluida en los datos de LIS, aquí no se consideran cuestiones relevantes como la tecnología, el comercio o la descentralización. En segundo lugar, una muestra más amplia y variada de países podría enriquecer el análisis. En tercer lugar, la mayoría de nuestros análisis son ejercicios contables acerca de cuánto pueden influir los cambios en la distribución territorial de la población en las tendencias recientes de desigualdad.

En cuanto al segundo objetivo de esta tesis doctoral, hemos prestado atención a las consecuencias derivadas de los shocks de desempleo (en la UE) sobre un índice sintético de privación material. Así, en el Capítulo 2, hemos intentado establecer relaciones de causalidad entre los shocks de desempleo y los cambios en los niveles de privación material. Centrándonos en la experiencia reciente de la UE, utilizamos una combinación de países europeos para construir una unidad de control sintética que reproduzca, de la manera más aproximada posible, la evolución real de la variable de resultado en los países afectados por el evento y durante el periodo previo al shock. Un elemento novedoso del capítulo es nuestra propuesta de definición de shock de desempleo. La falta de consenso en la literatura nos ha llevado a proponer una definición específica que podría ser utilizada en otros estudios. La utilización del doble criterio de la tasa de crecimiento de la tasa de paro y su nivel en un momento dado nos ha permitido identificar dos países en los que se produjeron dichos shocks: España y Grecia. Sin embargo, éste es un criterio relativo en el que la identificación de los países afectados por un evento depende de la gravedad del problema involucrado. Afortunadamente, a través

de nuestros análisis de sensibilidad hemos podido utilizar criterios más estrictos en la definición de estos shocks, lo que ha servido para delimitar más claramente los países afectados por estos shocks de los menos afectados por el aumento del desempleo.

Nuestros resultados muestran que en los países donde los criterios propuestos confirman la existencia de un shock de desempleo se produjo un aumento significativo en el nivel de privación material. Aun teniendo en cuenta los límites naturales para establecer relaciones causales, estos resultados refutan el supuesto tradicional de la baja sensibilidad de las medidas de privación material ante cambios en el ciclo económico. Esta conclusión es válida cuando se utilizan otros métodos para identificar el efecto observado. Para cubrir un período de pretratamiento más amplio, hemos extendido la serie combinándola con datos del PHOGUE. Incluso a costa de reducir el número de países analizados, el efecto del shock de desempleo sobre el nivel de privación material se mantiene. Lo mismo ocurre cuando se consideran otras medidas de privación material y también cuando se eliminan del análisis países relativamente similares a España y Grecia.

Nuestros resultados, en resumen, nos permiten anticipar cómo cambios drásticos en la tasa de desempleo pueden conducir a pérdidas rápidas de bienestar entre los hogares, que no se limitan a una mayor pobreza monetaria e ingresos insuficientes, sino que se extienden al bienestar material y las condiciones de vida. Los resultados obtenidos, que toman como marco temporal de referencia la denominada Gran Recesión en una zona de altos ingresos como la Unión Europea, podrían ser aún más severos ante incrementos aún mayores y más rápidos del desempleo como consecuencia de la crisis del COVID-19.

Con respecto al tercer y último objetivo, pretendíamos averiguar y cuantificar, en un conjunto de países de la OCDE, la contribución individual de diferentes determinantes a las diferencias de desigualdad, tanto dentro de las *grandes ciudades* como entre las *grandes ciudades* y otras áreas. En un intento por superar parcialmente los problemas

tradicionalmente asociados al binomio desigualdad y tamaño de la ciudad, el Capítulo 3 intentaba contribuir a la literatura con una mejor comprensión de esta relación. Para ello, utilizamos un enfoque metodológico que permite atender estas diferencias tanto desde una perspectiva dinámica como estática.

Uno de los principales resultados del capítulo es un notable aumento de la desigualdad de ingresos en las *grandes ciudades* de los países seleccionados. Esto sucede con ambos enfoques y para diferentes medidas de desigualdad. Estas diferencias están determinadas, principalmente, por un “efecto estructura” más que por cambios en la distribución en la composición de la población. La evidencia obtenida arroja algo de luz sobre la relevancia de la educación como uno de los determinantes más importantes que explican la desigualdad de ingresos.

Por lo que se refiere al enfoque dinámico, hemos examinado las tendencias de la desigualdad de ingresos en las denominadas *grandes ciudades* (más de 500.000 habitantes) durante un período cercano a las dos décadas. Los cambios en los retornos de la educación (*skill premium*) explican una gran parte del aumento de la desigualdad en Canadá – fundamentalmente, en la mitad superior de la distribución del ingreso – y los Estados Unidos, pero en Alemania la tendencia fue al revés. Los cambios en los rendimientos del resto de variables analizadas mostraron una imagen diferente. Con respecto a las variables de control relacionadas con geografía y vivienda, así como las relacionadas con el estado del mercado laboral, los retornos parecen describir un comportamiento más idiosincrático en cada país. En cuanto a las dotaciones, los cambios en el nivel educativo parecen jugar un papel central también en la configuración de la desigualdad de ingresos en las *grandes ciudades* a lo largo del tiempo. Hemos encontrado también que la mayoría de países comparten este resultado en toda la distribución. Además, en todos los países, la magnitud del impacto fue la mayor de todas las variables

de control. Por otro lado, en varios países también se detectó un efecto “desigualador” del tamaño del hogar.

En lo que respecta a la aproximación estática, hemos utilizado los datos más recientes disponibles para comparar las grandes ciudades con otras áreas. En línea con estudios previos, hemos encontrado que los retornos a la educación fueron mayores en las aglomeraciones más grandes de Estados Unidos. Sin embargo, se observó justo lo contrario en otros países, con ciudades más grandes con retornos más bajos para los trabajadores altamente cualificados. En el caso de los “efectos composición”, llegamos a las mismas conclusiones ya avanzadas para el enfoque dinámico, pero aquí los efectos son aún más fuertes. Los cambios en el nivel educativo desempeñan, nuevamente, un papel predominante en la justificación del aumento de la desigualdad de ingresos. Por otro lado, la evidencia encontrada revela el efecto negativo de la composición por edades en cinco de los seis países examinados.

En un escenario socioeconómico como el que vivimos hoy, lleno de incertidumbre, es fundamental arrojar algo de luz sobre la contribución particular de los principales determinantes de los cambios en la distribución del ingreso en las distintas áreas. Así, resulta especialmente relevante aclarar las razones que explican el reciente aumento de la desigualdad de ingresos en áreas donde reside un porcentaje cada vez mayor de la población. La COVID-19 ha surgido repentinamente como una peligrosa amenaza para las *grandes ciudades*. Para evitar que sus efectos se vuelvan más pronunciados en los grupos más vulnerables, los responsables del diseño y formulación de las políticas públicas deberían tomar como referencia la información más completa sobre la influencia de estos determinantes. Una respuesta adecuada en el momento actual no solo es crucial para mitigar los efectos devastadores de esta crisis, sino que también resulta crucial para anticipar las respuestas a futuras shocks.

El trabajo realizado abre la puerta a varias líneas futuras de investigación. Por un lado, con respecto al Capítulo 1, un análisis más detallado de los cambios de población con un mayor nivel de desagregación podría constituir una prometedora extensión del trabajo aquí descrito. Además, la mera inclusión de nuevos países y olas en la base de datos del LIS, que se efectúa de manera regular, amplía las posibilidades de profundizar en los resultados en los Capítulos 1 y 3. En cuanto al Capítulo 2, se abren oportunidades para replicar un ejercicio similar con metodologías alternativas de evaluación de impacto, como la técnica de diferencias en diferencias. Con el paso del tiempo, también podremos disponer de más datos y realizar pruebas de sensibilidad alternativas con bases de datos más extensas y una mayor cobertura temporal de las series. Sería el caso, por ejemplo, de la posible consideración de “*in time*” *placebos*. Por último, no podemos descartar la posibilidad de realizar algunas investigaciones relacionadas con la reciente crisis sanitaria y económica provocada por la COVID-19. Aplicando algunas de las técnicas de descomposición aquí expuestas a un grupo más amplio de países, podríamos evaluar los cambios que se están produciendo en los últimos meses y que muy probablemente afectarán y trastocarán el escenario socioeconómico principal aquí estudiado.

Conclusions

As stated in the introduction, this dissertation had three main aims. First, we wanted to know what the relevance of the territory was in explaining recent trends in income inequality in OECD countries. In other words, we wanted to know if this dimension was still important and how its contribution had evolved from the beginning of the 21st century to the present day. In response to that question, in Chapter 1 we implemented several complementary techniques using the LIS database and relying on the potential of inequality decompositions, useful tools for the correct design of redistributive policies. The idea was not only to provide a particular perspective on the same problem (the relevance of territory in the analysis) but offering a global and joint perspective by aggregating all of them.

The question we tried to answer in this chapter was concise: To what extent does territory drive the main changes observed in the recent evolution of income inequality? Despite not having found a strong pattern among the countries selected, the empirical specification of the DFL decomposition does seem to capture much of the change observed in income inequality during the years analysed. The 'region' variable, in particular, exerts a noteworthy effect. The evidence found reveals a generalized and interesting "reducing effect" in income inequality that is directly linked to territory. All this takes on great importance because high levels of spatial inequality are undesirable for the development and economic growth of any society. They can constrain progress, are often associated with crime problems, reveal institutional weaknesses, and even hinder social cohesion among their regions (Atkinson, 2015).

In short, we have shed some light on the transcendence of the territorial dimension in inequality by implementing an application not developed thus far: a counterfactual analysis for population weights in the context of international comparative analysis. In general terms, our results confirm and update those of previous works, as well as some of

the theoretical premises of the related literature. Income differences between regions have a small explanatory capacity of inequality in the selected countries while the changes in the territorial distribution of population affect inequality. Even with the challenges of providing conclusive evidence – the influence of social and institutional factors is very diverse in each country – we have obtained some interesting results that underscore the relevance of this determining factor and that could encourage the pursuit of future research on spatial inequality. These results are important both to inform the theory and to contribute to the design of public policies related to social cohesion. Regarding the first issue, we have confirmed that changes in the territorial distribution of population affect inequality. As stated by NEG models, these flows are related to spatial agglomeration effects, which, in turn, are at the root of regional economic disparities that can affect inequality in the income distribution.

A political implication of our results is that if population mobility is conceived as a possible tool to influence inter- and intra-inequality, it is necessary to coordinate this type of specific policies with other redistributive measures in order to optimize their joint impact. In any case, we need to remain cautious when drawing other policy implications, given the difficulty of going from an accounting exercise to the design of specific policies. In this regard, some of the limitations of our work may be interpreted as further promising extensions. One of these caveats is the very high weight of “other characteristics” suggesting that some relevant variables might have been omitted. Since we focus exclusively on the information included in LIS data, relevant issues such as technology, trade, or decentralization are not considered here. Second, a larger and more varied sample might enrich the analysis. Third, most of our analyses are accounting exercises of how much the changes in the territorial distribution of population may influence recent inequality trends.

As for the second goal of this dissertation, we focused on the consequences derived from unemployment shocks (in the EU) on a composite measure of material deprivation. In Chapter 2 we have tried to establish causality relationships between changes in material deprivation and unemployment shocks. Focusing on the EU experience, we use a combination of European countries to construct a synthetic control unit for each country that as much as possible resembles the actual evolution of outcome variables before the shock took place.

An important and novel element of our approach relates to our proposed definition of an unemployment shock. The lack of consensus in the literature has led us to propose a specific definition that could be used in other studies. The double criterion of the growth rate of the unemployment rate and its level has allowed us to differentiate two countries in which such shocks took place (Spain and Greece). However, this is a relative criterion in which the demarcation of countries affected by an event depends on the severity of the problem involved. Fortunately, through our sensitivity analyses we have been able to use more stringent criteria to define these shocks, which has served to more clearly delimit the countries affected by them and those that were not.

Our results show that in the countries for which the proposed criteria confirm the existence of an unemployment shock a significant increase in material deprivation occurred. Based on the natural limits for establishing causal relationships, these results refute the traditional assumption of the low sensitivity of material deprivation measures to changes in the economic cycle.

This conclusion holds when other methods are used to identify the observed effect. To cover a broader pretreatment period, we extended the series combining it with ECHP data. Even at the cost of reducing the number of countries analyzed, the effect of the unemployment shock on material deprivation remains. The same occurs when other

material deprivation measures are considered, and above all when countries relatively similar to Spain and Greece based on any of the criteria used to define the unemployment shock are removed from the analysis.

Our results, in short, allow us to anticipate how drastic changes in the unemployment rate can lead to rapid wellbeing losses among households, which are not limited to increased monetary poverty and insufficient income but also to material well-being and living conditions. Such results, derived from what happened in the so-called Great Recession in a high-income area such as the European Union, could be more severe in the face of even greater and rapid increases in unemployment such as those registered in these same countries due to the COVID-19 crisis.

Regarding the third and final goal, we intended to find out and estimate in a selected sample of OECD countries the individual contribution of different drivers to inequality differences both within *big cities* and between *big cities* and other areas. In an attempt to partially overcome the problems linked to the connection between inequality and city size, Chapter 3 tried to contribute to the literature with a better understanding of this relationship. For this purpose, we used a methodological approach that allows to look at these differences both from a dynamic and static perspective.

One of the main findings of the chapter is a remarkable increase of income inequality in big cities. This happens with both approaches and with different inequality measures. These differences are mainly determined by a *structure effect* rather than changes in the distribution of endowments. The evidence obtained shed some light about the relevance of education as one of the most important drivers explaining income inequality.

We examined income inequality trends in big cities for a period close to two decades. Changes in the returns to education (skill premium) accounts for a great amount of the increase in inequality in Canada and the United States, but in Germany the trend was just

the other way around. Changes in the returns to the rest of variables analyzed showed a different picture. With respect to geography and housing control variables, as well as those related to the labor market status, these returns seemed to describe a more idiosyncratic behavior in each country. As for the endowments, changes in educational attainment seemed to play a central role as well in shaping income inequality in big cities over time. In all countries, the magnitude of the impact was the largest of all the control variables. An unequalizing effect of household size was also observed in several countries.

We also used the most recent data available to compare big cities with other areas. In keeping with previous studies, we found that the returns to educational skills were higher in larger agglomerations of the United States. However, just the opposite was observed in other countries, with larger cities showing smaller returns for high-skilled workers. In the case of the composition effects, the same conclusion holds for the *dynamic approach* applied, but with even stronger effects. Thereby, changes in educational attainment played again a predominant role in justifying the growth in income inequality. On the other hand, the evidence shown brought to light a negative age-composition effect in five of the six countries examined.

We contribute here to this growing body of empirical evidence with a more accurate measurement of the effect of this and other potential drivers considering a larger number of countries and a double static and dynamic perspective. In a socio-economic scenario like the one we are living today, full of uncertainty, it is vital to shed some light on the particular contribution of the main drivers of the changes in the income distribution in the different areas. It is especially relevant to clarify the reasons explaining the recent rise in income inequality in areas where an increasing percentage of the population resides. COVID-19 has suddenly emerged as a dangerous threat to big cities. To prevent its effects

from becoming more marked on the most vulnerable groups, policymakers should take as reference the most complete information on the influence of these determinants. An adequate response at the present time is not only crucial to slim down the devastating effects of this crisis, but also to anticipate the responses to future shocks.

Finally, the lines of future research that this investigation opens are wide and varied. On the one hand, regarding Chapter 1, a more detailed analysis of population changes at a higher level of disaggregation should be a promising extension of the work here described. Moreover, the inclusion of new countries and waves in the LIS database, on a regular basis spreads the possibilities of deepening the analyses of Chapters 1 and 3. As for Chapter 2, the opportunities are open to replicate a similar exercise with alternative impact evaluation methodologies, such as the differences-in-differences technique. As time goes by, we will also be able to collect more data and apply alternative sensitivity tests requiring larger databases that allow to prolong some series. We are thinking, for instance, on *in-time* placebos. Finally, we cannot rule out the chances to carry out further research related to the recent health and economic crisis caused by COVID-19. By applying some of the decomposition techniques used in the different chapters to a wider group of countries, we could evaluate the changes that are taking place in recent months and that will most likely affect and disrupt the socioeconomic context analyzed here.

