Three essays on poverty and inequality

Claudia Samano-Robles

A thesis submitted for the degree of Doctor of Philosophy in

Economics

Institute for Social and Economic Research

University of Essex

June 2021

Summary

This thesis contains three essays that analyses income inequality and poverty.

Chapter 1 examines the impact of education on income inequality in 18 Latin American countries between 2000 and 2010. This period has raised interest in the academic community because inequality has fallen across the region, after several years of consistent high levels. Employing the novel technique proposed by Firpo et al., (2007) my research provides a detailed decomposition of inequality.

In Chapter 2 I examine and contrasts poverty and inequality measures using income and consumption data from Mexico between 1994 and 2014. I investigate how poverty and inequality measures have changed over time and compare results using income and consumption based definitions. Using data from the Household Income and Expenditure Surveys (ENIGH), I construct four measures of resources two income based and two consumption based. The results suggest that income and consumption measures of poverty can complement each other when evaluating certain policy programs.

In Chapter 3 we analyze the changes in the gender structure at the top of the income distribution in the United Kingdom over the last 19 years using administrative data from tax records. Despite the fact that women have increased their participation in the labour market over the past 20 years, they remained underrepresented at the very top of the income distribution.

Acknowledgments

I am extremely grateful to my supervisors Silvia Avram and Mike Brewer for all the comments and support through all this journey.

I would like to thank the Economic and Social Research Council (ESRC) for the scholarship and the opportunities to improve my academic career.

I am very thankful to my PhD examiners Olga Cantó Sánchez and Xavier Jara Tamayo for their comments that help to improve my research.

I am also grateful to the Institute of Social Research (ISER) and the University of Essex for the opportunity the be part of an incredible multicultural community where high academic standards are well combined with a friendly environment.

I am very thankful to my dearest Professors who were and are an inspiration for my academic career: Jim Davies, J. Mario Herrera Ramos, Jonathan Molinet Malpica, Miguel Székely Pardo, James Foster and Gerardo Esquivel Hernández.

Thanks to all my friends from the University of Essex: Cata, Andres, Abril, Fede, Victor, Caroline, Ipek, Duygu, Tatiana, David, Paulino, Greta, Adam, Enrico, Iva, Veridiana, Rumina, Franco, Yuliya, Liza, Laura, Andrea, Paloma, Daniel, Juliana, Joshua, Cara and Janice. For the coffee and great times in Colchester.

I would like to thank to my beloved parents Claudia and Ramiro for their courage and hard work. Special thanks to my brother Ramiro who has been always very supportive.

My deepest gratitude goes to Pedro José for being a unique and revolutionary human being who support me at all times and was always open to discuss my research. Most of all a huge thank you to my son Carlos Santiago, who arrived at the middle of the PhD and transformed my life.

Declarations

No part of this thesis has been submitted for another degree.

Chapter 1 is based on the version published as Sámano-Robles, C. (2018). "The impact of education on income inequality in Latin America between 2000 and 2010". Inequality, Taxation and Intergenerational Transmission (Research on Economic Inequality, Volume 26). Emerald Publishing Limited, 109-148.

Chapter 3 is co-authored with Professor Mike Brewer and an early version of part of this chapter was previously published as Brewer M. and Samano-Robles, C. (2019). "Top incomes in the UK analysis of the 2015-16 SPI", ISER Working Paper 2019-06. Chapter 1 and Chapter 2 are exclusively mine.

Contents

List of Figures

List of Tables

Introduction

Inequality has been a hot topic over the last years due to the increasing accumulation of income among the very rich (Alvaredo et al. (2018)). Historically, inequality in Latin America has been particularly high but in recent years, has been declining. In the last report from the World Inequality Database for Latin America, De Rosa et al. (2020) find that despite falling trends in some countries, inequality is still very high and most likely underestimated when using only survey data .

The main topics of this thesis are inequality and poverty. The first two Chapters focus on inequality and poverty in Latin America where over the period between 2000 and 2010 inequality has been declining in most of the countries in the region. The first Chapter analyses the effects of educational expansion on the declining levels of inequality in 18 countries in Latin America. In the second Chapter, I focus on the case of Mexico, where there is no agreement on the declining trend on income inequality. The disagreement is due to the problems of miss-measurement when using income surveys, specifically at the top of the income distribution. Researchers have shown that after adjusting incomes at the top, income inequality is higher (see Campos-Vazquez and Lustig

(2017)). However, to the best of my knowledge there is not yet evidence on how inequality and poverty measures may be affected by possible miss-measurement at the bottom of the income distribution. A growing literature in the US and the UK suggests income at the bottom of the distribution is underestimated and that consumption provides a more accurate measure of well-being (see Meyer and Sullivan (2003) and Brewer et al. (2017)). The information that consumption can provide to improve measurements of inequality and poverty has not been explored in Latin America, including Mexico. The second Chapter analyses different measures of inequality and poverty using income and consumption data to understand whether consumption measures of well-being provide different results compare to those using income data for Mexico. Finally, the third chapter of this thesis examines income inequality at the top of the income distribution. Using administrative data, we characterize the gender gap among the top income groups in the United Kingdom. This analysis has been motivated by the increasing evidence of under reporting of income among top income groups using survey data (see Burkhauser et al. (2018a)).

The theoretical links between education and income inequality have been extensively debated(see Salverda et al. (2009)). However, the empirical evidence is less clear, especially in less developed countries. Historically, Latin America has been characterized by high inequality levels. However, between 2000 and 2010 inequality has declined significantly throughout the region. At the same time, governments in the region have been expanding education: education expanded on average by one year over the same period. The first chapter investigates the links between educational expansion

and changes in income inequality in 18 countries from 2000 to 2010, using RIF based decomposition. I find that the expansion of education, keeping everything else constant increased inequality in the region, confirming the presence of the Paradox of Progress. However, once changes in returns to education are taken into account as well, the expansion of education contributes to a reduction in inequality in some countries while in others it increases inequality. The reason behind these heterogeneous effects of education could lie in the shape of the returns to education which is country specific and depends on educational policies implemented across different groups of the population. The policy implication from this Chapter are that the educational policies aiming to expand years of education should take into account the shape of the returns to education if they do not want to increase inequality levels.

The decline in income inequality in the region has brought the attention of researchers in Mexico to how well incomes at the top are measured in survey data and what would happen if they were adjusted based on administrative sources of information. However, to the best of my knowledge there is no comparable work looking at income miss-measurement at the bottom of the income distribution and if and how consumption data could be used to address it. The second chapter of this thesis uses income and consumption information to construct four measures of living standards (two income based and two consumption based) and compares inequality levels and trends using these four measures. We find evidence that there is also under-reporting a the bottom of the income distribution and that consumption data can complement the

information that income usually provides to measure inequality and poverty.

Finally, in the third chapter, motivated by the increase in income concentration among the top income groups in the United Kingdom, we analyze gender differences among the very rich. Very little is know about the gender gap among the top income groups using administrative data. We find that the rising share of women in the top income groups is driven by women with earned income and accompanied by an increasing share of top income women being aged 45-54 and living in London at the South East of England. The industries contributing the most to the rising female share of income have been "Real estate, renting and business" and "Financial intermediation".

Overall, we can conclude that policies that expand education in Latin America need to consider the specific group of the population they are targeting and the additional effects on inequality. Second, under-reporting of incomes in survey data is not an issue that affects only the top but also the bottom of the income distribution in Mexico and this could potentially be explained by the increase in informal jobs. The use of consumption data can help to provide additional information about households that are considered income and consumption poor. Finally, the availability of administrative data to analyze the gender gap among top income groups provides more detailed information on the characteristics of the industries and the women that remain under represented. Earned income is a key source of income for women to reduce the gap with respect to men. However, there is a concentration in specific industries and cities where

women can improve their income. As a result, the reduction in the gender gap is not a generalized phenomena across industries and cities in the UK.

References

Advani, A., Koenig, F., Pessina, L., and Summers, A. (2020a). Importing inequality: Immigration and the top 1 percent. *IZA Discussion Paper*, 13731.

Advani, A., Summers, A., et al. (2020b). Capital gains and uk inequality. *University of Warwick, Department of Economics*, No. 1260.

Alvaredo, F. (2017). Uk estimates of top income shares 2013-2014 and 2014-2015. *World Inequality Lab*, No. 201702.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). *World inequality report 2018*. Belknap Press.

Alvaredo, F. and Gasparini, L. (2013). Recent trends in inequality and poverty in developing countries. *Documentos de Trabajo del CEDLAS*.

Andersen, T. M. (2018). Human capital, inequality, and growth. *The Welfare State Revisited*, pages 109–122.

Atkinson, A. B. (2007). The distribution of top incomes in the united kingdom 1908–2000.

Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries, 1:82–140.

Atkinson, A. B., Casarico, A., and Voitchovsky, S. (2018). Top incomes and the gender divide. *The Journal of Economic Inequality*, 16(2):225–256.

Atkinson, A. B. and Piketty, T. (2007). *Top incomes over the twentieth century: a contrast between continental european and english-speaking countries*. oup Oxford.

Attanasio, O. and Székely, M. (1998). Household savings and income distribution in mexico.

Attanasio, O. P. and Székely, M. (2004). Wage shocks and consumption variability in mexico during the 1990s. *Journal of Development Economics*, 73(1):1–25.

Azevedo, J. P., Dávalos, M. E., Diaz-Bonilla, C., Atuesta, B., and Castañeda, R. A. (2013a). *Fifteen years of inequality in Latin America: how have labor markets helped?* The World Bank.

Azevedo, J. P., Inchauste, G., and Sanfelice, V. (2013b). Decomposing the recent inequality decline in latin america. Policy Research working paper WPS 6715, The World Bank Group.

Banks, J., Blundell, R., and Preston, I. (1994). Life-cycle expenditure allocations and the consumption costs of children. *European Economic Review*, 38(7):1391–1410.

Barros, R. P. d., Carvalho, M. d., Franco, S., and Mendonça, R. (2006). Uma análise das principais causas da queda recente na desigualdade de renda brasileira.

Battistón, D., García-Domench, C., and Gasparini, L. (2014). Could an increase in education raise income inequality?: Evidence for latin america. *Latin american journal of economics*, 51(1):1–39.

Blanchet, T., Chancel, L., and Gethin, A. (2019). How unequal is europe? evidence from distributional national accounts, 1980-2017. *WID. world Working Paper*, 6.

Blinder, A. S. (1973a). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455.

Blinder, A. S. (1973b). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Blundell, R., Browning, M., and Meghir, C. (1994). Consumer demand and the life-cycle allocation of household expenditures. *The Review of Economic Studies*, 61(1):57–80.

Blundell, R. and Preston, I. (1995). Income, expenditure and the living standards of uk households. *Fiscal Studies*, 16(3):40–54.

Boschini, A., Gunnarsson, K., and Roine, J. (2017). Women in top incomes: Evidence from sweden 1974–2013.

Bouillon, C., Legovini, A., and Lustig, N. (2005). Can education explain changes in income inequality in mexico? *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*.

Bourguignon, F., Ferreira, F. H., and Lustig, N. (2005). *The microeconomics of income distribution dynamics in East Asia and Latin America*. World Bank Publications.

Brewer, M., Etheridge, B., and O'Dea, C. (2017). Why are households that report the lowest incomes so well-off?

Brewer, M. and O'Dea, C. (2012). Measuring living standards with income and consumption: evidence from the uk. Technical report, ISER Working Paper Series.

Brewer, M., Sibieta, L., and Wren-Lewis, L. (2008). Racing away? income inequality and the evolution of high incomes. *Institute for Fiscal Studies Briefing Note*, 76.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018a). Survey undercoverage of top incomes and estimation of inequality: What is the role of the uk's spi adjustment? *Fiscal Studies*, 39(2):213–240.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018b). Survey undercoverage of top incomes and estimation of inequality: What is the role of the uk's spi adjustment? *Fiscal Studies*, 39(2):213–240.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018c). Top incomes and inequality in the uk: reconciling estimates from household survey and tax return data. *Oxford Economic Papers*, 70(2):301–326.

Burkhauser, R. V., Herault, N., Jenkins, S. P., Wilkins, R., et al. (2020). What accounts for the rising share of women in the top 1%? *IZA Discussion Paper*, 13359.

Bussolo, M., Maliszewska, M., and Murard, E. (2014). The long-awaited rise of the middle class in latin america is finally happening. (WPS 6912).

Bustos, A. (2015). Estimation of the distribution of income from survey data, adjusting for compatibility with other sources. *Statistical Journal of the IAOS*, 31(4):565–577.

Campos-Vázquez, R., Esquivel, G., and Lustig, N. (2014). The rise and fall of income inequality in mexico, 1989-2010. *Falling Inequality in Latin America: Policy Changes and Lessons*, page 140.

Campos-Vazquez, R. M. and Lustig, N. (2017). Labour income inequality in mexico: Puzzles solved and unsolved. Technical report, WIDER Working Paper.

Campos Vázquez, R. M. and Rodas Milian, J. A. (2019). Desigualdad en el ingreso: posibilidades de acción pública. *Economía UNAM*, 16(46):251–261.

Canavire-Bacarreza, G. and Rios-Avila, F. (2015). On the determinants of changes in wage inequality in bolivia. (835).

Cornia, G. (2012). Inequality trends and their determinants: Latin america over 1990-2011. Working Paper Series UNU-WIDER Research Paper WP2012/09, World Institute for Development Economic Research (UNU-WIDER).

Cortés, F. (2013). Medio siglo de desigualdad en el ingreso en méxico. *Economía unam*, 10(29):12–34.

Cortés, F. (2017). Desigualdad en el ingreso en méxico, 1963 a 2014. *2018) América Latina en la larga historia de la desigualdad. Una revisión de causas, efectos y políticas FLACSO, México, in printing.*

Cortés, F. and Vargas, D. (2017). La evolución de la desigualdad en méxico: viejos y nuevos resultados. *Revista de Economía Mexicana*, 2:39–96.

Cruces, G., Domenech, C., and Gasparini, L. (2014). Inequality in education: evidence for latin america. *Falling inequality in Latin America. Policy changes and lessons*, pages 318–339.

Cunningham, W. and Maloney, W. F. (2000). Measuring vulnerability: who suffered in the 1995 mexican crisis? *World Bank mimeo*.

Cutler, D. M. and Katz, L. F. (1992). Rising inequality? changes in the distribution of income and consumption in the 1980's. *The American Economic Review*, 82(2):546–551.

De Rosa, M., Flores, I., and Morgan, M. (2020). Inequality in latin america revisited: insights from distributional national accounts. *Technical Note*, (2020/02).

Del Castillo, M. (2015). La magnitud de la desigualdad en el ingreso y la riqueza en méxico: una propuesta de cálculo.

Del Castillo Negrete, M. (2015). La magnitud de la desigualdad en el ingreso y la riqueza en méxico. *México: cepal*.

Department for Work and Pensions (2020). Households below average income: An analysis of the uk income distribution: 1994/95-2018/19. Technical report, London:Department for Work and Pensions.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, pages 1001–1044.

Eika, L., Mogstad, M., and Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6):2795–2835.

Esquivel, G. (2015). Desigualdad extrema en méxico. *Concentración del poder económico y político*.

Ferreira, F. H., Firpo, S., and Messina, J. (2014). A more level playing field? explaining the decline in earnings inequality in brazil, 1995-2012.

Firpo, S., Fortin, N., and Lemieux, T. (2007). Decomposing wage distributions using recentered influence function regressions" unpublished manuscript.

Fortin, N., Lemieux, T., and Firpo, S. (2011a). Decomposition methods in economics. *Handbook of labor economics*, 4:1–102.

Fortin, N., Lemieux, T., and Firpo, S. (2011b). Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier.

Fortin, N. M., Bell, B., and Böhm, M. (2017). Top earnings inequality and the gender pay gap: Canada, sweden, and the united kingdom. *Labour Economics*, 47:107–123.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pages 20–37. Princeton University Press.

Gasparini, L., Cruces, G., Tornarolli, L., and Mejía, D. (2011a). Recent trends in income inequality in latin america [with comments]. *Economia*, 11(2):147–201.

Gasparini, L., Galiani, S., Cruces, G., and Acosta, P. (2011b). Educational upgrading and returns to skills in latin america: Evidence form a supply-demand framework 1990-2010. (127).

Gasparini, L., Galiani, S., Cruces, G., and Acosta, P. (2011c). Educational upgrading and returns to skills in latin america: evidence from a supply-demand framework, 1990-2010. *World Bank Policy Research Working Paper*, (5921).

Gasparini, L., Marchionni, M., and Escudero, W. S. (2005). Characterization of inequality changes through microeconometric decompositions: the case of greater buenos aires. *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, pages 47–81.

Gradin, C., Canto, O., and Del Rio, C. (2008). Inequality, poverty and mobility: Choosing income or consumption as welfare indicators. *investigaciones económicas*, 32(2):169–200.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2014). Marry your like: Assortative mating and income inequality. *American Economic Review*, 104(5):348–53.

Guvenen, F., Kaplan, G., and Song, J. (2020). The glass ceiling and the paper floor: Changing gender composition of top earners since the 1980s.

HM Revenue and Customs KAI Data, Policy and Coordination (2017). Survey of personal incomes public use tape documentation.

HM Revenue and Customs KAI Data, Policy and Coordination (2019). Survey of personal incomes, 2016-2017, public use tape, data collection.

ILO (2014). Informal employment in mexico: Current, situation, policies and challenges.

INEGI (2020). Estadisticas a proposito del dia del trabajo. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2020/trabajoNal.pdf.

Jann, B. (2015). PSHARE: Stata module to compute and graph percentile shares. Statistical Software Components, Boston College Department of Economics.

Juhn, C., Murphy, K. M., and Pierce, B. (1993a). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–42.

Juhn, C., Murphy, K. M., and Pierce, B. (1993b). Wage inequality and the rise in returns to skill. *Journal of political Economy*, 101(3):410–442.

Kaplanoglou, G. and Rapanos, V. T. (2018). Evolutions in consumption inequality and poverty in greece: The impact of the crisis and austerity policies. *Review of Income and Wealth*, 64(1):105–126.

Kim, J. (2016). Female education and its impact on fertility. *IZA World of Labor*.

López-Calva, L. F. and Lustig, N. (2010). *Declining Inequality in Latin America: A decade of progress?* Brookings Institution Press.

Lustig, N. and Lopez-Calva, Luis F.and Ortiz-Juarez, E. (2013a). Declining inequality in latin america in the 2000s: The cases of argentina, brazil, and mexico. *World Development*, 44:129 – 141.

Lustig, N. and Lopez-Calva, Luis F.and Ortiz-Juarez, E. (2013b). Deconstructing the decline in inequality in latin america. (WPS 6552).

Machado, J. A. and Mata, J. (2005a). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

Machado, J. A. and Mata, J. (2005b). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

McKay, A. and Lawson, D. (2003). Assessing the extent and nature of chronic poverty in low income countries: issues and evidence. *World Development*, 31(3):425–439.

McKenzie, D. J. (2003). How do households cope with aggregate shocks? evidence from the mexican peso crisis. *World Development*, 31(7):1179–1199.

McKenzie, D. J. (2006). The consumer response to the mexican peso crisis. *Economic Development and cultural change*, 55(1):139–172.

Meyer, B. D. and Sullivan, J. X. (2003). Measuring the well-being of the poor using income and consumption. Technical report, National Bureau of Economic Research.

Meyer, B. D. and Sullivan, J. X. (2009). Five decades of consumption and income poverty. Technical report, National Bureau of Economic Research.

Meyer, B. D. and Sullivan, J. X. (2011). Viewpoint: Further results on measuring the well-being of the poor using income and consumption. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(1):52–87.

Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. *Franco Modigliani*, 1(1):388–436.

Oaxaca, R. (1973a). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Oaxaca, R. (1973b). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Poterba, J. M. (1989). Lifetime incidence and the distributional burden of excise taxes.

Rodriguez-Castelan, C., Lopez-Calva, L. F., Lustig, N., and Valderrama, D. (2016). Understanding the dynamics of labor income inequality in latin america. (WPS 7795).

Salverda, W., Nolan, B., and Smeeding, T. M. (2009). *The Oxford handbook of economic inequality*. Oxford University Press.

Santaella, J. M. and Leyva, G. (2017). Quién se lleva los frutos del éxito en méxico?

Slesnick, D. T. (1993). Gaining ground: poverty in the postwar united states. *Journal of political Economy*, 101(1):1–38.

Székely, M. (2005). Pobreza y desigualdad en méxico entre 1950 y 2004. *El Trimestre Económico*, 72(288(4)):913–931.

Terhi, R. et al. (2018). Top incomes and income dynamics from a gender perspective: Evidence from finland 1995-2012. Technical report.

Vélez, C. E., Leibovich, J., Kugler, A. D., Bouillon, C. P., and Núñez, J. (2004). The reversal of inequality trends in colombia, 1978-95: A combination of persistent and fluctuating forces.

Yavorsky, J. E., Keister, L. A., Qian, Y., and Nau, M. (2019). Women in the one percent: Gender dynamics in top income positions. *American Sociological Review*, 84(1):54–81.

Chapter 1

The Impact of Education on Income Inequality in Latin America between 2000 and 2010

1.1 Introduction

Inequality in Latin America has been a hot topic because of its high levels and its recent homogeneous declining trend in the region.

The effects of the expansion of education on income inequality in Latin America have been recently under discussion, due to the so called "Paradox of Progress" (by Bourguignon et al. (2005)). This paradox is defined as the inequality-increasing effect of the expansion of education (keeping the returns constant). The authors argue that two mechanism are driving the phenomenon: the convexity of the returns to education, and

the fact that income sources tend to be a convex function of years of schooling. Therefore, if there is an increase in years of education we should expect an increase in inequality. Bourguignon et al. (2005) show evidence of the inequality-increasing effect of education in Argentina, Colombia, Indonesia, Malaysia and Mexico. Applying the same method, Battistón et al. (2014) finds that an increase in years of education increases inequality in the region under the presence of convexity in the returns to education in the labor market. The authors use data from 16 countries in Latin America over a period covering the 1990s up to the 2000s. However, Cruces et al. (2014) concludes that the expansion of education in the 2000s had an equalizing effect on earnings in 25 Latin American countries; they point out the equalizing effect of education has reached a turning point due to the gaps in the quality of education. A small number of studies have focused on analyzing the presence of this paradox across the region over the period of decreasing inequality. There is not yet conclusive evidence on the inequality increasing effects of education. This paper contributes to fill this gap in the literature on the effects of education on the decline in income inequality in Latin America.

My study contributes to a better understanding of the impact of education on the homogeneous decline in income inequality for 18 countries in Latin America. In particular this paper answer the following question: How has the expansion of education contributed to the decline in income inequality in Latin America? In order to answer this question, I implement the decomposition method proposed by Firpo et al. (2007) (hereafter FFL method) using 36 harmonized household income surveys from 18

Latin American countries[1].

My results show that education has an inequality-decreasing effect in four countries and an inequality-increasing effect in six countries. Further, the detailed decomposition shows that the changes in the returns to education are the driving component of these education effects, with changes in the returns to education reducing inequality in six countries and increasing it in only three countries. On the other hand, changes in the years of education ceteris paribus have an inequality-increasing effect in most of the countries. This effect is consistent with the presence of the Paradox of Progress in the region. The rest of the paper is organized as follows. Section 1.2 presents the literature review. The methodology is described in Section 1.3, Section 1.4 is devoted to the data description. Section 1.5 presents the results and Section 1.6 concludes.

[1]The surveys are part of the Socio-economic Database for Latin America and the Caribbean (SEDLAC) http://sedlac.econo.unlp.edu.ar/eng/

1.2 Literature Review

There is strong consensus on the evidence about decreasing inequality levels in most of the countries in Latin America over the last two decades[2]. A number of authors have been recently analyzing the factors behind the inequality decline in the region, however there is no consensus about the main determinants across countries. The following literature review is organized as follows: First I will present the literature that documents the link between education and inequality. Then, I will focus on the determinants of the fall in income inequality divided by the type of methodology that it is implemented: cross country analysis and decomposition techniques. The last subsection is devoted to the literature that implements the FFL decomposition method.

1.2.1 The effects of education on inequality

The Paradox of Progress

From a standard supply and demand framework, I expect that an increase in the supply of people with different education levels (keeping everything else constant) should reduce the relative wages between those with more years of education and the ones with no additional years. However, if the demand of people with more years of education increases, the reduction of relative wages is not possible, in fact, the most important component of disposable income in Latin America is labour income. Thus, the main channel through

[2]Gasparini et al. (2011a), Alvaredo and Gasparini (2013)

which changes in education impact on the inequality of household disposable income is their impact on earning and labour income, could increase. As a result, the wage dispersion will be higher and income inequality would increase.

Recently, another group of authors have found that the convexity of the returns to education could reverse the effects of education on inequality. Bourguignon et al. (2005) show evidence of the inequality-increasing effect of educational expansion in Argentina, Colombia, Indonesia, Malaysia and Mexico; a phenomenon named by the authors as the "Paradox of Progress". Using a counterfactual microsimulation technique, the authors simulate an increase in years of education, and find that this would increase inequality. According to the authors the mechanism is the following: First, the returns to education are a convex function of years of schooling. Second, the other incomes have to be a convex function of years of education. Under these conditions, an increase in the years of education will lead to an increase in inequality under the presence of increasing returns to education.

Gasparini et al. (2005) argues that several forces driving in different directions explained similar levels of inequality between 1986 and 1992 in Argentina, but the returns to education had an inequality-decreasing effect. However, between 1992 and 1998 all the determinants contribute to the rise in inequality, especially the increase in the returns to education. Bouillon et al. (2005) decompose the changes in income for Mexico between 1984 and 1994, a period of increasing inequality in the country. The authors find that changes in returns to education are the main component of the

increase in inequality, regardless of the expansion in years of education. The authors argue that the positive effect of education on inequality is due to the increase in the marginal returns to education.

Latter on, Battistón et al. (2014) using the same methodology as Bourguignon et al. (2005) and the SEDLAC dataset for 13 Latin American countries, note that education has an dis-equalizing effect between 1990s and 2000s. The authors suggest that this is due to the convexity of the returns to education, which in the case of Latin America has been related to the skill biased technological change. Gasparini et al. (2011c) analyze the evolution of returns to skills for 16 Latin American countries. They show that because of the increase in the supply of skilled workers, the returns to education declined for the last two decades. However, they argue that there is a lot of heterogeneity across countries.

On the other hand, Cruces et al. (2014) find that there is no correlation at the country level between having higher inequality in the distribution of education, and higher income inequality. The authors measure the impact of education inequality on income distribution between 2000 and 2010 in Latin America. They conclude that the expansion of education has not had a clear equalizing effect on income during the 1990s due to the market-oriented reforms. However, during 2000s it has an equalizing impact because of more pro-poor educational policies.

Finally, Vélez et al. (2004) analyzes the case of Colombia on the period between 1978 to 1995, when inequality shows a "U" shape. The authors identify two types of factors affecting inequality trends. The first type are persistent factors such as the years of education, the reduction in the size of the families, and the increase in female labor force participation. The second group are fluctuating factors that include the returns to education, sociodemographic characteristics and unobservable determinants. The authors argue that the expansion of education has an dis-equalizing effect in urban areas, while in rural areas, it has an equalizing effect.

Our conclusions from this literature review is that there is little consensus on the effects of education on inequality and on the role of the returns to education, and whether this role changed over time. The contribution of my paper is to study 18 countries in Latin America using harmonized micro-data, and to study the impact of the expansion of education on the decline in inequality between 2000 and 2010. This analysis will allow us to see if the Paradox of Progress is a pervasive phenomenon or not among the countries in the sample. In the next subsection, I review an additional channel through which education can effect income inequality: changes in household structure.

The effects of education on household structure

The expansion of education can impact household income through changes in the household structure. Changes in fertility rates and assortative mating affect family structure (see Andersen (2018)). According to Kim (2016), highly educated women delay pregnancy due to their good position in their labor market. Therefore education reduces their fertility and as a result per-capita disposable income within the household increases, as well as inequality. Also, there is evidence that more educated women prefer fewer children than less educated women. This preference increases the inequality with respect to other households with women with low education. Another channel is via assortative mating, when highly educated women marry highly educated men. If education is positively correlated with income, an increase in education will lead ceteris paribus to increased income inequality as differences between high and low educated couples increase (see Greenwood et al. (2014) and Eika et al. (2019)).

Before continuing with the methodology, it is necessary to review the literature on the determinants of the decline in income inequality in Latin America to understand the context of the current discussion and the relevance of this study. The following review is divided in to two groups of the literature, the first group does mainly a cross-country review from other authors and the second group implements a decomposition method to disentangle the determinants on income inequality in the region.

1.2.2 The decline in income inequality in Latin America

Over the last three decades, there is a strong consensus that inequality levels in most of the countries in Latin America have been receding[3]. However, there is yet no agreement about the main driving factors of such improvement. Most of the cross-country studies on income inequality tend to analyze the phenomenon by applying a decomposition method, while a few others implement a regression analysis[4]. The following two subsections include the review of works using a cross-country analysis and a decomposition method to estimate the determinants of the decline in income inequality in different countries in Latin America.

Cross-country analysis

Different explanations for the decline in income inequality have been proposed. Alvaredo and Gasparini (2013) argue that several forces are behind the downward trend. These include: improvement in macroeconomic conditions, higher employment levels, the expansion of education and social spending, particularly monetary transfers. However, they recognize that the empirical evidence about the main factors is still scarce. Cornia (2012) uses data from 18 Latin American countries spanning from 1990 to 2009. His work applies a country-level regression analysis, where the Gini index of the distribution of household disposable income per capita is the dependent variable and as independent variables are terms of trade, the rate of growth of GDP per-capita,

[3] see for instance Gasparini et al. (2011a), Alvaredo and Gasparini (2013)
[4] see Cornia (2012)

changes in exogenous factors such as the dependency rate and the activity rate, the distribution of human capital among workers, fiscal policies proxies by the ratio of direct to indirect taxes and public expenditure on social security/GDP, labor market policies like minimum wage, and macroeconomic policy. The author concludes that the causes of the decline of income inequality in the region were heterogeneous. The reversal of the skill premium appears to have played a central role in improving the distribution of income, but also the fiscal and labor market policies.

Lustig and Lopez-Calva (2013b) compile results from 11 countries. The authors suggest that there were two main driving factors behind the decreasing inequality levels: reduction in hourly labor income inequality and higher government transfers. According to the authors, the fall in hourly labor income inequality was explained by the reduction in the returns to education, either due to the increase in the supply of workers with higher levels of education or a reduction in the demand of workers with higher skills. In a similar fashion, Lustig and Lopez-Calva (2013a) analyze the cases of Argentina, Brazil and Mexico. The authors examine the role of the demand and supply of labor skills; institutional factors, such as minimum wages; unionization rates; and government transfers. Their results show that both labor and non labor income inequality had declined, and that there were two underlying phenomena in the three countries: the fall in the premium to skill labor, and larger and more progressive government transfers. The fall in the skill premium can be attributed to changes in the composition of demand and supply of labor by skill, and institutional factors such as rising minimum

wage and unionization. However, the relative strength of these factors varies substantially by country.

Gasparini et al. (2011a) document the recent trend in income inequality in Latin America and argue that there were several plausible factors behind the fall in inequality. They highlight: employment growth, changes in relative prices, realignment after reforms, cash transfer programs, and an increased in governments' concern over high and rising inequality levels.

López-Calva and Lustig (2010) argue that there were two main factors that drove the inequality decline: a reduction in the earnings gap between skilled and low skilled workers, and an increase in government transfers to the poor. They suggest that the expansion of basic education might have contributed to the gap reduction. However, there is a lot of controversy about the impact of the macroeconomic conditions and policy reforms on the income distribution. This is because it is not easy to separate, and measure the effects behind of such a complex variable like inequality.

Gasparini et al. (2011b) study the changes in wages using a supply and demand framework between 1990s and 2000s to explain the fall of wage skill premium. They find that can be partially attributed to the recent boom of commodity prices that could favor the unskilled workers. However, there are other forces playing a role in the changes. Lately, other authors such as Rodriguez-Castelan et al. (2016) have concentrated on the

explanation of the decline in labor income inequality in Latin America. They show that this decline is related with a reduction in college/primary education premium, the urban-rural earnings gap, as well as in high school/primary education premium. They find that there has been a decline in experience premium after 2002 in the region, that has been contributing to the decline in the returns to education.

Most of the literature presented in this subsection focus on the determinants of the decline in wage inequality in different countries while my research covers household income for different countries in the region using one methodology to decompose the effect of education.

Decomposition methods

This subsection reviews the works that use a decomposition technique in order to disentangle the main components of the decline in inequality in Latin America. I will pay special attention at the end of this subsection on to those works that are implementing the FFL decomposition which in all the cases it has been applied to specific countries in the region.

The following group of authors agree on the declines of the returns to characteristics as one of the main driving factors of the decrease in income inequality. Azevedo et al. (2013a) decompose inequality in labor income for 15 Latin American countries for the years 1995, 2000, 2005 and 2010. They apply the decomposition method proposed by

Juhn et al. (1993a) using the Gini and Theil index as statistics of interest. The authors attribute the decline of labor income inequality to a reduction in the returns to education and experience. In a similar way, but extending the analysis to different parts of the distribution, Azevedo et al. (2013b) use data from 14 Latin American countries for the years 2000 and 2010 applying the decomposition technique proposed by Barros et al. (2006) to determine the factors behind the decrease in earnings inequality. The authors suggest that the reduction in inequality can be attributed to an increase in earnings per-hour at the bottom of the income distribution. This decomposition method builds on the idea of Juhn et al. (1993a), but instead of estimating a Mincer model, they generate counterfactual distributions. This method belongs to the group of decompositions that go beyond the mean, which constitute a relevant advantage over other methods focusing on the mean. However, its main drawback is that it is not possible to measure the contribution of each covariate to the structure and composition effect. The heterogeneous results and the disadvantages of this decomposition methods leave unanswered questions and gaps in the literature that a more detailed decomposition technique such as the FFL method could answer.

The next group of authors implements the FFL decomposition method on a specific country: Motivated by the fall in labor income inequality in Brazil between 1995 and 2012, Ferreira et al. (2014) decompose the mean and the Gini index of the differences in earnings for the following periods 1995-2012, 1995-2003 and 2004-2012, and estimate the effects of the covariates on the changes in the Gini index and the mean. The FFL

decomposition method allows them to measure the size of the effect of four candidate exploratory factors: human capital, labor market institutions, demographic composition of the labor force, and spatial segmentation. They conclude that the fall in labour income inequality between 1995 and 2012 was driven mainly by the distribution of returns from the labor market rather than the composition effect. Also, the reduction in gender, racial and urban-rural gaps contributed to reduce inequality. Canavire-Bacarreza and Rios-Avila (2015) decompose the change in the labor income distribution for Bolivia between 2000 and 2012. They find that the main driving forces were: faster wage growth of low-paid jobs and wage stagnation of jobs that require higher education. However, they argue that a large proportion of the decline remains unexplained. Campos-Vázquez et al. (2014) analyze inequality trends for Mexico between 1989 and 2010, examining the role of demand and supply of labour skills, as well as institutional factors like minimum wages and unionization rate. They also consider the effect of cash transfers in explaining changes in inequality. Using the FFL method they decompose the changes in inequality of hourly wages and show that, during the period between 1994 and 2006, most of the changes occurred at the bottom of the income distribution, where the wages of low-skilled workers increased. Their decomposition results also suggests that in the period from 2006 to 2010 the effect of cash transfers contributed to the decline in income inequality. Recently Bussolo et al. (2014) have shown that the fall in inequality is expecting to continue in the region due to the decline in skill premia.

The FFL decomposition method provides a deep analysis of the driving factors of inequality. However, the authors have concentrated on specific countries. The potential of the FFL decomposition can be extended to more countries and periods of time. I applied the FFL method using household income surveys from 18 countries for the period between 2000 and 2010, when inequality was in evident decline. The detailed analysis that this decomposition provides, improves over the approach previously presented in Azevedo et al. (2013b), because I estimate the effect of education on the structure and composition components and not only the aggregate effect of these two components. The work of Ferreira et al. (2014) provides a detailed analysis using the FFL method only for the particular case of Brazil. My paper extends the analysis for 18 countries to compare the determinants behind the apparent homogeneous declining pattern of inequality in the region.

1.3 Methodology

In this section, I set out the decomposition method I use. This was first developed by Firpo et al. (2007) and it is usually called "the FFL method"'. The FFL method is a generalization of the Oaxaca-Blinder decomposition (Blinder (1973a) and Oaxaca (1973a)) and is intended to decompose summary statistics of an outcome variable than the mean. The approach belongs to the group of "distributional methods" according to the classification provided by Fortin et al. (2011a), and to the class of tools that allow for detailed decomposition. Distributional methods are characterized as those decomposition techniques that can decompose statistics other than the mean (such as the variance, Gini and quantiles). A detailed decomposition method is one that subdivide the structure and composition effect into the contribution of each covariate to the two effects. The main advantage of getting detailed information from the distribution is that it provides guidelines for policy recommendations that less detailed decomposition methods cannot have.

Within the group of distributional methods of decomposition there are other three types of decompositions similar to the FFL. The first one proposed by Juhn et al. (1993b), the second one by Machado and Mata (2005a) is based on counterfactual distributions. The JMPs decomposition is based on a residual imputation while the method by Machado and Mata (2005a) is a conditional quantile regression method. The third method was proposed by DiNardo et al. (1996), this method consist of using a

reweighting factor to estimate the structure component.

The residual imputation approach by Juhn et al. (1993b) is based on the strong assumption that residuals are independent of the covariates according to Fortin et al. (2011b). Also, this method does not allow to have a detail decomposition of the composition effects. This means we can not separate the contribution of each covariate into the composition effect.

The decomposition method proposed my Machado and Mata (2005b) originally have the same drawback, which is that is not possible to estimate a detailed decomposition of the composition effect. However, the authors suggest an unconditional reweighting approach to decompose the composition effect. Nonetheless, according to Fortin et al. (2011a) this approach does not provide a consistent effect.

Finally, the reweighting method by Juhn et al. (1993b) has the advantage of not being path dependent. However, it includes an interaction term that can be difficult to interpret.

The FFL method base on the RIF-regression is an straightforward extension of the Oaxaca-Blinder decomposition proposed by Oaxaca (1973b) and Blinder (1973b). This decomposition has the advantage of being path independent easy to interpret. Its main limitation is that it is not possible to have general equilibrium effects.

1.3.1 The FFL method

The FFL method is a two stage procedure: The first step is the estimation of a counterfactual distribution; this allows the total change in the statistic of interest to be split between the structure and composition effects. The second step involves the estimation of the contribution of each covariate to the structure and composition effects using the RIF-regression. The overall aim is to estimate the contribution of a set of covariates, X to the overall difference Δ_O^τ in some distributional statistic τ between two groups. In our application we take quantiles as the statistic of interest and the groups are time periods.

First Step: The estimation of weights to separate the structure and composition effect.

The first step in the methodology involves estimating three sets of weights. Following Firpo et al. (2007), suppose we have a random sample of $N = N_1 + N_0$ individuals, where N_1 and N_0 are the number of individuals in each group, where $i = 1, \ldots, N$ are the individuals. The probability that an individual i is in group 1 given a set of covariates $X = x$ is $p(x) = Pr[T = 1|X = x]$ called the "propensity score", where $T_i = 1$ if individual i is observed in group 1, and $T_i = 0$ if individual i is observed in group 0. Then, FFL method define three weighting functions as follow:

$$\omega_1(T) \equiv \frac{T}{p}; \qquad \omega_0(T) \equiv \frac{1-T}{1-p}; \qquad \omega_C(T,X) \equiv \left(\frac{p\,(X)}{1-p\,(X)}\right) \cdot \left(\frac{1-T}{p}\right)$$

The first two functions ($\omega_1(T)$ and $\omega_0(T)$) transform features of the marginal distribution of Y into features of the conditional distribution of Y_1 given $T = 1$ (conditional distribution of time one, given that households belong to time 1) and of Y_0 given $T = 0$ (conditional distribution of time zero, given that households belong to time 0). The third re-weighting function ($\omega_C(T, X)$) transforms features of the marginal distribution of Y into features of the counterfactual distribution of Y_0 given $T = 1$.

The counterfactual distribution cannot be directly identified, but the FFL shows that it is possible to estimate the counterfactual income distribution under the assumptions of ignorability and overlapping support. The ignorability assumption states that the distribution of the unobserved characteristics is the same across groups 1 and 0, once we condition on a vector of observed characteristics. The overlapping support ensures that no single observable or unobservable characteristic can identify the membership into one of the groups or periods of time. In our application this means that, if the ignorability and overlapping support conditions hold, then we can estimate a counterfactual income distribution that would be observed if households living in period t_0 had experienced the income structure observed in period t_1.

In practice, we can estimate $p(x)$ using a probit model, and having done this we can estimate the weights functions. With the estimated weights we can calculate the overall

decomposition:

$$\Delta_O^\tau = (\tau_1 - \hat{\tau_C}) + (\hat{\tau_C} - \tau_0) = \Delta_S^\tau + \Delta_X^\tau \tag{1.1}$$

We can define the terms similar in spirit to the Oaxaca-Blinder decomposition where Δ_O^τ is the overall income gap measured in terms of the distributional statistic τ, which for our case is quantile. The first term is the structure effect Δ_S^τ and the second term the composition effect Δ_X^τ, τ_1, τ_0 and $\hat{\tau_C}$ are the τ distributional statistics corresponding to the observed income distribution in time $t = 1$, $t = 0$, and with $t = C$ the estimated statistic from the counterfactual income distribution.

Second Step: Estimation of the Recentered Influence Function (RIF)

The second step is using OLS to estimate the linear projection of the recentered influence function (RIF) regression. This is known as the RIF regression and it is identical to standard OLS, except that the dependent variable is the recentered influence function of the statistic of interest.

The influence function of a quantile[5] is defined as: $IF\left(y; q_\tau\right) = \frac{\tau - \mathbf{1}\{y \leq q_\tau\}}{f_Y(q_\tau)}$. The Recentered Influence Function (RIF) of the quantile τ^{th} is $RIF\left(y; q_\tau\right) = q_\tau + IF(y; q_\tau) = q_\tau + (\tau - \mathbf{1}\left\{y \leq q_\tau\right\} / f_Y\left(q_\tau\right))$.

The linear approximation for the conditional expectation of the RIF is:

[5] $\tau - th$ quantile of the distribution F is defined as $Q(F, \tau) = inf\left\{y | F(y) \geq \tau\right\}$

$$\mathbf{E}\left[RIF\left(Y,\tau\right)|X\right]=X'\beta \tag{1.2}$$

where τ is the statistic of interest. The resulting regression coefficients are:

$$\hat{\beta}_t^{\tau}=\left(\sum_i^N \hat{\omega}(T_i)X_iX_i'\right)^{-1}\cdot\sum_{i=1}^N\hat{\omega}_t\left(T_i\right)\hat{RIF}\left(Y_i;\nu_t\right),\ t=0,1, \tag{1.3}$$

$$\hat{\beta}_C^{\nu}=\left(\sum_{i=1}^N\hat{\omega_C}\left(T_i,X_i\right)X_iX_i'\right)^{-1}\cdot\sum_{i=1}^N\hat{\omega_C}\left(T_i,X_i\right)X_i\hat{RIF}\left(Y_i;\nu_C\right) \tag{1.4}$$

where $\hat{\omega}_C\left(X_{i,0}\right)$ is the estimated weight from the first step.

With these estimated parameters we can express the overall structure and composition effects for any quantile τ as follows.

$$\hat{\Delta}_S^{\tau}=\mathbf{E}\left[X,T=1\right]\cdot\left(\hat{\beta}_1^{\tau}-\hat{\beta}_C^{\tau}\right), \tag{1.5}$$

$$\hat{\Delta}_X^{\tau}=\left(\mathbf{E}\left[X|T=0\right]\right)\cdot\hat{\beta}_0^{\tau}+\hat{R}^{\tau} \tag{1.6}$$

where $\hat{R}^{\tau}=\mathbf{E}\left[X|T=1\right]\cdot\left(\hat{\beta}_C^{\nu}-\hat{\beta}_0^{\nu}\right)$ is an approximation error due to the fact that the FFL decomposition is based on the first order approximation to the composition effect.

Crucially this allows us to estimate the contribution of each covariate to each of the structure and composition effect. For example, the contribution of education to the change in a particular income quantile can be expressed as follows:

$$\hat{\Delta}^{\tau}_{T_{edu}} = \Delta^{T}_{S_{edu}} + \Delta^{T}_{X_{edu}}, where \tag{1.7}$$

$$\hat{\Delta}^{\tau}_{S_{edu}} = \mathbf{E}\left[X_{edu}, T = 1\right] \cdot \left(\hat{\beta}^{\tau}_{1_{edu}} - \hat{\beta}^{\tau}_{C_{edu}}\right), \tag{1.8}$$

$$\hat{\Delta}^{\tau}_{X_{edu}} = \left(\mathbf{E}\left[X_{edu}|T = 0\right]\right) \cdot \hat{\beta}^{\tau}_{edu} + \hat{R}^{\tau} \tag{1.9}$$

1.3.2 Inference

The FFL decomposition method requires several steps, and this complicates statistical inference. Therefore, I implement a paired bootstrap sampling method in order to estimate confidence intervals. From the observed sample of each year and country, I obtained $B = 300$ bootstrap samples in order to estimate a set of parameters.

From the set of parameters $\hat{\theta}^{*}_{1} \ldots \hat{\theta}^{*}_{B}$ obtained from the B bootstrap replications I calculate the bootstrap confidence interval of $\hat{\theta}$ using the percentile method[6].

[6]The percentile method establish the confidence intervals directly from the bootstrap distribution of the parameters. Once the values of $\hat{\theta}^{*}_{B}$ are in ascending order, the confidence interval is the distance between the lower $\alpha/2$ and the upper $\alpha/2$

1.4 Data

The data used in this study are part of the Socio-Economic Database for Latin America and the Caribbean (SEDLAC). I use 36 harmonized cross-section household surveys from 18 countries. The surveys of Argentina, Uruguay and Venezuela are representative at urban level while the other 15 countries are representative at the national level. To make periods comparable across time, I use circa criteria for the years 2000 and 2010. Table 1.1 provides more details of the countries, years and surveys included in this research.

My interest is in the change in the distribution of household income. More precisely, I use per capita household income, defined as disposable income (net market income[7], plus cash transfers[8]) divided by the number of persons in the household,. I do not use a different equivalence scale because most of the relate literature in Latin America use per capita income (see Bourguignon et al. (2005), Gasparini et al. (2005), Battistón et al. (2014), Cruces et al. (2014), Vélez et al. (2004), Lustig and Lopez-Calva (2013b), Gasparini et al. (2011a), López-Calva and Lustig (2010) and Gasparini et al. (2011b)). Therefore, to have comparable results with the rest of the literature I use per capita income.

I use household income because it better reflects the well-being of the family

[7]Labor market income, capital income excluding social contributions and taxes.
[8]Social insurance, assisted program, etc.

members. Also, the expansion of education can impact household income through the changes in the structure of the households not just via the labor market, see the review in section 1.2.1. Labor income constitutes the largest component of total household income in Latina America, and so I expect results for household income to be similar to those obtained using earnings only.

The covariates can be organized in three groups: socio-demographic, labor market and education. Socio-demographic variables are: urban(dummy), household size(continuous), age(continuous) and gender(dummy). Labor market has one dummy variable that identifies if the head of the household works. The education variable is years of education of the head of the household (continuous). The use of variables from the head of the household is justified by the relevance of the family background (see section 1.2.1) on differences in income at different stages in life. Therefore the head of the household characteristics are a good approximation and control variables for the family background.

Table 1.2 has the descriptive statistics per country. Total per-capita income increased from 2000 to 2010, only two countries experience a decline: El Salvador and Venezuela with a decrease in income of 1.9 and 4.2 log points, respectively. On the other hand countries like Argentina, Bolivia and Ecuador experience an increase in income of two times the average over the same period: 2.3, 1.6 and 1.3 log points, respectively. Years of education increase by one year on average. In terms of labour market participation,

there was a reduction in 13 countries out of 18: Argentina, Panama, Peru, El Salvador and Uruguay experienced an increase. The three sociodemographic variables such as the household size, average age of the head of the household and the percentage of females as heads of the household, present the following changes: The average reduction in household size was very small. It went from 4.3 members in 2000 to 3.8 in 2010. The average age of the head of the household increased by 2 years. The proportion of females heads of the household increased by 6.2 percentage points.

1.5 Results

I present in this section my findings on the contribution of the expansion of education on the decline in income inequality. First, I start with the analysis of the re-centered influence function regression results in order to see the heterogeneous effect of education across the income distribution. Later, I will focus on the decomposition results to disentangle the total impact of education into structure and composition effect across income quantiles.

1.5.1 RIF Regressions

Before presenting the decomposition results, I analyze the results from the re-centered influence function regression for the variable of education. From equation 1.2, I estimate $IF(y_i; q_\tau)$ for the education variable. The estimates for the nine quantiles are reported in Figures 1.1 and 1.2[9]. We see a positive effect of education across the income distribution in both years 2000 and 2010. This implies that years of education have an inequality-increasing effect. As we move up in to the distribution, education increases its returns for household at the top quantiles. But, if we look at the change in household income between 2000 and 2010; the effect of education drops from 2000 to 2010 in most of the countries while in others such as Ecuador and Mexico it increases between 2000 and 2010. Nonetheless, those effects are not equally distributed along the

[9]The RIF regression coefficients for the rest of the variables are reported from Figure 1.18 to Figure 1.20

quantiles. For example, Nicaragua, Peru and Paraguay the drop was mainly at the top of the income distribution. El Salvador for instance, shows a fall in the bottom half of the income distribution. More importantly, the results indicate that education has differentiated effects on inequality in the region. For example, in Peru the reduction of the effects of education between 2000 and 2010 is bigger at the top end of the distribution, which contributes to the fall in inequality. The opposite situation is present in Costa Rica, where the effects of education across the distribution increases between 2000 and 2010 at the top end of the income distribution while decreases at the bottom, this contributes to an increase in inequality.

1.5.2 Empirical Findings on the Aggregate Decomposition of Income Inequality

Figure 1.3 shows the total change in inequality (Δ_0^τ), measured by the natural logarithm of the 90-10 ratio, between 2000 and 2010. Almost all of the 18 countries experienced a fall in inequality (in Colombia the change is not statistically significant). Bolivia, Dominican Republic, Peru and El Salvador are among the countries that experience the biggest fall in inequality. Costa Rica, Ecuador, Mexico and Uruguay experience a modest decline.

Figure 1.4 shows the decomposition of the log 90-10 ratio into the structure and composition effect. This is based on the equation (1.1). To estimate the reweighting

factor $\hat{w}_C(T_i, X_i)$, I use the set of covariates described in Section 4. For the decomposition estimation I use the same set of covariates. The decline in inequality for all of the countries has been driven by the structure effect (see Figure 1.5 and Table 1.16). On the other hand, the composition effect has a smaller contribution to the total change but increases inequality in 11 countries (see Figure 1.6).

1.5.3 Empirical Findings on the Education Effects on Income Inequality

The next step in the decomposition shows the total effect (structure and composition effect) of education on inequality, based on equations 1.7 and 1.9. Figure 1.7 shows that there are heterogeneous effects across countries. However, it is possible to divide the effects in to two groups. For the first group of countries, education has an inequality increasing effect: Colombia, Costa Rica, Ecuador, Mexico, El Salvador and Uruguay. The second group is composed of four countries, where education has an inequality reducing effect: Chile, Dominican Republic, Nicaragua and Peru.

Figure 1.8 divides the total effect of education on inequality into structure and composition effect. It is clear how some of the countries in the region are divided by the sign of the effect, specially the structure effect or the changes in the returns to education. The total effect (dot bars) of education is driven by the structure effect that contributes to the reduction of inequality for six countries in the sample. Not

surprisingly perhaps, the composition effect (dash bars) increases inequality in 14 countries.

Figure 1.9 and Figure 1.10 show the two effects with their respective confidence intervals. The composition effect (see Figure 1.9) increases inequality for most of the countries except in Dominican Republic, Argentina and El Salvador where the effect is not statistically significant. On the other hand, in Figure 1.10 the structure effect shows a different pattern. For six countries the structure effect contributes to the reduction in inequality, while in only three countries increases inequality (see also Table 1.17).

1.5.4 Discussion

The decomposition analysis of education shows that the changes in education have shaped the inequality levels in different directions in the region. These differentiated effects of education on the decline of income inequality allow us to identify in the 2000-2010 period two groups of countries: for one group, education has an inequality decreasing effect and in the second one education has an inequality increasing effect. Further, we find that the main component of the total effect of education is the structure effect rather than the composition component, where the latter has a positive effect for most of the countries.

The decomposition analysis presented in this work disentangles two relevant

components of the changes in education an their effects on income inequality. Most of the countries in Latin America have experienced an expansion of basic education over the last decade, among other policies targeting the educational sector. The results just presented in my work show that if we only consider the impact of the increased level of education (or the composition effect; see Figure 1.9) then we would conclude that education expansion increases inequality in all countries in the region. However, the structure effect which captures changes in the returns to education is playing a significant role within the total effect of education on inequality. Indeed, the structure effect plays such a significant role that in some countries it more than compensates the composition effect (see Figure 1.8), and in these countries the total effect of educational expansion is inequality-decreasing: Chile, Dominican Republic, Nicaragua and Peru.

My results are consistent with Ferreira et al. (2014) for the case of Brazil, where the authors find that the total effect of human capital reduces inequality in the period between 1995 to 2012. This study finds that the structure effect of human capital has an inequality decreasing effect, while the composition effect increases inequality. Also, for the case of Mexico my results are consistent with Campos-Vázquez et al. (2014). They show evidence of the paradox of progress in Mexico using the FFL method by decomposing the change in wages into structure and composition effect between 1998 and 2010. We find that the total effect of education is inequality increasing. The detailed analysis of the structure and composition effect shows that only the composition effect has a significant and inequality increasing effect. Finally, the work by

Battistón et al. (2014) where they analyze the effect of the expansion of education in 13 Latin American countries, they find that education increases inequality in the region between the 1990s and 2000s. However, they do not highlight the fact the changes in returns to education are the main component on the effects of inequality rather than the expansion of education considered alone. The changes in the returns compensate in some cases the increasing effects of the expansion of the years of education as it is the case of Chile, Dominican Republic, Nicaragua and Peru where the total effect of education is inequality decreasing.

My decomposition analysis shows that the Paradox of Progress is present in 14 out of 18 countries in the sample, because changes in the distribution of years of education have an inequality increasing effect (keeping the returns constant). However, the expansion of years of education is not the only way that education affect inequality. The structure effect, or the returns to education, is actually the main component of the total effect of education. The effect of the decline in the returns to education at the top half of the income distribution goes in the opposite direction, contributing to the reduction in inequality in six out of 18 countries . As a result, the total effect of education is inequality increasing in six countries and inequality decreasing in four countries.

The expansion of education

Figure 1.11 shows the average change in years of education in the 18 countries between 2000 and 2010. Years of education increase by one year, on average, across countries. The expansion of education across the income distribution benefited more the households at the bottom half of the income distribution in most of the countries (see graphs 1.12 and 1.13). For instance, in Argentina, Bolivia, Brazil and Chile the increase in years of education was mainly at the bottom half of the income distribution, while the top quantile does not experience a significant increase in years of education compare with the bottom quantiles. Dominican Republic and Honduras present a different pattern. The top half of the distribution experience a decline in years of education, while the bottom half increases between 2000 and 2010. Other countries observe an apparent homogeneous increase in years of education. For example Mexico and Costa Rica, where the increase in years of education is similar across the income distribution.

The correlation between the estimated effect of education on inequality and the observed expansion in years of education is shown in Figure 1.14. The correlation between this two variables is slightly negative, which means that positive changes in years of education are associated with the negative effects of educational expansion on inequality. Previously, we saw that the total effect of education was heterogeneous on inequality. Some of the countries experience a positive effect while other a reduction in inequality. Therefore, the weak correlation between the change in years of education and the estimated effect of education is consistent due to the contradictory forces

pushing inequality up and down. In addition to that, this correlations are in line with Bourguignon et al. (2005)'s argument about the expansion of education and the inequality levels. They said that regardless of the distribution of years of education their effects on inequality will vary depending on the mechanisms of transmission from education to income in each country.

Figure 1.15 shows the correlation between the estimated impact of education on inequality and the changes in the estimated returns to education. The graph shows a negative correlation between the two variables. This implies that bigger reductions in returns to education at the top quantiles with respect to the bottom are correlated with the inequality decreasing effect of education. This strong correlation is consistent with our results due to the inequality decreasing effect of the structure component.

1.6 Conclusions

I estimate the effects of education on income inequality in 18 Latin American countries. Using standardized household income surveys from 2000 to 2010, and applying the FFL decomposition method I decompose the change in income inequality into structure and composition effects across the income distribution. I focus on the impact of education on income inequality: I find that education decreases inequality in Chile, Dominican Republic, Nicaragua and Peru, while for Colombia, Costa Rica, Ecuador, Mexico, El Salvador and Uruguay, it increases inequality. In addition to that, if we consider only the impact of changes in the distribution of years of education (keeping the returns constant), their effect on inequality in all the countries has been increasing, a finding that is consistent with the so called "Paradox of Progress".

I further show that the returns to education decline between 2000 and 2010, specially a the top end of the income distribution for most of the countries. The association between the fall in the returns to education and the estimated effects on inequality is significant and negative. This imply that the inequality-decreasing effect of education is correlated with the reduction in the returns to education at the top of the income distribution, relative to the bottom.

All in all, the results suggest that the paradox of progress is a generalized phenomena in Latin America. However, the structure effect also plays a significant role

in the effects of education on inequality. The countries where the total effect of education is inequality-decreasing experience an inequality-increasing effect from the changes in years of education, but at the same time an inequality-decreasing effect from the changes in their returns. The changes in the returns to education more than compensate the inequality-increasing effect or phenomenon of the "Paradox of Progress". As a result, the total effect of education contributes to the reduction in income inequality in Chile, Dominican Republic, Nicaragua and Peru.

Therefore, policies that consider the expansion of education have to take into account their distributional effects. The changes in returns to education play an important role in the effects of education on inequality levels, as well as the particular circumstances of each country.

References

Alvaredo, F. and Gasparini, L. (2013). Recent trends in inequality and poverty in developing countries. *Documentos de Trabajo del CEDLAS*.

Andersen, T. M. (2018). Human capital, inequality, and growth. *The Welfare State Revisited*, pages 109–122.

Azevedo, J. P., Dávalos, M. E., Diaz-Bonilla, C., Atuesta, B., and Castañeda, R. A. (2013a). *Fifteen years of inequality in Latin America: how have labor markets helped?* The World Bank.

Azevedo, J. P., Inchauste, G., and Sanfelice, V. (2013b). Decomposing the recent inequality decline in latin america. Policy Research working paper WPS 6715, The World Bank Group.

Barros, R. P. d., Carvalho, M. d., Franco, S., and Mendonça, R. (2006). Uma análise das principais causas da queda recente na desigualdade de renda brasileira.

Battistón, D., García-Domench, C., and Gasparini, L. (2014). Could an increase in

education raise income inequality?: Evidence for latin america. *Latin american journal of economics*, 51(1):1–39.

Blinder, A. S. (1973a). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455.

Blinder, A. S. (1973b). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Bouillon, C., Legovini, A., and Lustig, N. (2005). Can education explain changes in income inequality in mexico? *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*.

Bourguignon, F., Ferreira, F. H., and Lustig, N. (2005). *The microeconomics of income distribution dynamics in East Asia and Latin America*. World Bank Publications.

Bussolo, M., Maliszewska, M., and Murard, E. (2014). The long-awaited rise of the middle class in latin america is finally happening. (WPS 6912).

Campos-Vázquez, R., Esquivel, G., and Lustig, N. (2014). The rise and fall of income inequality in mexico, 1989-2010. *Falling Inequality in Latin America: Policy Changes and Lessons*, page 140.

Canavire-Bacarreza, G. and Rios-Avila, F. (2015). On the determinants of changes in wage inequality in bolivia. (835).

Cornia, G. (2012). Inequality trends and their determinants: Latin america over

1990-2011. Working Paper Series UNU-WIDER Research Paper WP2012/09, World Institute for Development Economic Research (UNU-WIDER).

Cruces, G., Domenech, C., and Gasparini, L. (2014). Inequality in education: evidence for latin america. *Falling inequality in Latin America. Policy changes and lessons*, pages 318–339.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, pages 1001–1044.

Eika, L., Mogstad, M., and Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6):2795–2835.

Ferreira, F. H., Firpo, S., and Messina, J. (2014). A more level playing field? explaining the decline in earnings inequality in brazil, 1995-2012.

Firpo, S., Fortin, N., and Lemieux, T. (2007). Decomposing wage distributions using recentered influence function regressions" unpublished manuscript.

Fortin, N., Lemieux, T., and Firpo, S. (2011a). Decomposition methods in economics. *Handbook of labor economics*, 4:1–102.

Fortin, N., Lemieux, T., and Firpo, S. (2011b). Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier.

Gasparini, L., Cruces, G., Tornarolli, L., and Mejía, D. (2011a). Recent trends in income inequality in latin america [with comments]. *Economia*, 11(2):147–201.

Gasparini, L., Galiani, S., Cruces, G., and Acosta, P. (2011b). Educational upgrading and returns to skills in latin america: Evidence form a supply-demand framework 1990-2010. (127).

Gasparini, L., Galiani, S., Cruces, G., and Acosta, P. (2011c). Educational upgrading and returns to skills in latin america: evidence from a supply-demand framework, 1990-2010. *World Bank Policy Research Working Paper*, (5921).

Gasparini, L., Marchionni, M., and Escudero, W. S. (2005). Characterization of inequality changes through microeconometric decompositions: the case of greater buenos aires. *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, pages 47–81.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2014). Marry your like: Assortative mating and income inequality. *American Economic Review*, 104(5):348–53.

Juhn, C., Murphy, K. M., and Pierce, B. (1993a). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–42.

Juhn, C., Murphy, K. M., and Pierce, B. (1993b). Wage inequality and the rise in returns to skill. *Journal of political Economy*, 101(3):410–442.

Kim, J. (2016). Female education and its impact on fertility. *IZA World of Labor*.

López-Calva, L. F. and Lustig, N. (2010). *Declining Inequality in Latin America: A decade of progress?* Brookings Institution Press.

Lustig, N. and Lopez-Calva, Luis F.and Ortiz-Juarez, E. (2013a). Declining inequality in latin america in the 2000s: The cases of argentina, brazil, and mexico. *World Development*, 44:129 – 141.

Lustig, N. and Lopez-Calva, Luis F.and Ortiz-Juarez, E. (2013b). Deconstructing the decline in inequality in latin america. (WPS 6552).

Machado, J. A. and Mata, J. (2005a). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

Machado, J. A. and Mata, J. (2005b). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

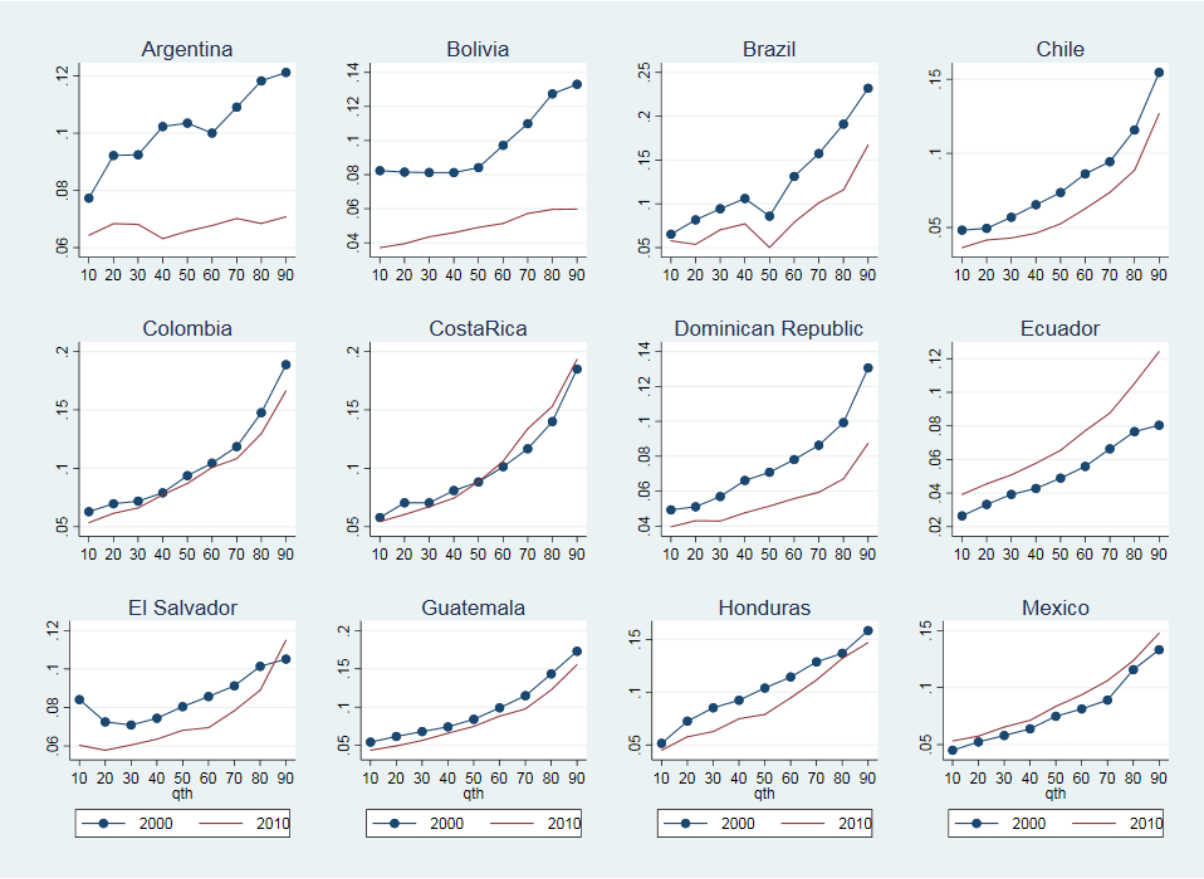
Oaxaca, R. (1973a). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Oaxaca, R. (1973b). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Rodriguez-Castelan, C., Lopez-Calva, L. F., Lustig, N., and Valderrama, D. (2016). Understanding the dynamics of labor income inequality in latin america. (WPS 7795).

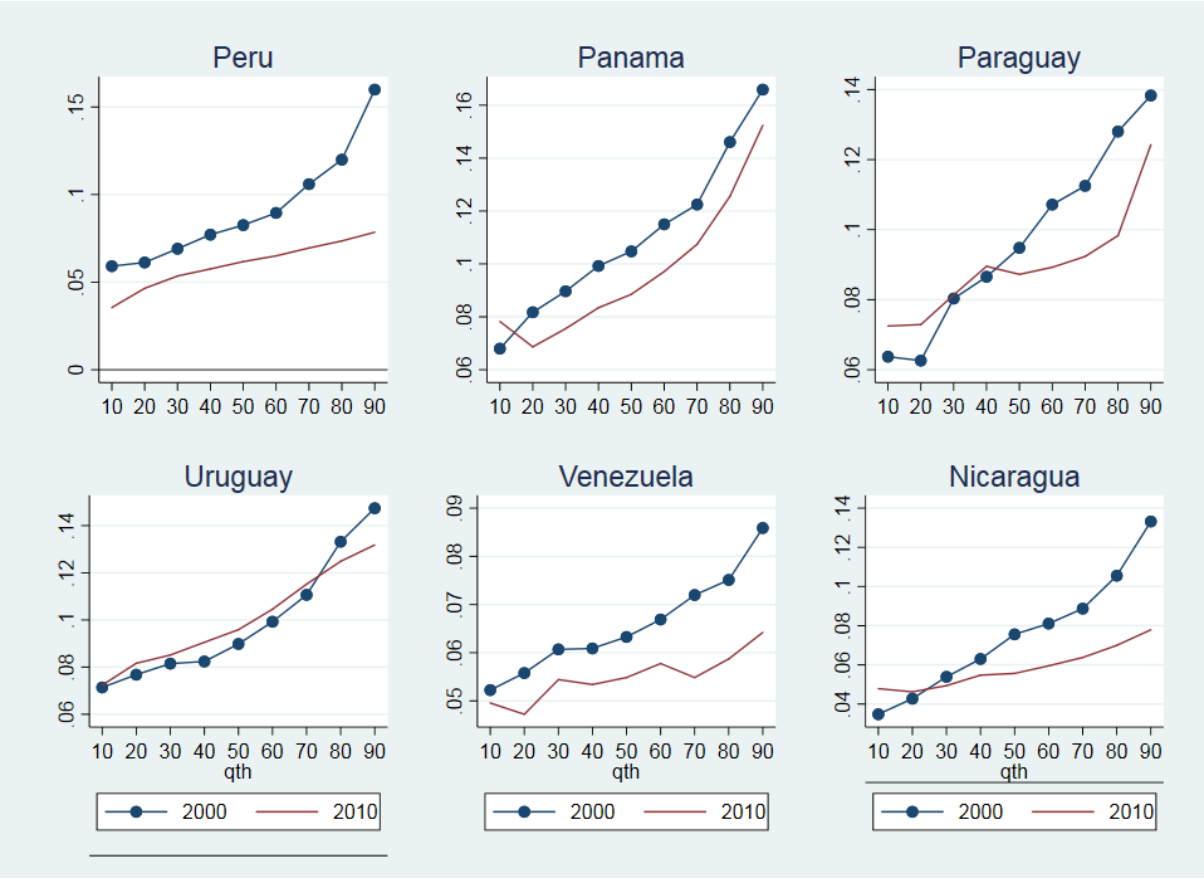
Vélez, C. E., Leibovich, J., Kugler, A. D., Bouillon, C. P., and Núñez, J. (2004). The reversal of inequality trends in colombia, 1978-95: A combination of persistent and fluctuating forces.

1.7 Figures



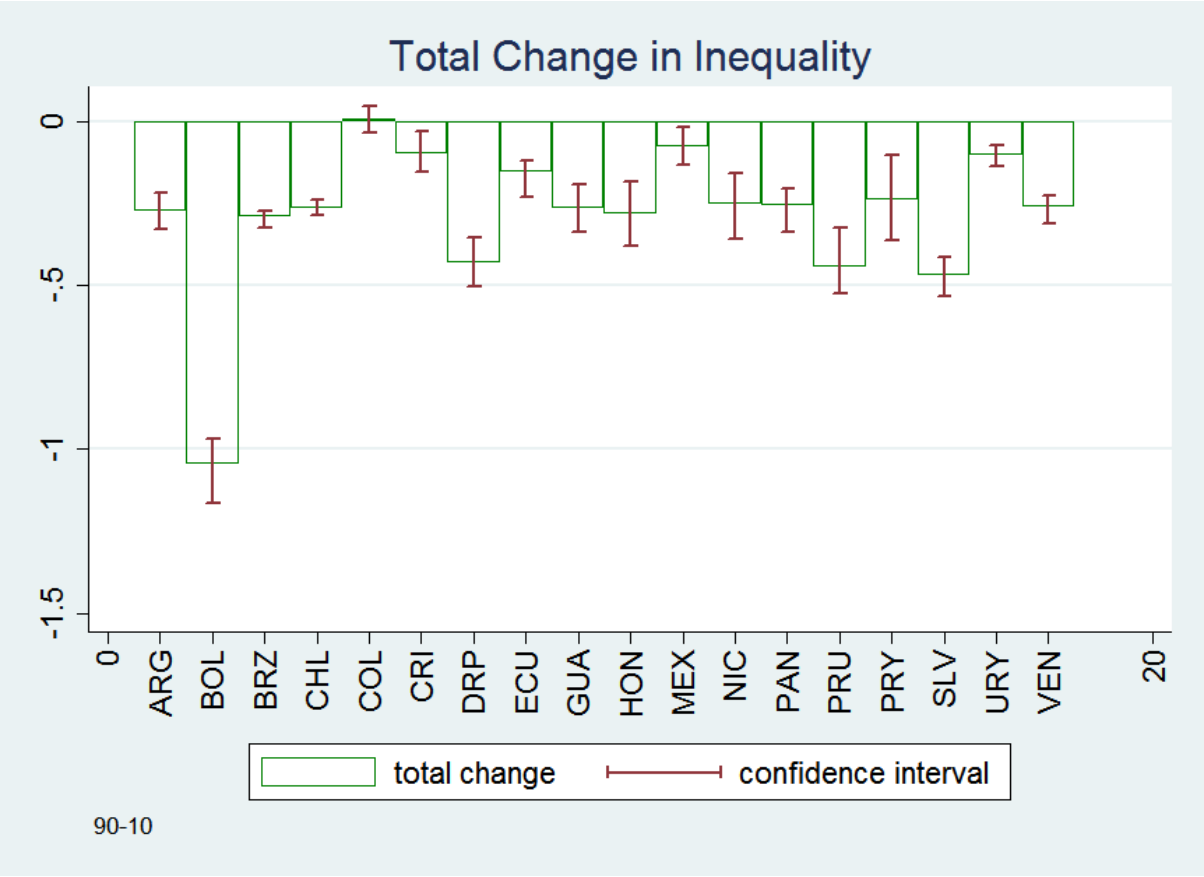
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.1: Unconditional Quantile Regression: Education cont.



Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

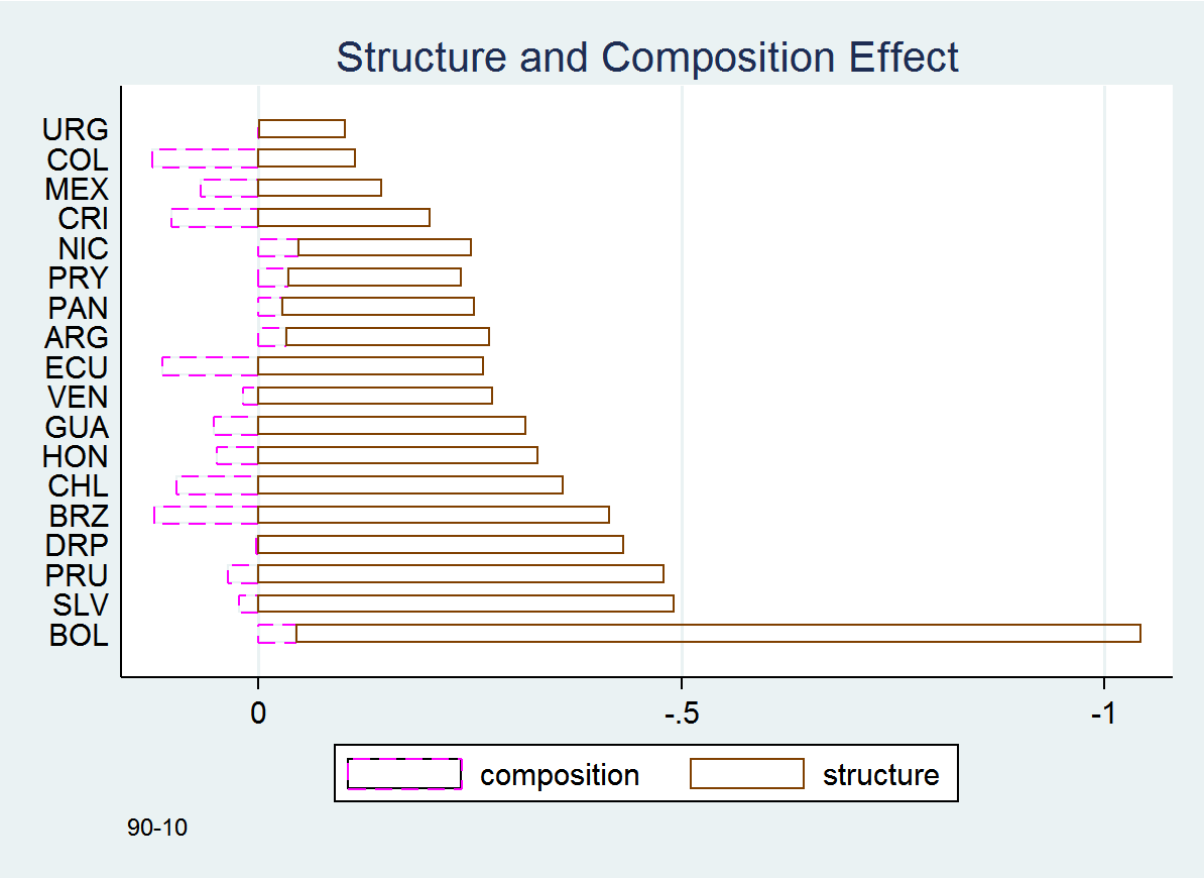
Figure 1.2: Unconditional Quantile Regression: Education



Notes: Bootstrapped 95% confidence intervals with 300 replications.

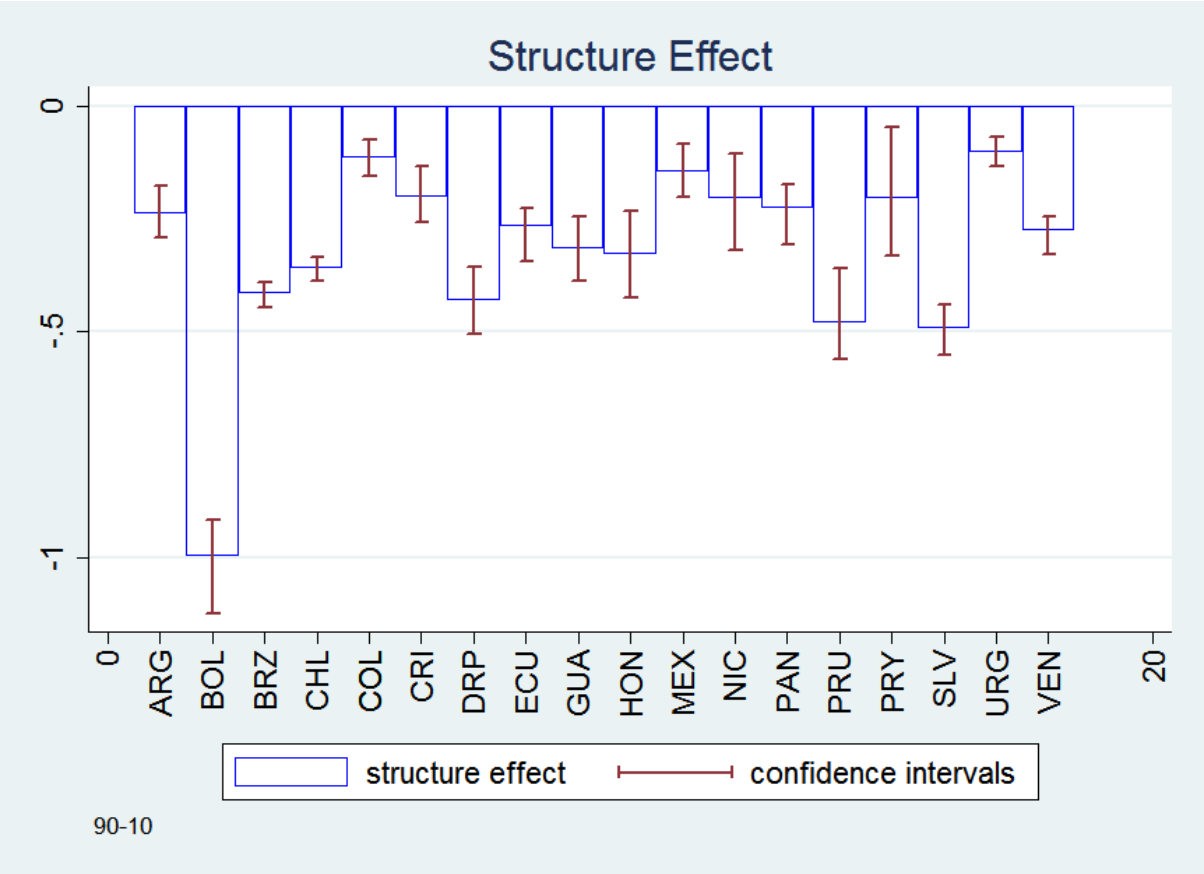
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.3: Total change in inequality



Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

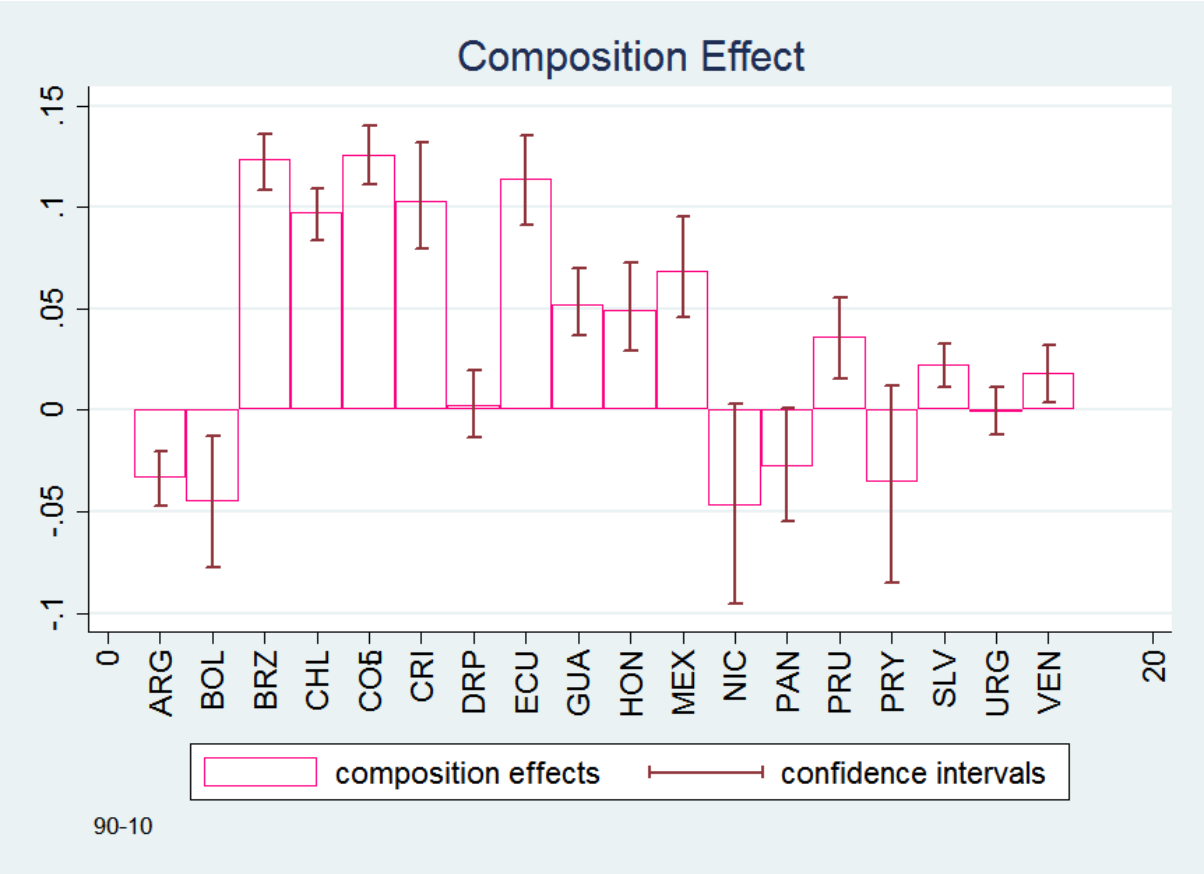
Figure 1.4: Structure effect and Composition effect



Notes: Bootstrapped 95% confidence intervals with 300 replications.

Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

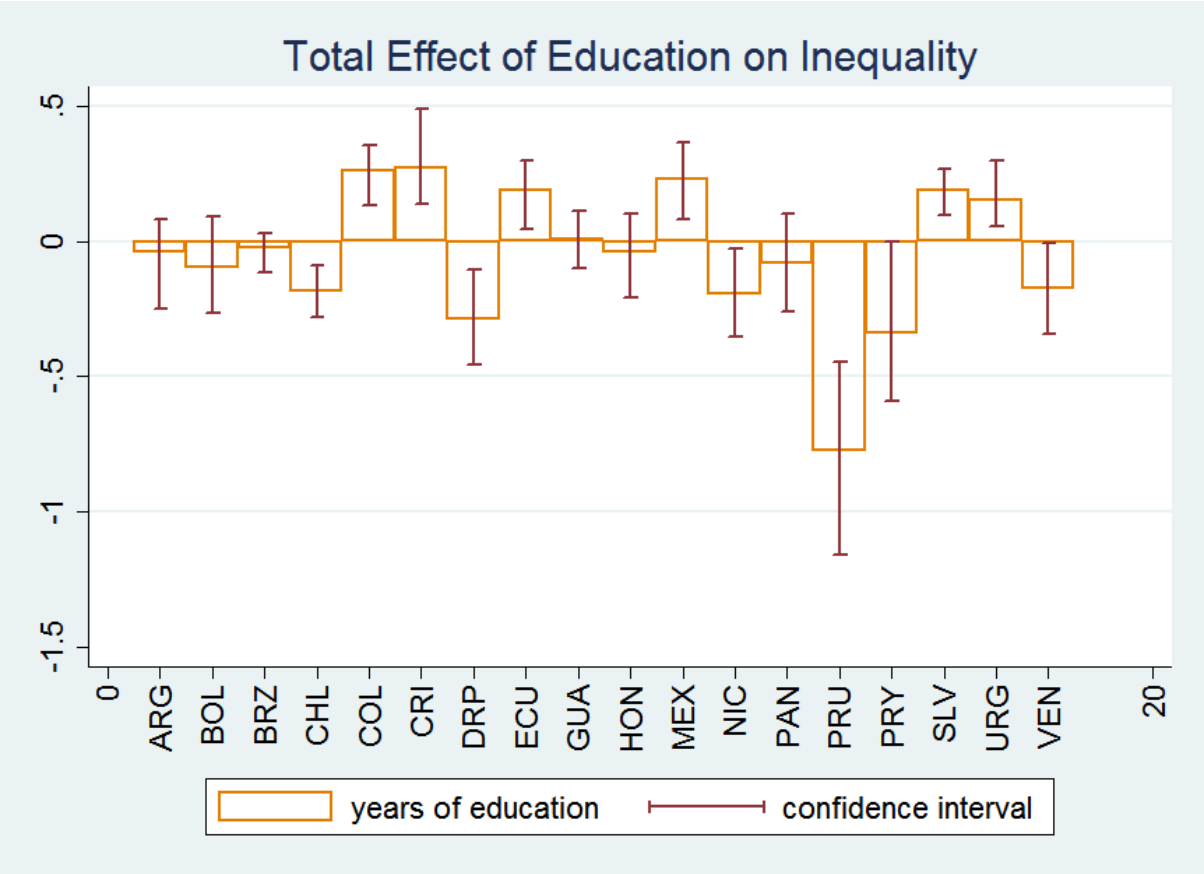
Figure 1.5: Structure Effect



Notes: Bootstrapped 95% confidence intervals with 300 replications.

Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

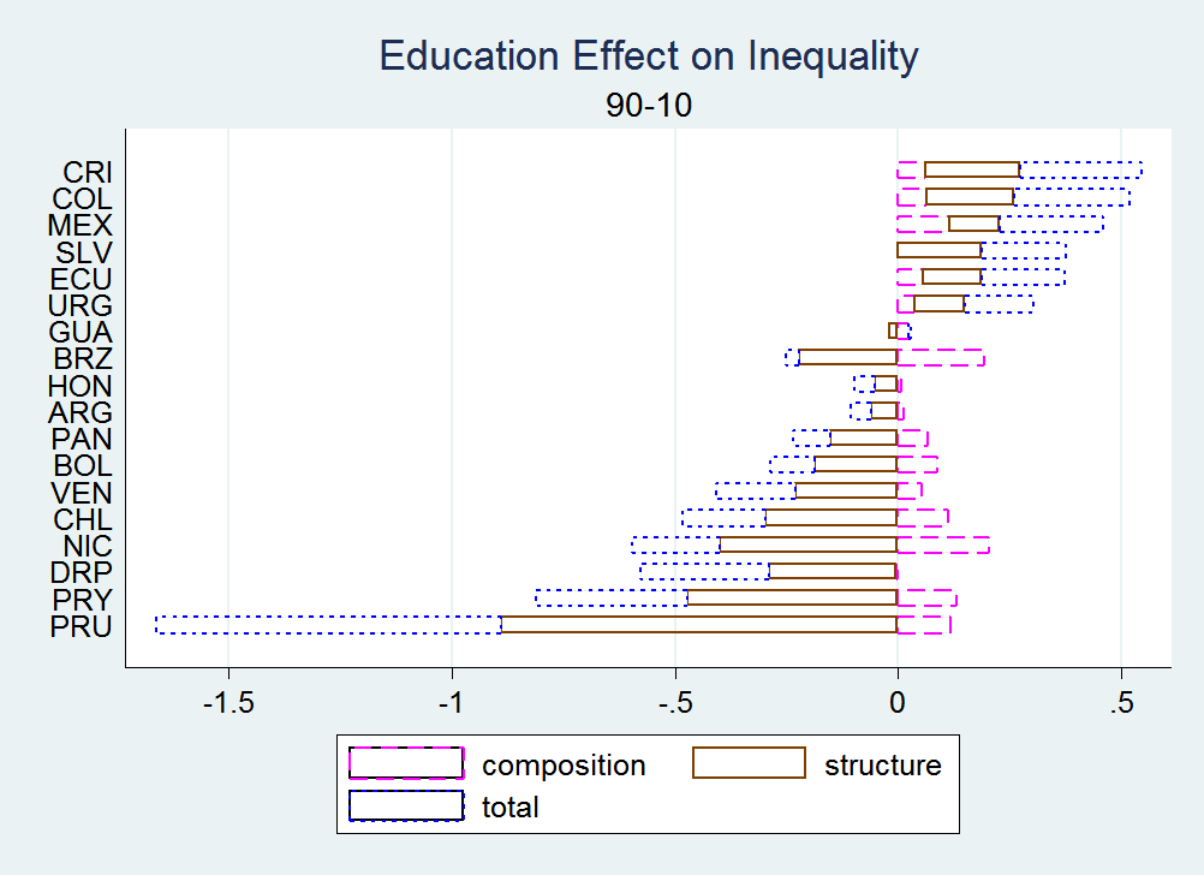
Figure 1.6: Composition Effect



Notes: Bootstrapped 95% confidence intervals with 300 replications.

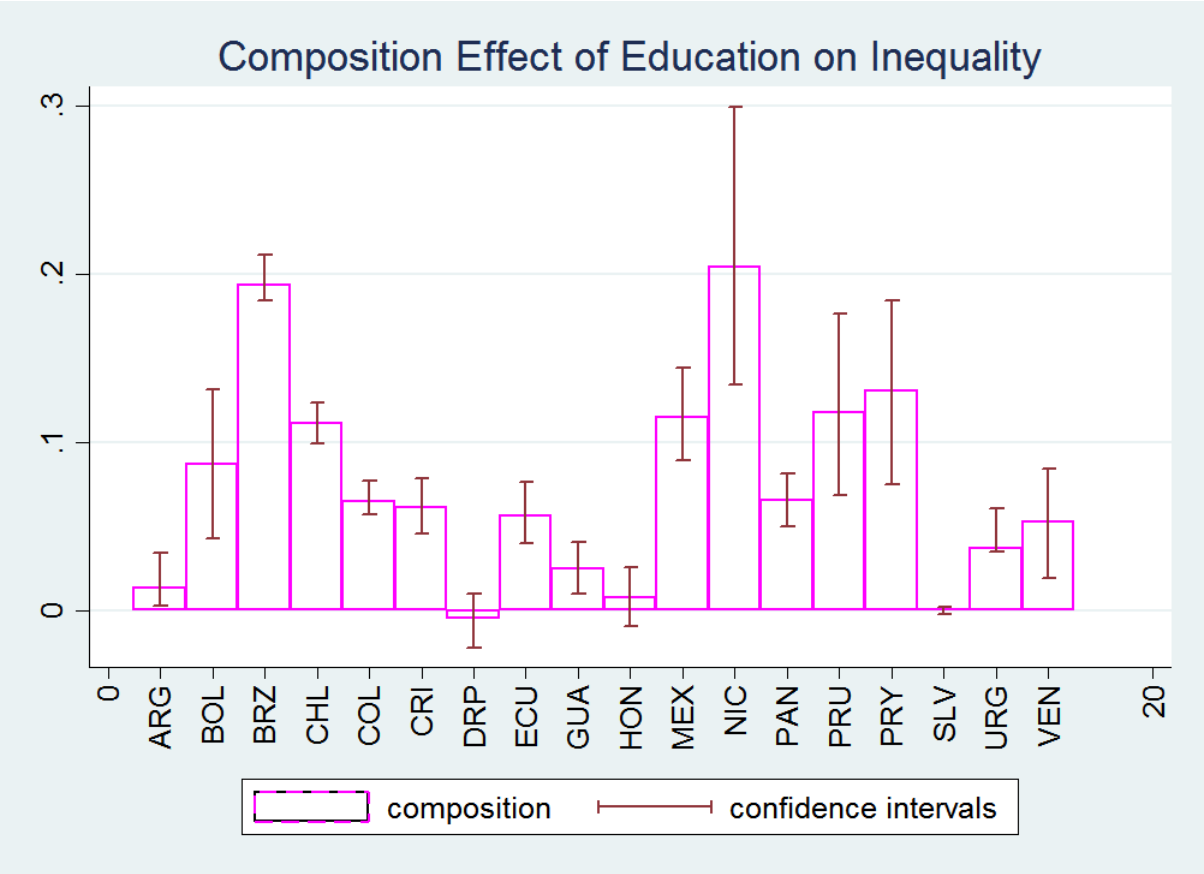
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.7: Total effect of education on inequality



Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

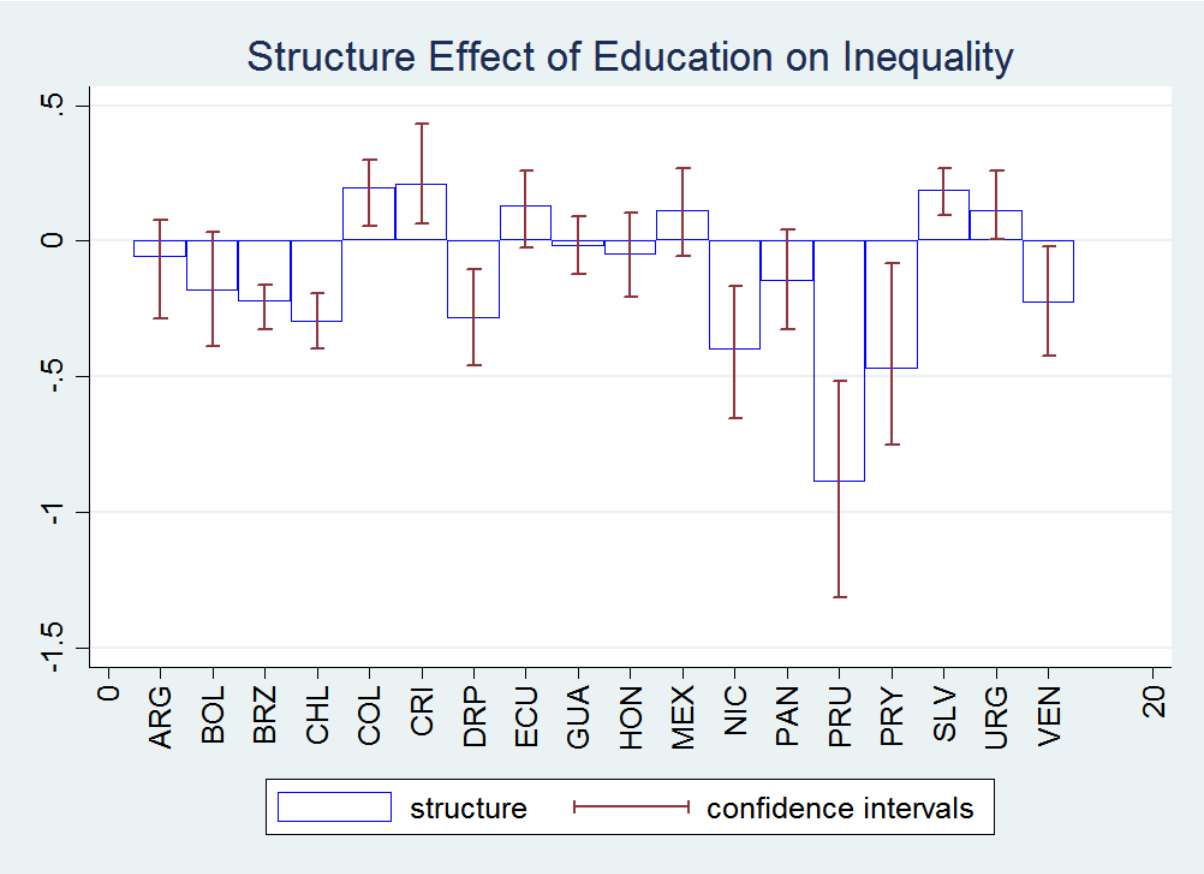
Figure 1.8: Education: structure and composition effect



Notes: Bootstrapped 95% confidence intervals with 300 replications.

Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

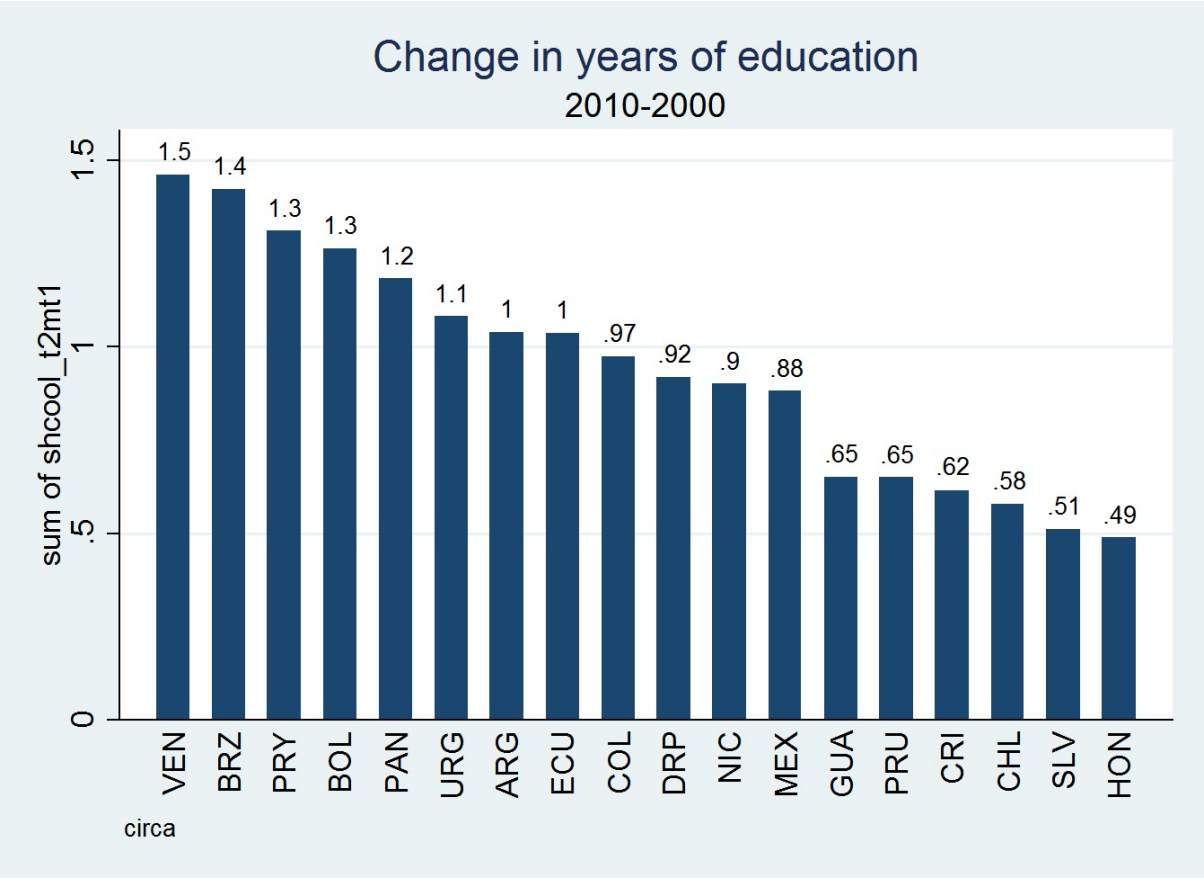
Figure 1.9: Education: Composition Effect



Notes: Bootstrapped 95% confidence intervals with 300 replications.

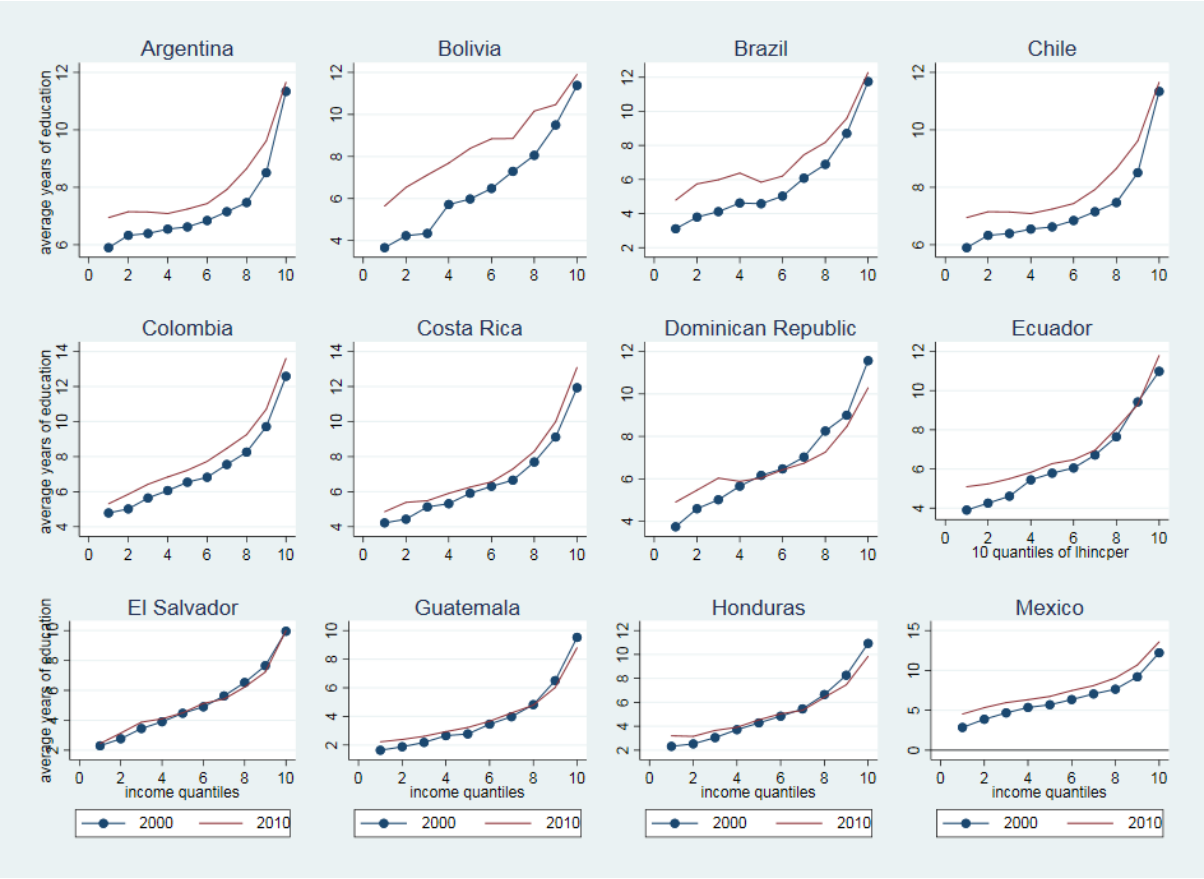
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.10: Education: Structure Effect



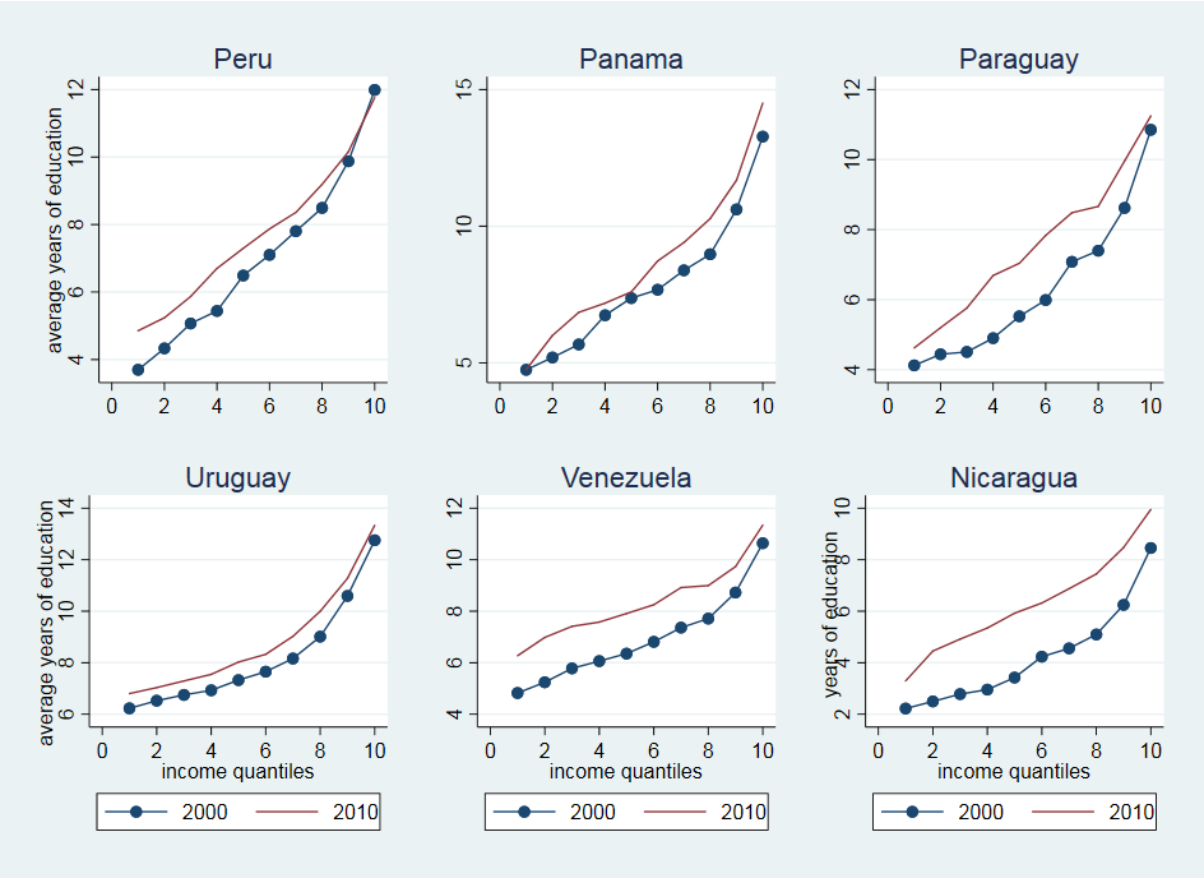
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.11: Average Change in Years of Education



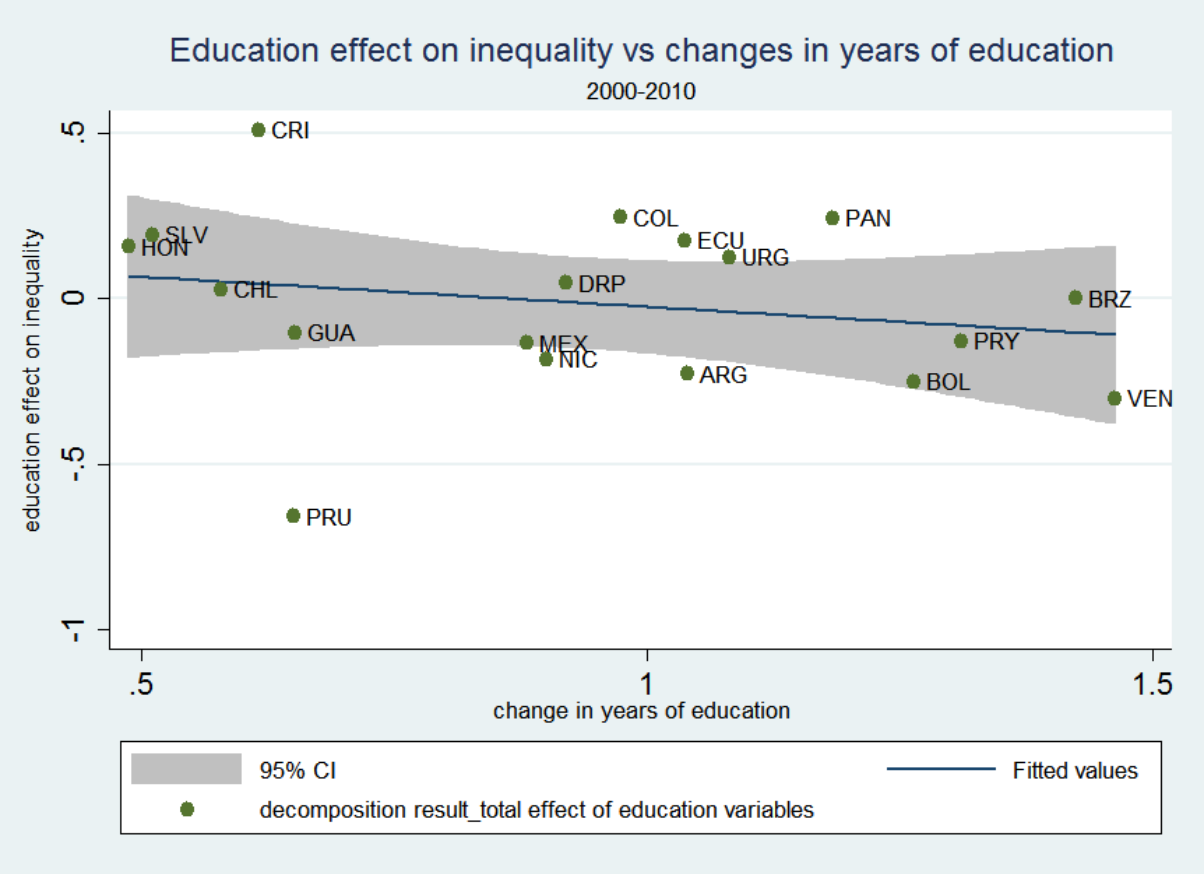
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.12: Distribution of years of Education 2000 vs 2010 cont.



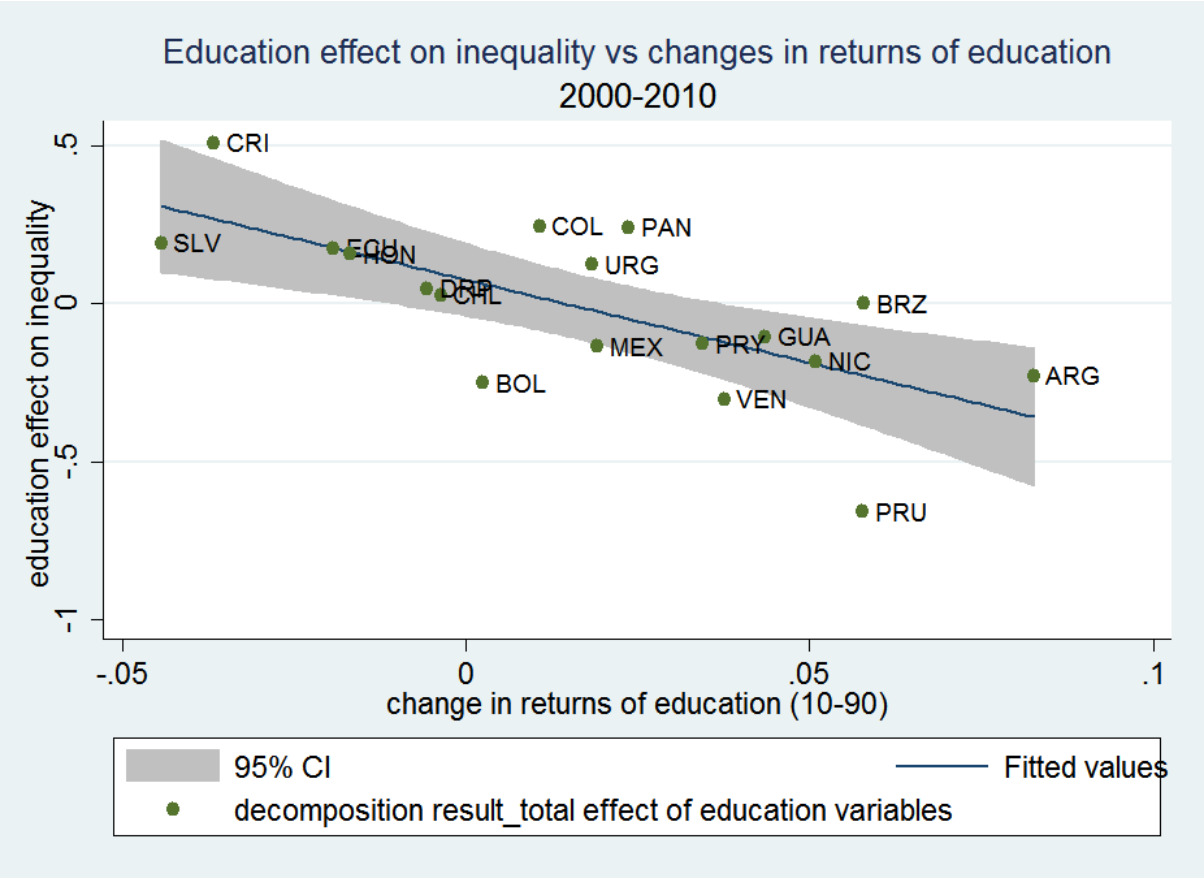
Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.13: Distribution of years of Education 2000 vs 2010



Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.14: Education effect on inequality vs changes in years of education



Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Figure 1.15: Education effect on inequality vs changes in returns of education

1.8 Tables

code	country	Circa 2000	Circa 2010	Coverage	Name of survey	Acronym
ARG	Argentina	2002	2012	urban	Encuesta Permanente de Hogares-Continua	EPH-C
BOL	Bolivia	2001	2011	national	Encuesta de Hogares	EH
BRZ	Brazil	2001	2011	national	Pesquisa Nacional por Amostra de Domicilios	PNAD
CHL	Chile	2000	2009	national	Encuesta de Caracterizacion Socioeconomica Nacional	CASEN
COL	Colombia	2001	2011	national	Gran Encuesta Integrada de Hogares	GEIH
CRI	Costa Rica	2001	2009	national	Encuesta de Hogares de Propositos Multiples	EHPM
DRP	Dominican Republic	2000	2011	national	Encuesta Nacional de Fuerza de Trabajo	ENFT
ECU	Ecuador	2000	2010	national	Encuesta de Empleo, Desempleo y Subempleo	ENEMDU
SLV	El Salvador	2002	2010	national	Encuesta de Hogares de Propositos Multiples	EHPM
GUA	Guatemala	2000	2011	national	Encuesta Nacional de Condiciones de Vida	ENCOVI
HON	Honduras	2001	2011	national	Encuesta Permanente de Hogares de Propositos Multiples	EPHPM
MEX	Mexico	2000	2010	national	Encuesta Nacional de Ingreso y Gasto de los Hogares	ENIGH
NIC	Nicaragua	2001	2009	national	Encuesta Nacional sobre Medicion del Nivel de Vida	EMNV
PAN	Panama	2002	2012	national	Encuesta de Hogares	EH
PRY	Paraguay	2002	2011	national	Encuesta Permanente de Hogares	EPH
PRU	Peru	1999	2010	national	Encuesta Nacional de Hogares	ENAHO
URG	Uruguay	2001	2011	urban	Encuesta Continua de Hogares	ECH
VEN	Venezuela	2000	2011	urban	Encuesta de Hogares por Muestreo	EHM

Table 1.1: Household surveys used form SEDLAC harmonization

year	2000								2010							
country	observations	mean log income	mean household size	mean age	% female *hh*	mean years of education *hh*	% no education *hh*	% working *hh*	observations	mean log income	mean household size	mean age	% female *hh*	mean years of education *hh*	% no education *hh*	% working *hh*
ARG	21,148	5.1	3.7	48.7	29.4	9.3	2.0	63.3	34,938	7.5	3.2	50.4	36.7	10.4	1.3	66.1
BOL	5,845	5.2	4.3	44.2	20.1	6.7	14.0	89.5	8,851	6.7	3.8	46.4	23.6	8.6	8.8	86.9
BRZ	112,594	5.2	3.4	44.4	28.5	5.9	17.7	73.9	117,796	6.3	3.0	46.8	38.3	7.3	15.5	69.8
CHL	65,036	11.2	3.9	50.3	21.4	7.3	7.8	68.9	71,460	11.8	3.4	53.7	30.2	8.1	0.0	63.2
COL	32,104	11.7	4.1	46.7	30.0	7.4	6.9	74.0	57,921	12.7	3.5	47.8	37.6	8.2	6.5	73.7
CRI	10,332	10.5	4.0	46.1	24.0	6.7	8.5	77.1	13,244	11.6	3.6	47.6	29.6	7.3	6.3	73.8
DRP	5,696	7.6	4.0	46.8	28.2	6.7	14.4	73.9	8,191	8.6	3.6	49.0	31.8	6.8	14.0	71.2
ECU	13,963	3.4	4.4	47.9	19.1	6.5	11.2	83.5	20,676	4.7	4.0	52.9	25.2	7.1	10.6	79.4
GUA	7,276	6.0	5.2	44.5	18.2	4.0	34.8	91.0	13,482	6.5	4.9	45.4	20.5	4.1	30.7	83.6
HON	7,157	6.6	4.9	46.3	26.7	5.2	22.7	80.7	6,867	7.4	4.6	48.5	33.0	5.3	19.1	77.9
MEX	10,108	7.0	4.2	45.3	18.7	6.5	13.8	81.2	27,655	7.6	3.9	48.5	24.8	7.8	10.0	77.7
NIC	4,191	6.2	5.4	46.6	27.4	4.2	31.6	79.1	6,515	7.3	4.7	46.9	37.3	6.3	18.7	76.5
PAN	13,308	4.6	4.1	47.5	23.3	7.9	8.8	75.2	12,298	5.4	3.7	49.8	29.9	8.7	7.4	76.5
PRU	3,517	5.1	5.2	48.3	19.6	7.1	12.7	84.2	21,496	5.9	4.3	50.5	23.4	7.8	9.9	87.9
PRY	3,789	12.4	4.6	46.1	25.4	6.4	5.8	79.3	4,894	13.4	4.0	49.0	31.2	7.5	3.8	76.7
SLV	16,479	6.4	4.4	48.6	32.9	5.2	26.1	71.7	21,166	4.5	4.0	47.9	34.1	5.2	21.1	73.2
URG	18,461	8.3	3.1	54.0	32.0	8.2	1.9	60.8	46,669	9.2	2.8	53.7	40.2	8.9	1.2	66.6
VEN	16,809	11.1	4.8	47.5	30.0	7.0	13.9	73.6	37,217	6.9	4.1	49.1	38.2	8.3	8.1	72.6

Source: Author's calculation based on SEDLAC (CEDLAS and World Bank)homogenized household income surveys.

Table 1.2: Summary statistics

1.9 Appendix

Aggregate Decomposition Results

Country	Value in 2000 (circa)		Value in 2010 (circa)		Total Change		Structure Effect		Composition Effect	
Argentina	2.348		2.074		-0.275		-0.241		-0.033	
	2.296	2.392	2.052	2.101	-0.327	-0.217	-0.293	-0.177	-0.047	-0.020
Bolivia	3.185		2.141		-1.044		-0.999		-0.045	
	3.110	3.291	2.092	2.173	-1.166	-0.968	-1.124	-0.917	-0.078	-0.013
Brazil	2.577		2.285		-0.292		-0.415		0.123	
	2.560	2.596	2.263	2.296	-0.324	-0.272	-0.448	-0.392	0.109	0.136
Chile	2.222		1.958		-0.264		-0.361		0.097	
	2.205	2.239	1.943	1.972	-0.288	-0.242	-0.388	-0.335	0.084	0.109
Colombia	2.341		2.351		0.010		-0.116		0.125	
	2.303	2.375	2.326	2.367	-0.035	0.048	-0.158	-0.075	0.111	0.140
Costa Rica	2.290		2.190		-0.100		-0.203		0.103	
	2.241	2.340	2.163	2.242	-0.155	-0.030	-0.258	-0.134	0.079	0.132
Dominican Republic	2.435		2.004		-0.431		-0.433		0.002	
	2.376	2.491	1.957	2.050	-0.504	-0.354	-0.506	-0.358	-0.014	0.020
Ecuador	2.329		2.175		-0.153		-0.267		0.114	
	2.291	2.395	2.140	2.196	-0.230	-0.121	-0.346	-0.228	0.091	0.135
Guatemala	2.469		2.204		-0.265		-0.317		0.052	
	2.408	2.524	2.163	2.240	-0.339	-0.192	-0.389	-0.246	0.037	0.070
Honduras	2.928		2.646		-0.282		-0.331		0.049	
	2.864	2.997	2.576	2.710	-0.379	-0.186	-0.424	-0.233	0.029	0.073
Mexico	2.376		2.298		-0.078		-0.146		0.068	
	2.322	2.419	2.271	2.328	-0.134	-0.018	-0.204	-0.084	0.046	0.095
Nicaragua	2.339		2.087		-0.252		-0.205		-0.047	
	2.255	2.428	2.033	2.141	-0.358	-0.158	-0.321	-0.106	-0.096	0.003
Panama	2.726		2.469		-0.256		-0.228		-0.028	
	2.684	2.780	2.422	2.508	-0.337	-0.206	-0.309	-0.174	-0.055	0.001
Peru	2.674		2.231		-0.444		-0.480		0.036	
	2.565	2.751	2.207	2.258	-0.527	-0.324	-0.562	-0.359	0.016	0.055
Paraguay	2.754		2.513		-0.241		-0.206		-0.035	
	2.651	2.851	2.450	2.580	-0.362	-0.102	-0.334	-0.047	-0.085	0.012
El Salvador	2.623		2.153		-0.470		-0.492		0.022	
	2.583	2.679	2.126	2.185	-0.532	-0.416	-0.553	-0.440	0.011	0.033
Uruguay	2.114		2.010		-0.104		-0.103		-0.001	
	2.090	2.141	1.991	2.026	-0.136	-0.073	-0.135	-0.068	-0.012	0.011
Venezuela	2.083		1.823		-0.260		-0.278		0.018	
	2.061	2.129	1.800	1.845	-0.312	-0.229	-0.329	-0.247	0.004	0.032

Boostrap confidence interval with 300 repetitions

Figure 1.16: Aggregate Decomposition Results

Education Effects on Inequality

Country	Total Effect		Structure Effect		Composition Effect	
Argentina	-0.046		-0.060		0.014	
	-0.248	0.081	-0.286	0.077	0.003	0.034
Bolivia	-0.099		-0.186		0.087	
	-0.266	0.093	-0.391	0.033	0.043	0.131
Brazil	-0.029		-0.223		0.193	
	-0.116	0.026	-0.327	-0.162	0.184	0.211
Chile	-0.186		-0.298		0.111	
	-0.279	-0.089	-0.396	-0.195	0.099	0.123
Colombia	0.259		0.194		0.065	
	0.130	0.355	0.056	0.299	0.057	0.077
Costa Rica	0.273		0.211		0.062	
	0.134	0.486	0.064	0.431	0.046	0.078
Dominican Republic	-0.289		-0.284		-0.005	
	-0.458	-0.106	-0.458	-0.103	-0.022	0.010
Ecuador	0.187		0.131		0.056	
	0.045	0.297	-0.027	0.258	0.040	0.076
Guatemala	0.005		-0.019		0.025	
	-0.099	0.111	-0.121	0.092	0.010	0.041
Honduras	-0.044		-0.052		0.008	
	-0.208	0.100	-0.208	0.103	-0.009	0.026
Mexico	0.230		0.114		0.115	
	0.080	0.365	-0.055	0.268	0.089	0.144
Nicaragua	-0.197		-0.400		0.204	
	-0.355	-0.029	-0.653	-0.167	0.134	0.299
Panama	-0.085		-0.151		0.066	
	-0.258	0.099	-0.328	0.043	0.050	0.081
Peru	-0.773		-0.891		0.118	
	-1.158	-0.443	-1.313	-0.516	0.068	0.176
Paraguay	-0.341		-0.471		0.130	
	-0.588	0.000	-0.754	-0.082	0.075	0.184
El Salvador	0.188		0.187		0.000	
	0.096	0.267	0.096	0.267	-0.002	0.002
Uruguay	0.151		0.114		0.038	
	0.056	0.295	0.005	0.257	0.035	0.060
Venezuela	-0.178		-0.231		0.053	
	-0.340	-0.007	-0.424	-0.022	0.019	0.085

Boostrap confidence interval with 300 repetitions

Figure 1.17: Education Effects on Inequality

Decomposition Changes in Inequality from 2000 to 2010: 90-10 differential (cont.)

90-10 quantile difference	Argentina	Bolivia	Brazil	Chile	Colombia	Costa Rica
Value in 2000 (circa)	2.348	3.185	2.577	2.222	2.341	2.290
	2.296 2.392	3.110 3.291	2.560 2.596	2.205 2.239	2.303 2.375	2.241 2.340
Value in 2010 (circa)	2.074	2.141	2.285	1.958	2.351	2.190
	2.052 2.101	2.092 2.173	2.263 2.296	1.943 1.972	2.326 2.367	2.163 2.242
Total Change	**-0.275**	**-1.044**	**-0.292**	**-0.264**	**0.010**	**-0.100**
	-0.327 -0.217	-1.166 -0.968	-0.324 -0.272	-0.288 -0.242	-0.035 0.048	-0.155 -0.030
Aggregate decomposition of Total Change						
Structure Effect	-0.241	-0.999	-0.415	-0.361	-0.116	-0.203
	-0.293 -0.177	-1.124 -0.917	-0.448 -0.392	-0.388 -0.335	-0.158 -0.075	-0.258 -0.134
Composition Effect	-0.033	-0.045	0.123	0.097	0.125	0.103
	-0.047 -0.020	-0.078 -0.013	0.109 0.136	0.084 0.109	0.111 0.140	0.079 0.132
Total Effect						
urban		0.425	0.066	0.170	0.052	0.072
		0.254 0.596	-0.012 0.120	0.126 0.201	-0.106 0.211	0.002 0.131
size	-0.206	0.100	-0.016	-0.118	-0.087	0.016
	-0.318 -0.080	-0.143 0.311	-0.079 0.083	-0.175 -0.036	-0.187 0.026	-0.120 0.189
age	-0.309	-0.351	-0.352	-0.289	0.125	-0.608
	-0.548 -0.144	-0.629 0.003	-0.475 -0.245	-0.441 -0.136	-0.048 0.281	-0.828 -0.302
female	0.027	0.073	0.023	0.008	0.026	-0.048
	-0.002 0.053	0.022 0.122	0.002 0.042	-0.009 0.022	-0.002 0.052	-0.092 -0.008
years of schooling	-0.046	-0.099	-0.029	-0.186	0.259	0.273
	-0.248 0.081	-0.266 0.093	-0.116 0.026	-0.279 -0.089	0.130 0.355	0.134 0.486
work	-0.049	-0.544	-0.170	-0.041	-0.118	-0.128
	-0.129 0.018	-0.870 -0.232	-0.226 -0.123	-0.105 0.019	-0.197 -0.042	-0.297 0.022
constant	0.251	-0.716	0.134	0.201	-0.259	0.319
	-0.039 0.698	-1.386 -0.100	-0.055 0.330	-0.034 0.442	-0.519 0.035	-0.176 0.697
residual	0.057	0.069	0.053	-0.009	0.012	0.004
Detailed Decomposition of the Composition Effect						
urban		-0.261	0.000	-0.020	0.006	0.011
		-0.317 -0.211	-0.002 0.002	-0.023 -0.017	0.001 0.009	0.004 0.016
size	-0.053	0.011	-0.043	0.014	-0.007	0.010
	-0.064 -0.039	-0.012 0.036	-0.047 -0.035	0.009 0.021	-0.019 0.005	-0.005 0.024
age	-0.011	0.018	0.042	0.018	0.015	0.025
	-0.015 -0.003	0.002 0.033	0.037 0.048	0.010 0.027	0.011 0.020	0.016 0.035
female	0.013	-0.012	-0.002	-0.002	0.009	0.009
	0.007 0.019	-0.023 -0.002	-0.007 0.003	-0.006 0.004	0.003 0.016	0.002 0.018
years of schooling	0.014	0.087	0.193	0.111	0.065	0.062
	0.003 0.034	0.043 0.131	0.184 0.211	0.099 0.123	0.057 0.077	0.046 0.078
work	0.000	-0.013	-0.010	0.002	0.000	0.004
	-0.002 0.003	-0.025 -0.005	-0.013 -0.008	-0.002 0.006	-0.001 0.000	-0.003 0.011
residual	0.004	0.124	-0.057	-0.027	0.038	-0.017
Detailed Decomposition of the Structure Effect						
urban		0.685	0.066	0.191	0.047	0.061
		0.479 0.897	-0.011 0.122	0.144 0.222	-0.111 0.204	-0.003 0.120
size	-0.153	0.089	0.027	-0.132	-0.080	0.006
	-0.257 -0.041	-0.131 0.278	-0.033 0.119	-0.187 -0.055	-0.164 0.021	-0.127 0.168
age	-0.298	-0.369	-0.395	-0.308	0.110	-0.632
	-0.544 -0.128	-0.662 0.000	-0.522 -0.284	-0.466 -0.150	-0.065 0.268	-0.852 -0.323
female	0.015	0.085	0.025	0.009	0.016	-0.058
	-0.021 0.046	0.025 0.141	0.000 0.049	-0.012 0.027	-0.018 0.048	-0.109 -0.012
years of schooling	-0.060	-0.186	-0.223	-0.298	0.194	0.211
	-0.286 0.077	-0.391 0.033	-0.327 -0.162	-0.396 -0.195	0.056 0.299	0.064 0.431
work	-0.049	-0.530	-0.160	-0.042	-0.118	-0.132
	-0.132 0.018	-0.844 -0.225	-0.213 -0.114	-0.103 0.013	-0.196 -0.043	-0.295 0.015
constant	0.251	-0.716	0.134	0.201	-0.259	0.319
	-0.039 0.698	-1.386 -0.100	-0.055 0.330	-0.034 0.442	-0.519 0.035	-0.176 0.697
residual	0.053	-0.056	0.109	0.017	-0.026	0.020

Boostrap confidence interval with 300 repetitions

Figure 1.18: Decomposition Results cont.

Decomposition Changes in Inequality from 2000 to 2010: 90-10 differential (cont.)

90-10 quantile difference	Dominican Republic	Ecuador	Guatemala	Honduras	Mexico	Nicaragua
Value in 2000 (circa)	2.435	2.329	2.469	2.928	2.376	2.339
	2.376 2.491	2.291 2.395	2.408 2.524	2.864 2.997	2.322 2.419	2.255 2.428
Value in 2010 (circa)	2.004	2.175	2.204	2.646	2.298	2.087
	1.957 2.050	2.140 2.196	2.163 2.240	2.576 2.710	2.271 2.328	2.033 2.141
Total Change (90-10)	**-0.431**	**-0.153**	**-0.265**	**-0.282**	**-0.078**	**-0.252**
	-0.504 -0.354	-0.230 -0.121	-0.339 -0.192	-0.379 -0.186	-0.134 -0.018	-0.358 -0.158
Aggregate decomposition of Total Change						
Structure Effect	-0.433	-0.267	-0.317	-0.331	-0.146	-0.205
	-0.506 -0.358	-0.346 -0.228	-0.389 -0.246	-0.424 -0.233	-0.204 -0.084	-0.321 -0.106
Composition Effect	0.002	0.114	0.052	0.049	0.068	-0.047
	-0.014 0.020	0.091 0.135	0.037 0.070	0.029 0.073	0.046 0.095	-0.096 0.003
Total Effect						
urban	0.174	0.183	0.017	0.218	-0.013	-0.199
	0.066 0.287	0.082 0.257	-0.044 0.088	0.110 0.339	-0.146 0.096	-0.358 -0.041
size	0.176	-0.025	0.157	-0.295	-0.102	-0.105
	-0.026 0.360	-0.156 0.103	-0.009 0.314	-0.505 -0.116	-0.241 0.039	-0.362 0.152
age	-0.359	0.148	0.077	-0.042	-0.082	-0.196
	-0.639 -0.040	-0.077 0.359	-0.185 0.282	-0.405 0.284	-0.292 0.131	-0.574 0.133
female	0.020	0.001	-0.009	-0.021	0.010	0.035
	-0.040 0.074	-0.030 0.030	-0.046 0.033	-0.082 0.043	-0.018 0.044	-0.040 0.098
years of schooling	-0.289	0.187	0.005	-0.044	0.230	-0.197
	-0.458 -0.106	0.045 0.297	-0.099 0.111	-0.208 0.100	0.080 0.365	-0.355 -0.029
work	0.159	0.051	0.128	-0.214	0.078	0.195
	-0.004 0.333	-0.093 0.219	-0.119 0.354	-0.425 0.019	-0.077 0.233	-0.043 0.398
constant	-0.327	-0.701	-0.624	0.118	-0.194	0.164
	-0.852 0.212	-1.045 -0.262	-1.078 -0.162	-0.509 0.663	-0.604 0.260	-0.400 0.784
residual	0.015	0.003	-0.016	-0.002	-0.004	0.052
Detailed Decomposition of the Composition Effect						
urban	0.018	0.024	0.006	0.030	-0.121	-0.097
	0.009 0.028	0.014 0.032	0.001 0.012	0.013 0.044	-0.139 -0.100	-0.132 -0.061
size	0.011	0.002	0.010	-0.011	0.001	0.000
	-0.003 0.024	-0.013 0.014	0.003 0.019	-0.023 -0.001	-0.007 0.010	-0.031 0.032
age	0.019	0.047	0.007	0.036	0.045	0.003
	0.007 0.031	0.026 0.070	0.003 0.014	0.023 0.053	0.031 0.061	
female	-0.003	-0.001	0.003	-0.009	0.002	-0.025
	-0.011 0.003	-0.009 0.011	-0.002 0.007	-0.020 0.001	-0.008 0.009	-0.046 -0.002
years of schooling	-0.005	0.056	0.025	0.008	0.115	0.204
	-0.022 0.010	0.040 0.076	0.010 0.041	-0.009 0.026	0.089 0.144	0.134 0.299
work	0.008	0.001	0.026	-0.013	0.007	0.004
	0.002 0.016	-0.007 0.008	0.006 0.045	-0.024 -0.006	0.001 0.014	-0.001 0.012
residual	-0.047	-0.016	-0.024	0.009	0.020	-0.136
Detailed Decomposition of the Structure Effect						
urban	0.156	0.158	0.011	0.188	0.108	-0.102
	0.055 0.262	0.066 0.225	-0.047 0.078	0.081 0.307	-0.035 0.225	-0.291 0.081
size	0.165	-0.027	0.146	-0.284	-0.102	-0.105
	-0.022 0.342	-0.143 0.089	-0.014 0.300	-0.484 -0.113	-0.234 0.032	-0.328 0.108
age	-0.378	0.101	0.070	-0.078	-0.127	-0.199
	-0.665 -0.050	-0.147 0.326	-0.196 0.278	-0.459 0.263	-0.349 0.098	-0.577 0.133
female	0.023	0.002	-0.012	-0.012	0.009	0.060
	-0.043 0.084	-0.038 0.037	-0.053 0.035	-0.078 0.059	-0.027 0.050	-0.038 0.145
years of schooling	-0.284	0.131	-0.019	-0.052	0.114	-0.400
	-0.458 -0.103	-0.027 0.258	-0.121 0.092	-0.208 0.103	-0.055 0.268	-0.653 -0.167
work	0.151	0.050	0.103	-0.200	0.071	0.190
	-0.006 0.324	-0.086 0.212	-0.128 0.311	-0.403 0.028	-0.080 0.223	-0.043 0.386
constant	-0.327	-0.701	-0.624	0.118	-0.194	0.164
	-0.852 0.212	-1.045 -0.262	-1.078 -0.162	-0.509 0.663	-0.604 0.260	-0.400 0.784
residual	0.062	0.019	0.008	-0.011	-0.024	0.188

Boostrap confidence interval with 300 repetitions

Figure 1.19: Decomposition Results cont.

Decomposition Changes in Inequality from 2000 to 2010: 90-10 differential

90-10 quantile difference	Panama	Peru	Paraguay	El Salvador	Uruguay	Venezuela
Value in 2000 (circa)	2.726	2.674	2.754	2.623	2.114	2.083
	2.684 2.780	2.565 2.751	2.651 2.851	2.583 2.679	2.090 2.141	2.061 2.129
Value in 2010 (circa)	2.469	2.231	2.513	2.153	2.010	1.823
	2.422 2.508	2.207 2.258	2.450 2.580	2.126 2.185	1.991 2.026	1.800 1.845
Total Change (90-10)	**-0.256**	**-0.444**	**-0.241**	**-0.470**	**-0.104**	**-0.260**
	-0.337 -0.206	-0.527 -0.324	-0.362 -0.102	-0.532 -0.416	-0.136 -0.073	-0.312 -0.229
Aggregate decomposition of Total Change						
Structure Effect	-0.228	-0.480	-0.206	-0.492	-0.103	-0.278
	-0.309 -0.174	-0.562 -0.359	-0.334 -0.047	-0.553 -0.440	-0.135 -0.068	-0.329 -0.247
Composition Effect	-0.028	0.036	-0.035	0.022	-0.001	0.018
	-0.055 0.001	0.016 0.055	-0.085 0.012	0.011 0.033	-0.012 0.011	0.004 0.032
Detailed Decomposition of the Total Effect						
urban	-0.095	0.233	-0.152	0.247		
	-0.169 -0.006	0.056 0.419	-0.312 0.035	0.149 0.348		
size	0.578	0.331	-0.140	-0.184	-0.119	-0.010
	0.445 0.718	-0.014 0.617	-0.397 0.123	-0.332 -0.045	-0.222 -0.034	-0.179 0.210
age	-0.598	-1.042	-0.163	0.276	-0.122	-0.123
	-0.863 -0.341	-1.674 -0.544	-0.609 0.343	0.060 0.479	-0.282 0.073	-0.517 0.249
female	0.046	-0.053	-0.034	0.092	0.022	0.021
	0.005 0.087	-0.132 0.011	-0.118 0.047	0.044 0.142	-0.003 0.050	-0.031 0.077
years of schooling	-0.085	-0.773	-0.341	0.188	0.151	-0.178
	-0.258 0.099	-1.158 -0.443	-0.588 0.000	0.096 0.267	0.056 0.295	-0.340 -0.007
work	0.031	0.285	0.028	-0.035	-0.119	-0.137
	-0.080 0.173	-0.072 0.680	-0.253 0.284	-0.140 0.069	-0.167 -0.038	-0.342 0.046
constant	-0.141	0.643	0.589	-1.074	0.080	0.439
	-0.612 0.261	-0.274 1.812	-0.211 1.318	-1.436 -0.645	-0.264 0.306	-0.211 1.058
residual	0.007	-0.069	-0.029	0.020	0.003	-0.272
Detailed Decomposition of the Composition Effect						
urban	-0.007	0.021	-0.050	0.059		
	-0.014 -0.001	0.003 0.041	-0.078 -0.033	0.046 0.073		
size	-0.009	0.058	-0.045	-0.019	-0.054	-0.002
	-0.016 -0.001	0.000 0.105	-0.078 -0.015	-0.029 -0.010	-0.066 -0.045	-0.022 0.028
age	0.024	0.051	0.038	-0.004	0.001	0.004
	0.014 0.034	0.027 0.084	0.013 0.068	-0.008 -0.001	0.000 0.003	-0.009 0.020
female	-0.009	0.008	-0.003	-0.007	0.014	-0.004
	-0.018 0.000	-0.005 0.030	-0.018 0.020	-0.012 -0.002	0.009 0.023	-0.020 0.013
years of schooling	0.066	0.118	0.130	0.000	0.038	0.053
	0.050 0.081	0.068 0.176	0.075 0.184	-0.002 0.002	0.035 0.060	0.019 0.085
work	0.000	-0.023	-0.003	0.002	-0.006	-0.001
	-0.003 0.001	-0.042 -0.006	-0.015 0.004	0.000 0.005	-0.011 -0.002	-0.004 0.002
residual	-0.091	-0.196	-0.101	-0.009	0.008	-0.033
Detailed Decomposition of the Structure Effect						
urban	-0.088	0.213	-0.101	0.189		
	-0.162 0.003	0.045 0.385	-0.281 0.101	0.103 0.275		
size	0.588	0.273	-0.095	-0.165	-0.065	-0.008
	0.460 0.719	-0.012 0.503	-0.338 0.144	-0.306 -0.032	-0.161 0.013	-0.151 0.184
age	-0.622	-1.093	-0.201	0.280	-0.122	-0.127
	-0.894 -0.361	-1.758 -0.571	-0.671 0.324	0.065 0.482	-0.282 0.069	-0.534 0.257
female	0.056	-0.061	-0.030	0.099	0.008	0.024
	0.005 0.102	-0.162 0.016	-0.133 0.065	0.050 0.147	-0.024 0.041	-0.045 0.095
years of schooling	-0.151	-0.891	-0.471	0.187	0.114	-0.231
	-0.328 0.043	-1.313 -0.516	-0.754 -0.082	0.096 0.267	0.005 0.257	-0.424 -0.022
work	0.031	0.308	0.031	-0.037	-0.112	-0.136
	-0.080 0.175	-0.065 0.721	-0.244 0.283	-0.143 0.068	-0.164 -0.028	-0.339 0.046
constant	-0.141	0.643	0.589	-1.074	0.080	0.439
	-0.612 0.261	-0.274 1.812	-0.211 1.318	-0.306 -0.032	-0.264 0.306	-0.211 1.058
residual	0.098	0.127	0.072	0.029	-0.005	-0.239

Boostrap confidence interval with 300 repetitions

Figure 1.20: Decomposition Results

Chapter 2

Measuring Inequality and Poverty with Income and Consumption in Mexico, 1994 - 2014

2.1 Introduction

The use of income or consumption data to measure living standards has been under debate in the recent literature. This is due to the different trends obtained when measuring poverty and inequality and their implications in terms of policies. Some authors argue that consumption is preferable over income based on two arguments. First, consumption is a better measure of long-term living standards due to the presence of smoothing using savings, credit or borrowing to offset temporary income shocks. Second, consumption appears to be a better measure of well-being for people at

the bottom of the distribution because it tends to be less vulnerable to under-reporting than income.

In Mexico however, very little is known about the information that consumption data can provide about household well-being. This paper aims to fill this gap in the literature. I estimate inequality and poverty measures using spending and consumption data, and compare them with the traditional measures based on income. In addition, I contrast the changes in the distribution at three points in time covering two decades: 1994, 2002 and 2014 . This period has been characterized by high levels of income inequality during the 1990s and falling inequality after the 2000s. Specifically, I ask the following questions: Are consumption- or spending-based measures of inequality and poverty similar to income-based measures? How different are households identified as income-poor to those that are consumption-poor?

To answer the first question, I start with standard measures of inequality and poverty in 1994, 2002 and 2014 using four definitions: cash income, broad income, spending and consumption. Then, I contrast the changes along the distribution using growth incidence curves (GIC). To look at poverty, I analyze the extent of the overlap between households identified as income-poor and those identified as consumption-poor, and I examine correlations between different socio-demographic characteristics and the probability of being income-poor or consumption-poor. In this work, I define poverty as the households in the bottom twenty percent of the income or consumption distribution. In this way, my measures are comparable to official income-based measures of poverty.

The rest of this chapter is structured as follows. In section 2.2, I describe the economic context and the inequality and poverty trends in Mexico between 1994 and 2014. Section 2.3 presents the literature review for other countries and Mexico using income and consumption measures of inequality and poverty. Section 2.4 describes the data and the construction of income and consumption definitions to measure poverty and inequality. In section 2.5, I present the results. Conclusions are presented in section 2.6.

2.2 The Economic Context in Mexico

To help the interpretation of the results, this section provides background information about economic, poverty and inequality trends in Mexico.

The period of analysis is 1994 to 2014. At the start of this period, Mexico experienced an economic crisis. Cunningham and Maloney (2000) estimate that household income fell 30 percent and the crisis affected households at the top of the income distribution more. In 2002, the economy slowly started to recover, until the economic crisis in the US hit Mexican growth rates in 2009. Since 2009, the economy has slowly recovered. In 2009, economic growth was -5.3 percent. Later on, between 2010 and 2014 economic growth averaged 3.3 percent per year (see Figure 2.1).

2.2.1 Poverty in Mexico

Poverty has been extensively documented in Mexico during this period, with the Consejo Nacional de Evaluacion de la Politica de Desarrollo Social (CONEVAL) reporting the official measures of poverty. The first official measure of poverty was published by the government in 2002 and has been used since by CONEVAL. Poverty is measured at the national and the state level. The main source of information is the Socioeconomic Conditions Module of the National Survey of Household Income and Expenditure (MCS-ENIGH, Modulo de Condiciones Socioeconomicas de la Encuesta Nacional de Ingreso y Gasto de los Hogares), a household survey carried out every two

years starting with 2008 by the National Institute of Statistics and Geography (INEGI, Instituto Nacional de Geografia y Estadistica). Between 2002 and 2006 the official measure of poverty was based only on income. From 2008, CONEVAL published a new multidimensional poverty measure. This new measure considers income and six other dimensions: education lag, access to health services, access to social security, access to food, housing quality and space, access to basic housing services and degree of social cohesion.

CONEVAL has three poverty lines based on income. The first one is calculated as the income needed to afford a basic food basket. The second one considers the basic food basket and other services such as health and education. The third poverty line adds to the previous one other services such as housing, transportation and clothes. I will be referring to the first official poverty line only to compare the results in this work. The first official poverty measures indicates that poverty fell slightly between 1994 and 2014. During this period poverty fell 0.6 percentage points from 21.2 percent in 1994 to 20.6 in 2014. The fall was a bit larger between 1994 and 2002: 1.2 percentage points[1].

2.2.2 Income inequality in Mexico

Between 1994 and 2014 income inequality declined (see Figure 2.2). In 1994, the Gini index was 0.53; later, in 2014, it fell to 0.49[2]. According to Székely (2005), in the first

[1] see https://coneval.org.mx/Medicion/EDP/Paginas/Datos-censales.aspx
[2] Socio-Economic Database for Latin American and the Caribean (SEDLAS).

part of this period, inequality was very high. Then, between 1994 and 1996, inequality fell due to a general fall in living standards at all income levels during the crisis. After 1996 and before 2004, inequality remained very high because of an expansion of the high income and middle classes. Over the 2000s, inequality has been declining. However, there is a debate about the determinants of the observed declined in Mexico. Cortés (2017) argues that despite inequality falling after 2002, it is not possible to identify structural changes. Social policies, and in particular the Progresa/ Oportunidades program that targets the poorest households in rural areas were not enough to significantly alter inequality trends. While the program helped reduce inequality among the poor, it could not reach all households at the very end of the income distribution. The author emphasizes that the income share of the households in the bottom percentile has remained the same between 2002 to 2014.

The debate on the decline of income inequality has extended to questioning the quality of the data and specifically the data coming from very rich households. Campos-Vazquez and Lustig (2017) show that after imputing values for households at the top of the income distribution, inequality did not decline between 2006 and 2017. Also Cortés (2013) argues that there is not enough evidence to say that inequality has been declining since 2002. Even, when the income measures used are similar, there is no agreement about trends in inequality in Mexico.

In addition, researchers have compared measures of income from National Accounts

to those provided by the household income surveys, such as ENIGH. Bustos (2015) argues there are two main issues with ENIGH: under-reporting of income by households, and a truncation effect due to the exclusion of households with very high incomes. He proposes a statistical adjustment to the distribution of income in ENIGH to make it compatible with information from National Accounts. After the adjustment, he finds that the official poverty measure does not change much, but inequality increases enormously. He estimates a Gini value of 0.803 in 2012 instead of 0.44 without the adjustment.

Del Castillo (2015) propose a framework to adjust for under-reported income in order to provide better measures of income inequality. They suggest to estimate inequality after adjusting incomes at the top of the distribution. They also find that this adjustment increases inequality substantially. The income share owned by the top ten percent of the distribution increases by from 35% to 65%, so the Gini coefficient goes from 0.45 to 0.68.

Similarly, Del Castillo Negrete (2015) and Esquivel (2015) show how inequality changes when they adjust and incorporate incomes from the very rich households. After that, several works have shown various methodologies to adjust the differences Cortés and Vargas (2017), Campos Vázquez and Rodas Milian (2019) and Santaella and Leyva (2017). In general they show that inequality is higher when considering missing incomes at the top from survey data.

Since this work focuses on the differences in using income and consumption definitions, I use only data from ENIGH. I am not adjusting the top incomes using any other source of information as the latest literature has done.

2.3 Measuring Living Standards: Income vs Consumption

2.3.1 Income vs Consumption

Traditionally, measures of inequality and poverty around the world have been constructed using income data. As a result most of the efforts haven been concentrated on improving the collection of income data. On the other hand, economic theory suggests consumption might be a better measure of life-time resources than current income. At least two economic theories, the permanent income hypothesis by Friedman (1957) and life-cycle models by Modigliani and Brumberg (1954) suggest consumption better captures the life time resources of households. Both theories assume that individuals can borrow or use savings when they experience temporary income shocks to preserve their consumption level. As a result, consumption will vary less over time compared to income. The advantage of consumption compare to income as a measure of living standards has been documented by Poterba (1989), Cutler and Katz (1992) and Slesnick (1993). If we assume all the population has the option to smooth consumption with savings and/or can access credit markets, consumption data should reflect better long term living standards.

A growing literature has pointed out the possibility of miss-measurement of income at the bottom of the distribution and the advantages of using consumption data over

income to measure poverty. The work by Meyer and Sullivan (2009) suggest that consumption not only better reflects the long-term resources of a family , but it also captures the effects of savings and dis-saving, the ownership of durable goods and access to credit in the US. They conclude that a consumption based poverty measure would capture better changes in well-being and the effects of anti-poverty government policies.

Meyer and Sullivan (2011) and Meyer and Sullivan (2003) show for the US that consumption data captures better the well-being of the poor compared to measures based on income. They highlight the fact that expenditure exceeds income among the poor, suggesting a problem of income under-reporting. Brewer and O'Dea (2012) and Brewer et al. (2017) show that, in the UK households with the lowest incomes spend more on average than households located at higher quantiles of the income distribution. They conclude that this mismatch is likely due to under-reporting of income at the bottom rather than the over-reporting of expenditure or consumption smoothing. Finally, they document that low consumption is better correlated with other measures of living standards than low income.

Evidence from Greece by Kaplanoglou and Rapanos (2018) suggest that the use of consumption data is preferable to income data due to do the extreme under-reporting of income in the household surveys.

On the other hand, consumption may also suffer from specific shortcomings as a measure of living standardsBlundell and Preston (1995) suggest that consumption habits

strongly depend on the life cycle stage. Intertemporal smoothing assumptions at the heart of permanent income and life cycle models, might not hold Therefore, consumption data might not reflect long term resources. For the case of Mexico, the literature has shown that it is not clear that individuals and household population can smooth consumption when there is a permanent income shock using savings or credit (see McKenzie (2006)). For the particular case of Mexico, McKenzie (2006) shows that during the economic crisis of 1995 there was no evidence of consumption smoothing against the decline of household income. They show that households reallocated their consumption. Households reduce consumption on durable goods in order to keep their food consumption at the same level.

Banks et al. (1994) argue that when a family is expecting a child they would tend to change their preferences and needs and all this could be reflected in their consumption patterns. The status in the labor market could be another source of changes in consumption(see Blundell et al. (1994)). In addition, **?** suggest that another drawback of using consumption data to measure well being is related to changes in real interest rates where real interest rate could influence their inter-temporal substitution possibilities. Another disadvantage in using consumption data is that when we construct a consumption definition, the estimation of the value of durable goods and housing needs to be imputed The imputations can result in an inaccurate measure of consumption (see Gradin et al. (2008)).

To sum up, there are pros and cons in using income or consumption data to measure inequality and poverty. The aims of this research is to compare measures of inequality

and poverty and analyze if they provide similar information or they complement each other, regardless of the presence or not of consumption smoothing in Mexico.

2.3.2 Related Literature for Mexico

There is very little literature for Mexico comparing assessments of poverty and inequality using consumption and income. There is, though, a small literature that uses data on income and consumption to assess how well households adjust their consumption after an income shock. For example, Attanasio and Székely (2004) use consumption and income data to investigate how households adjust their consumption after shocks to their income. They find that during the 1990s households are not able to insure idiosyncratic risks either through savings or through other assets. In another paper, Attanasio and Székely (1998) show that savings are concentrated among households with high levels of education, while households with low levels of education have very low savings. This makes them more vulnerable when they have an income shock. The authors argue that most of the income earners in this group work in the informal sector. Other authors such as McKenzie (2003) also analyze how households deal with shocks in income using the same data. They find that during the economic crisis of 1994, household income and consumption decreased simultaneously, a pattern inconsistent with consumption smoothing. They found that households changed the composition of their consumption, reducing health care and allocating more resources to food.

2.4 Data and Definitions

2.4.1 Data

I use the National Income and Expenditure Household Survey for Mexico (ENIGH) for three years 1994, 2002 and 2014. The survey collects detailed data on the structure and distribution of income and household expenditure on durables and non durables. The survey is representative at the national level and for rural and urban areas and considers the household as the unit of observation[3].

The reported income refers to the six months prior to the month of interview while expenditure data is collected via a diary. The 1994 survey was collected between September and December. The information from 2002 and 2014 was collected between August and November. The income data and socio-demographic characteristics are collected via an interview that is taking place for seven days. Expenditure data is collected with a diary that is left at the house for seven days and is filled by the household members.

The data on income and expenditure as it is presented by INEGI[4] is quarterly. I transformed into average monthly data.

[3]https://www.inegi.org.mx/programas/enigh/tradicional/2014/.

[4]Instituto Nacional de Estadística Geografía e Informática. Institute of National Statistics, Geography and Informatics.

2.4.2 Income and expenditure measures

I construct two consumption definitions. Spending includes expenses on food, clothing and footwear, housing, cleaning, health, transportation, education and recreation, personal, transfers to expenditure, auto consumption, in-kind remuneration[5] and in-kind transfers[6]. Consumption is defined as all items of spending except that on mortgage interest or of the purchase of vehicles plus the imputed rent for owner-occupiers, and the consumption flow from vehicles. The estimated rent of the house is a self-reported variable in the survey. To estimate the flow from vehicles, I follow the same methodology as Brewer and O'Dea (2012). For each year, I estimate deciles of the expenditure on non-durables[7]. Then, I estimate the average expenditure on vehicles in each decile of non-durable spending and separately according to the number of cars owned by the household; and I use these averages as the estimated consumption flow from vehicles.

I construct two measures of income, the first - cash income - includes labor income, income from own business, transfers, and other income. The second measure of income is known as 'broad income' which is defined as cash income plus the estimated rent of the house for homeowners less mortgage interest payments and the estimated flow from vehicles for car owners. The estimated rent of the house and the estimated flow from

[5]This is an estimation of the value of products or services that workers receive as part of their compensation.

[6]in-kind transfers from other households and institutions.

[7]Nondurables are all the expenses minus expenses on vehicles acquisition, maintenance, rent of the house and estimated rent of the house

vehicles are included in the broad definition of income because they can be considered to generate an additional flow of income.

Finally, income and consumption values are expressed in 2014 pesos, and all concepts are expressed on a per-capita basis. [8]

[8] I use the same price index for all four measures. Brewer and O'Dea (2012) argue that, if one is comparing time trends, then different price indices should be used if one is adding implicit consumption of housing and vehicles to a measure of cash spending or cash income, but we do not do this here. For that reason, we do not compare explicitly growth in the two income-based measures, or compare growth in the two consumption-based measures.

2.5 Results

In what follows, I compare measures of living standards. I compare broad income with consumption, and cash income with cash spending; the main difference between the two comparisons is that the first pair of concepts includes the implicit income accruing to owner occupiers and the consumption flow from vehicles, while the second does not.

2.5.1 Changes in average levels of living standards over time

Figure 2.3 shows that, from 1994 to 2014, average growth in living standards was positive irrespective of which definition is used. Spending and consumption grew more than cash income and broad income respectively. The average growth in each sub-period shows the differences in pattern between 1994 -2002 and 2002 -2014. The first subperiod (1994 - 2002) reflects the impact of the economic crisis on income, consumption and spending growth. The average growth of consumption was -1.15 percent, while broad income grew by just 0.19 percent. Cash income grew by 4.82 and spending grew 3.39 percent. In the second subperiod characterized by economic recovery, consumption and spending grew more than their income counterparts: consumption grew by 7.30 percent and broad income by 5.13 percent; spending grew by 6.46 percent and cash income by 3.66 percent.

2.5.2 Inequality

Summary indices

Table 2.1 shows summary statistics of inequality for the four definitions, as well as two standard measures of inequality. On average, cash income, broad income, spending and consumption inequality increased in 2014 with respect to 2002 and 1994, after a decline in 2002 due to the economic crisis. However, inequality levels using income and consumption are different. Inequality levels measured by the Gini index using spending and consumption are below those using income measures in all three years. Also, the decline in income inequality between 1994 and 2002 was faster than the decline in spending and consumption inequality. Between 2002 and 2014, cash income inequality measured using the Gini increased faster than spending inequality, while consumption inequality declined slightly in the same period.

The Theil index shows a similar pattern to the Gini index between 1994 and 2002, all the income and consumption measures declined. Then, from 2002 to 2014 only income measures increased while spending and consumption declined during the same period. Income and consumption provide different perspectives of inequality.

Another measure of inequality is the share of total resources going to the bottom decile, and we show this in Figure 2.4. Consumption and spending shares are always above income shares. This reflects the lower inequality levels using consumption data compared with income data. All four income and consumption measures point to an

increase in the share going to the bottom decile between 1994 and 2002. Nonetheless, the trends differ between cash income and spending over the second subperiod. There was a slight reduction in the cash income share going to the bottom decile, while according to the other three definitions the bottom decile increased its share between 2002 and 2014. Contrasting within income and consumption definitions, the main difference is between broad income and cash income between 2002 and 2014. The bottom decile's share of broad income increased from 1.4 to 1.5 percent between 2002 and 2014. At the same time, cash income remained constant around 1.4 percent. This result shows that the income flow from housing and vehicles might have contributed to the increased share going to the bottom decile. However, consumption and spending definition show very similar trends over the entire period.

Growth Incidence Curves

This section present the results from the Growth Incidence Curves (GIC). The GIC displays the complete picture of the quantile specific rates of income and consumption growth. This analysis will let us see the differences in growth between income and consumption definition across the distribution.

Figure 2.5 shows Growth Incidence Curves (GICs) for 1994-2014. They show that growth has been stronger at the bottom for both income and consumption. Households in the bottom half of the distribution experienced positive growth in all four income

and consumption concepts: broad income, cash income, spending and consumption. After the 50th percentile, growth becomes negative in the case of consumption and broad income, while cash income and spending growth remains positive up until the 70th decile. At the very bottom, broad income and consumption grew more than cash income and spending. Between 1994 and 2014, growth in income, consumption and spending were inequality decreasing, consistent with the decline in inequality that started in the mid 2002s.

I next divide the period into two subperiods. Figure 2.6 shows growth rates between 1994 and 2002, a period of economic crisis. They show an inequality-decreasing pattern, as in the complete period. However, consumption grew less than broad income, and spending grew less than cash income. The bottom half of the distribution is characterized by low growth rates in consumption and broad income. At the top, the four growth rates converge to each other. Generally, households in the top half of the distribution experienced bigger losses in their living standards compared to households at the bottom.

Figure 2.7 shows the GICs for the second subperiod starting in 2002 and ending in 2014. During this time, the economy started to recover in the aftermath of the 1994 crisis. Growth rates remain inequality-decreasing, but with different patterns compared with the previous subperiod. Consumption grew more than broad income for households in the bottom half. Spending also grew more than cash income up to the 70

percentile where the two growth rates converge. In this period of recovery, households increased their consumption compared with the previous subperiod.

The possible role of informality in explaining under-reporting of income

Looking over both periods, then, we can see that, during the economic crisis, households at the bottom appear to have reduced their consumption and spending. However, over the second sub-period, they increased their consumption and spending above increases to their income. This two patterns suggest that households at the bottom of the distribution smooth consumption given changes in their income. However, we usually think that consumption smoothing is more common among high income households than those at the bottom of the distribution. Instead of consumption smoothing, then, the patterns we see could be due to under-reporting of income at the bottom of the distribution. Stronger growth in consumption relative to income suggests that under-reporting of income at the bottom might be significant. One possible explanations is the increase in informal employment in Mexico. Households at the bottom employed in the informal sector might enjoy higher levels of income than those reported to the ENIGH survey and this higher income is reflected in their consumption levels.

The size of the informal sector has been documented by the Institute of National Statistics (INEGI) in Mexico. In their latest report dated April 2020 INEGI (2020),

informal employment[9] is estimated to account for 56.2 percent of total employment over the last quarter of 2019. In 2014 the estimated percentage was 57.8 percent. According to that report, informal labor is concentrated among people between 15 and 19 years old and 60 years old and more. The latest International Labor Organization (ILO) report ILO (2014) suggests that informal employment declines during times of economic growth and increases in crises. However, after the 2009 crisis, informal employment remained high in Mexico. This increase in informal employment might have allowed households at the bottom to increase their consumption over and above increases in reported income. A similar view is taken by Cortés (2017) who argues that low-income households have been able to maintain their income level due to the increase in the informal sector.

[9]Informal employment is defined as employment in non-agricultural firms that are not legally register as firms according to the law. The economic unit operates using household resources and does not keep accounting records, it also includes self-employed people in the agricultural sector as well as those without social security.

2.5.3 Poverty

Poverty analysis in developing countries such as Mexico tends to use absolute definitions. Using an absolute approach however is problematic when comparing income and consumption based definitions of poverty due to the need to define equivalent poverty thresholds. To avoid this complication, this analysis will use a relative measure. Given that official measures of poverty identify roughly 20 per cent of the population as poor, I define the poor as belonging to the bottom quintile of the income or consumption distributions respectively. This allows for easy identification of the poor in a comparable way using both income and consumption based measures.

Figures 2.8 and 2.9 show the percentage of households that are identified as either income- or consumption-poor and the percentage of households that are both income- and consumption-poor. Figure 2.8 shows that in 1994 and 2002 approximately four percent of the households are only broad income poor but not consumption poor and around 16 percent are both broad income and consumption poor. The overlap between households that are income poor and those who are consumption poor is thus significant but not complete. Between 2002 and 2014 the percentage of households that are both broad income and consumption poor decreased by 2.3 percent, thus reducing the overlap between income and consumption based definitions of poverty. Figure 2.9 compares the poverty overlap using the second group of definitions: cash income and spending. Approximately five percent of households are considered cash income poor only or spending poor only in 1994 and 2002. In 2014, the overlap between households

identified as cash income poor and spending poor again declined from 15.46 to 13 percent.

It is not entirely clear what is driving the fall in the overlap of income and consumption measures of poverty. As I discussed above, one possibility is that income under-reporting increased, but there is little research on income under-reporting at the bottom of the distribution in Mexico.

Regardless of the reason, it is clear that there is not a perfect overlap between households that are income poor and households that are consumption poor. The following subsections, therefore, compare poverty profiles generated using income and consumption definitions. I examine correlations between different socio-demographic characteristics and the probability of being income-poor, consumption-poor or being poor according to both definitions.

Cash Income and Spending

Previous results show there is not a complete overlap between the income and consumption definitions of poverty. This is important from a policy perspective as anti-poverty programs may not be targeted appropriately if only income is used to identify the poor. To better understand differences between income and consumption poor households, I run a series of regressions to establish which characteristics most closely correlate with being only income poor, only consumption poor or both.

I estimate a multinomial regression model where the dependent variable takes three values: ”1” for households that are cash income poor but not spending poor, ”2” for households that are spending poor but not cash income poor and ”3” for households both cash income poor and spending poor. The base outcome is the value of ”0” households that are non poor. The predictors are several sociodemographic characteristics: females household headship, living in a rural area, the number of members in the house, the education of the household head (secondary education or less),the age of the household head (60 years or more) and the tenancy status. I then interpret the characteristics that predict (for example) the category 2 as being those that are associated with being spending poor but not with being income poor.

Figure 2.10 shows the effect of having a female household headship on the probability of being cash income or spending poor. In 1994 and 2002, the effect was negative and very close to zero possibly because in those years the number of female headed households was small. However, in 2014, the effect of female headship on both probabilities increased. Households where the head of the household is female are more likely to be both income poor and consumption poor. This result agrees with the official poverty report which estimates that 9.6 percent of women were living below the extreme poverty line in 2014. The gender of the head of the household has impacted the probability to be income poor and consumption poor in a similar way over time.

Figure 2.11 shows that living in a rural area has a positive effect on the probability of being both spending poor and cash income poor, while having little effect on the probability of being poor according to only one measure. The effect of living in a rural area on the probability of being both income and consumption poor declined slightly in 2014.

Figure 2.12 shows the effect of the number of people living in the house on the probability of being poor using income and spending definitions. The effects on both the probability of being cash income poor and spending poor are positive and have similar trends in 1994 and 2002. However, in 2014 the effect on the probability of being spending poor increased while the effect on the probability of being cash income poor remained at the same level as in 1994 and 2002 and very close to zero. This suggests that large households may have additional needs that are not well captured by an income based measure of poverty. Large households may have a larger number of earners or members receiving income. However, their spending is also likely to be higher. Therefore their situation appears different when using a consumption based measure of poverty.

Figure 2.13 shows the effects of the head of the household having at most 12 years or education on the probability of being cash income poor or spending poor. As expected, households headed by less educated individuals are more likely to be poor, according to both income and consumption definitions. Estimated effects are very similar across the

three years suggesting that the effect of the head's education has not changed much over time. This is despite an expansion of education in Mexico between 2000 and 2010 which implies stronger selection effects in 1994 compared to 2014.

Figure 2.14 shows the effects of having a head of household aged 60 or more on the probability of being cash income poor or spending poor. Comparing the effects on the probability to be cash income poor and spending poor, the trends are very similar. In all cases, the effect of having an older head declined. However, households with a head aged 60 or more are significantly less likely to be income poor but more likely to be spending poor. This suggest that households headed by an older individual might have lower consumption relative to their income. This pattern could be showing how insecurity might be affecting consumption. On the other hand, the income and spending of elderly households improved as evidenced by the falling effects on the probability of being poor (both income and consumption). However, the increase in the informal employment has been concentrated in this age group and younger individuals (15 to 19 years old).

Finally, Figure 2.15 shows the effect of renting a house on the probability of being cash income poor or spending poor. Trends in the impact of this variable differ. Households who were renting were at the same risk of being cash income poor as home-owners in 1994 and 2002. By 2014 , they were more likely to be poor. However, their probability of being spending poor declined during the same period.

The differences in trends between income and consumption definitions shows that consumption data captures aspects of vulnerability that income alone does not. Sociodemographic characteristics such as the number of people living in the house and the effects of living in a house that is being rented show different trends in marginal effects on income and spending poverty. In 2014, households with many people were more likely to be spending poor than cash income poor. Renters were more likely to be spending poor than income poor in 1994, whereas in subsequent years the income poverty risk increased considerably while the spending poverty risk declined slightly.

The effects of the other socioeconomic variables have similar trends. The effects of female headship increased both in the case of income and spending poverty. The effect of the head's education remained relatively constant. Households headed by an older individual on the other head experienced a decline of both their income and spending poverty risk.

Looking across the coefficients it is possible to group the sociodemographic characteristics into two groups. In the first group the effects of the sociodemographic variables on the probability of being income or consumption poor increased by 2014. For the second group, the effects declined or remained the same through the time.

The first group includes female household headship, the number of household

members and tenancy status. This group shows that despite there are differences between income and consumption poverty. Consumption data complements the information that income data provides in terms of poverty.

The second group includes sociodemographic characteristics that have a similar effect on the probability of being income or consumption poor. The higher probability of being consumption poor compared to income poor could be explained by the increase in informal employment that has been concentrated in younger and older groups. The additional income that informal employment is providing might be boosting consumption levels but is not properly captured by income data.

To provide some context on these results it is useful to compare them from other developing countries where similar characteristics are associated with chronic poverty. McKay and Lawson (2003). The authors present studies with evidence on the positive correlation between illiteracy and chronic poverty for China and India. In the same line they present the cases of Peru and Pakistan where the increase in the years of education reduce the probability of being chronically poor. In addition to that, they show research with evidence on the increase in the household size and the presence of people from third generation positively correlated with being chronically poor for Pakistan, rural China and South Africa. For the location they mention evidence for Uganda, Vietnam, China where chronic poverty is prevalent in rural areas.

My results show that in 1994 four socio-demographic characteristics are more related with the probability of being consumption poor than income poor: having 60 years of age or more as the head of the household, living in a house where the family pay a rent and households living in a rural area and the number of people living in a house. For 2014 the age of the head of the household and the number of people keep identifying better households that are consumption or spending poor. In addition to that, being the head of the household with twelve years of education slightly identifies more spending poor households than those considered income poor. Most of these characteristics are related with chronic poverty rather than transient poverty according to the literature (see McKay and Lawson (2003)). Specifically, having 60 years of age or more as the head of the household, households living in a rural area and the number of people living in a house and being the head of the household with twelve years of education.

I conclude that poverty definitions based on income and spending provide different information about households long term living standards. The two definitions complement each other, providing a fuller picture of the characteristics of households that can be considered poor than either the income definition alone. If we take into account the view of long term living standards then the consumption measures may give have a broader picture of poverty.

Broad Income and Consumption

The following group of graphs show the effects of the covariates on broad income and consumption definitions. From Figure 2.17 to Figure 2.21 all the figures show a similar pattern compare to the effects of the same covariates on cash income and spending definitions. Therefore, adding the income flow from housing and cars into the cash income and spending definition does not change the probabilities of being income poor or consumption poor given the observable characteristics.

2.6 Discussion and Conclusions

This paper examines trends in poverty and inequality in Mexico using income and consumption data for three years: 1994, 2002 and 2014. I find that growth in each sub period shows different patterns. Between 1994 and 2002 cash income and spending grew by 4.82 and 3.39 percent respectively while broad income and consumption grew only 0.18 and -1.15 percent. Then, in the second sub-period between 2002 and 2014 both income and consumption displayed positive growth showing a recovery after the economic crisis. The average growth rates suggest the income flow from housing and vehicles included in the broad income and consumption definitions was negatively affected during the economic crisis. On the other hand, cash income and spending definitions that do not include the income flow from vehicles and housing had positive average growth rates in both sub-periods.

The analysis of inequality using income and consumption has shown that while the trends are very similar, the levels differ. Consumption and spending inequality is generally lower than income inequality. This stylized fact is consistent with the lower variability of consumption compared to income. Cash income and broad income show a similar pattern in inequality levels with a drop in 2002, followed by an increase in 2014. However, consumption and spending measures show a different story. Consumption inequality declined between 2002 and 2014 while spending inequality increased. Finally, the share of resources going to the bottom decile declined between 2002 and 20014 in

the case of income but increased in the case of spending. These results are in line with the recent debate on the trends in inequality levels. Therefore there is not consistent evidence that inequality is still declining when looking at different income and consumption definitions.

Most of the literature for Mexico has focus on correcting the top incomes in order to get a more real approximation of income inequality. However, consumption data can correct the estimation at the bottom of the distribution. Therefor income and consumption can complement the analysis of inequality. Further research needs to be done in order to compare income and consumption measures at the top.

The GIC analysis shows differences in trends between the two sub-periods. First, between 1994 and 2002 -a period of economic crisis-, consumption and spending grew less than broad income and cash income in households in the bottom half of the distribution. This trend changed in the second sub-period, when the economy recovered. Between 2002 and 2014, consumption and spending grew more than broad income and cash income in the bottom half of the distribution. During the economic crisis, households at the bottom appear to have reduced their consumption and spending. However, over the second sub-period, they increased their consumption and spending above increases to their income. Stronger growth in consumption relative to income suggests that under-reporting of income at the bottom might be significant. One possible explanation is the increase in the informal employment in Mexico. Households

at the bottom employed in the informal sector might enjoy higher levels of income than those reported and this higher income is reflected in their consumption levels.

The poverty analysis shows that the overlap between households that are income poor and those who are consumption poor is significant but not complete. Between 2002 and 2014 the percentage of households that are both broad income and consumption poor decreased by 2.3 percent. It is not entirely clear what is driving the fall in the overlap of income and consumption measures of poverty. As before, one possibility is that income under-reporting increased for both income definitions. However, most of the literature so far has concentrated on adjusting the incomes for those at the top of the distribution.

Finally, I examined correlations between different socio-demographic characteristics and the probability of being income poor, consumption poor or being poor according to both definitions. I find that having 60 years old and more as the head of the household and the increasing number of people living in the house are socio-demographic characteristics more related with consumption poverty rather than income.

The results show that using consumption and spending to measure inequality and poverty give a complementary perspective of inequality and poverty. These differences are important from a policy perspective and combining information from both income and consumption can provide a more accurate picture of living standards.

References

Attanasio, O. and Székely, M. (1998). Household savings and income distribution in mexico.

Attanasio, O. P. and Székely, M. (2004). Wage shocks and consumption variability in mexico during the 1990s. *Journal of Development Economics*, 73(1):1–25.

Banks, J., Blundell, R., and Preston, I. (1994). Life-cycle expenditure allocations and the consumption costs of children. *European Economic Review*, 38(7):1391–1410.

Blundell, R., Browning, M., and Meghir, C. (1994). Consumer demand and the life-cycle allocation of household expenditures. *The Review of Economic Studies*, 61(1):57–80.

Blundell, R. and Preston, I. (1995). Income, expenditure and the living standards of uk households. *Fiscal Studies*, 16(3):40–54.

Brewer, M., Etheridge, B., and O'Dea, C. (2017). Why are households that report the lowest incomes so well-off?

Brewer, M. and O'Dea, C. (2012). Measuring living standards with income and consumption: evidence from the uk. Technical report, ISER Working Paper Series.

Bustos, A. (2015). Estimation of the distribution of income from survey data, adjusting for compatibility with other sources. *Statistical Journal of the IAOS*, 31(4):565–577.

Campos-Vazquez, R. M. and Lustig, N. (2017). Labour income inequality in mexico: Puzzles solved and unsolved. Technical report, WIDER Working Paper.

Campos Vázquez, R. M. and Rodas Milian, J. A. (2019). Desigualdad en el ingreso: posibilidades de acción pública. *Economía UNAM*, 16(46):251–261.

Cortés, F. (2013). Medio siglo de desigualdad en el ingreso en méxico. *Economía unam*, 10(29):12–34.

Cortés, F. (2017). Desigualdad en el ingreso en méxico, 1963 a 2014. *2018) América Latina en la larga historia de la desigualdad. Una revisión de causas, efectos y políticas FLACSO, México, in printing.*

Cortés, F. and Vargas, D. (2017). La evolución de la desigualdad en méxico: viejos y nuevos resultados. *Revista de Economía Mexicana*, 2:39–96.

Cunningham, W. and Maloney, W. F. (2000). Measuring vulnerability: who suffered in the 1995 mexican crisis? *World Bank mimeo.*

Cutler, D. M. and Katz, L. F. (1992). Rising inequality? changes in the distribution of income and consumption in the 1980's. *The American Economic Review*, 82(2):546–551.

Del Castillo, M. (2015). La magnitud de la desigualdad en el ingreso y la riqueza en méxico: una propuesta de cálculo.

Del Castillo Negrete, M. (2015). La magnitud de la desigualdad en el ingreso y la riqueza en méxico. *México: cepal*.

Esquivel, G. (2015). Desigualdad extrema en méxico. *Concentración del poder económico y político*.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pages 20–37. Princeton University Press.

Gradin, C., Canto, O., and Del Rio, C. (2008). Inequality, poverty and mobility: Choosing income or consumption as welfare indicators. *investigaciones económicas*, 32(2):169–200.

ILO (2014). Informal employment in mexico: Current, situation, policies and challenges.

INEGI (2020). Estadisticas a proposito del dia del trabajo. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2020/trabajoNal.pdf.

Kaplanoglou, G. and Rapanos, V. T. (2018). Evolutions in consumption inequality and poverty in greece: The impact of the crisis and austerity policies. *Review of Income and Wealth*, 64(1):105–126.

McKay, A. and Lawson, D. (2003). Assessing the extent and nature of chronic poverty in low income countries: issues and evidence. *World Development*, 31(3):425–439.

McKenzie, D. J. (2003). How do households cope with aggregate shocks? evidence from the mexican peso crisis. *World Development*, 31(7):1179–1199.

McKenzie, D. J. (2006). The consumer response to the mexican peso crisis. *Economic Development and cultural change*, 55(1):139–172.

Meyer, B. D. and Sullivan, J. X. (2003). Measuring the well-being of the poor using income and consumption. Technical report, National Bureau of Economic Research.

Meyer, B. D. and Sullivan, J. X. (2009). Five decades of consumption and income poverty. Technical report, National Bureau of Economic Research.

Meyer, B. D. and Sullivan, J. X. (2011). Viewpoint: Further results on measuring the well-being of the poor using income and consumption. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(1):52–87.

Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. *Franco Modigliani*, 1(1):388–436.

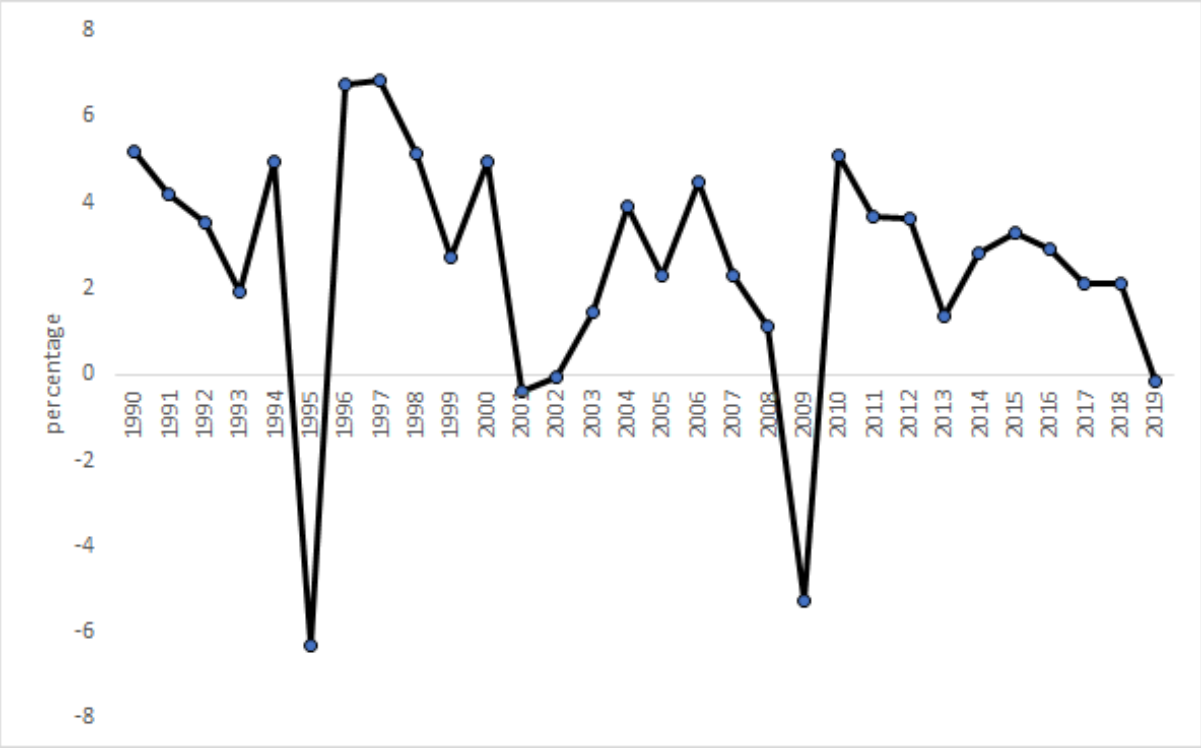
Poterba, J. M. (1989). Lifetime incidence and the distributional burden of excise taxes.

Santaella, J. M. and Leyva, G. (2017). Quién se lleva los frutos del éxito en méxico?

Slesnick, D. T. (1993). Gaining ground: poverty in the postwar united states. *Journal of political Economy*, 101(1):1–38.

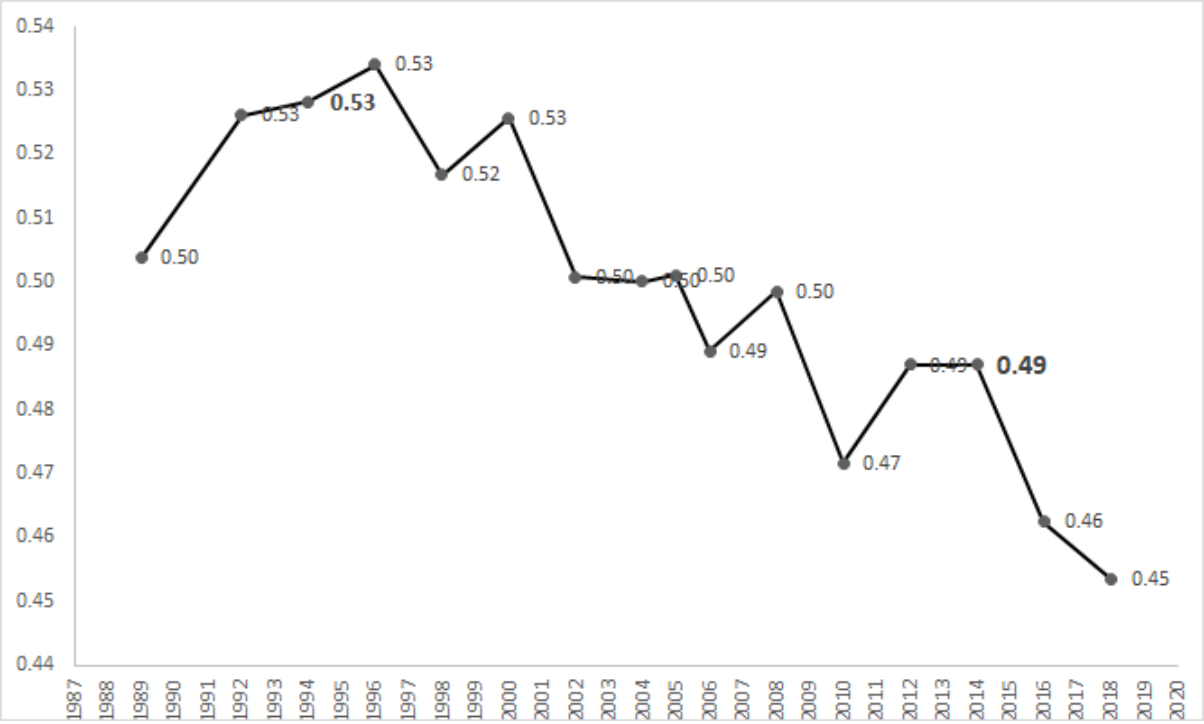
Székely, M. (2005). Pobreza y desigualdad en méxico entre 1950 y 2004. *El Trimestre Económico*, 72(288(4)):913–931.

2.7 Figures



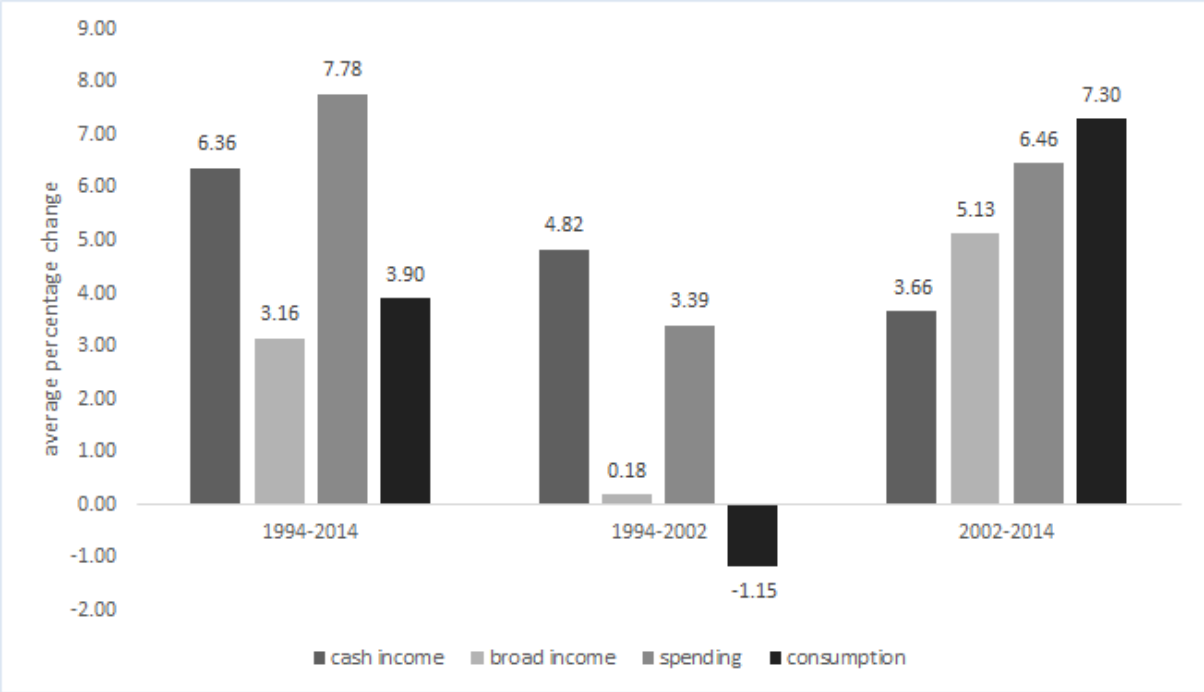
Source: The World Bank Group. World Bank national accounts data.

Figure 2.1: GDP growth (annual percentage)



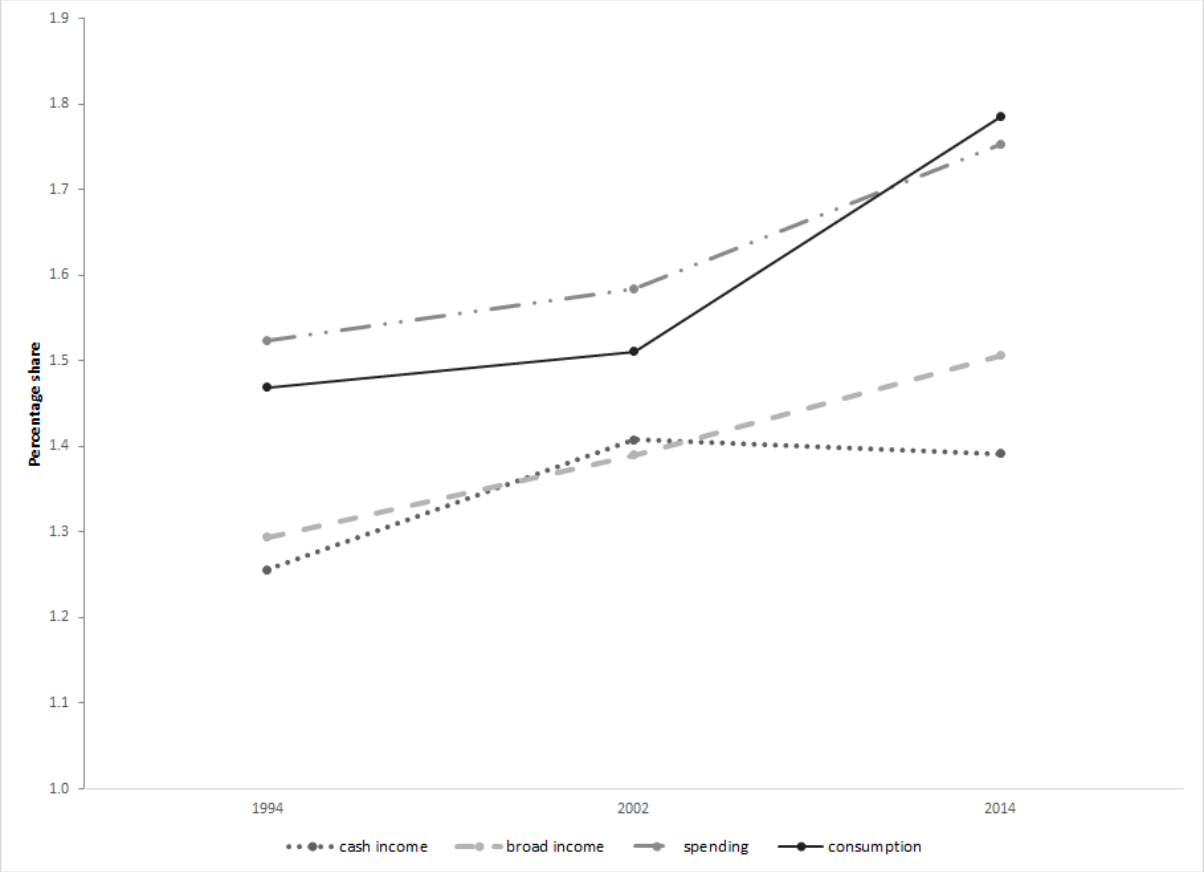
Note: The Gini index calculated using per capita income.
Source: Data Socio-Economic Database for Latin American and the Caribean SEDLAC (CEDLAS and The World Bank) Downloaded from https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/ December 2020.

Figure 2.2: Income Inequality in Mexico Gini index)



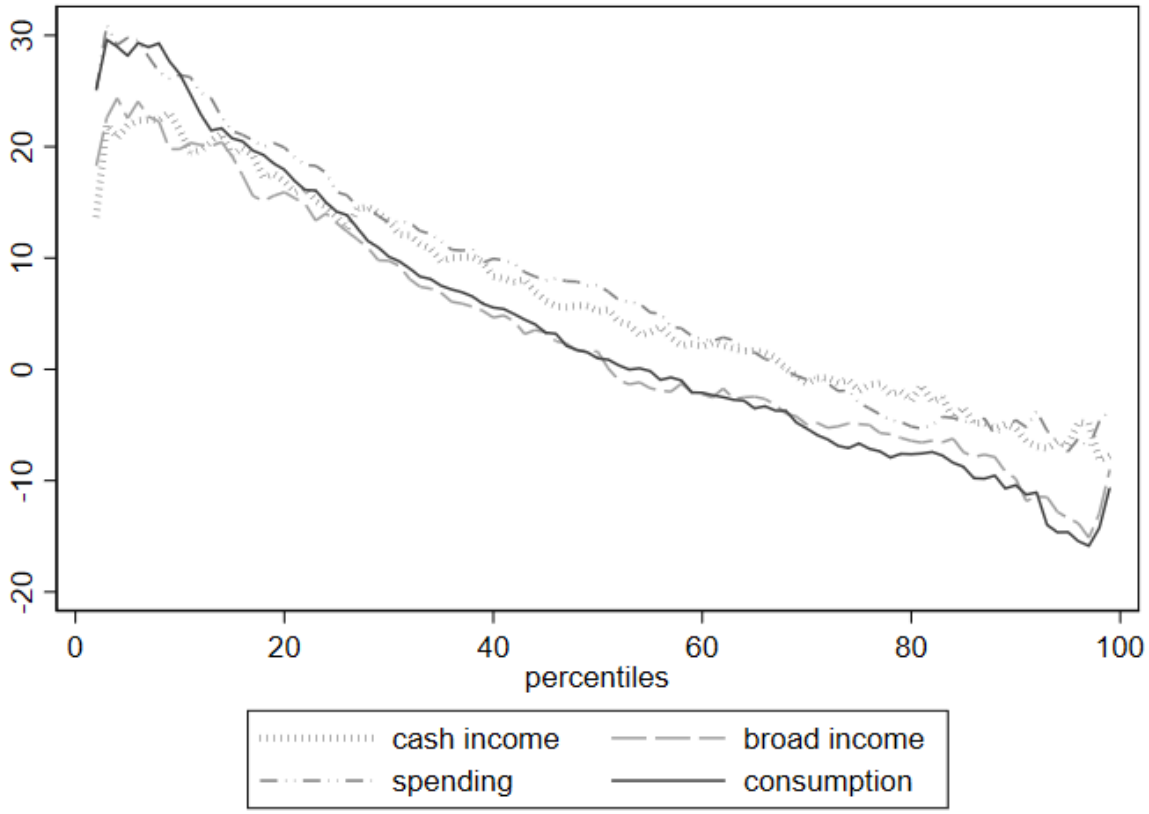
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.3: Average Growth Rates 1994 - 2014



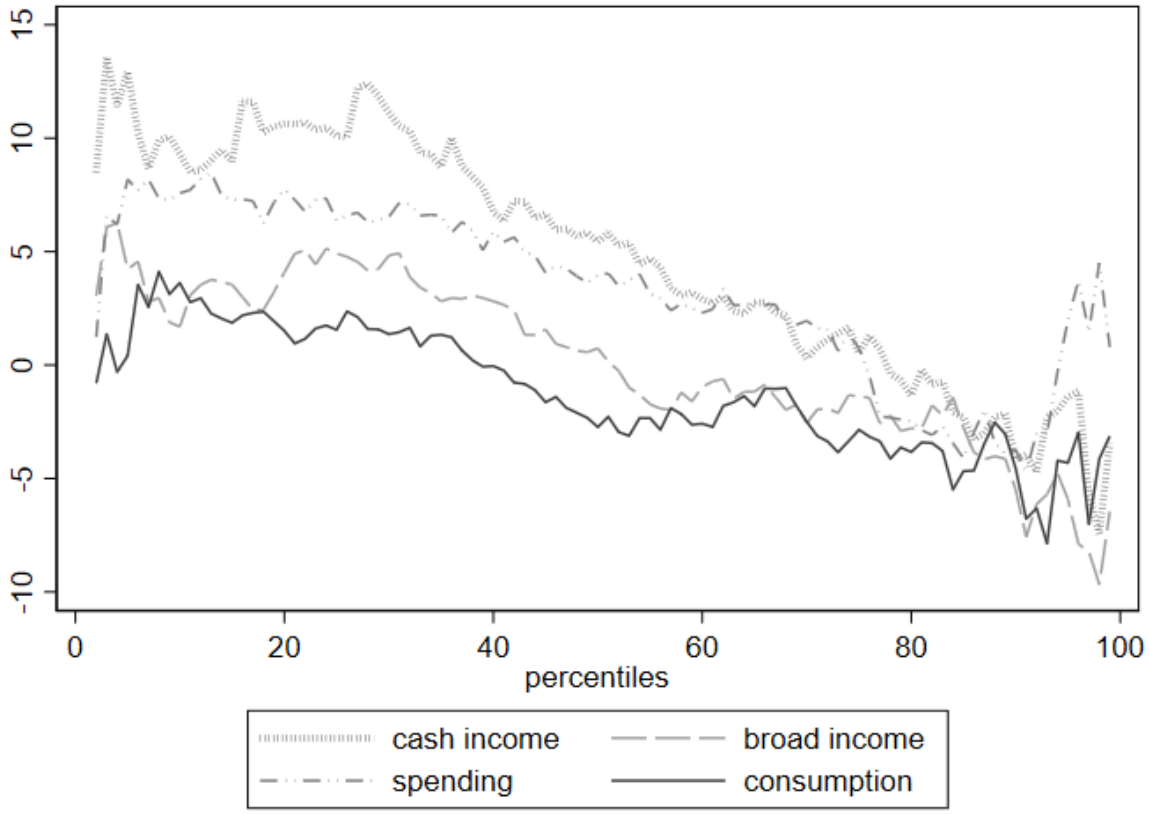
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.4: Percentage Share of the Bottom Decile



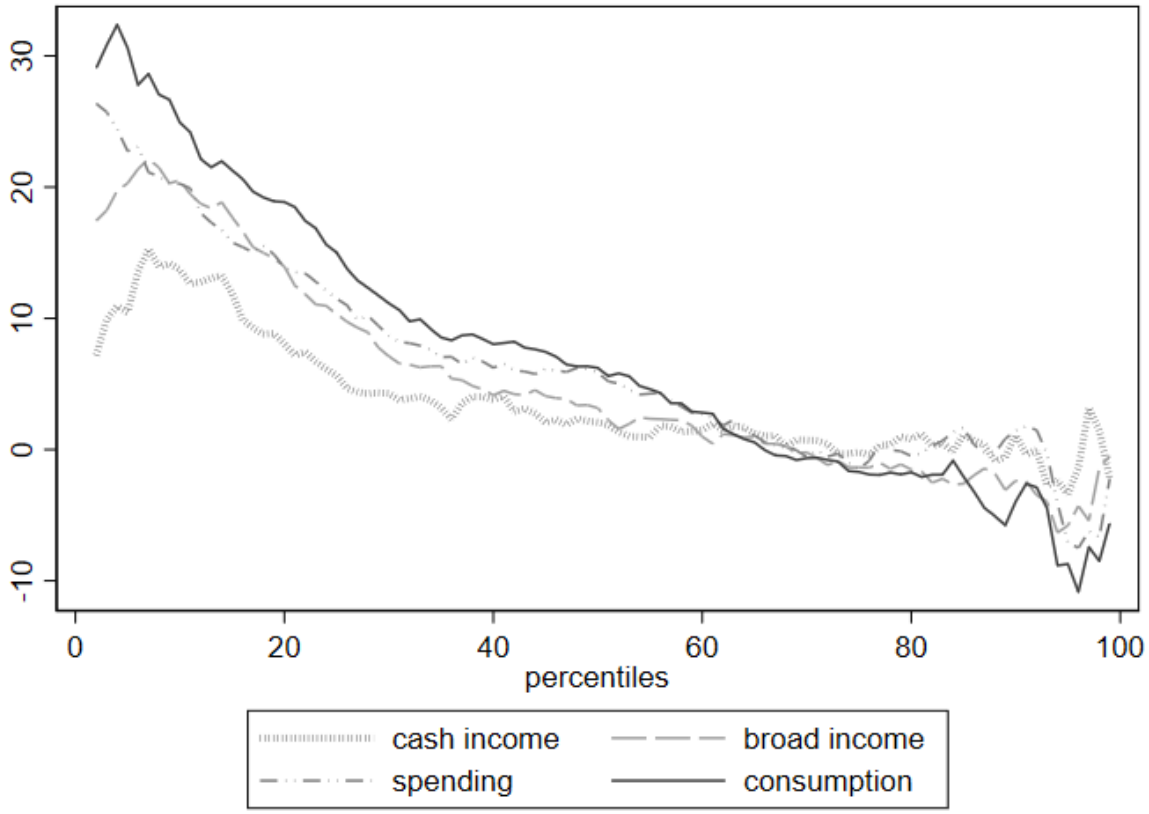
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.5: Growth Incidence Curves 1994 - 2014



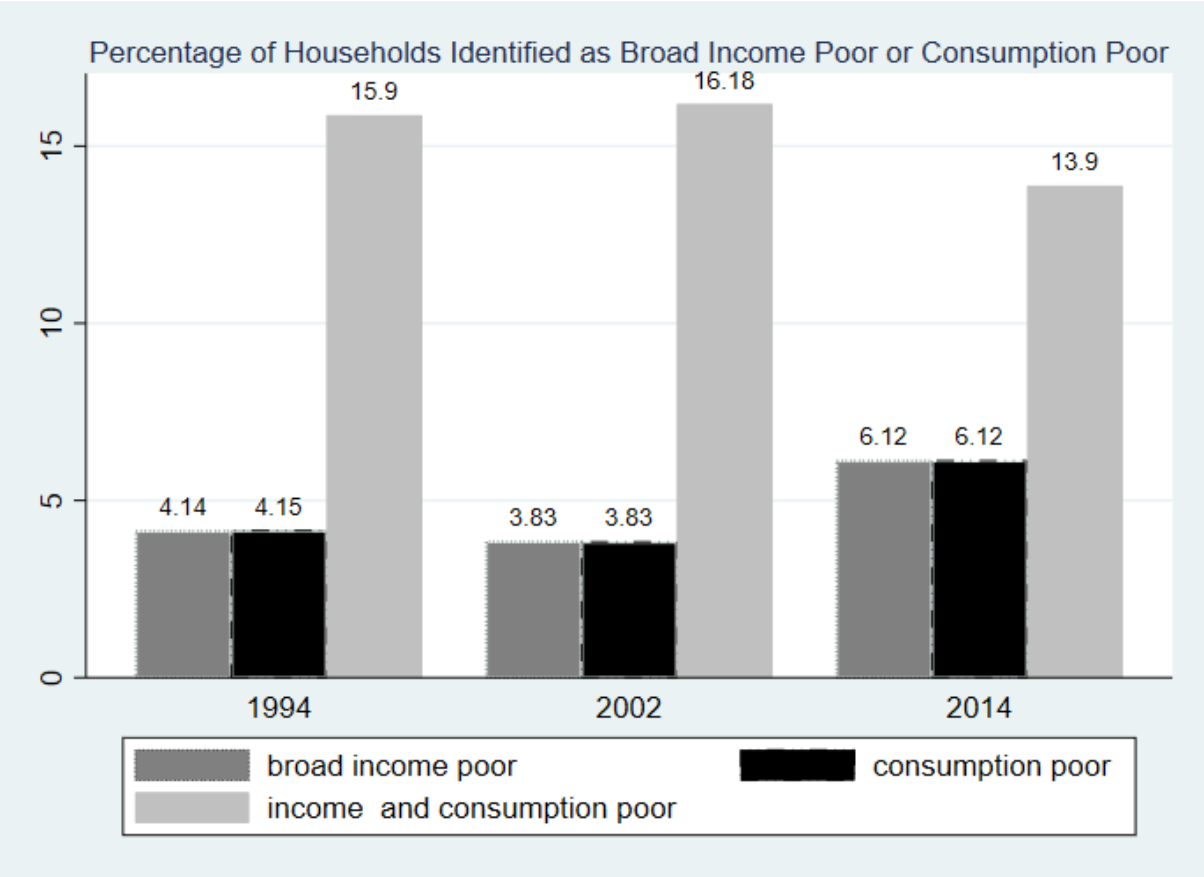
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.6: Growth Incidence Curves 1994 - 2002



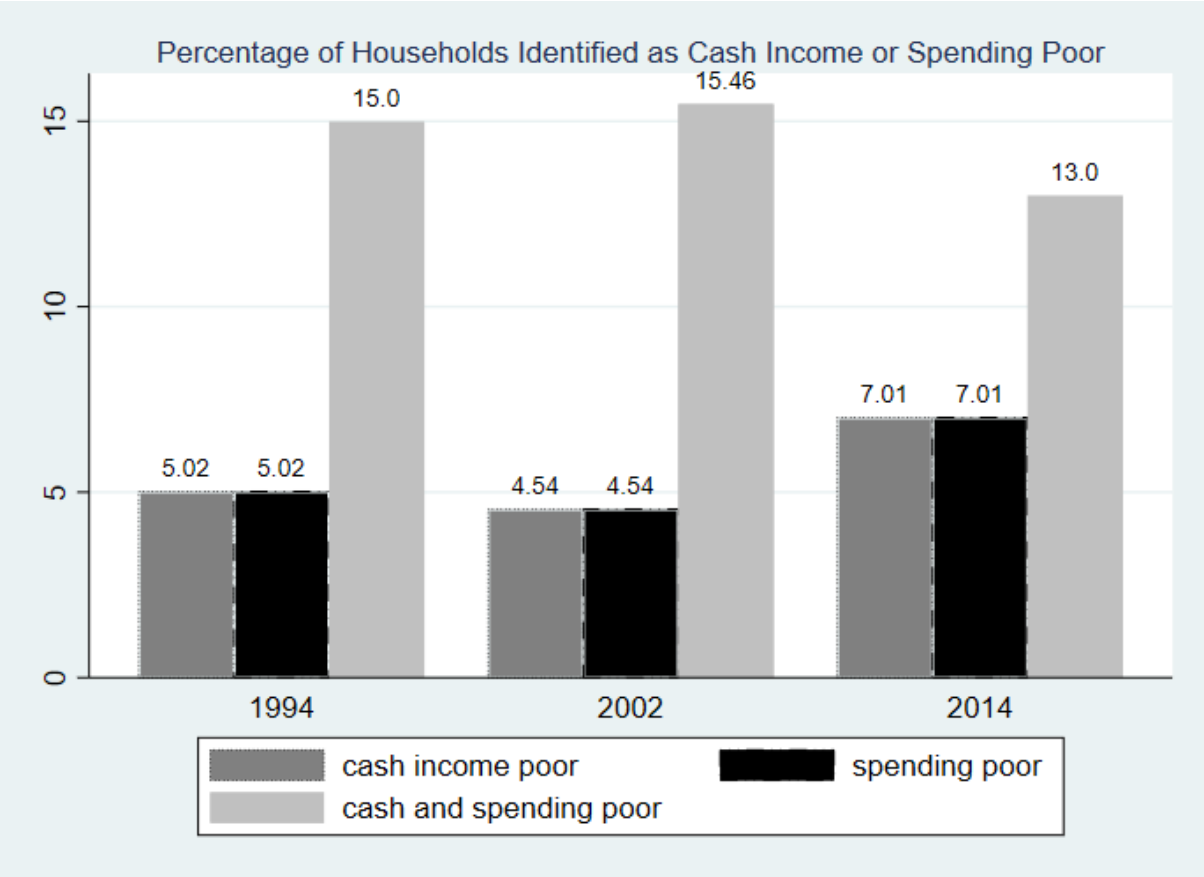
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.7: Growth Incidence Curves 2002 - 2014



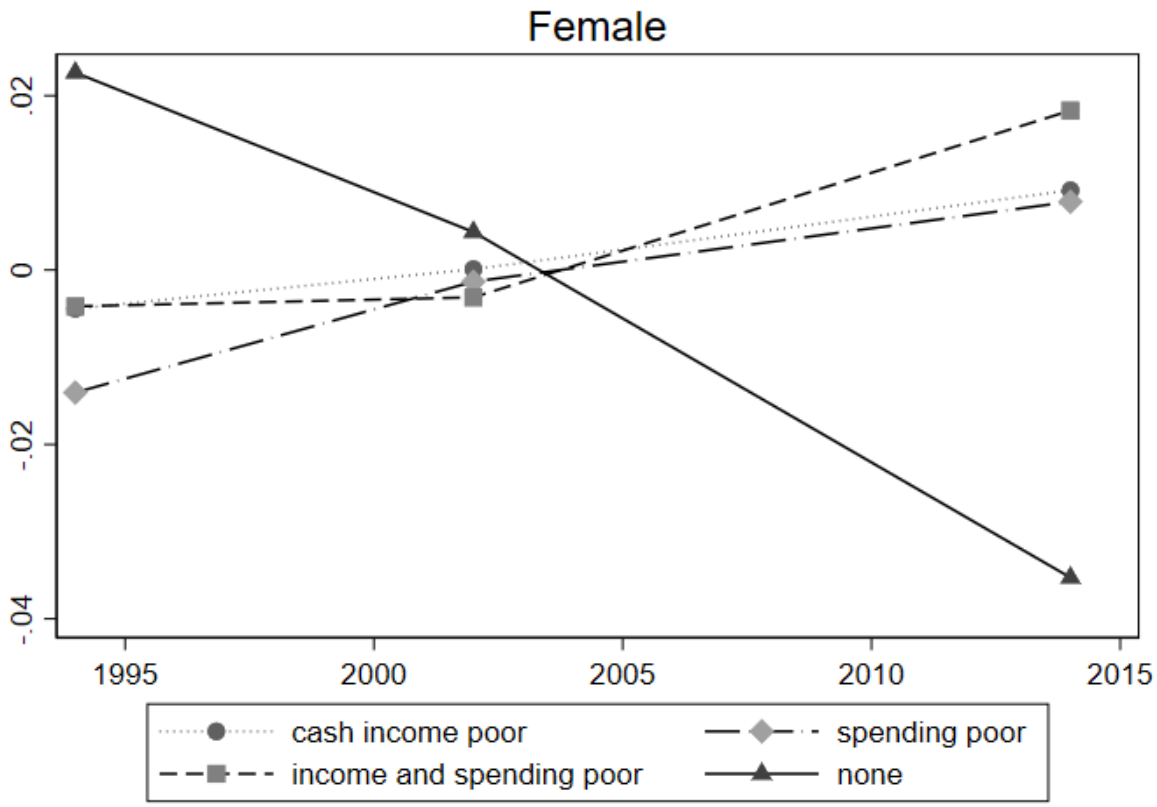
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.8: Broad Income vs Consumption



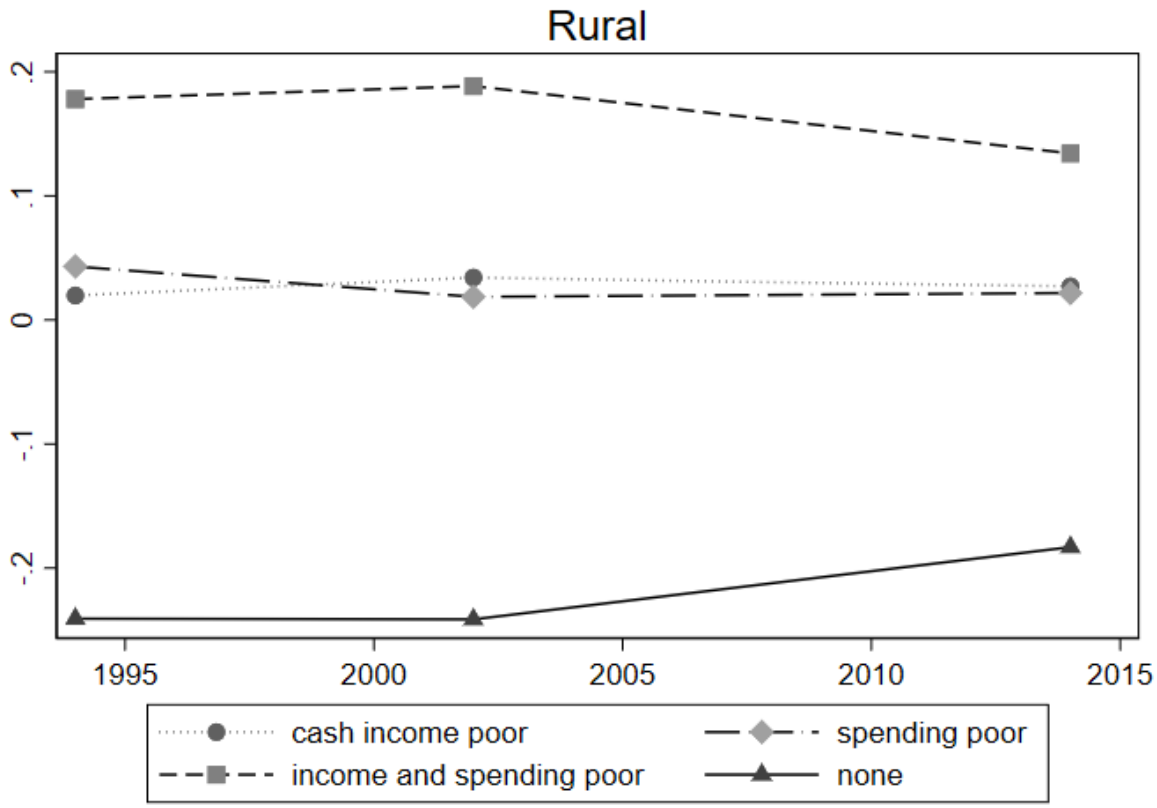
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.9: Cash Income vs Spending



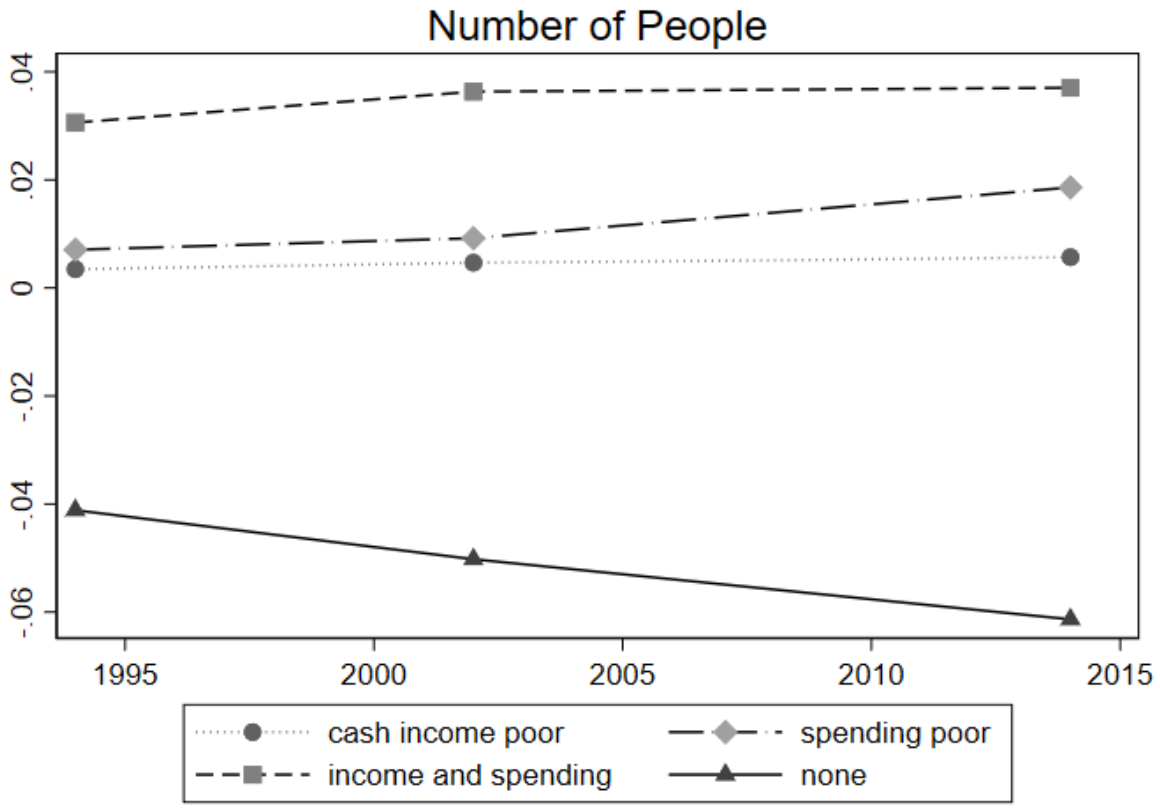
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.10: Average Marginal Effects of being Female as the head of the household on Cash Income and Spending poor Households



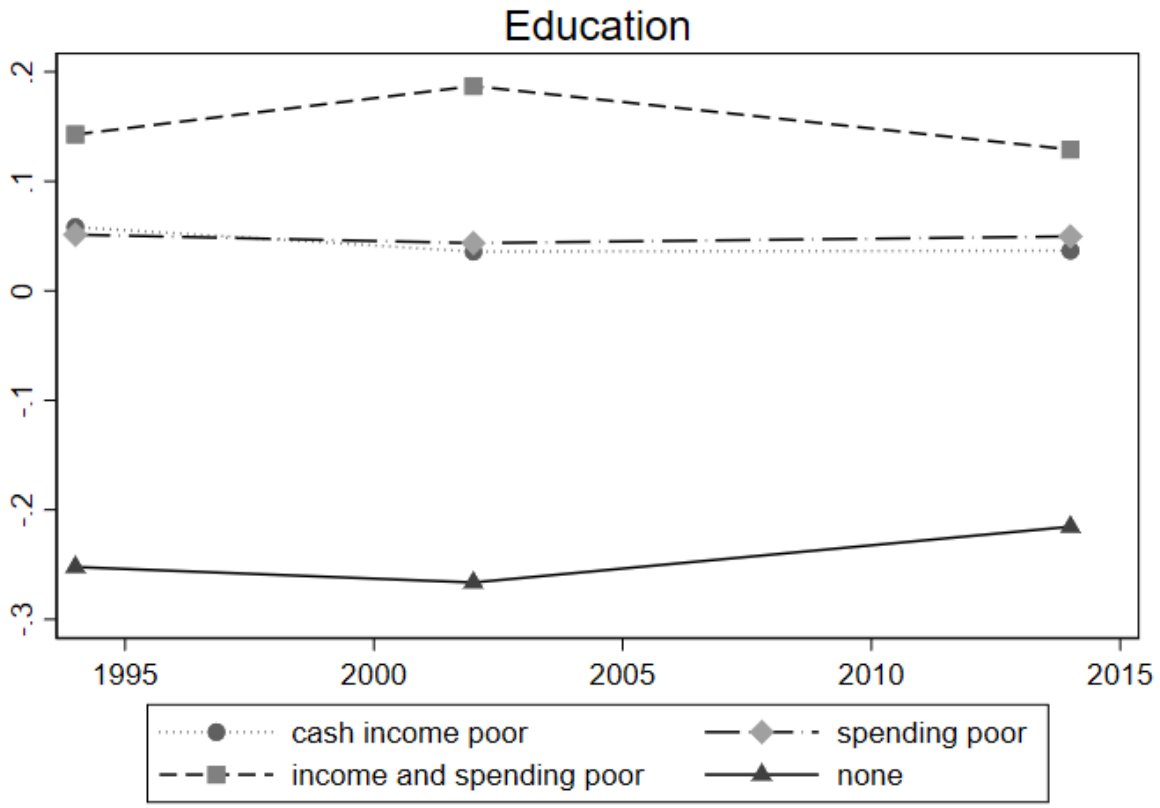
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.11: Average Marginal Effects of Rural on Cash Income and Spending poor Households



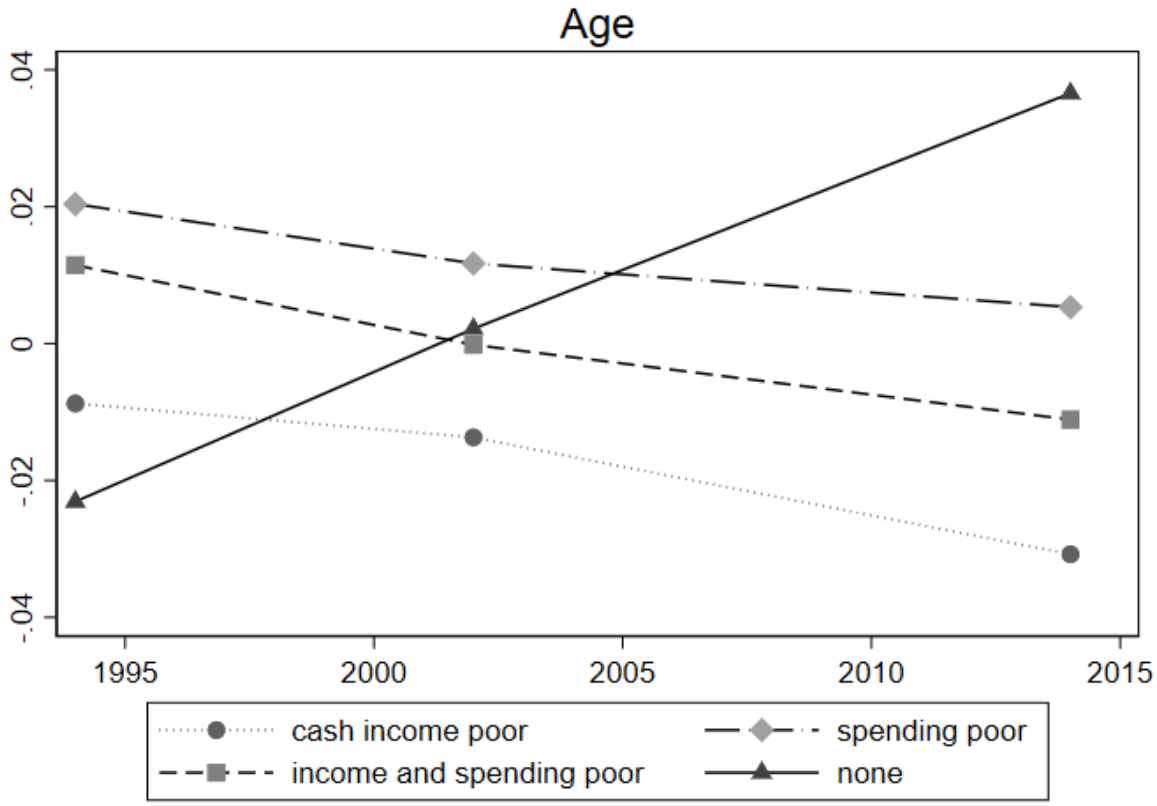
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.12: Average Marginal Effects of Number of People on Cash Income and Spending poor Households



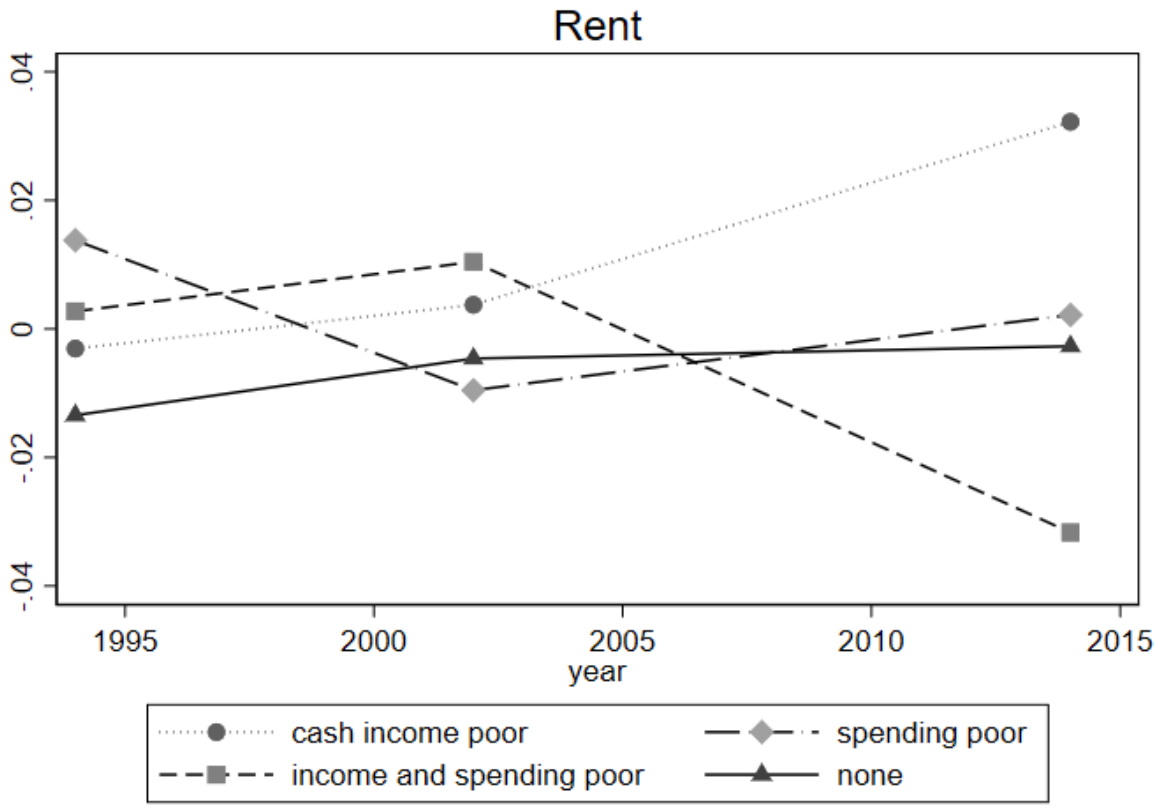
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.13: Average Marginal Effects of the Head of the Household Twelve years of education on Cash Income and Spending poor Households



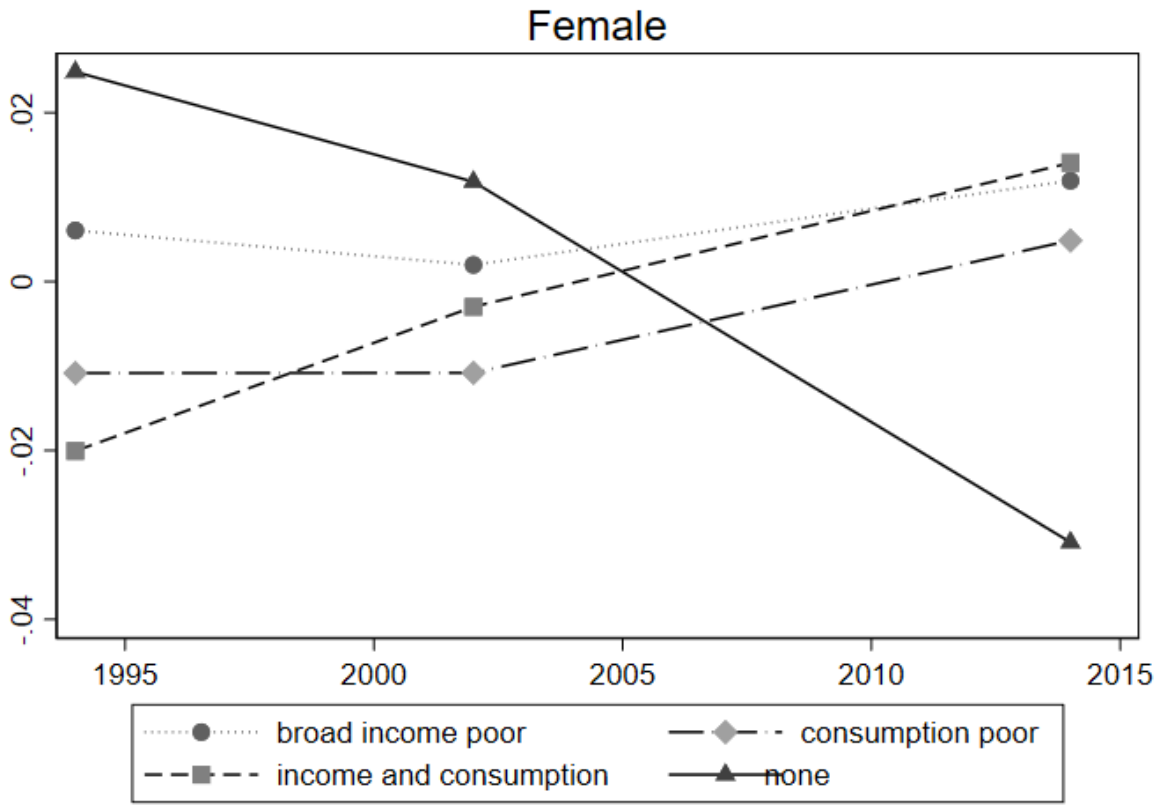
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.14: Average Marginal Effects of Having 60 years of age or more as the Head of the Household on Cash Income and Spending poor Households



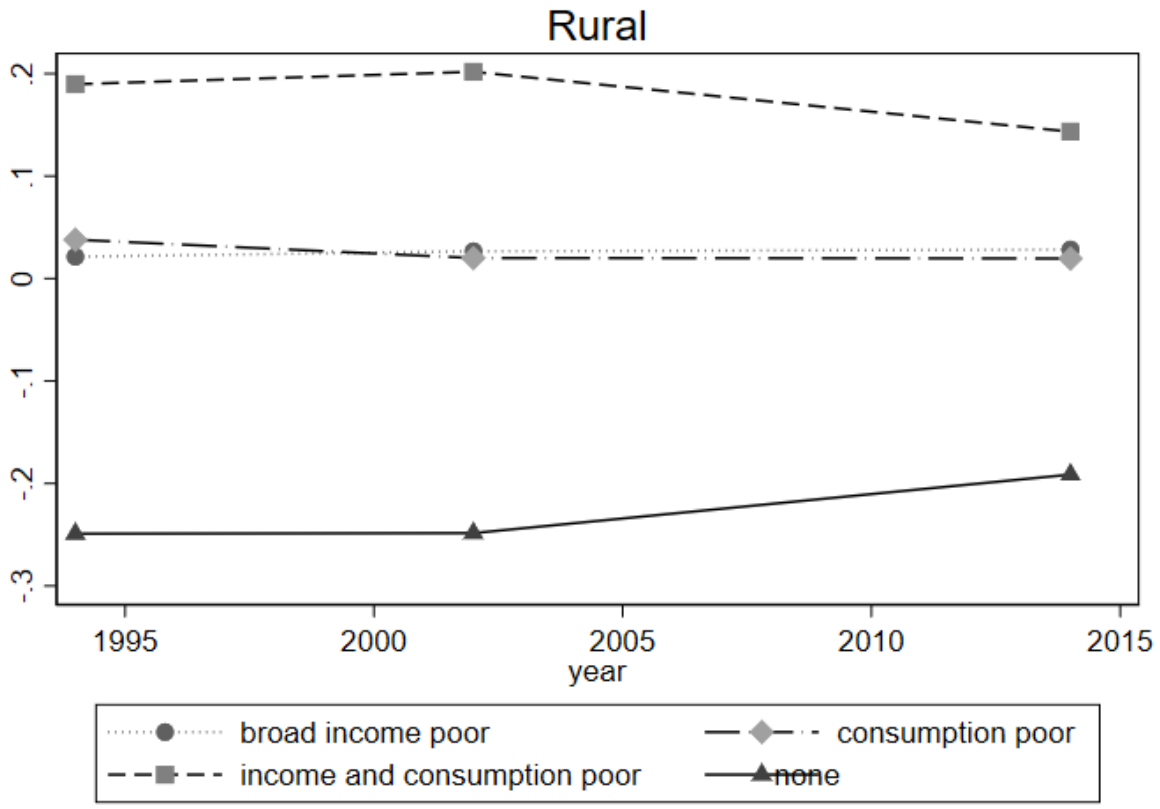
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.15: Average Marginal Effects of Renting a house rather than being the owner on Cash Income and Spending poor Households



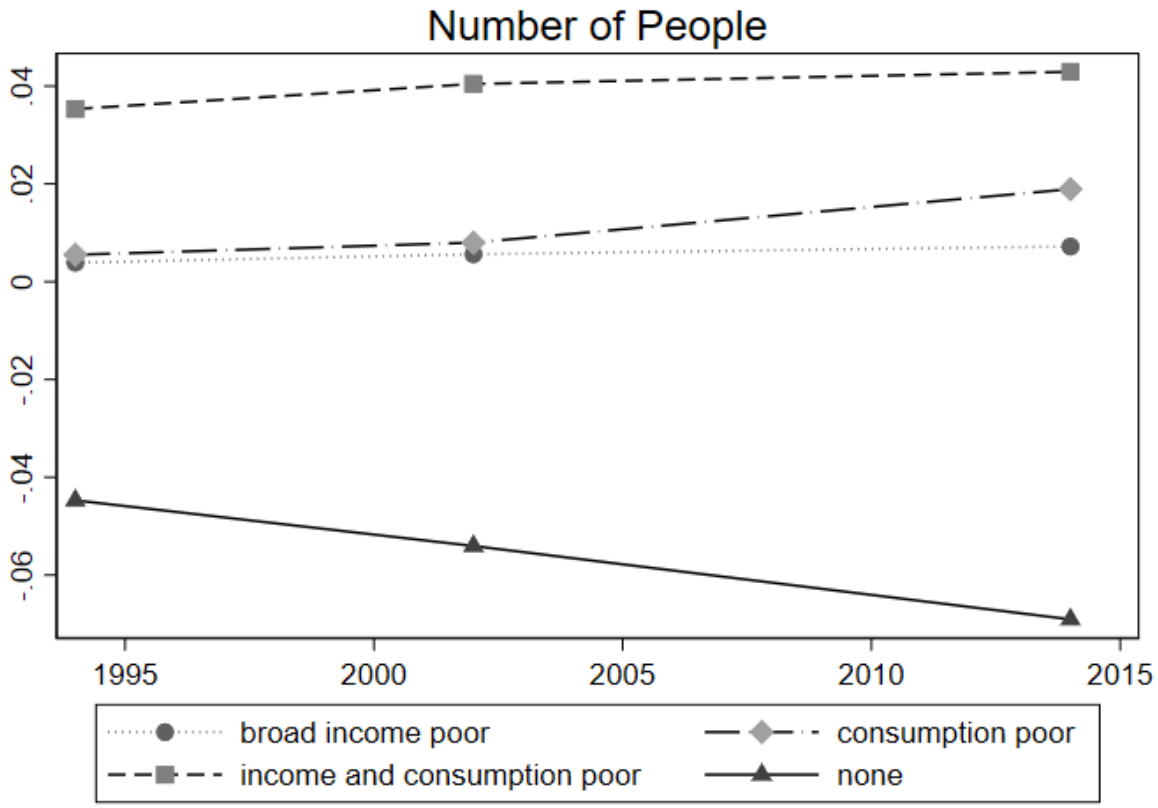
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.16: Average Marginal Effects of being Female as the head of the household on Broad Income and Consumption poor Households



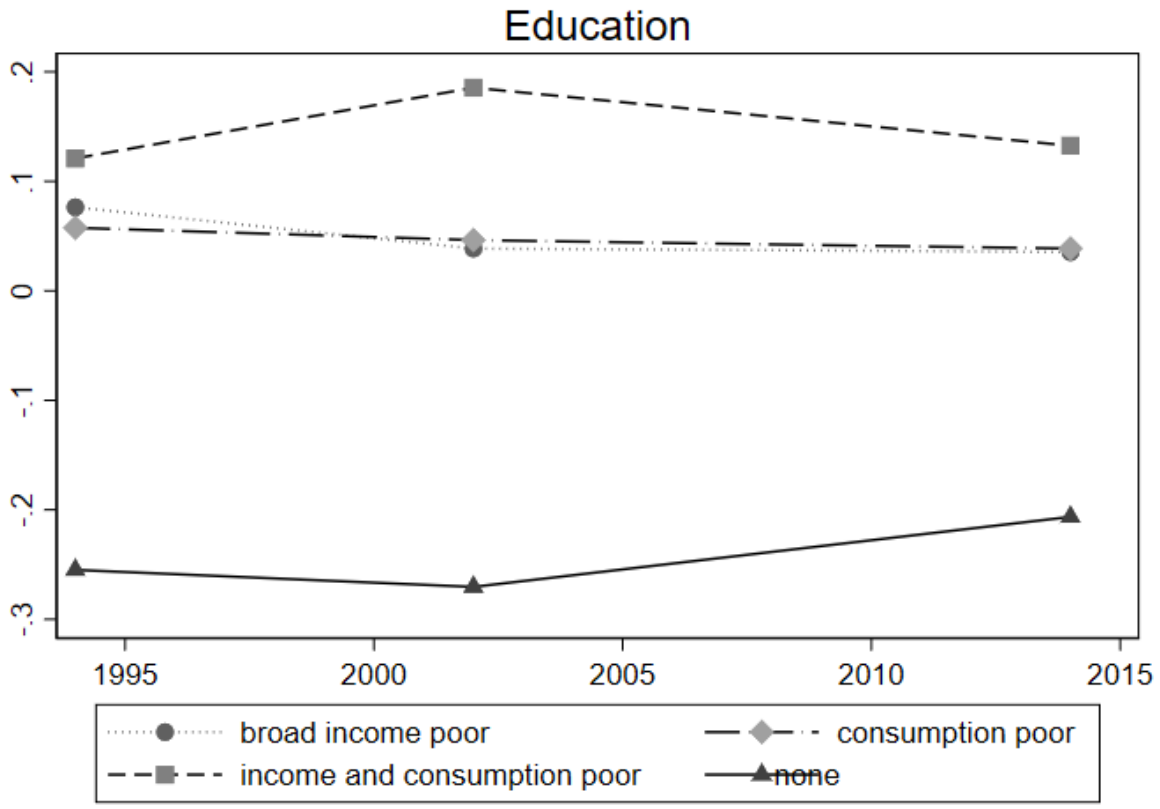
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.17: Average Marginal Effects of Rural on Broad Income and Consumption poor Households



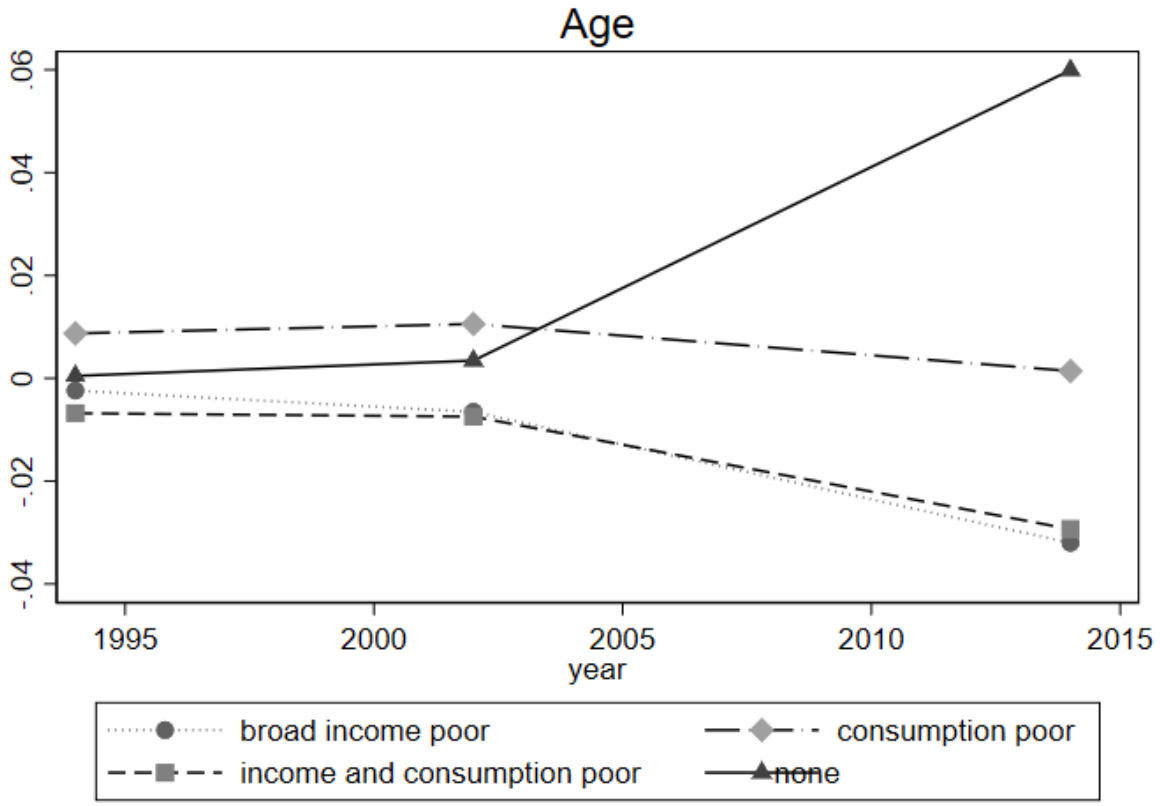
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.18: Average Marginal Effects of Number of People on Broad Income and Consumption poor Households



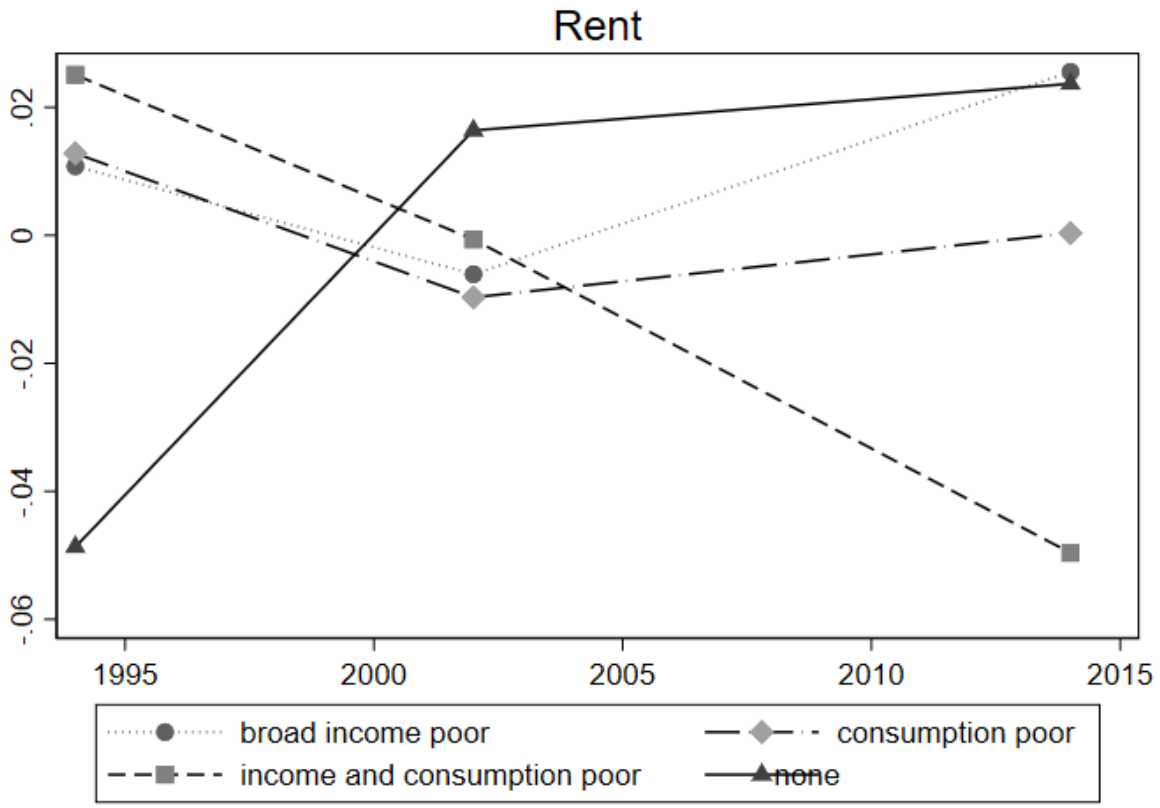
Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.19: Average Marginal Effects of the Head of the Household Twelve years of education on Broad Income and Consumption poor Households



Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.20: Average Marginal Effects of Having 60 years of age or more as the Head of the Household on Broad Income and Consumption poor Households



Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Figure 2.21: Average Marginal Effects of Renting a house rather than being the owner on Broad Income and Consumption poor Households

2.8 Tables

				percentile (pesos per month)						
1994	**Mean**	**Gini**	**Theil**	**5**	**10**	**25**	**50**	**75**	**90**	**95**
cash income	3,848	0.549	0.617	487	693	1,168	2,113	4,084	7,967	12,082
broad income	4,721	0.547	0.599	615	851	1,410	2,612	5,002	10,059	15,661
spending	3,222	0.512	0.503	497	673	1,103	1,912	3,561	6,747	10,053
consumption	4,096	0.518	0.509	617	813	1,353	2,438	4,520	8,680	13,559
2002										
cash income	3,624	0.511	0.492	538	739	1,258	2,187	4,021	7,450	11,669
broad income	4,330	0.517	0.508	631	849	1,443	2,561	4,821	9,172	14,204
spending	3,168	0.498	0.483	526	700	1,143	1,946	3,496	6,397	10,102
consumption	3,875	0.509	0.501	613	822	1,350	2,325	4,284	8,033	12,602
2014										
cash income	3,803	0.518	0.560	599	832	1,322	2,233	4,005	7,487	11,374
broad income	4,456	0.510	0.535	762	1,020	1,586	2,635	4,752	8,960	13,318
spending	3,272	0.485	0.481	647	847	1,276	2,054	3,459	6,402	9,315
consumption	3,924	0.482	0.469	795	1,022	1,544	2,461	4,199	7,745	11,343

					percentiles share (%)					
1994	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
cash income	1.2	2.2	3.0	3.9	4.9	6.2	8.0	10.8	16.0	43.8
broad income	1.3	2.2	3.0	3.9	4.9	6.3	8.0	10.7	16.1	43.6
spending	1.5	2.5	3.4	4.3	5.4	6.7	8.4	11.2	16.5	40.1
consumption	1.4	2.4	3.3	4.2	5.3	6.6	8.4	11.1	16.4	40.7
2002										
cash income	1.4	2.5	3.5	4.4	5.4	6.8	8.5	11.2	16.2	40.1
broad income	1.4	2.4	3.3	4.3	5.3	6.6	8.4	11.2	16.6	40.4
spending	1.6	2.7	3.6	4.6	5.6	6.8	8.6	11.0	15.9	39.6
consumption	1.5	2.6	3.5	4.4	5.4	6.7	8.5	11.1	16.3	40.0
2014										
cash income	1.5	2.7	3.5	4.3	5.3	6.5	8.2	10.7	15.5	41.8
broad income	1.6	2.7	3.6	4.4	5.4	6.6	8.2	10.7	15.7	41.1
spending	1.9	3.1	3.9	4.8	5.7	6.9	8.4	10.7	15.4	39.3
consumption	1.9	3.1	3.9	4.8	5.7	6.9	8.4	10.8	15.6	38.9

Source: Author's calculations using Encuesta Nacional de Ingreso y Gasto (ENIGH) from Mexico.

Table 2.1: Percentile shares by income and consumption definition

Chapter 3

The Gender Gap in Top Incomes in the UK, 1997/8 to 2016/7

3.1 Introduction

It is well known now that the share of total income going to the top of the income distribution has risen sharply over the last three decades in many countries, and especially the English-speaking countries (see Alvaredo et al. (2018), for example), with the UK having the second-highest share of income going to the top 1% amongst comparable developed countries. Within that rising share of top incomes, Atkinson et al. (2018) shows that, in many countries, the share of top income that is accrued by women is also rising, but that women remain under-represented in top income groups. They show that, for example, in 2011 (the most recent year of data used), the share of women in the top 1% in the UK was 17.8 %, and in the top 0.1%, it was only 9.2%.

This paper provides new analysis of top income shares in the UK from 1997/8 to 2016/7, making use of a sample of administrative tax records for the UK made available by HM Revenue and Customs (HMRC), and investigates in greater detail than Atkinson et al. (2018) the reasons for the rise in the female share of top incomes, as well as extending their analysis with five years' more data.

The data-set that we use, known as the Public Use Tape (PUT) version of the Survey of Personal Incomes (SPI)[1], is relatively easy for researchers to access. Access to the full SPI, by contrast, which lies behind some recent work on top income shares in Advani et al. (2020b) and Advani et al. (2020a), can only be done in a data-lab located in London, and access has not been possible at all since March 2020 when the coronavirus pandemic hit the UK. Like us, Atkinson et al. (2018) used the SPI PUT, but crucially, unlike them, we make full use of the so-called *composite cases* in order to uncover the link between incomes at the very top and occupation and gender. As we explain in Section 3.3.2, these composite cases reflect a form of data aggregation that HMRC performs on sample records on individuals at the very top of the income distribution, so as to reduce the risk of identification. The aggregation has a very limited impact on estimates of top income shares and the female share of top incomes, but after the aggregation it is not possible to use the PUT micro-data alone to analyse the distribution of industry sector or region of residence containing those in the highest income groups. However, HMRC publish supplementary information on the characteristics of the individuals whose records were aggregated that we exploit so as to

[1]See HM Revenue and Customs KAI Data, Policy and Coordination (2019) and earlier years.

estimate these relationships in full. Specifically, we use decomposition methods to show the industries within which those on top incomes work, and how this varies between men and women, and the extent to which this can explain the rising female share of top incomes.

Our work extends that of Atkinson et al. (2018) by using the composite cases, and extending it by five more years, from 2011 to 2016. We find that the female share of top incomes in the UK continued to rise after 2011, although this rise is more evident among the top 1% than the top 10%. The share of women in the top 10 to 1% group reached 30% by 2016; among the top 1 to 0.1%, it rose from 13.2% in 1996 to 18.9% in 2016. However, there has been little change in the female share in the top 0.1% since 1997, although there are signs of a rising trend in the most recent years. These changes are overwhelmingly driven by earned income, rather than income from investments. They are accompanied by changes in the ages of women in top income groups, and their region of residence, with and increasing concentration of women in the top income groups being aged 45-54, and an increasing fraction living in London or the South East of England. The decomposition by industry shows that the rise in the female share of top incomes is overwhelming driven by increased female shares of income within individual industries. In the top 10 to 1%, and in the top 1 to 0.1% the industries contributing the most to the rising female share of income have been "Real estate, renting and business" and "Financial intermediation", both of which have seen an increased female of share of income within them, and have become more important industries as a share of all women with high incomes. As noted,

the rise in the female income share outside the top 0.1% has slowed slightly since 2009, particularly between the top 10 to 1%: this is due to a fall in the importance of "Health and social work" and "Education", two industries with above-average female income shares.

The rest of this paper is organized as follows. Section 3.2 presents the literature review, Section 3.3 discusses the PUT data and our methods, Section 3.3 presents our new results on top income shares and female income shares through to 2016, Section 3.4 presents the shift-share analysis exploring the role of within and between industry changes in explaining the rising female top income shares. Section 3.5 concludes.

3.2 Literature Review

Our work extends a small literature on gender gaps in top incomes (including Atkinson et al. (2018), already mentioned). Most of the literature are disproportionately concentrated in the US, UK and Nordic countries. The reason is that research in this area depends on having access to high quality data from the top of the distribution. It is possible to separate the literature into two groups. The first group focuses on the growing importance of earned income in explaining the growth in top incomes among women, and the change in gender gaps at the top. The second group explores why women at the top of the income distribution in some countries still receive more of their income from unearned sources than do men. In the first group there is evidence for the US where Guvenen et al. (2020) use administrative data to analyze the dynamics of

earnings and gender differences among top earners. They find that between the 1980s and the 2010s women remain underrepresented in the top percentile groups. They show that the Finance and Insurance industry has increased its relevance among the top earners. However, in terms of the gender composition there has been little change. They find that the increase in women representation has nothing to do with their concentration in high earnings industries like Finance and Insurance. Their decomposition results show that industry composition does not account for the increase in women share among the top earnings groups. In the UK, Brewer et al. (2008) was one of the first to use the SPI PUT to look at the gender split of top incomes in the UK, but without exploiting the information in the composite cases. More recently, Burkhauser et al. (2020) decompose the increase of the share of women at the top 1% in the UK between 1999 and 2015. The authors use the Survey of Personal Incomes (SPI) and the Family Resources Survey (FRS) to implement a Oaxaca-Blinder decomposition for women and men at the top 1% separately. They find that the increase in the time spent in full-time education accounts for most of the increase in the share of women in the top 1%.

In the second group of the literature are studies from Nordic countries such as Boschini et al. (2017). The authors analyze gender differences in top incomes in Sweden from 1974 to 2013. They find that women's participation in top incomes has been increasing, but they are still a minority. They also find that women at the very top derive their income primarily from capital gains. A similar work for Finland by Terhi

et al. (2018) analyze the top income distribution from a gender perspective for Finland between 1995 and 2012. Using administrative data they show that the capital income share for women is larger than for men among the top one percent. Also, they show that among the top ten percent of the income distribution the share of women is less than 30 % and has not been improving during the period of analysis.

Our work is also relevant to the small part of the very large literature on gender gaps in labour market outcomes that focuses on high earners. For example, Fortin et al. (2017) ask to what extent the gender pay gap within those on the highest earnings is driving average pay gaps between men and women, and they look at the UK and other countries. Yavorsky et al. (2019) focus on the income of the households at the top 1% in the US between 1995 and 2016. Using the the Survey of Consumer finances the authors analyze how women's income contributes to the total household income and the associated characteristics. They find that there has been no closing of the gender gap among the top 1%.

3.3 Data and methods

3.3.1 Data

Public debate and academic research on income inequality in the UK mostly uses the *Households Below Average Income* (HBAI) series (a misleading name, as it covers the whole of the income distribution) (see Department for Work and Pensions (2020)). HBAI is the name of both a micro-data-set and a report released each year by statisticians in the Department for Work and Pensions, part of the UK government. The HBAI data-set in turn is derived from answers to a large, nationally-representative, government-run, household survey: the Family Resources Survey (FRS).

It has long been suspected that the FRS does not give an accurate impression of the circumstances of those with very high incomes in the UK. As a result, the HBAI data-set has, for a number of years, featured a correction to the highest incomes (affecting less than the richest 2%). However, the nature of the adjustment means that it is not possible to use the corrected HBAI data to analyse the characteristics of individuals with the highest incomes because the adjustment imputes the same level of disposable income to all households that report an income above a certain threshold. And Burkhauser et al. (2018b) and Burkhauser et al. (2018a) argue that, even after the SPI adjustment, the HBAI series may be under-estimating the income of those with very high incomes in the UK.

An alternative to using data from household surveys to study the circumstances of those with the very highest incomes is to use data from administrative sources on

incomes declared for tax purposes. This was pioneered by Thomas Piketty, Emmanuel Saez and the late Anthony Atkinson; their work was first collected together in Atkinson and Piketty (2007), and is now available at the World Inequality Database (WID) at https://wid.world). The advantages of using data from tax authorities to learn about the income of the very rich are that there is lot of data (because everyone has to pay taxes); data is available for a lot of countries over very long time periods (including countries and time periods where no estimates are available from household surveys); the information is usually confirmed against what employers and financial institutions think to be the case, and there are penalties for getting it wrong. On the other hand, tax authorities care only about the sort of income that is taxable (and so this will certainly exclude unrealised capital gains (but see Advani et al. (2020b)), but also sources of cash income that do not need to be declared because they are not liable to income tax), and they know about income only if is declared to them. Tax registers typically contain little information on demographic characteristics, and in countries with individual-level tax systems like the UK, it is usually not possible to link taxpayers who are married to each other.

We follow this line of work, using data from the Public Use Tape version of the Survey of Personal Incomes (HM Revenue and Customs KAI Data, Policy and Coordination (2019)). The SPI is a stratified random sample of taxpayers (plus some non-taxpayers) that over-samples those on higher incomes; see HM Revenue and Customs KAI Data, Policy and Coordination (2017) or Burkhauser et al. (2018b) for more details. Atkinson (2007) showed how to combine data from the SPI with other information (essentially,

estimates of the total UK population and of total economy-wide income) to estimate top income shares, and for a number of years, the SPI, with the corrections implemented by Atkinson (2007), was used as the source of data on the top income shares held by the WID (see Alvaredo (2017)). However, the headline series for the top 1% for the UK at the WID is now based on a DINA approach: see Blanchet et al. (2019).

SPI data is available for most years since 1995, although we do not use 1995 and 1996 for our decomposition. The documentation available for 1995 is not enough to fully expand the composite with the information on industry. For 1996 the sample is very small, and there are very few observations of women with very high incomes.[2]

3.3.2 Further details on processing the Survey of Personal Incomes and estimating top income shares

The SPI is a stratified random sample of taxpayers (plus some non-taxpayers) that over-samples those on higher incomes; see HM Revenue and Customs KAI Data, Policy and Coordination (2017) or Burkhauser et al. (2018b) for more details. To preserve anonymity amongst those on very high incomes (and particularly when grossing factors fall below 2, meaning that a very rich individual has a greater than 50% chance of being included in the sample), HMRC combines the information on certain individuals into what they call *composite cases*. The procedure for doing this is described in HM Revenue and Customs KAI Data, Policy and Coordination (2017). In practice, it means that these individuals are removed from the sample and replaced with a single composite case that is assigned

[2]We thank James Browne for useful discussions on this issue.

the average values of all financial variables, and the total grossing weight of the now-removed individuals. Information on the categorical variables is set to missing (-1), but information on the region of residence and the main industry of the individuals behind the composite case is published in an annex (HMRC always condition on sex and age-band when constructing composite cases). This procedure, then, does not alter total weighted income of the sample. We expanded the composite cases in order to have additional information on the gender, industry and region of very high income groups.

Control totals for 2015-16 and 2016-17

At the time of writing, estimates of top income shares and levels that have been calculated directly from the SPI were available from the WID up to 2014/15. To estimate total income for 2015/16, we followed the process in Alvaredo (2017), taking data from more recent versions of the UK Blue Book. We were not able to replicate exactly Alvaredo's values for 2014/5, and so our estimate for future years is obtained by multiplying Alvaredo's estimate for 2014/15 by our estimated growth rate in total income since 2014/15.

Process for dealing with composite cases

In order to expand the composite cases we follow the SPI documentation and replace replace each composite observation with a number of synthetic individual observations (as many as were combined into the composite case), and then we assign these synthetic individuals values of region of residence and industry as specified in the SPI annex. Of

course, we have no way of knowing what is the joint distribution of region and industry, and we do not report that in this paper, but we can use these composite cases to look at how income is related to region and industry.

3.3.3 Methods

Process for estimating top income shares and levels

We estimated all top income shares using the Stata command pshare Jann (2015), accounting for the grossing weights that are supplied with the SPI.[3] Following this procedure, and using the control totals provided by the WID, we also estimate the top incomes shares for the years available in the WID. Table 3.1 shows that we are not able to reproduce exactly the series at WID, but any differences are small; the mean difference between our estimated top 10% share and that for the WID for the years 1995 to 2015 is 0.096% (not percentage points, so we are out by less than 1 in a 1000); for the top 1% and top 0.1%, it is 0.229% and 0.417% respectively. Estimates for the top 0.01% (which are not available at WID) are likely to be subject to small sample bias (see Jann, 2016 for simulation results on this). When analysing the characteristics of those in various income centiles, and to work out the levels of income needed to be in various centiles, we work directly from the discrete distribution of income implied by the micro-data (rather than, e.g., using the micro-data to estimate a continuous income distribution function). So we define "the top x%" as the richest N observations where

[3]pshare allows for the total income to be specified as a parameter, rather than being calculated from the data. It does not allow one to do the same for the total population, and so we increased the sampling weights of those in the bottom half of the distribution until the sum of the weights equalled the known population control total.

the sum of the weights for the first N observations was strictly less than x% of the population (as given by the control totals), and the sum of the weights of the first N+1 observations was greater than or equal to x% of the population.[4].

Inference

Estimated standard errors for income shares were produced by the Stata command pshare (Jann, 2016). The sample size of the SPI has increased over time, and this explains the general fall over time in the size of the estimated standard errors seen in Figure 3.1. Estimated standard errors for the fraction of income that is earned, were computed using a bootstrap method. The SPI is a stratified sample with widely-varying sampling probabilities, but the Public Use Tape does not contain information that would allow researchers to identify the separate strata. As an approximation, we placed all observations with the same value of FACT, the grossing weight, into the same pseudo-strata. For each year, we then drew 999 stratified bootstrap samples using these pseudo-strata. The drawback to this method is that, for some years of the SPI, the grossing variable FACT can take some unusual values. In 2004/5, for example, there are 596 unique values of FACT, and so we have 596 pseudo-strata, 10.9% of which have 1 observation, and 23.1% of which have 10 or fewer observations (the 1-observation-strata account for 0.03% of the weighted population, and the strata with 10 or fewer

[4]The SPI comes with a set of grossing factors (FACT) which can be thought of as how many individuals in the UK are represented by the single entry in the micro data-set. No information is given on how these are calculated, but we assume they are akin to the inverse of the probability each individual had of being sampled. What is slightly unusual is that these grossing factors are not integers. Rounding or truncating these weights so that they become integers did not seem appropriate, and so we worked with non-integer weights throughout.

observations contain 0.12% of the weighted population (unweighted, these are 0.01% and 0.08%)).

3.3.4 Shift-share

This section describes the shift-share analysis that we use to decompose the change in the female income share into a term reflecting the change in the income share of each industry, and a term reflecting changes in gender participation within industry. Let's denote by TI_{ft} the total income share of women at time t. Using a standard shift-share decomposition, the change in TI_{ft} can be expressed as.

$$\Delta TI_{ft} = \sum_j \alpha_{fj} \Delta TI_{jt} + \sum_j \alpha_j \Delta TI_{fjt} \tag{3.1}$$

where TI_{jt} is the total income share in sector j, TI_{fjt} denotes the total income share of women in sector j, and $\alpha_{fj} = (TI_{fjt} + TI_{fjt-1})/2$ and $\alpha_j = (IT_{jt} + IT_{jt-1})/2$ are decomposition weights. The first term in equation (1) represents the change in female income share that is attributable to changes in the industry structure of the economy (between-industry component), while the second term reflects changes in the female income participation within each industry. This will account for the changes in incomes for women and how the changes in industry have contributed.

3.4 Results: top income trends

3.4.1 Recent Trends in Top Incomes

Figure 3.1 shows the new estimates of top income shares, including new estimates for 2015/16 and 2016/17. The share of pre-tax income that goes to the richest 1% of adults was at its lowest level in 1978, at slightly under 6%. Like the Gini coefficient, this measure of inequality rose through the 1980s, but it then continued to rise through the 1990s and the 2000s: in fact, the share of income going to the top 1% grew by more between 1990 and 2009 than it did in the 1980s. The share of income going to the top 0.1% went up by a half between 1996 and 2009, to reach 6.5%, or 65 times as much as in a world where income was shared equally (and the richest 0.01% of adults in 2016/17 had just over 2.0% of income, or 200 times as much as they would have if all income was shared equally). Top income shares have rise so much over time, in 2009 (the least-equal year on record, according to this data) the richest 0.1 % had a larger share of national income than did the richest 1% in 1979 (the most-equal year). Top income shares did fall back considerably in 2009, after the financial crisis. The estimates since 2010 are missing for a couple of years, but by 2016/17, the share of income going to the top 0.1% was of 5.46% among the highest (after 2009/10), and the top 1% share was 13.8%).[5]

Figure 3.2 shows what fraction of top incomes are from earned income (as opposed to

[5]As a measure of inequality, these top income shares are telling a different story from the usual statistics that are based on household-level disposable income derived from surveys, where the Gini coefficient and the 90:10 have hardly changed since the early 1990s, and are lower now than immediately before the financial crisis (see Cribb et al., 2018; DWP, 2019). It is beyond the scope of this paper to explore these differences thoroughly: they could be due, for example, in differences between the distribution of before- and after- tax income; differences in individual- and household-level income, or differences in the accuracy of the income measures.

income from financial investments). In 2016/17, the vast majority of income in the top 10 to 1%, and even in the top 1 to 0.1%, is from earned income. Within the top 0.1 % income from financial assets come less than 20% of total income.

3.4.2 Gender Composition and Income Shares of Top Incomes

Figure 3.3 shows the fraction of adults in various top income groups who are female, beginning in 1996. It is clear that women are under-represented, especially at the very top of the income distribution, but that the female share is rising over time. The share of women between the top 10 to 1 % increased by around 8 percentage points between 1996 and 2016 to reach 30%. Among the top 1 to 0.1 %, the female share is lower, rising from 13.2% in 1996 to 18.9% in 2016. However, there has been little change in the female share in the top 0.1% (in the richest 0.05 percent, for example, the share in 1996/97 and 2016/17 was broadly the same, at around 11%), although there are signs of a rising trend in the most recent years.

Figure 3.4 show that women's income share has risen specially among the top income groups between the 10 to 0.1 percent. For the very top income groups between the 0.1 to 0.05 percent the trend is not clear. Between 2012 and 2015 it seems that the share increased by 4.4 percentage points. However by 2016 it dropped 0.4%.

Figures 3.5, 3.6 and 3.7 present the earned income by gender among the top 10 to 1 percent, 1 to 0.1 percent and the 0.1 percent. Between the top 10 to 1 in Figure 3.5 is clear that the earned income share for women is smaller than the one for men. On average the differences was of 4.8% during the period. However, the difference has been

declining. In 1996 the earned income share for women was 89% while for men 95%, by 2016 the earned income share dropped to 86 percent approximately while for men was 89%. The difference was reduced by 1.3 percentage points. For the income group between the 1 to 0.1%, the earned income share for both men and women is smaller than the previous income group as it is shown in Figure 3.6. The earned income share for women is on average 78% and for men 88% for the complete period. Nonetheless, the difference in the shares between men and women is bigger than in the previous income group. On average the difference is 10%. However, it has been declining: in 1996 the difference in share between men and women was 18% and by 2016 it was 6%. The last income group of those in the top 0.1 percent the earned income share it is even smaller than in the previous two income groups for both men and women. As it is shown in Figure 3.7 the earned income share for women is on average 69% while for men is 82% for the complete period. The differences in the earned income share are wider than in the previous income groups: on average the difference is 13%. The earned income share for women is lower than that of men among this income group, and lower than that of other income groups. It seem that the differences in shares have been declining but not as fast as in the top 10 to 1% group. In 1996 the earned income share for women was 58% and by 2016 was 78%. This increase it is not observed in the previous income groups.

In all income groups, then, women have a smaller share of earned income than men. But the difference has been declining over the period in question. For women in the top

1 percent, the fraction of income that is earned is higher now than it was in 1996, but the opposite is true for those outside the top 1 %.

Overall then, the fraction of women at the top of the UK income distribution has been rising, except at the very top, and has been caused by a rise in the number of women with high levels of earnings, rather than investment, income. In the top 10 to 1 %, the rise in the female share slowed down after the late 2000s; in the top 1 %, the share has continued to rise even in the most recent years.

3.4.3 Industry Composition of Top Incomes

The SPI data classifies every adult with the industry from which they earn most of their income, and the industry shares are shown in Table 3.3, 3.4, 3.5 and 3.6.

Table 3.3 shows that, within the top 10 to 1 %, there was a fall in the share between 1997 and 2016 in those who work in "Manufacturing", "Public administration" and "Education" by 8.1, 3.8 and 2.7 percentage points respectively. At the same time there were increases in the shares in "Real estate, renting and business", "Financial intermediation" and "Health and social work" of 8.5, 1.6 and 1.7 percentage points. Table 3.4 shows a very similar pattern for the top 1 to 0.1 %: a decline in "Manufacturing" and "Wholesale and retail", and an increase in "Financial intermediation" and "Real estate, renting and business" (note that over half of those in the top 1 to 0.1 % are concentrated in those two industries in 2016). For the top 0.1 %, Tables 3.5 and 3.6 confirm that the main changes were a rise in "Financial

intermediation" and in "Manufacturing". The rise in the share of the Financial intermediation industry has been documented by Guvenen et al. (2020) for the US. They show that by 2012 one-third of the workers in the top 0.1 % were in the finance and insurance industry, while in the 1980's the health care industry accounted for the largest share.

3.4.4 Distribution of women by age groups

Figures 3.8, 3.9 and 3.10 show the percentage share of women in different age bands. Within the top 10 to 1% (Figure 3.8) and the top 1 to 0.1% (Figure 3.9) we can see that the proportion of women between the age 45 and 54 has been increasing and at the same time younger women between 25 to 34 years old has been declining. Among women in the top 1 to 0.1 this is more evident: in 1997 women between 45 to 54 years old represented around 28%, but 2016 their share was up to 35% while women between the ages 25 to 34 had a fall in their share by seven percentage points between 1997 and 2016. In the top 0.01%, this pattern is even more pronounced: women between the ages 45 and 54 years old represent approximately 44% in 2016, while back in 1997 their proportion was 27%.

3.4.5 Regional distribution of women in the top 1%

Table 3.2 shows the regional distribution of women in the top 1%. More than fifty percent of the women form the top one percent are concentrated in London and South

East and this concentration has increased over time. In 1997 London concentrated 31% of the women and the South East around 20%. By 2016 the percentage of women in London increased by 8.5 percentage points while in South East 0.23 percentage points.

Women from top income groups are concentrated not only in specific industries such as Financial intermediation but also in two specific regions: London and South East.This concentration has increased over the years. In addition to that, we show that women in the top income groups are between 45 and 54 years old and that their proportion has been increasing at expenses of younger women. Despite the fact that women share has increased they are still under-represented at the very top. The next section will show the results form the shift and share decomposition to find out the contribution the contribution of specific industries to the increase in the women share among the top income groups.

3.5 Results of the shift-share decomposition

Here we present the results from the decomposition. First, we present the overall decomposition results, and then we discuss what the shift-share says about the importance of particular industries.

3.5.1 Overall results of the decomposition

This subsection shows the results from the general decomposition in Tables 3.7 to 3.10. All tables shows the female share of total income in the initial year and the final year (as plotted in Figure 3.4), and then show the contribution of the changing female share of income within individual industries, and the contribution of the changing importance of different industries, according to the formula in Section 3.3.4. The decomposition is carried out over the period from 1997 to 2016, and then split into the sub-periods before and after 2009; this choice of year is partly governed by the change in the way that industry is defined in the underlying data in 2009, but also corresponds to the period before and after the financial crisis (SPI data is not available for 2008). Figure 3.11, 3.12 and 3.13 show the key results graphically.

As was clear from Figure 3.4, the female share of income changed much more between 1997 and 2009 than between 2009 and 2016 for the top 10 to 1% (rising by 6.4 percentage points out of the total change of 7.3 percentage points). For the top 1 to 0.1% and top 0.05 per cent, the majority of the change happened in the first period (3.1 out of 5.1 percentage points, and 1.7 out of 2.9 percentage points). For the top 0.1 to 0.05%, the female share fell in the first period and rose in the second period.

For the four income groups, and for all time periods, we always observe that the within-industry effect is considerably bigger than the changing industry shares, and that the within-industry effect always takes the same sign as the overall change in the female share of income. In detail, among the top 0.1 to 0.05% and among the top 0.05 %, the effect of changing industry shares has been negative for the complete period (see Figure

3.11) and the subperiod 1997-2009 (see Figure 3.12). For the 2009-2016 subperiod, Figure 3.13 shows that the changing industry share contribution was positive only for the top 0.1 to 0.05%. On the other hand, Figure 3.12 shows that most of the increase in female income share happened before 2009. For the 2009-2016 period, the growth in the female income share was driven by the contribution of the within-industry change, while the industry share contribution had a negative effect on the overall increase in the income share, except for the Top 0.1 to 0.05% income group (see Figure 3.13).

Because the industry classifications changed in 2009, the results for the full period might be affected by our mapping from the post-2009 classification to the pre-2009 classification. Tables 3.9 and 3.10 show that this makes no difference to the overall decomposition results over the 2009-2016 period, at least to the level of accuracy shown here.

3.5.2 Contribution of different industries

Tables 3.11 to 3.16 show the detailed decomposition results, over the whole period and the two sub-periods, but only for the top 10 to 1% and the top 1 to 0.1% groups (small sample sizes in individual industries make some of the results for the top 0.1% somewhat erratic).

Over the whole period, the rising female share of income in the top 10 to 1% group was due to rising female share of income in individual industries. This was particularly marked in "Public administration" (from 17.8 to 30.9%), "Education" (from 40.7 to 51.1%), "Other services" (from 24.2 to 33.4%) and "Financial intermediation" (23.7 to

32.6%), but the decomposition identifies the rising female shares in “Real estate, renting and business” (21.0 to 29.3%) and “Manufacturing” (8.3 to 14.1%) as being the two more important industries, given those industries’ greater share of top incomes in 1997. The biggest changes in industry shares over the whole period were the rise in the importance of “Real estate, renting and business” and a decline in “Manufacturing”. Combining those two factors, the industries identified by the shift-share as explaining the most of the overall change over the period were “Real estate, renting and business” ,“Health and Social Work” and “Financial intermediation”.

In the top 1 to 0.1%, there were large increases in the female share of income between 1997 and 2016 in “Public administration” (10.3 to 22.6%), “Health and social work” (11.9 to 24.6%), “Hotels and restaurants” (14.4 to 23.5%) and “Utilities” (7.8 to 15.0%), but the decomposition identifies the rising female shares in “Real estate, renting and business” (12.3 to 19.4%) and “Financial intermediation” (10.9 to 17.4%) as being the two more important industries, given those industries’ greater share of top incomes in 1997. The biggest changes in industry shares over the whole period were the rise in the importance of “Financial intermediation” and “Real estate, renting and business” and a decline in “Manufacturing” and “Wholesale and retail”. Combining those two factors, the industries identified by the shift-share as explaining the most of the overall change over the period were “Financial intermediation” and “Real estate, renting and business”.

For both top income groups, the 1997-2009 period saw the most change in the overall female share of income. In the first sub-period, this was driven by "Real estate, renting and business" and "Health and social work" in the top 10 to 1% group, and "Financial intermediation" and "Health and social work" in the top percent. In the more recent 2009-2016 period, the small rise in female share of income in the top 10 to 1% group was driven by "Real estate, renting and business", which became more important and more female-dominated, but this was almost entirely offset by a negative contribution from "Education", reflecting its declining importance over this period and the fact that the female income share is greater than a half in this income bracket. The rise in female share of income in the top 1 to 0.1%t group in the 2009-2016 period was due to "Real estate, renting and business" and "Financial intermediation", with an offsetting downward pressure from "Health and social work", which declined in importance and has a relatively high female share of income.

Our results show that administrative data contains very rich information that help to characterize gender gap trends in a more accurate way comparing with survey data. This characterization based on a fully expanded administrative data could be potentially combined with survey data. Burkhauser et al. (2020) use both administrative data from SPI and survey data from Family Resources Survey to get additional sociodemographic characteristics from the top 1% and implement a regression-based decomposition. However, they did not use a fully expanded version of the SPI as we did in this paper because their aim was basically exploring in detail the

additional information that survey data can provide for top income groups. Our goal was different: we concentrate on getting the most detailed information that administrative data can provide. Potentially, administrative and survey data could be matched statistically to examine other dimensions of gender inequality at the top of the income distribution. However, this goes beyond the scope of our paper and can be set as a future research.

3.6 Conclusion

This paper has contributed to the literature on the gender gaps in top incomes, analysing a publicly-available sample of administrative data from the UK, but having taken steps to (safely) undo some of the anonymisation that is performed to data on the highest income individuals.

We show that women are under-represented at the top of the UK income distribution, but that this under-representation has been falling over the past two decades in the top 10 to 0.1%, and within the last decade within the top 0.1%. The rising share of women in top income groups is driven by women with earned income and accompanied by an increasing share of top income women being aged 45-54 and living in London or the South East of England. The decomposition by industry, which would not be possible without our undoing of some of the anonymisation, shows that the rise in the female share of top incomes is overwhelming driven by increased female shares of income within industries. In the top 10 to 1%, and in the top 1 to 0.1%, the industries contributing the most to the

rising female share of income have been "Real estate, renting and business" and "Financial intermediation", both of which have seen an increased female of share of income within them, and have become more important industries as a share of all women with high incomes. The rise in the female income share outside the top 0.1% has slowed slightly since 2009, particularly at lower top incomes: this is due to a fall in the importance of "Health and social work" and "Education", two industries with above-average female income shares.

We have also provided new estimates of top income shares for years not yet covered by the WID, and we have shown how, although full sampling and stratification details are not provided by the data owners, a bootstrap can be used to estimate confidence intervals around key statistics.

All in all we show that under-representation of women among top income groups in the UK has been falling. However, specific industries and regions have concentrate the improvements. Therefore, it remains necessary to implement specific policies to help to reduce the concentration in "Real estate, renting and business" and "Financial intermediation" industries. In this way, women from all the industries can have the same opportunities to increase their income and reduce the gender gap. Additionally, the concentration in London and the South East of England impose geographic barriers for those women living in different areas. More ambitious policies across regions need to be implemented in order to make other regions economically attractive for women. Policies that address geographical imbalances and the large dependence on the financial sector

would benefit women income across the distribution not just women from the top income groups.

References

Advani, A., Koenig, F., Pessina, L., and Summers, A. (2020a). Importing inequality: Immigration and the top 1 percent. *IZA Discussion Paper*, 13731.

Advani, A., Summers, A., et al. (2020b). Capital gains and uk inequality. *University of Warwick, Department of Economics*, No. 1260.

Alvaredo, F. (2017). Uk estimates of top income shares 2013-2014 and 2014-2015. *World Inequality Lab*, No. 201702.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). *World inequality report 2018*. Belknap Press.

Atkinson, A. B. (2007). The distribution of top incomes in the united kingdom 1908–2000. *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, 1:82–140.

Atkinson, A. B., Casarico, A., and Voitchovsky, S. (2018). Top incomes and the gender divide. *The Journal of Economic Inequality*, 16(2):225–256.

Atkinson, A. B. and Piketty, T. (2007). *Top incomes over the twentieth century: a contrast between continental european and english-speaking countries*. oup Oxford.

Blanchet, T., Chancel, L., and Gethin, A. (2019). How unequal is europe? evidence from distributional national accounts, 1980-2017. *WID. world Working Paper*, 6.

Boschini, A., Gunnarsson, K., and Roine, J. (2017). Women in top incomes: Evidence from sweden 1974–2013.

Brewer, M., Sibieta, L., and Wren-Lewis, L. (2008). Racing away? income inequality and the evolution of high incomes. *Institute for Fiscal Studies Briefing Note*, 76.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018a). Survey undercoverage of top incomes and estimation of inequality: What is the role of the uk's spi adjustment? *Fiscal Studies*, 39(2):213–240.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018b). Top incomes and inequality in the uk: reconciling estimates from household survey and tax return data. *Oxford Economic Papers*, 70(2):301–326.

Burkhauser, R. V., Herault, N., Jenkins, S. P., Wilkins, R., et al. (2020). What accounts for the rising share of women in the top 1%? *IZA Discussion Paper*, 13359.

Department for Work and Pensions (2020). Households below average income: An analysis of the uk income distribution: 1994/95-2018/19. Technical report, London:Department for Work and Pensions.

Fortin, N. M., Bell, B., and Böhm, M. (2017). Top earnings inequality and the gender pay gap: Canada, sweden, and the united kingdom. *Labour Economics*, 47:107–123.

Guvenen, F., Kaplan, G., and Song, J. (2020). The glass ceiling and the paper floor: Changing gender composition of top earners since the 1980s.

HM Revenue and Customs KAI Data, Policy and Coordination (2017). Survey of personal incomes public use tape documentation.

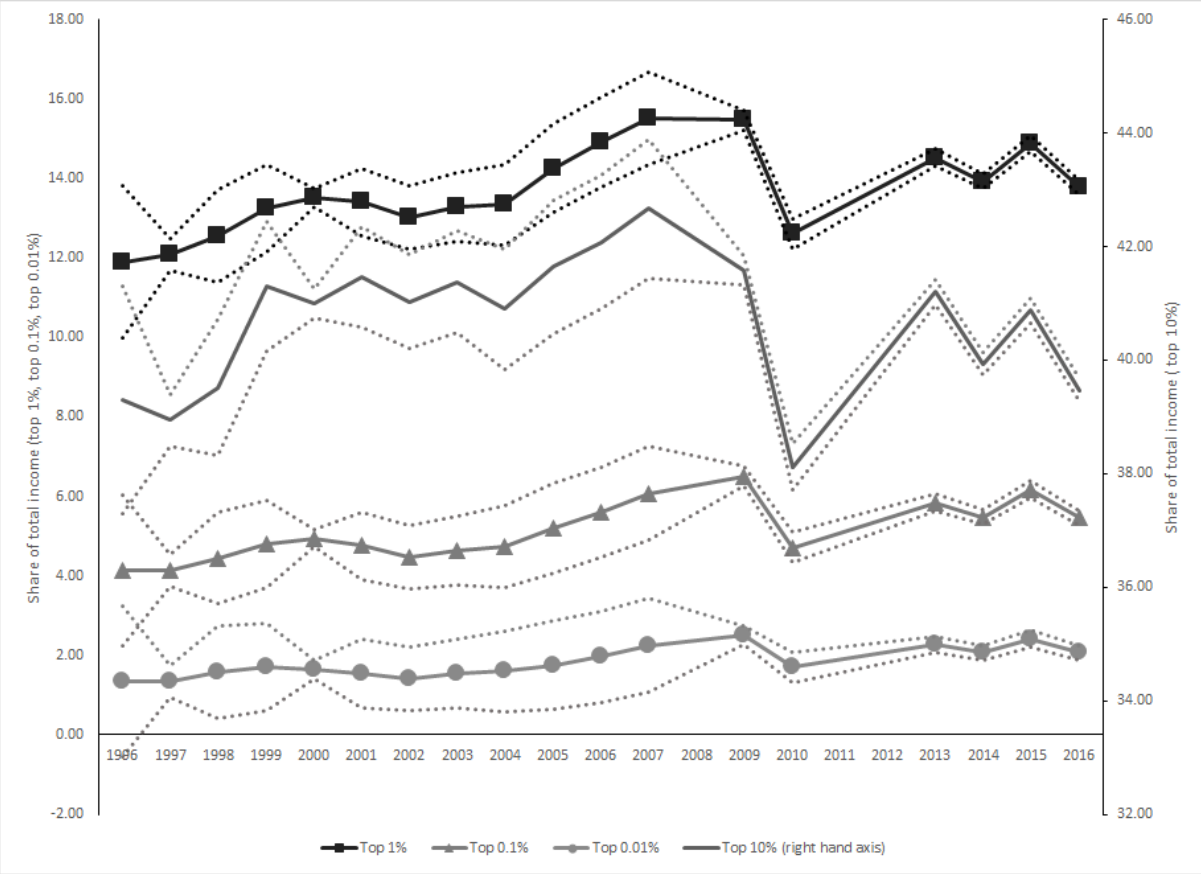
HM Revenue and Customs KAI Data, Policy and Coordination (2019). Survey of personal incomes, 2016-2017, public use tape, data collection.

Jann, B. (2015). PSHARE: Stata module to compute and graph percentile shares. Statistical Software Components, Boston College Department of Economics.

Terhi, R. et al. (2018). Top incomes and income dynamics from a gender perspective: Evidence from finland 1995-2012. Technical report.

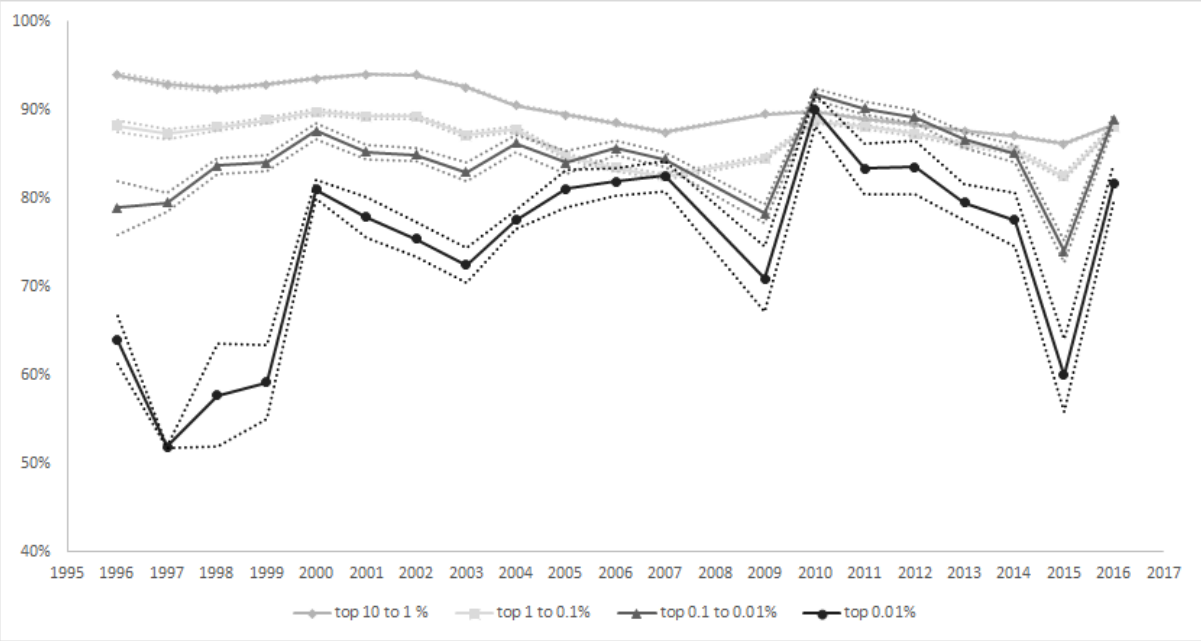
Yavorsky, J. E., Keister, L. A., Qian, Y., and Nau, M. (2019). Women in the one percent: Gender dynamics in top income positions. *American Sociological Review*, 84(1):54–81.

3.7 Figures



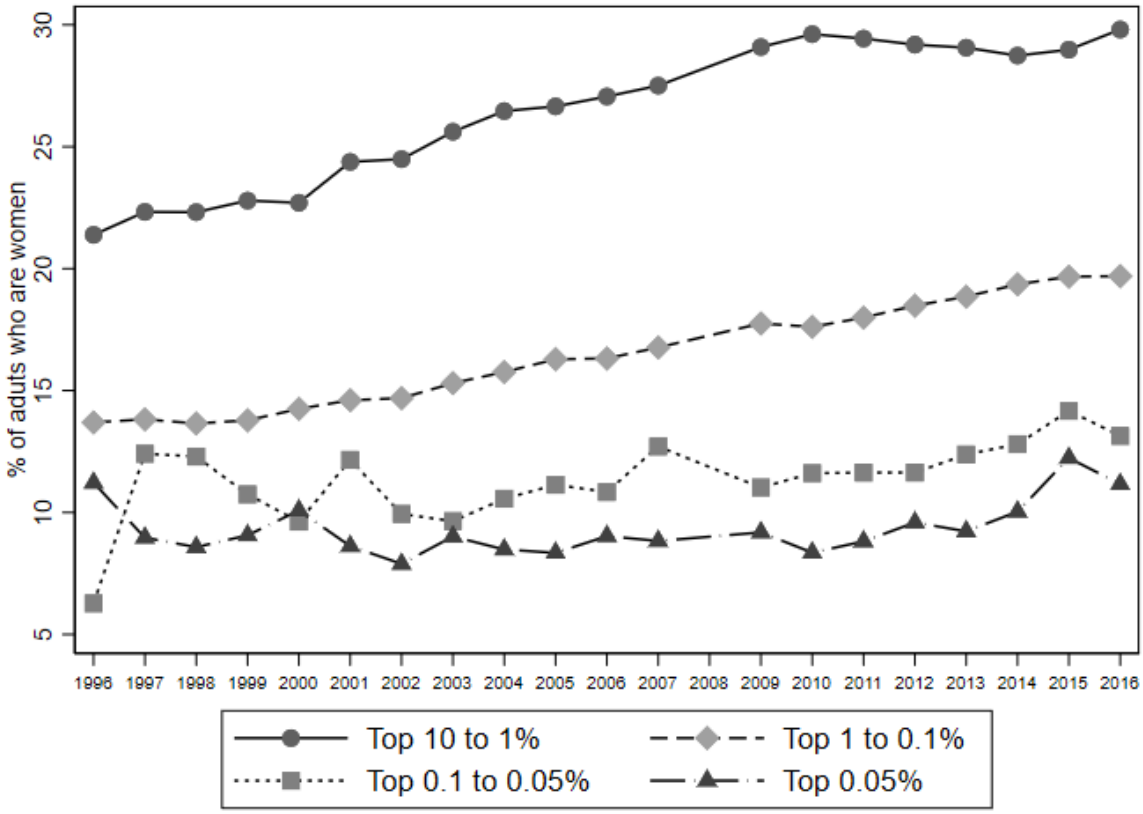
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.1: Top Income Shares in the UK 1996/7 to 2016/7



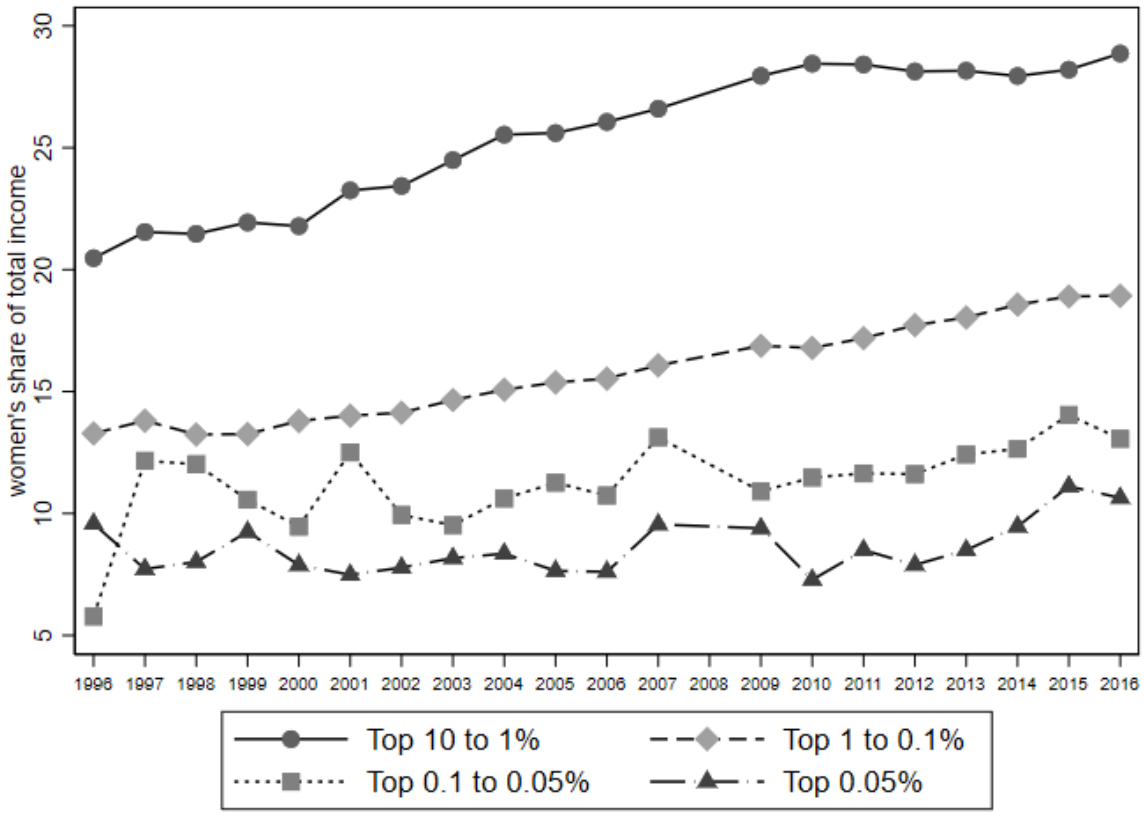
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.2: Fraction of top incomes in the UK that is earned, 1996/7 to 2016/7



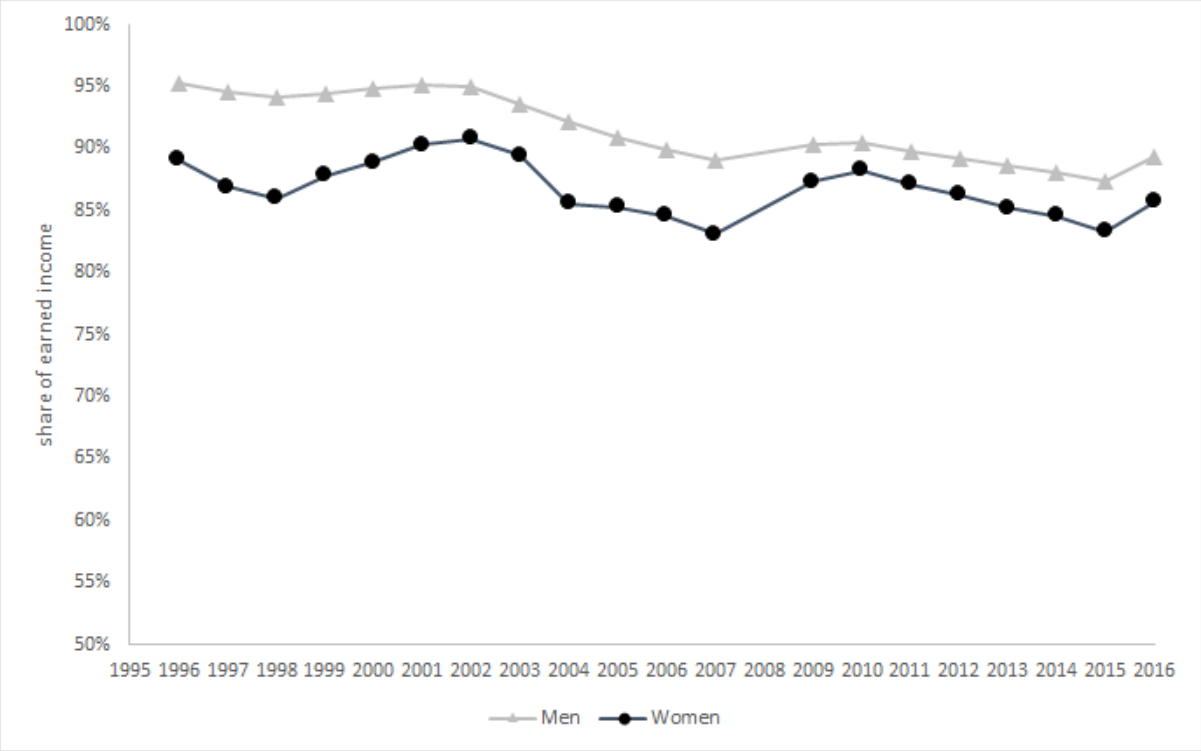
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.3: Share of women in top income groups in the UK, 1996/7 to 2016/7



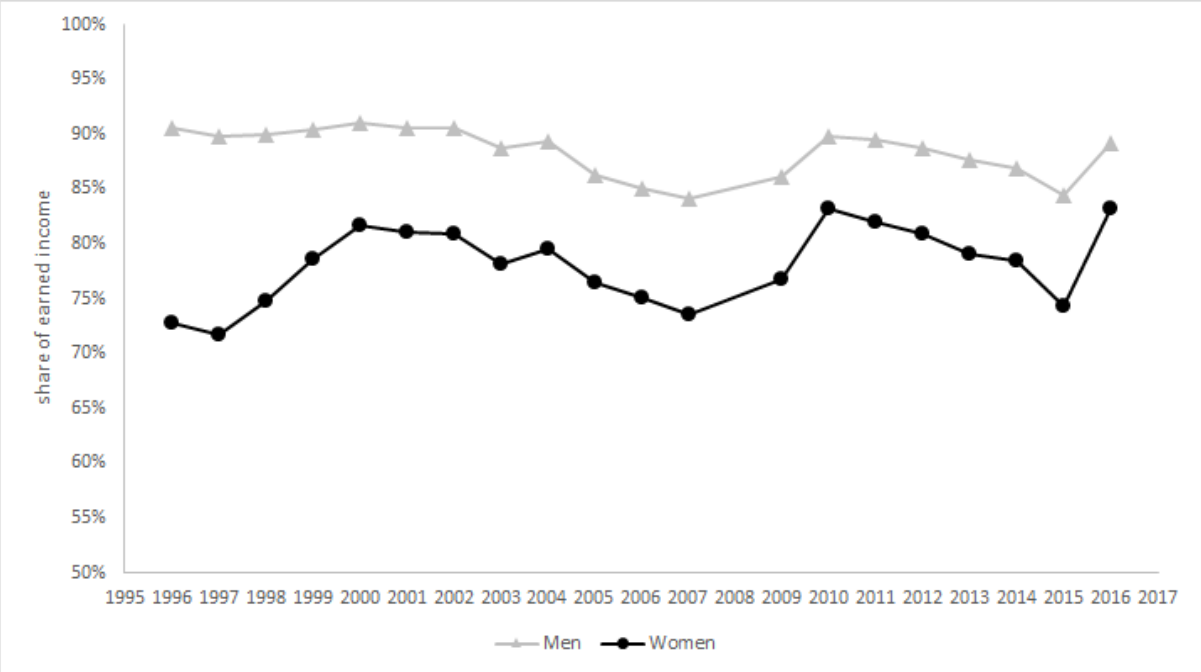
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.4: Women's share of total income by top income groups in the UK, 1996/7 to 2016/7



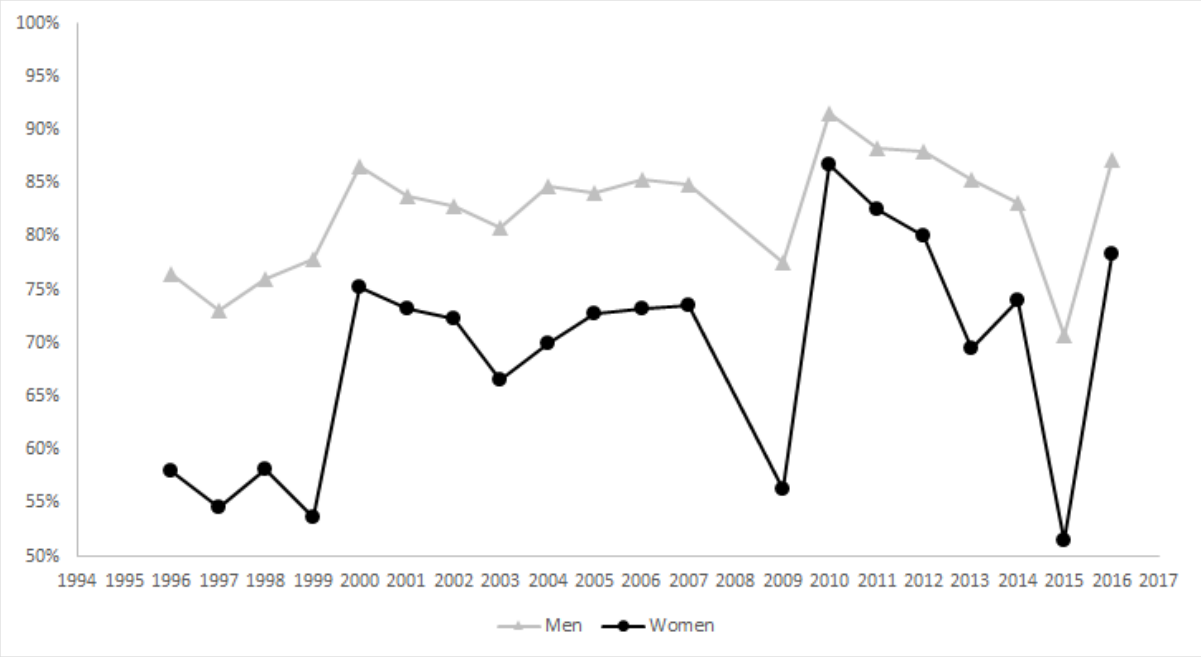
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.5: Earned income share for men and women in the top 10 to 1% in the UK, 1996/7 to 2016/7



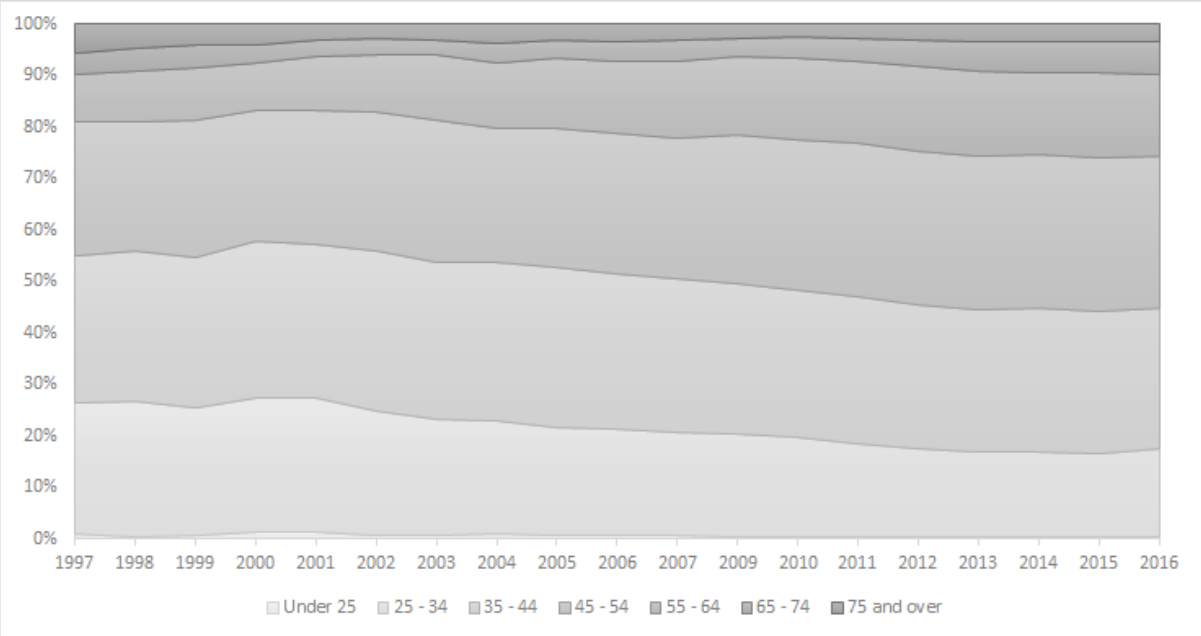
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.6: Earned income share for men and women in the top 1 to 0.1% in the UK, 1996/7 to 2016/7



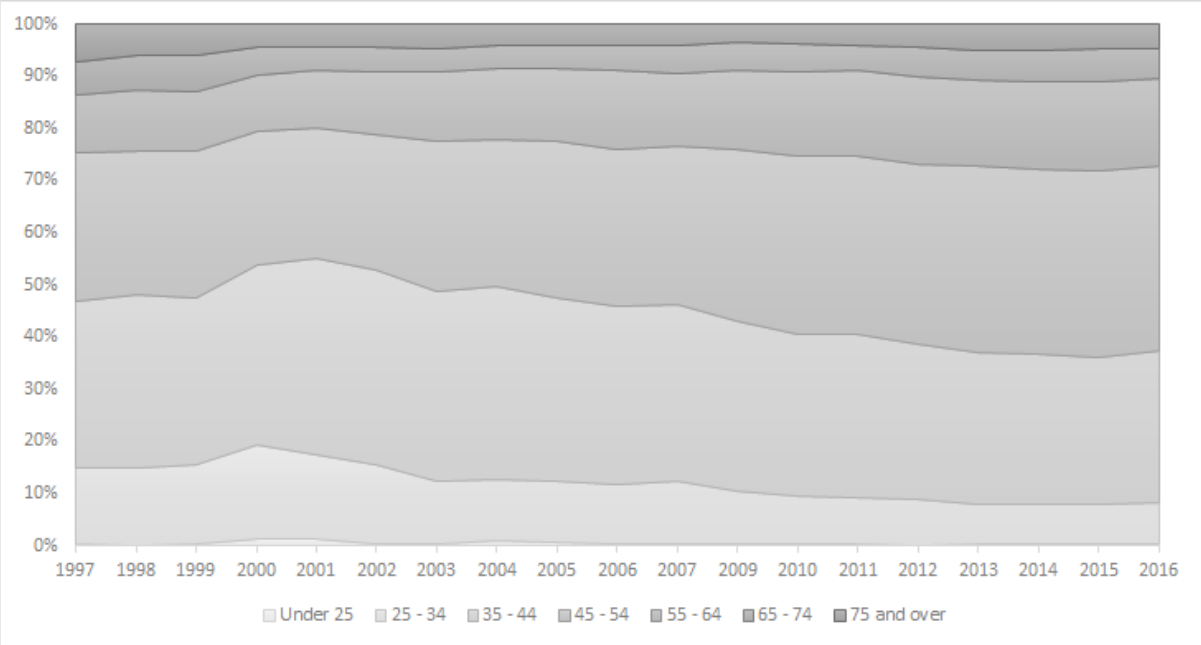
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.7: Earned income share for men and women in the top 0.1% in the UK, 1996/7 to 2016/7



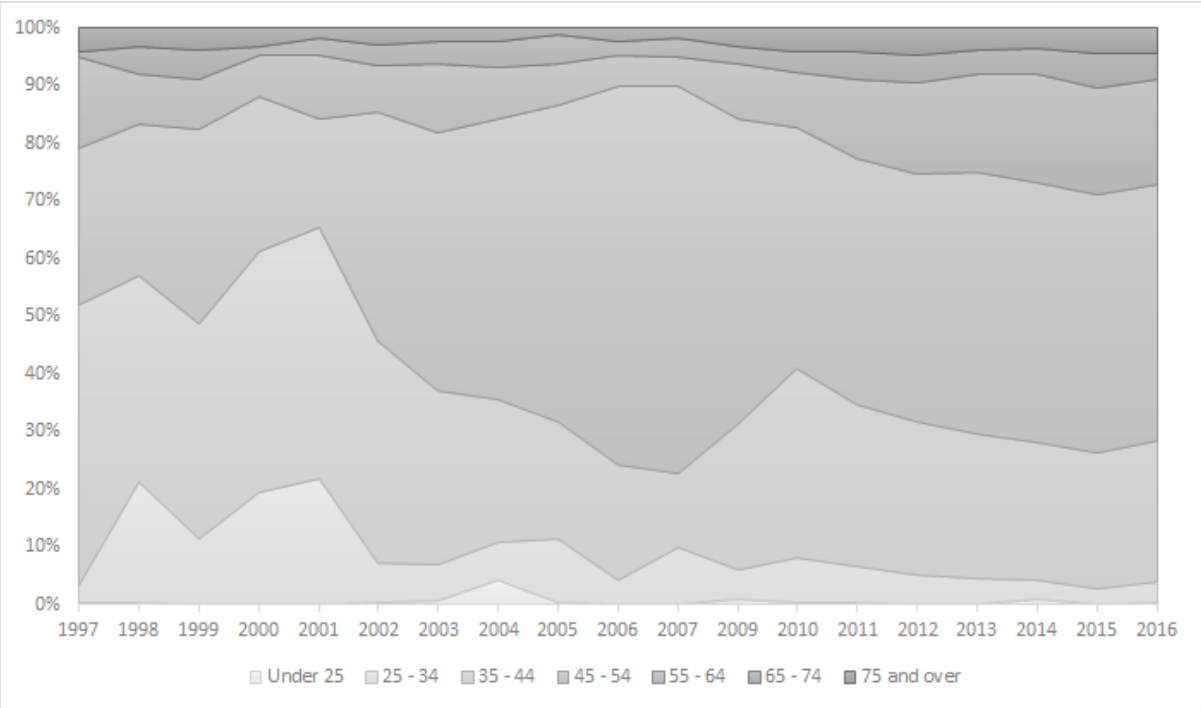
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.8: Share of women by group of age in the top 10 to 1% in the UK, 1997/8 to 2016/7



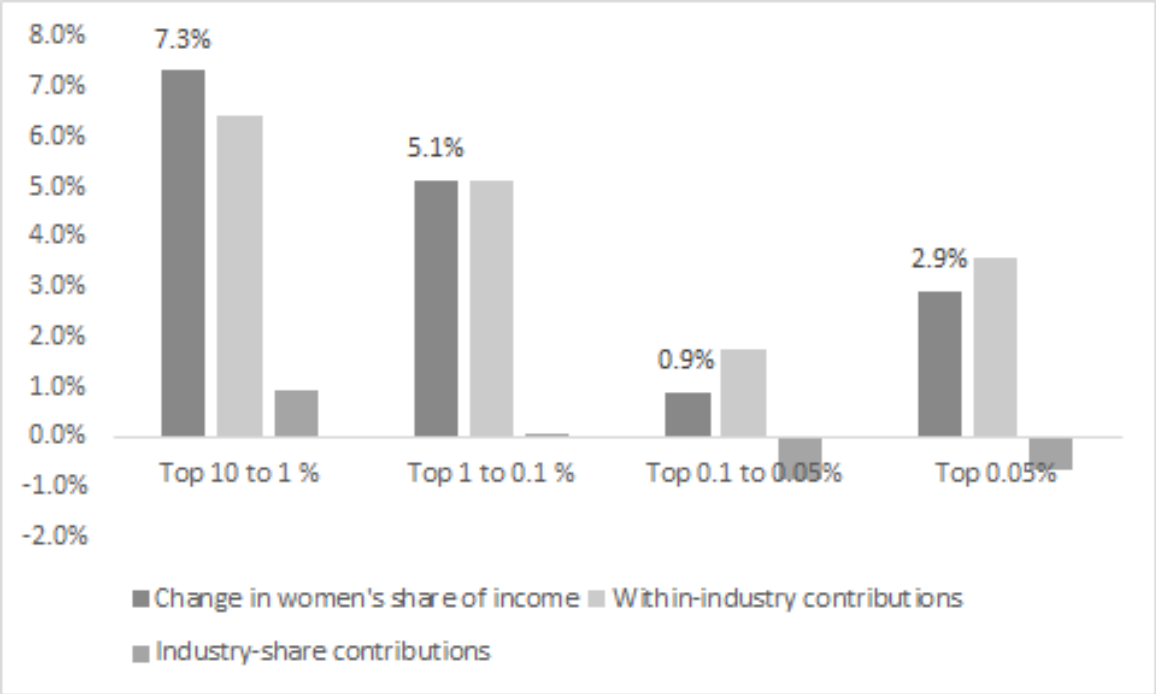
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.9: Share of women by group of age in the top 1 to 0.1% in the UK, 1997/8 to 2016/7



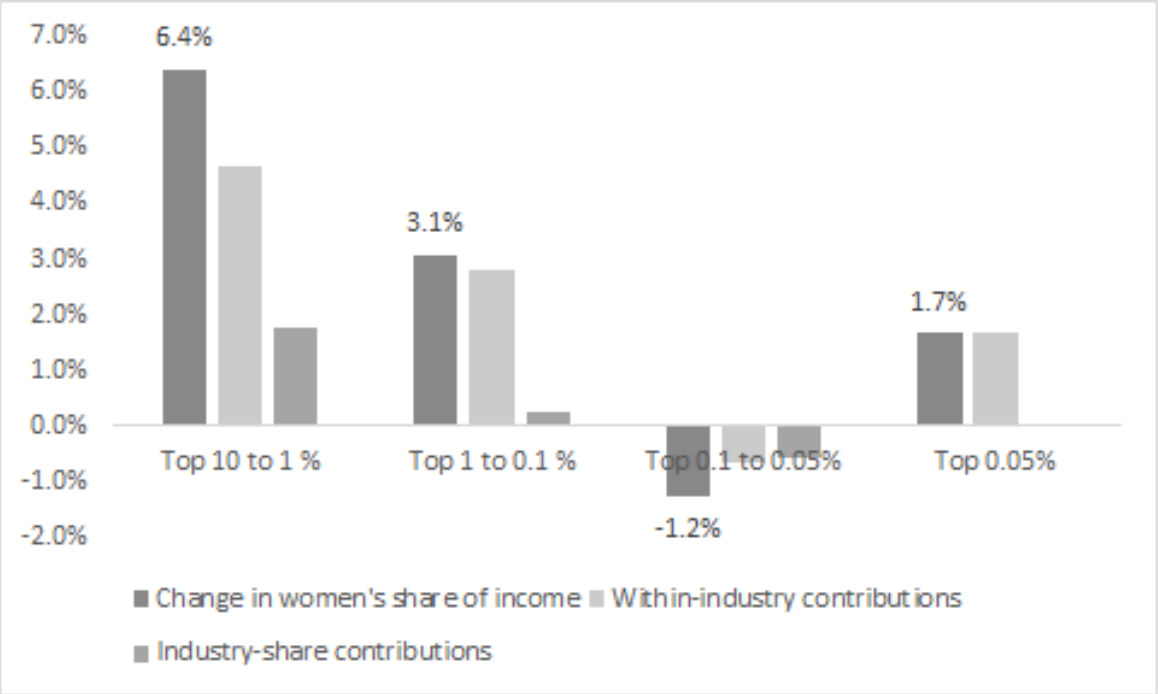
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.10: Share of women by group of age in the top 0.1% in the UK, 1997/8 to 2016/7



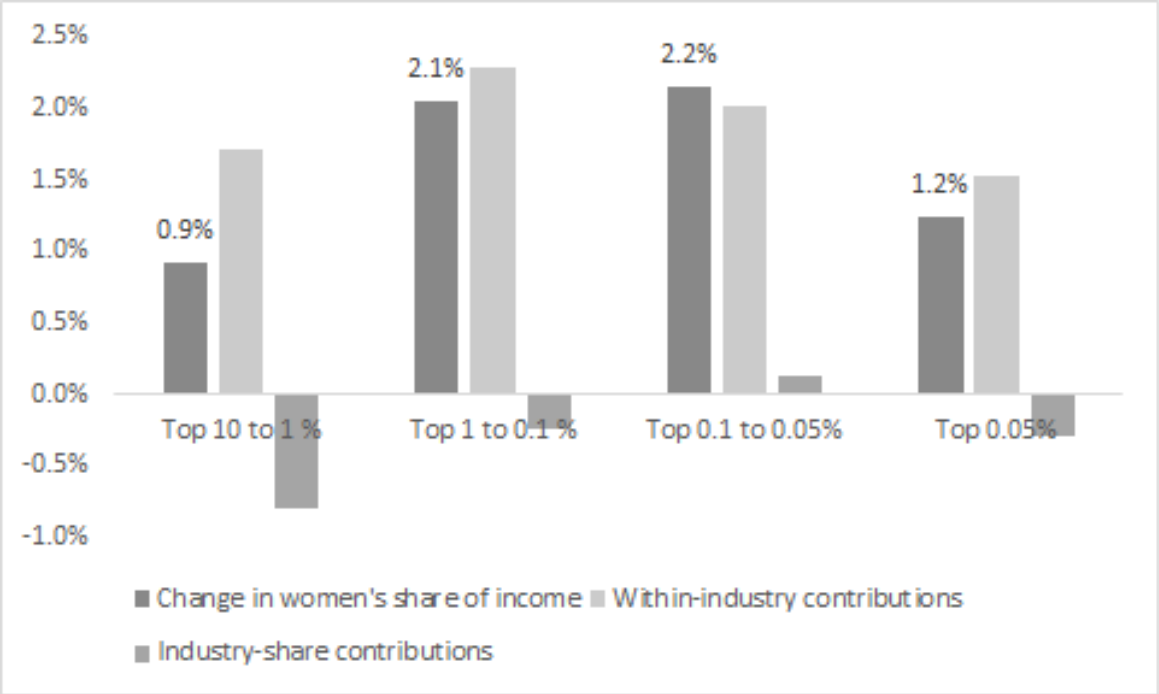
Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.11: Shift-share results, 1997 - 2016



Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.12: Shift-share results, 1997 - 2009



Source: SPI: authors' calculations using the SPI and the methods described in Section 3.

Figure 3.13: Shift-share results, 2009 - 2016

3.8 Tables

year	Top 10%		Top 1%		Top 0.1%		Top 0.01%
	SPI	WID	SPI	WID	SPI	WID	SPI
1996	0.39296	0.39300	0.11902	0.11901	0.04125	0.04130	0.01355
1997	0.38940	0.38940	0.12074	0.12071	0.04149	0.04150	0.01343
1998	0.39516	0.39470	0.12558	0.12530	0.04451	0.04440	0.01586
1999	0.41308	0.41329	0.13265	0.13239	0.04809	0.04795	0.01707
2000	0.40992	0.40984	0.13514	0.13508	0.04943	0.04936	0.01645
2001	0.41465	0.41411	0.13403	0.13386	0.04753	0.04753	0.01540
2002	0.41029	0.41011	0.13016	0.13027	0.04469	0.04487	0.01410
2003	0.41376	0.41404	0.13286	0.13239	0.04636	0.04571	0.01544
2004	0.40896	0.40828	0.13339	0.13300	0.04730	0.04711	0.01599
2005	0.41640	0.41609	0.14262	0.14224	0.05196	0.05177	0.01756
2006	0.42080	0.41990	0.14916	0.14820	0.05607	0.05548	0.01960
2007	0.42672	0.42615	0.15511	0.15440	0.06076	0.06050	0.02255
2009	0.41578	0.41528	0.15469	0.15420	0.06503	0.06460	0.02505
2010	0.38118	0.38083	0.12605	0.12550	0.04716	0.04660	0.01693
2013	0.41198	0.41290	0.14528	0.14530	0.05846	0.05841	0.02271
2014	0.39935	0.39990	0.13923	0.13880	0.05472	0.05480	0.02064
2015	0.40875		0.14883		0.06177		0.02416
2016	0.39469		0.13765		0.05456		0.02072

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.1: Top income shares in the UK a comparison of our estimates vs those at the World Inequality Database

Region	1997	2000	2009	2013	2016
North East	1.41	1.02	1.35	1.11	0.89
North West	6.15	6.63	5.80	5.34	5.19
Yorkshire and Humberside	4.51	3.53	4.21	4.01	3.38
East Midlands	4.78	4.26	4.19	4.21	3.72
West Midlands	4.22	4.29	4.73	4.25	4.02
Eastern	11.89	10.88	10.03	9.91	9.99
London	31.17	36.74	35.28	36.48	39.72
South East	19.87	20.24	19.19	19.56	20.10
South West	8.20	6.34	5.19	5.55	5.38
Wales	1.75	1.42	1.95	1.71	1.49
Scotland	4.71	3.55	5.34	5.09	4.47
Northern Ireland	0.84	0.95	1.30	1.12	1.15
Other/abroad	0.52	0.16	1.46	0.00	0.49
Unknown	0.00	0.00	0.00	1.67	0.01

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.2: Share of women by Region from the top 1 %

Industry	1997	2000	2009	2013	2016
Agriculture	0.009	0.007	0.005	0.005	0.004
Mining	0.008	0.006	0.005	0.007	0.006
Manufacturing	0.168	0.148	0.083	0.089	0.086
Utilities	0.011	0.010	0.013	0.014	0.015
Construction	0.047	0.058	0.057	0.050	0.056
Wholesale and retail	0.092	0.093	0.079	0.081	0.083
Hotels and restaurants	0.010	0.012	0.009	0.009	0.010
Transport, storage and comms	0.056	0.050	0.050	0.056	0.058
Financial intermediation	0.061	0.068	0.074	0.074	0.077
Real estate, renting & business	0.163	0.197	0.210	0.229	0.248
Public administration	0.072	0.050	0.049	0.040	0.033
Education	0.085	0.064	0.083	0.073	0.058
Health and social work	0.055	0.050	0.078	0.077	0.072
Other services	0.025	0.029	0.021	0.020	0.021
Other	0.003	0.003	0.005	0.007	0.007
Not in work	0.126	0.118	0.164	0.143	0.139
Claimants	0.001	0.000	0.000	0.000	0.000
International organisations		0.000	0.000	0.000	0.000
Unknown			0.000		
Missing	0.008	0.038	0.015	0.026	0.027

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.3: Industry shares of total income, top 10-1%

Industry	1997	2000	2009	2013	2016
Agriculture	0.016	0.006	0.005	0.004	0.004
Mining	0.008	0.007	0.007	0.010	0.007
Manufacturing	0.105	0.093	0.053	0.052	0.048
Utilities	0.005	0.004	0.006	0.006	0.004
Construction	0.027	0.034	0.034	0.027	0.030
Wholesale and retail	0.117	0.093	0.079	0.078	0.076
Hotels and restaurants	0.011	0.011	0.008	0.007	0.008
Transport, storage and comms	0.036	0.031	0.032	0.034	0.038
Financial intermediation	0.128	0.144	0.181	0.191	0.204
Real estate, renting & business	0.273	0.283	0.259	0.276	0.302
Public administration	0.009	0.005	0.009	0.006	0.005
Education	0.012	0.012	0.016	0.014	0.011
Health and social work	0.072	0.069	0.101	0.085	0.066
Other services	0.025	0.026	0.019	0.018	0.019
Other	0.007	0.004	0.008	0.009	0.009
Not in work	0.137	0.128	0.148	0.147	0.137
Claimants	0.000	0.000	0.000	0.000	0.000
International organisations		0.000	0.000	0.000	0.000
Unknown			0.000		
Missing	0.013	0.051	0.034	0.034	0.031

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.4: Industry shares of total income, top 1-0.1%

Industry	1997	2000	2009	2013	2016
Agriculture	0.013	0.005	0.008	0.004	0.006
Mining	0.005	0.004	0.010	0.005	0.005
Manufacturing	0.085	0.057	0.036	0.043	0.032
Utilities	0.004	0.003	0.006	0.004	0.003
Construction	0.030	0.021	0.049	0.022	0.018
Wholesale and retail	0.087	0.066	0.071	0.067	0.051
Hotels and restaurants	0.021	0.004	0.007	0.005	0.009
Transport, storage and comms	0.016	0.019	0.019	0.022	0.026
Financial intermediation	0.190	0.280	0.269	0.308	0.317
Real estate, renting & business	0.298	0.328	0.286	0.304	0.332
Public administration	0.001	0.003	0.003	0.002	0.003
Education	0.002	0.005	0.005	0.006	0.001
Health and social work	0.035	0.024	0.020	0.012	0.013
Other services	0.024	0.026	0.024	0.025	0.023
Other	0.008	0.004	0.007	0.011	0.011
Not in work	0.146	0.117	0.127	0.114	0.108
Claimants	0.000	0.000	0.000	0.000	0.000
International organisations		0.000	0.000	0.000	0.000
Unknown			0.000		
Missing	0.036	0.035	0.054	0.045	0.042

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.5: Industry shares of total income, top 0.1-0.05%

Industry	1997	2000	2009	2013	2016
Agriculture	0.011	0.008	0.009	0.006	0.009
Mining	0.001	0.005	0.006	0.004	0.002
Manufacturing	0.123	0.060	0.037	0.024	0.022
Utilities	0.003	0.008	0.005	0.004	0.002
Construction	0.017	0.012	0.034	0.023	0.025
Wholesale and retail	0.069	0.053	0.077	0.062	0.042
Hotels and restaurants	0.006	0.006	0.004	0.003	0.006
Transport, storage and comms	0.018	0.018	0.020	0.017	0.021
Financial intermediation	0.222	0.321	0.297	0.343	0.353
Real estate, renting & business	0.230	0.257	0.228	0.260	0.263
Public administration	0.001	0.001	0.002	0.000	0.001
Education	0.003	0.004	0.006	0.003	0.004
Health and social work	0.011	0.010	0.010	0.003	0.004
Other services	0.056	0.051	0.044	0.051	0.059
Other	0.008	0.008	0.013	0.011	0.016
Not in work	0.173	0.128	0.132	0.121	0.113
Claimants	0.000	0.000	0.000	0.000	0.000
International organisations		0.000	0.000	0.000	0.000
Unknown			0.000		
Missing	0.048	0.050	0.075	0.064	0.060

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.6: Industry shares of total income, top 0.05%

Income group	Women's share of income, 1997	Women's share of income, 2016	Change in women's share of income	Within-industry contributions	Industry-share contributions
Top 10 to 1 %	0.215	0.289	0.073	0.064	0.009
Top 1 to 0.1 %	0.138	0.189	0.051	0.051	0.000
Top 0.1 to 0.05%	0.122	0.131	0.009	0.018	-0.008
Top 0.05%	0.077	0.106	0.029	0.036	-0.006

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.7: Shift-Share results 1997-2016

Income group	Women's share of income, 1997	Women's share of income, 2009	Change in women's share of income	Within-industry contributions	Industry-share contributions
Top 10 to 1 %	0.215	0.279	0.064	0.047	0.017
Top 1 to 0.1 %	0.138	0.169	0.031	0.028	0.003
Top 0.1 to 0.05%	0.122	0.109	-0.012	-0.007	-0.006
Top 0.05%	0.077	0.094	0.017	0.017	0.000

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.8: Shift-Share results 1997-2009

Income group	Women's share of income, 2009	Women's share of income, 2016	Change in women's share of income	Within-industry contributions	Industry-share contributions
Top 10 to 1 %	0.279	0.289	0.009	0.017	-0.008
Top 1 to 0.1 %	0.169	0.189	0.021	0.023	-0.002
Top 0.1 to 0.05%	0.109	0.131	0.022	0.020	0.001
Top 0.05%	0.094	0.106	0.012	0.015	-0.003

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.9: Shift-Share results 2009-2016 consistent definition of industry

Income group	Women's share of income, 2009	Women's share of income, 2016	Change in women's share of income	Within-industry contributions	Industry-share contributions
Top 10 to 1 %	0.279	0.289	0.009	0.017	-0.008
Top 1 to 0.1 %	0.169	0.189	0.021	0.023	-0.003
Top 0.1 to 0.05%	0.109	0.131	0.022	0.020	0.002
Top 0.05%	0.094	0.106	0.012	0.015	-0.003

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.10: Shift-Share results 2009-2016 post 2009 definition of industry

Industry	Women's share of income in this industry, 1997	Women's share of income in this industry, 2016	Industry share of income, 1997	Industry share of income, 2016	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.191	0.233	0.009	0.004	0.000	-0.001	-0.001
Mining	0.078	0.114	0.008	0.006	0.000	0.000	0.000
Manufacturing	0.083	0.141	0.168	0.086	0.010	-0.011	-0.002
Utilities	0.102	0.153	0.011	0.015	0.001	0.001	0.001
Construction	0.059	0.098	0.047	0.056	0.002	0.001	0.003
Wholesale and retail	0.177	0.262	0.092	0.083	0.008	-0.002	0.005
Hotels and restaurants	0.298	0.341	0.010	0.010	0.000	0.000	0.000
Transport, storage and comms	0.144	0.179	0.056	0.058	0.002	0.000	0.002
Financial intermediation	0.237	0.326	0.061	0.077	0.005	0.005	0.011
Real estate, renting & business	0.210	0.293	0.163	0.248	0.014	0.025	0.038
Public administration	0.178	0.309	0.072	0.033	0.009	-0.012	-0.002
Education	0.407	0.511	0.085	0.058	0.009	-0.014	-0.005
Health and social work	0.492	0.568	0.055	0.072	0.004	0.010	0.014
Other services	0.242	0.334	0.025	0.021	0.002	-0.001	0.001
Other	0.596	0.564	0.003	0.007	0.000	0.002	0.002
Not in work	0.281	0.252	0.127	0.139	-0.004	0.003	-0.001
Missing	0.157	0.262	0.008	0.027	0.001	0.005	0.006

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.11: Shift-Share results for the top 10 to 1%

1997-2016

Industry	Women's share of income in this industry, 1997	Women's share of income in this industry, 2016	Industry share of income, 1997	Industry share of income, 2016	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.372	0.187	0.016	0.004	-0.003	-0.002	-0.005
Mining	0.042	0.088	0.008	0.007	0.000	0.000	0.000
Manufacturing	0.090	0.150	0.105	0.048	0.006	-0.009	-0.002
Utilities	0.078	0.150	0.005	0.004	0.000	0.000	0.000
Construction	0.097	0.105	0.027	0.030	0.000	0.000	0.001
Wholesale and retail	0.125	0.192	0.117	0.076	0.008	-0.008	0.000
Hotels and restaurants	0.144	0.235	0.011	0.008	0.001	-0.001	0.000
Transport, storage and comms	0.074	0.142	0.036	0.038	0.002	0.000	0.003
Financial intermediation	0.109	0.174	0.128	0.204	0.008	0.013	0.021
Real estate, renting & business	0.123	0.194	0.273	0.302	0.019	0.006	0.025
Public administration	0.103	0.226	0.009	0.005	0.001	-0.001	0.000
Education	0.259	0.276	0.012	0.011	0.000	0.000	0.000
Health and social work	0.119	0.246	0.072	0.066	0.009	-0.001	0.008
Other services	0.167	0.191	0.025	0.019	0.001	-0.001	0.000
Other	0.597	0.471	0.007	0.009	-0.001	0.001	0.000
Not in work	0.219	0.201	0.137	0.137	-0.003	0.000	-0.002
Missing	0.155	0.171	0.013	0.031	0.000	0.003	0.003

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.12: Shift-Share results for the top 1 to 0.1%

1997-2016

Industry	Women's share of income in this industry, 1997	Women's share of income in this industry, 2009	Industry share of income, 1997	Industry share of income, 2009	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.191	0.218	0.009	0.005	0.000	-0.001	-0.001
Mining	0.078	0.100	0.008	0.005	0.000	0.000	0.000
Manufacturing	0.083	0.137	0.168	0.083	0.009	-0.012	-0.003
Utilities	0.102	0.121	0.011	0.013	0.000	0.000	0.000
Construction	0.059	0.081	0.047	0.057	0.001	0.001	0.002
Wholesale and retail	0.177	0.242	0.092	0.079	0.006	-0.003	0.003
Hotels and restaurants	0.298	0.310	0.010	0.009	0.000	0.000	0.000
Transport, storage and comms	0.144	0.168	0.056	0.050	0.001	-0.001	0.000
Financial intermediation	0.237	0.308	0.061	0.074	0.004	0.004	0.008
Real estate, renting & business	0.210	0.265	0.163	0.210	0.009	0.012	0.021
Public administration	0.178	0.244	0.072	0.049	0.005	-0.006	-0.001
Education	0.407	0.503	0.085	0.083	0.008	-0.001	0.007
Health and social work	0.492	0.582	0.055	0.078	0.005	0.013	0.018
Other services	0.242	0.296	0.025	0.021	0.001	-0.001	0.000
Other	0.596	0.508	0.003	0.005	0.000	0.001	0.001
Not in work	0.281	0.246	0.127	0.164	-0.004	0.009	0.005
Missing	0.157	0.209	0.008	0.015	0.000	0.002	0.002

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.13: Shift-Share results for the top 10 to 1%

1997-2009

Industry	Women's share of income in this industry, 1997	Women's share of income in this industry, 2009	Industry share of income, 1997	Industry share of income, 2009	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.372	0.238	0.016	0.005	-0.002	-0.003	-0.005
Mining	0.042	0.100	0.008	0.007	0.000	0.000	0.000
Manufacturing	0.090	0.119	0.105	0.053	0.003	-0.006	-0.003
Utilities	0.078	0.140	0.005	0.006	0.000	0.000	0.001
Construction	0.097	0.097	0.027	0.034	0.000	0.001	0.001
Wholesale and retail	0.125	0.158	0.117	0.079	0.004	-0.006	-0.002
Hotels and restaurants	0.144	0.232	0.011	0.008	0.001	-0.001	0.000
Transport, storage and comms	0.074	0.130	0.036	0.032	0.002	-0.001	0.001
Financial intermediation	0.109	0.161	0.128	0.181	0.007	0.009	0.015
Real estate, renting & business	0.123	0.166	0.273	0.259	0.012	-0.002	0.009
Public administration	0.103	0.201	0.009	0.009	0.001	0.000	0.001
Education	0.259	0.218	0.012	0.016	0.000	0.001	0.000
Health and social work	0.119	0.197	0.072	0.101	0.006	0.006	0.011
Other services	0.167	0.153	0.025	0.019	0.000	-0.001	-0.001
Other	0.597	0.494	0.007	0.009	-0.001	0.001	0.000
Not in work	0.219	0.194	0.137	0.148	-0.003	0.002	-0.001
Missing	0.155	0.141	0.013	0.034	0.000	0.003	0.003

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.14: Shift-Share results for the top 1 to 0.1%

1997-2009

Industry	Women's share of income in this industry, 2009	Women's share of income in this industry, 2016	Industry share of income, 2009	Industry share of income, 2016	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.218	0.233	0.005	0.004	0.000	0.000	0.000
Mining	0.100	0.114	0.005	0.006	0.000	0.000	0.000
Manufacturing	0.137	0.141	0.083	0.086	0.000	0.000	0.001
Utilities	0.121	0.153	0.013	0.015	0.000	0.000	0.001
Construction	0.081	0.098	0.057	0.056	0.001	0.000	0.001
Wholesale and retail	0.242	0.262	0.079	0.083	0.002	0.001	0.003
Hotels and restaurants	0.310	0.341	0.009	0.010	0.000	0.000	0.001
Transport, storage and comms	0.168	0.179	0.050	0.058	0.001	0.001	0.002
Financial intermediation	0.308	0.326	0.074	0.077	0.001	0.001	0.002
Real estate, renting & business	0.265	0.293	0.210	0.248	0.006	0.011	0.017
Public administration	0.244	0.309	0.049	0.033	0.003	-0.005	-0.002
Education	0.503	0.511	0.083	0.058	0.001	-0.013	-0.012
Health and social work	0.582	0.568	0.078	0.072	-0.001	-0.003	-0.004
Other services	0.296	0.334	0.021	0.021	0.001	0.000	0.001
Other	0.508	0.564	0.005	0.007	0.000	0.001	0.001
Not in work	0.246	0.252	0.164	0.139	0.001	-0.006	-0.005
Missing	0.209	0.262	0.015	0.027	0.001	0.003	0.004

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.15: Shift-Share results for the top 10 to 1%

2009-2016

Industry	Women's share of income in this industry, 2009	Women's share of income in this industry, 2016	Industry share of income, 2009	Industry share of income, 2016	This industry's contribution to term A	This industry's contribution to term B	Total contribution from this industry
Agriculture	0.238	0.187	0.005	0.004	0.000	0.000	0.000
Mining	0.100	0.088	0.007	0.007	0.000	0.000	0.000
Manufacturing	0.119	0.150	0.053	0.048	0.002	-0.001	0.001
Utilities	0.140	0.150	0.006	0.004	0.000	0.000	0.000
Construction	0.097	0.105	0.034	0.030	0.000	0.000	0.000
Wholesale and retail	0.158	0.192	0.079	0.076	0.003	-0.001	0.002
Hotels and restaurants	0.232	0.235	0.008	0.008	0.000	0.000	0.000
Transport, storage and comms	0.130	0.142	0.032	0.038	0.000	0.001	0.001
Financial intermediation	0.161	0.174	0.181	0.204	0.002	0.004	0.006
Real estate, renting & business	0.166	0.194	0.259	0.302	0.007	0.008	0.016
Public administration	0.201	0.226	0.009	0.005	0.000	-0.001	-0.001
Education	0.218	0.276	0.016	0.011	0.001	-0.001	0.000
Health and social work	0.197	0.246	0.101	0.066	0.005	-0.008	-0.004
Other services	0.153	0.191	0.019	0.019	0.001	0.000	0.001
Other	0.494	0.471	0.009	0.009	0.000	0.000	0.000
Not in work	0.194	0.201	0.148	0.137	0.001	-0.002	-0.001
Missing	0.141	0.171	0.034	0.031	0.001	0.000	0.001

Source: authors' calculations using the SPI and the methods described in Section 3.

Table 3.16: Shift-Share results for the top 1 to 0.1%

2009-2016

Conclusions

There has been an increasing interest among policy makers and academics in inequality trends and poverty around the world. However, there is not yet conclusive evidence on their determinants, so as to allow policy makers to designs the best policies to reduce inequality and poverty levels. Motivated by the puzzling situation in Latin America where high levels of inequality and poverty are persistent but their trends over the last years have been declining, the first two chapters of this thesis focus on the impact of education on the declining trends of inequality in the region and a comparison of different measures of poverty and inequality in Mexico using income and consumption data. Finally, the third Chapter focuses on the gender gap in the United Kingdom among top income groups. The concentration of income during the last years has been a hot topic, however very little is known about the income differences between men an women in high income groups.

In the first Chapter of this thesis, I used survey data from 18 countries in Latin America for the period between 2000 and 2010. Employing the novel decomposition method based on RIF regressions, I found that the expansion of education increases inequality in most of the countries. However, if we take into account the changes in

returns to education, it reduces inequality in most countries. Returns to education fell between 2000 and 2010 across the region and are the main component of the effects of educational expansion on inequality. As a result, policies aiming to expand education need to consider their effects on inequality. In order to do that, governments need to analyse the shape of the returns to education to estimate the additional effects on inequality that could undermine the positive effects of the educational expansion for those at the bottom of the income distribution.

In the second Chapter of this thesis, I examine and contrast poverty and inequality measures using income and consumption data from Mexico between 1994 and 2014. Using data from Household Income and Expenditure Surveys, I found that growth was pro-poor, whether income or consumption are used to calculate inequality. Both definitions follow each other very closely, and during the period of economic crisis (1994-2002) consumption grew less than income. However, during the period of recovery (2002-2014), consumption grew more than income. The poverty analysis shows that households identified as income poor and consumption poor do not completely overlap and the overlap decreased by 2014. These results suggest there might be under-reporting at the bottom of the income distribution and one possible explanation is the increase in informal employment. Finally, I argue that consumption definitions of inequality and poverty are complements of income based definitions rather than substitutes. The complementary between income and consumption data might help policy makers to form a more complete picture of poverty and the characteristics of the

poor.

Finally, in the third Chapter, we analyzed the changes in the gender structure at the top of the income distribution in the United Kingdom over the last 20 years using the Survey of Personal Incomes. We show that women are under-represented at the top of the income distribution, but this under-representation has been falling over the past two decades. The rising share of women in the top income groups is driven by women with earned income and accompanied by an increasing share of top income women being aged 45-54 and living in London at the South East of England. The industries contributing the most to the rising female share of income have been "Real estate, renting and business" and "Financial intermediation" both of which have seen an increased female share of income within them, and have become more important industries as a share of all women with high incomes. The rise in the female income share outside the top 0.1 percent has slowed slightly since 2009, particularly at the lower top incomes: this is due to the fall in the importance of "Health and social work" and "Education", two industries with above-average female income share. We conclude that the reduction in the gender gap among top income groups is not a generalized phenomena across the country and industries. It remains necessary to implement policies that spread better opportunities to women across the industries and country.