# A STUDY OF THE CAUSES AND CONSEQUENCES OF REGIONAL INCOME INEQUALITY IN CANADA: A SPATIAL PANEL DATA APPROACH

Yannick Marchand
Department of Geography
McGill University
Montreal, Canada

November 2017

A thesis submitted to McGill University in
partial fulfilment of the requirements of
the degree of Master of Arts

Abstract

From 1981 to 2011, Canada, like many other OECD countries, experienced a sharp increase in income inequality. Using income data from the 20% sample of the long-form censuses and the national household survey, this thesis investigates the changing structure and nature of income inequality across Canadian regions. Analysis of these sub-national geographies of inequality suggest an increasing degree of spatial clustering (both high and low values) of inequality along an east-west divide as well as across the urban-rural spectrum, especially within large metropolitan areas and their surroundings. In fact, over the period of study, I find that large metropolitan areas are the main drivers of the observed increase in inequality at the national level. Levels of inequality in medium and small urban areas are also increasing, whereas on the other hand, for most rural areas inequality appears to remain stable over the last 30 years.

To understand the causes of such variations in regional levels of inequality, I develop and estimate a series of spatial panel error regression models. Results suggest that differences in the precariousness of regional labour markets, along with a weak manufacturing sector and increasing inequalities in terms of educational attainment all contribute to more unequal distributions of income. I also identify important differences in these causes across urban and rural regions.

Finally, I investigate the effects of income inequality on regional economic growth. Using the standard approach relying on cross-sectional models, the consistent pattern I find is that regions with initially higher levels of inequality do subsequently experience greater average annual growth rates over the long-run (from 1981 to 2011). In contrast, the short-/medium-term responses are different. Results from fixed-effects models point to a negative relationship between inequality and growth. Moreover, across both types of models, I find significant differences between urban and rural regions.

# Résumé

De 1981 à 2011, le Canada, comme plusieurs pays membres de l'OCDE, a connu une augmentation marquée des inégalités de revenu. À partir des données de l'échantillon 20% du recensement long et de l'Enquête nationale auprès des ménages, cette thèse a pour but d'observer les changements dans la structure et la nature des inégalités de revenu au niveau des régions Canadiennes. L'analyse des inégalités régionales suggère la présence grandissante de pochettes (« clusters ») d'inégalité (avec des valeurs élevées et basses) selon une division est-ouest, mais aussi selon le spectrum urbain-rural, et ce particulièrement à l'intérieur des grandes régions métropolitaines et leurs environs. En fait, au cours des trente années de la période d'étude, j'observe que l'accroissement des inégalités au Canada est principalement lié à la montée des inégalités dans les grandes régions métropolitaines. Les niveaux d'inégalités à l'intérieur des moyennes et petites régions urbaines sont aussi grandissants, alors que les niveaux d'inégalités en milieux ruraux sont demeurés stable au cours de la période d'étude.

Avec comme objectif de mieux comprendre les causes des variations dans les niveaux d'inégalités du revenu à travers les régions Canadiennes, j'ai développé et estimé une série de modèles de régression panel spatiale d'erreur. Les résultats suggèrent, entre autres, que les différences au niveau de la précarité du secteur d'emploi, la faiblesse du secteur manufacturier et les inégalités grandissantes dans les niveaux d'éducation favorisent l'accroissement des niveaux d'inégalités du revenu. Il existe cependant, d'importantes différences dans les causes des variations de l'inégalité à travers les régions urbaines et rurales.

Finalement, je me suis questionnés à savoir quels effets peuvent avoir des niveaux d'inégalités élevés sur la croissance économique des régions à court, moyen et long terme. En utilisant une approche standard qui s'appuie sur des modèles transversaux, il est possible de conclure que les régions qui possédaient des niveaux d'inégalités plus élevé au départ sont aussi les régions qui connaissent la plus grande croissance économique dans les trente années qui ont suivies. Par contre, cette relation ne s'est pas validée pour le court et moyen terme, alors que les modèles à effets fixes révèlent une relation négative entre inégalité et croissance économique. Une de fois de plus, ces résultats varient selon la division urbaine et rurale.

# ACKNOWLEDGEMENTS

My first thanks go to my supervisor Sébastien Breau, who with great patience and resourceful advice led me to better myself every step of the way. The writing of this thesis was a three-year journey, during which I had to pause when I discovered the joys of parenthood. However, the support I had from Sébastien and other members of the department, in particular Professors Oliver Coomes, Kevin Manaugh and Raja Sengupta, as well as my lab mates Diana Burbano and Alicia Cavanaugh, kept me going throughout my graduate studies at McGill. I cannot also forget the important group of individuals that work on the 7th floor who helped guide me through administrative issues; here, my thanks go to Nancy Secondo, Paula Kestleman, June Connolly and Catherine Klempa. Special thanks also to Joseph Varcirca whose assistance with my computer ensured it could handle the ambitious analytic models of estimaton in this thesis (especially in Chapter 4).

I also want to thank the staff at the Quebec Inter-University Center for Social Statistics (QICSS) laboratory and summer school, especially Danielle Forest who is the Statistics Canada analyst at the McGill-Concordia Research Data Center. And to Professor Jean Dubé at the Université Laval, who introduced me to spatial panel data models during the summer courses and helped me solidify the foundations of some of the statistical methods used in this thesis.

The funding provided by my supervisor Sébastien Breau, the Bradbury award and the QICSS grants was also important for the realization of this project

Lastly, and most importantly, I could not have completed this thesis without the support of my dear Marilie. You are the best partner and mother a child could have. And to Loïc, my son, hopefully one day you will read this and understand why I spent all this time on the computer.

# Table of contents

## Contents

## List of tables and figures

# CHAPTER 1: INTRODUCTION

In a speech back in 2013, President Obama remarked that he “believed [inequality was] the defining challenge of our time” (Obama, 2013). What he was referring to was the fact that by 2010, the total income share of the top decile of the income distribution had reached 50% in the United States, a level of inequality not seen in more than half a century (Piketty, 2014). The United States, of course, is not alone to experience such a phenomenon. In its 2011 report, *Divided We Stand*, the OECD showed that all but 5 of its 35 member countries have faced rising levels of inequality since the mid-1980s. Canada is among the majority group of OECD countries having faced such an increase. Moreover, as the report shows, Canada experienced one of the steepest increases in inequality over time (OECD, 2011).

Inequality is thus now at the top of the agenda for many world leaders (WEF 2015), and political deciders are keen for more empirical research on inequality to guide their choices of appropriate corrective public policies. Given recent advances in statistical techniques and available datasets, researchers are setting out to assess changes in the level of inequality in different places and at different times, all the while trying to identify the causes and consequences of those changing patterns. This thesis project seeks to add to this literature.

Building on research carried out by Breau (2015), this thesis proposes to examine patterns of income inequality at the regional level in Canada. Its contribution to the literature is three-fold. *First*, it provides a long-term perspective describing patterns of inequality from 1981 to 2011 (the latest year of data available to researchers) while emphasizing the contrast in urban vs. rural trajectories in the distribution of income. *Second*, spatial panel regression models are developed and estimated to shed light on the main determinants of differences in regional patterns of inequality. *Third*, standard cross-sectional models are used to investigate the effects of income inequality on regional economic growth. More specifically, the research carried out in the thesis is guided by the following key questions:

*(1) What are the trends and patterns of income inequality across Canadian regions (defined as census divisions) from 1981 to 2011?*

*(2) Do those trends differ along an urban and rural divide?*

*(3) What are the factors associated with increasing or decreasing regional levels of inequality?*

*(4) Do these factors have the same effects on rural and urban regions?*

*(5) What are the effects of income inequality on regional economic growth?*

To properly address these questions, I structured my thesis into four core chapters. After a short overview of the literature, I begin with a presentation of the methodology in Chapter 2. A first set of descriptive statistics is then presented in Chapter 3. Chapters 4 and 5 follow with a presentation of results from spatial panel regression and cross-sectional models.

The thesis findings point to the changing face of regional income inequality in Canada over the 1981 to 2011 period. In Chapter 3, I identify and track changes in spatially clustered regions of high and low-income inequality. An east-west divide is found, along with an important divide between urban and rural areas. The rise in inequality in Canada over the past 30 years comes mainly from increasing inequality in large metropolitan areas. More recently, small and medium cities are also witnessing a surge in inequality. In contrast, rural areas in the country remain more equal places.

In Chapter 4, I investigate what might lead some regions to have higher levels of income inequality than others. I find that regions with more precarious labour markets tend to have higher levels of inequality. The lack of jobs in the manufacturing and public administration sectors also lead to higher inequality. On the socio-demographic front, regions with polarized educational attainment levels tend to be more unequal. And high proportions of seniors and young individuals in a region also tend to put pressure on the active labor force that leads to higher inequality. These factors behind regional income inequality vary along the urban-rural spectrum.

While I wish to shed light on the causes behind the changing structure and nature of regional income inequality, in Chapter 5 I also explore what are the consequences of inequality by exploring the impacts of unequal income distributions on the region's economic growth. In the long term, I find that regions with initially higher levels of inequality experience higher levels of growth. However, the relationship not appear to hold over the short- and medium-term horizons, where changes in the level of economic development and inequality levels tend to move in opposite directions.

## 1.1 A first-cut literature review: Setting the scene

Income inequality has been increasing in most of OECD countries since the late 1970s and early 1980s (OECD 2011). Bound and Johnson (1992), Katz and Murphy (1992), Levy and Murnane (1992), Murphy and Welch (1992) were among the first economists to document a pattern of rising inequality in the US. All of these studies were carried out at the national level. In Canada, the same patterns of rising inequality were observed a few years later by Atkinson (2008), Brozowski et al. (2010), Fortin et al. (2012), Frenette et al. (2007), Heisz (2007), and Green and Kesselman (2006). Since then, a broad range of studies have been published highlighting the trends in inequality, their roots and potential causes. To help situate the present study on regional income inequality in Canada in the field of research on inequality, I first begin by presenting some of the work that has been done at the national level before proceeding to work the sub-national level and work on comparing urban and rural studies.

### 1.1.1 National-level studies

The bulk of the work on inequality has been carried out at the national-level. Here, the economic transition from an industrial to a post-industrial society tends to be associated with higher levels of income inequality (Hammet, 2003; Saez, 2006). In OECD countries, Cornia (1999) and Woods (1994) state that market liberalisation in the context of ongoing globalisation leads to the further development of a primarily knowledge-based workforce. In turn, this new conjecture accentuates the positive correlation between incomes and education, skill and talent (Juhn, Murphy and Pierce, 1993). Although there is no overarching theory of inequality, I find in the literature three distinct sets of explanations (Atkinson, 1975; Lydall, 1979; Osberg, 1981). The first relates to *structural* variables, where particular attention is paid to supply and demand factors (ties to labor market variables). The second is more *socio-demographic* in nature and looks at the link between human capital variables and income. The last set of explanations is tied to *institutional* factors and considers policies, social and political institutions to be the main drivers of inequality. All three approaches together provide us with a multi-dimensional conceptual framework to study inequality (Chakravorty, 1996; Bourne, 1993; Breau, 2015) as most variables associated with inequality can be linked to one of these approaches (see section 1.2). Overall, while national-level studies successfully identify some of the key factors of income inequality, they overlook important sub-national differences in the regional trajectories of inequality and the potential determinants of these differences.

### 1.1.2 Sub-national level studies

The number of sub-national level studies of inequality has grown over the years as researchers increasingly recognize the importance of spatial heterogeneity in levels of inequality within countries (see Cavanaugh and Breau 2017). In the United States where, most sub-national studies are found, there is a wide range of state-level studies (Blank and Card, 1993; Karoly and Klerman, 1994; Topel, 1994; Partridge, Rickman, Levernier, 1996) as well as county-level studies (Bartik, 1994, 1996; Persky and Tam 1994; Borjas, Freeman, Katz 1996; Levernier, Partridge, Rickman, 1998).

In Canada, the number of such studies is more limited but includes, at the provincial-level, work by Finnie (2001), Gray et al. (2004) and Breau (2007), and regional-level work by Breau (2015). Within these sub-national studies, two scales of analysis are emphasized: urban and rural.

The urban scale features prominently given that inequality tends to be more pronounced within larger metropolitan areas (Chakravorty, 1996; Madden, 2000; Glaeser et al., 2009; Bourne, 1993; Mitchell and Soroka, 1993; Soroka, 1999; Bolton and Breau, 2012; Cloutier, 1997; Baum-Snow and Pavan, 2012; Korpi 2008; Essletzbichler 2015; Florida and Mellander 2016). This also includes intra-urban (or neighbourhood-level) studies of inequality (Chen, Myles, Picot, 2012; Myles et al., 2000; Breau et al. 2017).

As for the rural scale, it has received far less attention. Data availability may have played a role here. The most interesting findings in the United States points towards identifying variables that might contribute to higher inequality level in the city while having the opposite effect in rural areas (Lobao et al., 1999; McLaughlin, 2002; Levernier et al., 1998; Peters, 2013). Given that there exist important variations among rural spaces in terms of income inequality (Breau, 2014; Peters, 2012), research on inequality could benefit and learn from those rural areas that are less prone to rising income inequality by identifying the reasons for such resilience. Rural areas are socially, economically and culturally diverse (Cloke, Marsden et Mooner, 2006; Freshwater, 2001; Fuguitt, 1989; Jean, 2012). I should revisit the role of structural, institutional, and socio-demographic factors in explaining levels of income inequality in rural areas.

## 1.2 Potential contributors to income inequality in regions roots in the literature

An important aspect of this thesis is the identification of the determinants of inequality across regions. As stated earlier, there are three main sets of explanations behind the recent rise in inequality.

I begin with the *structural* variables that represent the labor market and economic characteristics of regions. The argument for the inclusion of an economic development variable such as mean wages traces back to the work of Kuznets (1955) and his theory of the inverted U-shaped pattern of inequality. This theory suggests that initially, the relationship between inequality and economic development would be positive, until a certain level of development is reached beyond which inequality would plateau and then eventually decline.

The median wages squared variable also comes into play when modeling the expected negative relationship between inequality and a more advanced state of economic development. Hence, median wages squared is expected to be negative. Yet, Bluestone and Harrison (1988) suggest the possibility of levels of inequality rising again after having 'plateaued', as the situation is exemplified by the United States in the 1980s. In the event that median wage square coefficient is positive, results from Canadian regions from 1981-2011 would also point to the great U-turn theorized by Harrison and Bluestone (1988) (and, more recently described by Brankovic 2016 as 'Kuznets waves').

In addition to these two variables, the unemployment rate has a history of positive association with levels of inequality. MacPhail (2000) found positive linkages between Canada's high unemployment rate of the 1980s and then higher levels of inequality. Breau (2015) finds that between 1996 and 2006, a more unequal distribution of income is found in regions with higher unemployment rates. In this case, joblessness is of interest for two more reasons. First, the unemployment rate says something about the overall health of a region's economy, which influences the distribution of income. Likewise, unemployment could be hurting disproportionately low-income workers as suggested by Erksoy (1994), thus deepening the level of inequality. For the period 1981 to 2006, the unemployment rate is expected to hold a significant positive relation with the inequality measure. Incorporating 2011 might change the sign and significance of the variable as recessions tend to increase unemployment rates and lower inequality level.

In the OECD’s (2011) broad ranging report on inequality, Canada is highlighted as one of the most unequal societies among member countries once part-timers and self-employed are accounted for in the Gini coefficient of earnings inequality. The reasoning for including such variables in the analysis is that it connects back to the literature on precarious employment and how the latter effects low-income earners in particular. The link between levels of inequality and the percentage of workforce employed in part-time jobs and self-employed could be positive or negative. Part-time workers could earn high hourly wages that combined with other fruitful sources of revenue. Self-employed workers could represent highly successful businesspersons as well as middle class entrepreneurs.

I am also interested in the industrial composition of a region. Breau (2015), Peters (2014) and McLaughlin (2002) among others, turn to the industrial mix of regions when explaining some of the variations in levels of inequality. Overall, the significance of specific industrial sectors and their impact across regions and time varies from one study to another. Diverse results are interesting in that it opens the door for an analysis of income distributions trends through more detailed industrial sectors. Notable results in the Canadian regional context are that a higher percentage of workers in manufacturing sector tends to contribute to lower levels of inequality (Breau, 2015). In another study evaluating the link between innovation and inequality in metropolitan areas in Canada, Breau et al. (2014) produce evidence of a positive association between more unequal distribution of income and Knowledge Intensive Business Sector (KIBS). Given manufacturing and KIBS sector outcomes, this project continues the breakdown of the economy into 13 more industrial sectors, as will be seen in Chapter 2.

As I try to predict the effect of each industrial sector on the income distribution, the safest bet is to expect the manufacturing to reaffirm already proven effects of equalizing income distributions across regions, as in Breau (2015). Nonetheless, the main reasoning behind manufacturing jobs reducing inequalities is the provision of sufficient numbers of ‘good’ paying jobs, which could arguably be the case of other sectors of the economy as well, such as public administration, utilities and the education and health sectors.

On the other hand, employment in the finance, insurance and real estate (FIRE) sectors could potentially contribute to inequality. High paying jobs concentrated in space is part of the reasoning, in a similar way to what the knowledge intensive business services (KIBs) sector did in Canada from 1996 to 2006 (Breau, 2014). Additionally, Johnson and Beale (1997) and

Marcouiller (1997) point to employment in the arts and entertainment sector to be a proxy for the level of amenities in a region. In turn, regions with high level of amenities tend to attract high income earners (Marcouiller, 2004), which could link the sector of arts and entertainment to growing income inequalities. Reinforcing this last idea, Goulding (2006) describes growing amenities in non-metropolitan areas such as ski resorts, lake houses, scenic landscapes as an invitation to income inequality.

Of the jobs that could hold a mixed effect on inequality level, are the sectors of construction, transportation, communications, retail, wholesale and business services. Without significant results in the literature, I included those sectors for exploratory purposes.

The second important set of variables are the *socio-demographic* variables. Since the 1950s, the Canadian labour market has seen remarkable changes, one of which is the steady increase of women in the workforce (Statistics Canada, 2016). In 1981, 35% of workers were women, a proportion that jumps to 48% in 2011[1]. In previous studies, female labour force participation has had mixed effects on household level inequality (Breau, 2015). These mixed results revolve around the problem of where women contribute to the distribution of incomes when either (i) adding a second revenue to an already well off household or by (ii) lowering inequality when bringing home a much needed second stream of revenue. However, the effect of an increasing number of female workers poses yet another puzzle with regard to individual level inequality. Are women disproportionally occupying low paying jobs? What about the gender pay gap? With a dataset that goes back to 1981, there is potential of gaining insights on the relation between female participation rates and inequality.

The variable reflecting the percentage of visible minorities living in a particular census division is meant to investigate the economic inequalities faced by ethnic groups in Canada. By definition, the visible minorities variable is the sum of census recorded non-white immigrants and aboriginal persons. It suits the regional analysis well, as immigrants make up most of the proportion of visible minorities in urban areas, whereas aboriginals hold the same role in rural areas. Building from previous studies, I expect the variable of visible minorities to be positively associated with inequality for two reasons. On the one hand, Picot (2008) and Mok (2009) point to widening income gaps between immigrants and Canadian-born. At the same time, the

[1] Authors computation from the datasets.

economic hardships faced by aboriginal people do not seem to fade away between 1980 to 2011 (Pendakur, 2001; 2011). Both of these dynamics suggest that I should expect a positive relationship between visible minorities and inequality.

The educational ratio is designed to reflect the polarization of educational attainment among workers active in the labor force. The ratio is the sum of the percentage of workers with a university degree (bachelor and above) and percentage of workers with no high school diploma. Assumptions on the relationship between income levels and educational attainment (i.e. human capital theory) means that a large education ratio is expected to be associated with high levels of inequality. The education ratio proved to be a significant contributor of inequality in the Canadian context from 1996 to 2006 (Breau, 2015) and comparable results are reached across regions of the European union from 1995 to 2000 (Rodriguez-Pose and Tselios, 2009)

The proportions of individuals below age 15 and above age 65 in a region are also potential drivers of inequality. Young and senior population shares could increase inequality since a greater dependence ratio puts additional pressure on the active labor force (Breau, 2015). Yet, the percentages of young and senior individuals have been linked to both increasing and lowering inequality level in the literature, notably Breau (2015) and Nielsen and Alderson (1997). With data going back to 1981 to reflect longer demographic trends, the goal is to make a significant contribution to the debate.

This research will also examine the impact of a third set of inequality drivers, *institutional* variables. In particular, I will look at the relationship between inequality and minimum wages, transfers to persons and unionization rates.

In the end, the thesis thus adopts a multi-dimensional approach in order to test and validate explanations of regional inequality using a new and more expanded panel dataset. In the next chapter, I turn my attention to the development of this regional dataset.

# Chapter 2: Methodology

## Introduction

In this chapter, I begin by presenting the work behind the development of the panel dataset which is unprecedented in the literature on Canadian regional inequality. In describing the methodological aspects of the data development phase, I make sure to address potential issues of time period poolability in order to justify the robustness of the data upon which chapter 3, 4 and 5 analysis rely.

Sub-section 2.1 begins by describing the data development process with a discussion of both the advantages and limitations of working with the 20% long form sample of the census. An argument is also made for pooling the 2011 National Household survey with previous census data. This is followed by a discussion of the need to build a time consistent geography from the original census files as the boundaries of regions (i.e., census divisions) do change from one census cycle to another. Finally, a discussion of the concordance between the Standard Industrial Classification and North American Industrial Classification closes the sub-section.

Once the longitudinal and spatial properties of the dataset are legitimised, section 2.2 presents an overview of the dataset. Here, the indicators of inequality used for the analysis are discussed, as are the labour market and economic variables, industrial variables, socio-demographic and institutional variables.

The division between rural and urban spaces is central to how the dataset will be analysed throughout this project. Section 2.3 is dedicated to presenting the chosen methodology for defining how a region is classified as either rural or urban.

## 2.1 Developing a longitudinally and spatially consistent dataset for studying regional patterns of inequality in Canada: How and why it works

### 2.1.1 Long form census micro data: Advantages

The bulk of the data used for the analysis are derived from the 20 percent sample long form census of 1981, 1986, 1991, 1996, 2001 and 2006, and the 2011 National Household Survey. There are many advantages of working with the long form census micro-datasets. To begin with, recall that the long form census is mandatory (and distributed to every fifth household) which means that it yields very large samples (a minimum of four million individual records) given that its response rates are consistently over 90% in any given census year. Moreover, population weights are built into the dataset so that the 20% sample results reflect the trends observed for the Canadian population as a whole. A mandatory survey, most importantly gives us confidence in the quality of the income variables as the survey sensitive groups of low and high incomers are less prone to not respond or answer the questions at hand. Moreover, it is important to note that the long form census in Canada reports income variables that are not top coded as is the case in other countries (notably the US) where it is well documented that top coding has the effect of artificially lowering the level of inequality (Fichtenbaum and Shahidi, 1998; MacPhail, 2000; Moore & Pacey, 2003).

Another important feature of the long form census is the similarities and consistency in the methodological approach behind each survey. In other words, the same variables are collected every five years. This includes detailed information on the geography of individuals which can be in turn used to develop a consistent panel of regions (defined as Census Divisions in this thesis; see section 2.1.4 for more details). When and if some variables are updated due to methodological considerations and presented differently through different census cycles, the changes are traceable. Table 1, which is constructed from the census public codebooks, highlights the lineage across the six census cycles and the NHS from 1981 to 2011 and shows the pathways of all variables used in the present analysis.  There are a few methodological changes from the census bureau worth noting. In terms of the geography available, enumerations areas (EAs) are set forth in 1996 and not previously recorded. On the demographic front, the way to account for visible minorities changes considerably after 1981; whereas the 1981 (and earlier censuses) asked respondents to select one or multiple ethnic origins from a list of 15 ethnicities, starting in 1986 that list was expanded to include 118 possible ethnic groups. Finally, in terms of the labour market variables, the industrial classification system used for categorizing workers into different industrial groups (i.e., based on similar production processes or similar products), changes four times during the seven censuses span cover in this project.

**Table 1**. Census variables used for 1981, 1986, 1991, 1996, 2001, 2006 censuses and the 2011 NHS

| | Census variables | 1981 | 1986 | 1991 | 1996 | 2001 | 2006 | 2011 |
|---|---|---|---|---|---|---|---|---|
| Geography | Census divisions | PCD | PCD | PCD | PCD | PCD | PCD | PCD |
| | Census subdivisions | PCSD | PCSD | PCSD | pcsd_new | PCSD | PCSD | PCSD |
| | Province | Pr | Pr | Pr | PR | PR | PR | PR |
| | Enumeration areas | | | | Pedea | PRCDDA | PRCDDA | PRCDDA |
| Demography & Citizenship | Single years of age | Age | Age | Age | Age | AGE | Age | AGECONT |
| | Sex | Sex | Sex | Sex | Sex | SEX | Sex | Sex |
| | Year of immigration | ImmYear | ImmYear | ImmYear | YRIM | YRIM | YRIM | YRIM |
| | Aboriginal identity | AborFlp | ETO38 | DeAbP | AbDerr | ABDERR | AbDerr | AbDerr |
| | Visible minority | Ethor | VisMin | VisMin | DVisMin | DVISMIN | DVisMin | DVisMin |
| Labour market ativity & income | Class of worker | CowD | CowD | COWD | COWD | COWD | COWD | COWD |
| | Full time or part time work | FPTIM | FPTIM | FPTim | FPTim | FPTIM | FPTim | FPTim |
| | Industry Classification | Industry | IND80 | Ind80 | IND80 | IND80 | NAICS02 | NAICS07 |
| | Labour force activity | LFTag | LFTag | LFTag | LFTag | LFTAG | LFTag | LFTag |
| | Total income | TotInc | TotInc | TotInc | TotInc | TOTINC | TotInc | TotInc |
| | Wages and salaries | Wages | Wages | Wages | Wages | WAGES | Wages | Wages |
| Education & Schooling | Highest certificate, diploma or degree | Dgreer | DgreeR | DgreeR | DgreeR | DGREER | HCDD | HCDD |
| | Highest level of schooling | HLOS | HLOS | HLOS | HLOSR | HLOSR | HCDD | HCDD |
| | Total years of schooling | Ps_OtR + Ps_UvR | TotYrs | TotYrs | TotYrSr | TOTYRSR | | |
| | Weighting variables | CompW5 | CompW5 | COMPW5 | CompW2 | COMPW2 | CompW2 | CompW2 |

*The table is use for the SAS coding: the variables are comparable through census years if listed on the same row, but some coding might have been necessary to make them concord.

*Empty cell implies that the variable was not measure for the census year, and it could not be constructed from any other variable.

*First column is based on 2006 variables descriptions.

Thus, the survey design gets refined somewhat between censuses, creating some of the changes listed above. While all of these changes can be retraced over time they do forced some methodological compromises. These are properly addressed in the following sub-sections. Overall, the Canadian long form census, to its advantage, be forged into a longitudinally consistent regional a panel dataset.

### 2.1.2 Long form census micro data: Limitations

Working with the long form census micro-dataset presents two major challenges: access to the confidential information and divulgation of the analytical results. In Canada, researchers can access the census long form micro-datasets through Statistics Canada's network of Research Data Centers (RDCs), which is an initiative supported by the Social Sciences and Humanities Research Council of Canada, the Canada Foundation for Innovation and the Canadian Institutes of Health Research. RDCs located in the province of Quebec are in turn managed by the Quebec Inter-University Center for Social Statistics (QICSS) and access to its lab facilities is granted only after a vetting process. Upon project approval, data are supplied to the researchers on individual private server. In submitting a project proposal to the RDCs, the guidelines must clearly identify all variables required for the analysis[2].

Moreover, because the micro-datasets are not publicly available, all research findings and output must be reviewed and approved through a rigorous disclosure protocol. To ensure the integrity and confidentiality of the census micro-datasets, the disclosure rules enforced by Statistics Canada do require some methodological compromises. For instance, rules concerning the minimum number of observations required for certain geographical units impose certain restrictions on the potential geographic scale of analysis. And for studies of inequality analysis, which requires working with income variables, the disclosure requirements are even more demanding. A minimum of 40 private households in each geographic area is required. In addition, the highest earner in the area must hold less than a certain percentage of the area's total income.

As the goal of this project is to explore the dynamics of regional income inequality in Canada, the census division level is the most suitable proxy for regions in Canada which is large enough (in terms of aggregation) to meet the disclosure rules for most of the variables of interest[3]. Where finer spatial scales could have been disclosed in certain urban areas, this is not the case for rural Canada which have smaller population figures. The census division scale allowed for data of all regions to be disclosed, which is judged quintessential to this project.

### 2.1.3 The National Household Survey

Before proceeding with the details of the data development exercise, it is important to say a few words on the change from the Census to the National Household Survey (NHS). In 2011, the mandatory national census was replaced by the NHS which was now to be answered on a voluntary basis. The response rate for the NHS was slightly below 70%, but the number of respondents is 9% higher than the 6,136,517 respondents in 2006. In deploying the NHS in 2011, Statistics Canada argued that it would mitigate the risk of sampling error by distributing the survey to every three households instead of every five. Looking back, however, Statistics

[2] The variable list for this particular project is presented above in sub-section 2.1.1. Table 2.1.
[3] Presented in subsection 2.2.

Canada's chief statistician warned potential NHS data users on how the formula for the non-mandatory survey could lead to the underrepresentation of certain groups, namely recent immigrants, low-income individuals and those living in small communities (Smith, 2015). Statistics Canada addressed the issue of underrepresentation using estimation models based on 2006 data. However, they specified that they could not completely reduce some of the volatility inherent with estimates for smaller populations (Smith, 2015) which means that in a scenario where low and high income groups are underrepresented, inequality measures could potentially be skewed downward.

As we will see in chapter 3, trends in the data point to a significant drop in income inequality from 2006 to 2011. Whether these trends are imputable to the major change in surveying methodologies between the two years is a question worth asking. Also, the underrepresentation of low and high income groups might come into effect. Overall, answering this question would require further analysis at the national scale as recent comparative work by Rheault et al. (2015) using income tax data for the province of Quebec suggests that the reliability of the NHS income data at very fine levels of spatial disaggregation can be problematic.

Furthermore, when explaining potential inconsistencies in income variables between the 2006 long form census and the 2011 NHS, Statistics Canada issued in a statement, and I quote: ''Any significant change in the methods of a survey can affect the comparability of data over time. [...] The significance of any quality shortcomings depends, to some extent, on the intended use of the data.'' (Statistics Canada, 2015). As we will see later in chapter 2, the sensitivity of the income variables is one problem among others that has to be addressed in order to ensure the poolability of the NHS with the long form censuses. Additionally, one must also consider important macroeconomic trends that happened between the two surveys. The 2008 recession could be responsible for significant parts of the drop in inequality measures. Veall (2012) points to falling top income shares in 2008 and 2009 for Canada, as is traditionally expected after recessions. Overall, the importance of including the 2011 data in this study and the macroeconomic context into which the NHS came to be still provides more incentives to pool the datasets together.

### 2.1.4 Geographic concordance

Building a panel dataset of 284 census divisions over seven census cycles requires a time consistent geography. Spatial reconfigurations of cross-sectional geographic units represent a challenge for longitudinal analysis (Goodchild, Anselin, & Deichmann, 1993; Martin et al. 2002; Puderer, 2008). To ensure longitudinal consistency, a methodology was developed to create concordant census divisions' boundaries for a 30-year period based on the most recent delimitations possible.

Census divisions (CDs) represent groups of neighbouring municipalities joined together for the purposes of regional planning and managing common services (Statistics Canada, 2015). CDs are administrated under provincial jurisdiction and their boundaries are subject to change over time through annexation, partial annexation, dissolution or creation of CDs or smaller geographic units. In 1981, there were 266 CDs whereas in 2011 their number had increased to 293. Although mainly small geographic variations are found from one census year to another, comparisons over the broader 30-year horizon reveal much more variation in CD boundary changes. This is confirmed in Table 2 which shows that from 1981 to 2011, nearly half the CDs were subject to delimitations changes.

**Table 2**. Number of census divisions, by province, where at least one boundary change occurred between 1981 and 2011

| Code | Province | # of CDs | # of changes | % changes |
|---|---|---|---|---|
| 10 | Newfoudnland and Labrador | 10 | 1 | 10% |
| 11 | Prince Edward Island | 3 | 0 | 0% |
| 12 | Nova Scotia | 18 | 0 | 0% |
| 13 | New Brunswick | 15 | 1 | 7% |
| 24 | Quebec | 97 | 95 | 98% |
| 35 | Ontario | 49 | 5 | 10% |
| 46 | Manitoba | 23 | 1 | 4% |
| 47 | Saskatchewan | 18 | 0 | 0% |
| 48 | Alberta | 18 | 7 | 39% |
| 59 | British Columbia | 27 | 5 | 19% |
| 60 | Yukon | 1 | 0 | 0% |
| 61 | Northwest Territories | 5 | 5 | 100% |
| | **Total:** | **284** | **120** | **42%** |

Indeed, 42% of all CDs saw part of their boundaries redrawn at one point in time. Much of these boundary changes occurred in the province of Quebec, where a major redesign took place between 1981 and 1986 under the auspices of the ''municipalité régional de compté'', a new administrative geographic unit defined by the province upon which federal census divisions are based. Overall, the proportion of census divisions with boundaries not consistent in time is too high, at 42%, for researchers interested in carrying out comparative and longitudinal analyses of regional inequality by pooling multiple census years together.

To address this boundary issue, the methodology developed in this thesis relies on a Geographic Information System software, ArcMap 10, to overlay layers of census division delimitations in a given year according to those as defined by the boundaries of CDs in 2011. Statistics Canada, through its Standard Geographical Classification reports, records all divergences and changes to CD boundaries from one census cycle to another. This information was crucial in identifying all relevant boundary changes over the 1981 to 2011 period and to recreate a consistent set of $n$ = 284 CD boundaries using the smaller census subdivision (CSD) units.

However, most changes in boundaries -- though not all -- were retraceable or compatible at the CSD level; in all, there were 98 such minor incongruences. Table A1 in the appendix provides further details and explanations of why those incongruences could not be addressed. In sum, non-respect of geographic consistency through time when building a panel dataset of individual aggregated to a particular geographic unit can lead to significant problems and biases when making statistical inferences (Goodchild, Anselin, & Deichmann, 1993; Martin et al. 2002; Puderer, 2008). From my point of view, solving the issue of geographical consistency, as any variable consistency, is key to a robust analysis and presentation of the dynamics of income inequality between regions in Canada.

### 2.1.5 Industrial classification concordance

Just as spatial boundaries of regions may change over time, so may the classification for workers employed in various industries. In Canada, census responses to industry of work information from 1981 to 2001 were coded under the Standard Industrial Classification (SIC-1980). This is a four-digit code originally established in the U.S. (back in 1937) that classifies workers into 286 mutually exclusive industries and is designed to be exhaustive of the entire labour force. Codes refer to the general nature of the business carried out in the establishment where the person works.

The Standard Industrial Classification was eventually replaced by the North American Industry Classification System (NAICS) in 1997. This new classification is organized in a hierarchical structure that can be read as a broad two-digit, 20 industrial sector codes, down to a four-digit level of detail for 324 industrial sectors[4]. The original NAICS version of 1997 was first implemented in the Canadian census in 2001. NAICS was further revised in 2002 and 2007 to reflect changes of an evolving world economy, leaving the censuses of 2001 and 2006, as well as the 2011 NHS, with three different NAICS versions. Fortunately, Statistics Canada did produce and make publicly available concordances files between these three classifications.

To recap, the censuses of 1981, 1986, 1991, 1996 and 2001 are based on the Standard Industrial Classification whereas the censuses of 2001 and 2006 and the 2011 NHS use three different versions of the NAICS. Similar to the geographic concordance problem, industrial classifications need to be time consistent for longitudinal comparisons to be made. If it is not a problem to trace the changes of the three NAICS versions, the same cannot be said of comparing the SIC and NAICS. The two classifications are known to be incomparable due to differences in their conceptual framework. A firm that fell under a particular SIC code might be classified under an arguably unrelated NAICS code.

To deal with this issue, I turn to industry aggregates presented in subsection 2.2. Briefly, the 286 SIC codes and the 324 NAICS codes are aggregated to a maximum of 17 industry sectors based on general consistency in categorization over time. By using industry aggregates, the conceptual differences between NAICS and SIC are minimized. The aggregation, in other words, enables the poolability of industrial variable from 1981 to 2011 in one panel dataset. Accordingly, the bulk of the variables of interest to regional income inequality can be presented knowing that potential poolability issues have been addressed.

## 2.2 Overview of the dataset

Table 3 presents an overview of the contents of the dataset constructed from the raw micro-data files, from the variable names, their definitions and sources, along with their predicted effect on the distribution of income (as per the theoretical framework presented in chapter 1). Therefore, a plus sign (+) means the variable is expected to be associated with higher regional levels of inequality and vice-versa.

[4] Note that NAICS can be disaggregated to a 6-digit code, but only the 4 digit codes are available in census data.

While Table 3 provides an overview of the dataset, the remainder of section 2.2 offers more details on each set of variables. It starts with a conceptual discussion on how inequality is measured using four different distribution metrics (i.e. the four dependent variables to be used in this study) before proceeding to a discussion of how the independent variables identified in Chapter 1 are constructed.

**Table 3.** Overview of the data

| Variable | Definition | Effect | Sources |
|---|---|---|---|
| **Section 2.2.1: Inequality measures** | | | |
| Gini coefficient of total income | | | Census long-form |
| Theil index of total income | Generalized entropy index ($\alpha = 1$) | | Census long-form |
| Gini coefficient of wage | | | Census long-form |
| Theil index of wage | Generalized entropy index ($\alpha = 1$) | | Census long-form |
| **Section 2.2.2: Labour market and economic variables** | | | |
| Median wage | Median wages in 2002$ | + | Census long-form |
| (Median wage)$^2$ | | - | Census long-form |
| Unemployment rate | % Unemployed | + | Census long-form |
| Part-time worker (%) | | + | Census long-form |
| Self-employed worker (%) | | -/+ | Census long-form |
| **Section 2.2.3: Industrial variables** | | | |
| *Secondary sector (%)* | | - | Census long-form |
| Manufacturing (%) | | - | Census long-form |
| Resource-intensive industries (RII) (%) | i.e. Access to abundant natural resources | - | Census long-form |
| Labour-intensive industries (LII) (%) | i.e. Labour costs | - | Census long-form |
| Scale-intensive industries (SII) (%) | i.e. Length of production runs | - | Census long-form |
| Differentiated goods (DG) (%) | i.e. Tailoring product to highly varied demand characteristics | - | Census long-form |
| Science-based industries (SBI) (%) | i.e. Rapid application of scientific advance | - | Census long-form |
| Construction (%) | | -/+ | Census long-form |
| *Tertiary sector (%)* | | -/+ | Census long-form |
| Transportation (%) | | -/+ | Census long-form |
| Communication (%) | | -/+ | Census long-form |
| Utilities (%) | | - | Census long-form |
| Retail (%) | | -/+ | Census long-form |
| Wholesale (%) | | -/+ | Census long-form |
| *Quaternary sector (%)* | | + | Census long-form |
| FIRE (%) | | + | Census long-form |
| Public administration (%) | | - | Census long-form |
| Education and health (%) | | - | Census long-form |
| Leisure and food (%) | | + | Census long-form |
| KIBS (%) | | + | Census long-form |
| Business services (%) | | -/+ | Census long-form |
| **Section 2.2.4: Socio-demographic variables** | | | |
| Female participation rate | % of female in the labour force | -/+ | Census long-form |
| Visible minorities (%) | | + | Census long-form |
| Education ratio | % High school + % University | + | Census long-form |
| Young (%) | % of population age < 15 | -/+ | Census long-form |
| Senior (%) | % of population age > 65 | + | Census long-form |
| **Section 2.2.5: Institutional variables** | | | |
| Minimun wage | 2002$, provincial level variable | - | HRSDC* |
| Transfers to persons (per capita) | 2002$, provincial level variable | - | Statistics Canada** |
| Unionization rate (%) | % of workers unionized, provincial level variable | - | Odesi*** |

* Human Resources and Skills Development Canada (HRSDC), customized search for general minimum wage rates in Canada

** Statistics Canada: Table 384-0004

*** <odesi> [Online]. Database by Ontario Council of University Libraries

### 2.2.1 Inequality measurements

Income inequality can be measured using several different metrics. This project uses income inequality metrics developed from census reported individual total income and wages. The Gini coefficient of total income is a perennial metric in studies of inequality. In addition to the Gini, the thesis will report the Theil index to add robustness to the analysis.

With regards to the latter, it is important to specify the difference in the income concept chosen for the analysis. In the census data, wages and salaries refer to gross wages before deductions for items such as income tax, pensions, employment insurance, etc. In contrast, total income is a broader concept that refers to total money income received including wages and salaries, child benefits, old age security pension and guaranteed income supplement, benefits from Canada or Quebec pension plan, benefits from employment insurance, dividends and interests on bonds, deposits, savings certificates and other investment income, and other money income. In both cases, it is worth recalling that the findings of this thesis are built around income and wage data collected at the individual level in gross values, for individuals aged between 15 and 65 with a minimum income above 0, in order to reflect active members of the labour force.

To be sure, there are other income concepts upon which the same analysis could have been performed, such as household or per capita income and after-tax or net income. When building inequality measures around gross income figures, I methodologically accept not to account for the progressive taxation in the Canadian federal system and how it can potentially redress some of the income inequality. Strictly speaking, inequality measures built from gross income figures rather than net figures produce artificially higher levels of inequality. Given that after-tax incomes are not available in census data, I counter the latter effect by keeping only the individuals that are within a certain age range with a wage or total income above 0, and by presenting results from inequality measures built from both wage and total income.

As mentioned earlier, the Gini coefficient of total income will act as the primary metric (i.e., benchmark) upon which results in this thesis will be reported. Although each metric has its own specifics with regards to distributional properties, they do tend to be highly correlated. Table 4 shows the correlation between four inequality metrics pooled over time ($t$=7) for my census divisions ($n$=284) from the top-left to the bottom-right quadrant. The Gini and Theil both show values with above 0.85 in terms of Pearson's correlation coefficients. Results shown in the diagonal of the bottom left quadrant also suggest a good fit between Gini coefficients of total income and wage, with a 0.76 coefficient. The same is true for the Theil wages and income correlation coefficient (which registers a value of 0.84). Given the high level of correlation between the Gini and Theil, it could be tempting to follow through with only one inequality measure. However, it is important to know how diverse, conceptually and mathematically, inequality can be measured.

**Table 4**. Pearson's r correlation coefficient between Gini coefficient and Theil index of total income and wage

| Pearson's r | | Income | | Wage | |
|---|---|---|---|---|---|
| | | Gini | Theil | Gini | Theil |
| Income | Gini | 1.0000 | | | |
| | Theil | 0.8689 | 1.0000 | | |
| Wage | Gini | 0.7662 | 0.6254 | 1.0000 | |
| | Theil | 0.7800 | 0.8467 | 0.8887 | 1.0000 |

At the more conceptual level, it is commonly accepted that in order for an inequality metric to be considered as such, it should follow a set of well-defined properties (Cowell, 2011). First, *anonymity* stipulates that a metric must depend on the attributes of the individual and not their identity. That is to say, an economy is as unequal if you switch the income of two individuals between them. According to the *population principle*, a metric should be independent of the size of the population it is applied to. A metric should also remain unchanged if every individual's income is doubled, as per the *scale independence* property. Finally, the *transfer principle* stipulates that a transfer from a rich person to a poor person leads to lower levels of inequality. All four metrics presented in this thesis follow the properties of anonymity, population principle, scale independence and transfer principle.

The Gini coefficient is the ratio of the area between the line of perfect equality and the Lorenz curve (Atkinson, 1975). The line of perfect equality is a 45-degree line while the Lorenz curve shows the proportion of overall income possessed by a given percentage of the population. As inequality increases the area between the two lines increases, rendering it easy to interpret inequality measures bounded between 0 and 1. A Gini coefficient of 1 would be one individual possessing all income (a situation of perfect inequality), and 0 if incomes were distributed equally across all individuals. Mathematically, the Gini coefficient is computed as follow:

$$Gini = \frac{1}{2n^2\overline{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|$$

where *n* is the number of individuals and *y* is the income of individual *i* or *j*.

One downside of the Gini coefficient comes from the inability to identify the whereabouts of inequality within the income distribution (Cowell, 1995). The arc of the Lorenz curve could be skewed upward or downward and in both cases produce the same area under the line of perfect equality. However, in the former the very rich are getting more than the rich, while in the latter some already unfortunate individuals start doing worst than already unfortunate individuals.

The Theil index is introduced in the analysis to resolve, at least partially, for the Gini coefficient's downside. The Theil is a general entropy index, built around a sensitivity parameter (α) to express the weight given to different parts of the income distribution. The Theil index, compared to the Gini coefficient, values the middle of the distribution with an α of 1. Mathematically, the Theil index can be specified as follows:

$$T_{\alpha=1} = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{\mu}\ln\left(\frac{x_i}{\mu}\right)$$

where *N* is the total number of individuals, *x* is the income of individual *i* and µ is the population mean income. Thus, reporting both the Gini coefficient and Theil index brings more robustness to the results presented in chapters 3, 4 and 5.

Now let us turn our attention to the variables identified in Chapter 1 that could potentially explain some of the variation in income inequality represented by the Gini coefficients and Theil indexes.

### 2.2.2 Labour market and economic variables

Recall that in in Chapter 1, labour market characteristics and levels of economic development are both identified as potential predictors of regional differences in levels of inequality. Identified in the literature as structural components, those predictors are quantitatively expressed in this project through the five following variables: median wages, median wages squared, the unemployment rate, the percentage of self-employed and part-time workers. Clarification on the definition and usage of each variables follows.

At the regional level, median wages are used as a proxy for general level of economic development. It corresponds to the median wages of workers active in the labor force. In the context of this project, labor force workers are those aged 25 to 65 and having earned more than 1000$ the year previous to the census. Nominal dollar values have been adjusted for inflation at 2002 dollars.

The unemployment rate shows the percentage of workers aged between 25 to 65 and having earned more than 1000$ the year before the census that are experienced or inexperienced, either looking or not looking for work. It also includes workers on temporary lay-offs, new job seekers, etc.

Part time workers correspond to workers who reported working 30 hours or less on the basis of all jobs held. The variable is calculated as a percentage of part-time workers to the labor force (those aged 25 to 65 and having earned more than 1000$ the year previous to the census.) Similarly, the self-employed worker variable is also expressed as a percentage of the labor force and is composed of persons who worked mainly for themselves, with or without paid help, operating a business, farm or professional practice, alone or in partnership.

In addition to paying attention to the different types of workers when measuring the level of inequality of a region, I also identified in Chapter 1 the need to account for the economic sectors to which those workers subscribe.

### 2.2.3 Industrial variables

As per the discussion of Chapter 1, the industrial mix of a region is a potential predictor of regional differences in levels of inequality. Long form census respondents were asked to select from a list of industries the one that refers to the general nature of the business carried out in the establishment where they worked, which translated into the industrial categories discussed in section 2.1.5. These industrial variables report the percentages of workers in the labor force that work in every industry defined in Table 5 below.

In the first column of Table 5, Group 1 classifies industries into four broad categories: primary, secondary, tertiary and quaternary sectors of the economy (the latter two consisting of services more broadly). This breakdown of the service sector into tertiary and quaternary is based on the work of Kenessey (1987). Adding the four percentages together sums to 100% of the labor force. In column 2, the Group 2 category breaks down the four economic sectors into more detailed and specific industry sectors. Similarly, Group 3 (column 4) is a detailed version of the components of the manufacturing variable defined in Group 2, but only sums up to the percentage of workers in the manufacturing industry.

**Table 5**. Overview of industrial sectors

<table>
<tr><th></th><th>Group 1</th><th>NAICS</th><th>Group 2</th><th>Group 3</th></tr>
<tr><td></td><td rowspan="2">Primary</td><td>11</td><td colspan="2">Agriculture, forestry, fishing and hunting</td></tr>
<tr><td></td><td>21</td><td colspan="2">Mining, quarrying, and oil and gas extraction</td></tr>
<tr><td></td><td rowspan="2">Secondary</td><td>31<br>33</td><td>Manufacturing</td><td>Resource intensive industries<br>Labour intensive industries<br>Scale intensive industries<br>Differentiated goods<br>Science-based industries</td></tr>
<tr><td></td><td>23</td><td colspan="2">Construction</td></tr>
<tr><td rowspan="12">Service sectors</td><td rowspan="5">Tertiary</td><td>48<br>49</td><td colspan="2">Transportation and warehousing</td></tr>
<tr><td>51</td><td colspan="2">Information and cultural industries</td></tr>
<tr><td>22</td><td colspan="2">Utilities</td></tr>
<tr><td>41</td><td colspan="2">Wholesale trade</td></tr>
<tr><td>44<br>45</td><td colspan="2">Retail trade</td></tr>
<tr><td rowspan="7">Quaternary</td><td>52<br>53</td><td colspan="2">Finance, insurrance, and real estate and rental and leasing</td></tr>
<tr><td>54</td><td colspan="2">Knowledge intensive business services</td></tr>
<tr><td>61<br>62</td><td colspan="2">Educational services, and health care and social assistance</td></tr>
<tr><td>71<br>72</td><td colspan="2">Arts, entertainment and recreation, and accomodation and food services</td></tr>
<tr><td>91</td><td colspan="2">Public administration</td></tr>
<tr><td>55<br>56<br>81</td><td colspan="2">Management of companies and enterprises, administrative support, waste management and remediation services, and other services (except public administration)</td></tr>
</table>

Note: Based on author compilation, NAICS2007 definitions

The Group 3 categorization comes from the desire to explore in more depth the effects of the manufacturing sector on levels of inequality. In order to do so, the manufacturing sector is divided into five manufacturing sub-types following a methodology developed by the OECD (1987). Table 6 list the five manufacturing subcategories along with a short definition, example and list of equivalent NAICS 2007 codes that constitute each.

**Table 6**. OECD classification of manufacturing industries

| Manufacture type | Resource-intensive | Labour-intensive | Scale-intensive | Differentiated goods | Science-based |
|---|---|---|---|---|---|
| Definition | Access to abundant natural resources | Labour costs | Lenght of production runs | Tailoring product to highly varied demand characteristics | Rapid application of scientific advance |
| Examples | Aluminium smelting | Clothing, footwear | Steel | Machine tools | Pharmaceuticals |
| NAICS codes | 3111 to 3119, 3121, 3122, 3161, 3169, 3211, 3212, 3219, 3221, 3241, 3273, 3274, 3279, 3313, 3314 | 3131 to 3133, 3141, 3149, 3151, 3152, 3159, 3162, 3322, 3323, 3324, 3325, 3326, 3329, 3371, 3372, 3379, 3399 | 3222, 3231, 3251, 3261, 3262, 3271, 3272, 3311, 3312, 3315, 3321, 3361 to 3363, 3365, 3366, 3369 | 3327, 3328, 3331 to 3336, 3339, 3342, 3343, 3346, 3351 to 3353, 3359 | 3252 to 3256, 3259, 3341, 3344, 3345, 3364, 3391 |

The OECD distinguishes the industries on the basis of the primary factors affecting the competitive process in each industry. Resource intensive industries are characterized by the importance of access to abundant natural resources. Labour intensive industries depend on labour costs. Scale intensive industries are defined by long cycles of production, while the differentiated goods industries produce in response to varied demand characteristics. Finally, the application of scientific advances is central to the so-called science-based industries. Overall, this gives us five more variables to work with in predicting variations in inequality across regions. The industrial mix, reflected in those variables, is given much relevance in this project as a potential determinant of inequality, as does the socio-demographic context in which those workers operates (more on this below).

### 2.2.4 Socio-demographic variables

As discussed in chapter one, over the 30-year period of study (1981-2011), the Canadian population underwent important social and demographic compositional changes that could potentially create regional differences in levels of inequality. Five socio-economic variables will try to capture the bulk of those changes: the female participation rate in the workforce, the percentage of visible minorities, the level of human capital acquired through education and the percentage of young and senior individuals.

More specifically, the female participation rate corresponds to the percentage of females in the labor force, those aged 25 to 65 and having earned more than 1000$ the year previous to the census. The visible minorities variable on the other hand corresponds to the percentage of the population in a census divisions that reported as Chinese, South Asian, Black, Filipino, Latin American, Southeast Asian, Arab, West Asian, Korean, Japanese, Aboriginal or another visible minority category (or individuals who report having multiple visible minority status).

The educational ratio acts as the variable responsible for capturing the level of human capital in a region. Of the many ways I could have accounted for educational capital, the education ratio was selected to reflect the polarization of educational attainment among workers active in the labor force. It is the sum of the percentage of workers with a university degree (bachelor and above) and percentage of workers with no high school diploma (this is a common approach in urban studies; see, for instance, Chakravorty 1996).

The last socio-demographic variables included the percentage of young and seniors in the population. They reflect the level of dependency of the population, best described as the amount of pressure the active workforce is under due to the number of individuals they 'economically' support. The percentage young variable represents those younger than 16 years old whereas the percentage seniors correspond to individuals above the age of 65.

### 2.2.5 Institutional variables

As was mentioned earlier, the bulk of the data comes from the quinquennial censuses. I also include in the analyses three institutional variables: the minimum wage, current transfers to person and unionization rates, all of which are reported at the provincial-level from varied sources. To be clear, provincially reported variables means that all census divisions within a province share a repeated value for minimum wages, transfers to person and unionization rates. Also, for all three variables, Nunavut, created in 1999, is given the Northwest Territories values prior to 2001.

The minimum wages data are retrieved from the publicly available Human Resources and Skills Development Canada (HRSDC) database. Nominal dollar values have been adjusted for inflation at 2002$ values.

Current transfers to person are also retrievable online via Statistics Canada Table 384-0004 and reflect transfers to persons at the federal, provincial and local levels. Transfer payments to persons that are accounted for include child tax benefits and/or credits, employment insurance benefits, old age security payments and the Canada & Quebec Pension Plans (CPP) (Statistics Canada, 2006). Again, nominal dollar values have been adjusted for inflation at 2002$.

Unionization rates represent the number of workers that agreed to a collective union agreement. The data come from multiple sources and necessitated the pooling together of three surveys: the Survey of Work History (SWH) provided 1981 data, the Labour Market Activity Survey (LMAS) did so for 1986 (note that the LMAS 1990 serves as a proxy for the inexistent 1991 unionization rate) while the Labor Force Survey contributed 2001, 2006, 2011 (and 1997 as a proxy for 1996).

Now that I have described the key variables used in the analyses, it is important to say more about how to measure regional inequality in Canada along an urban and rural divide. This first requires us to trace the line between what regions will be considered rural and which will be urban. This discussion, along with a short overview of Canada's urbanisation since 1981 will close Chapter 2.

## 2.3 Definition of rural

Key to the analysis presented in this research project is assessing the difference in levels of inequality between rural and urban regions across Canada. Here, there is no universally accepted definition of the distinction between the two. The existing literature on urban and rural classifications points to population size, density and proximity as appropriate factors to consider when making a distinction between urban and rural regions. Moreover, as du Plessis (2002) points out, the appropriate definition may well be determined by the question being asked. In this sub-section, I define a rural and urban gradient that is appropriate for the purposes of studying regional income inequality dynamics in Canada.

I begin by reviewing the work of the American sociologist Calvin Beale. In the mid-1970s, Beale, in partnership with the USDA's Economic Research Service developed the ''Beale codes'' (USDA ERS, 2016). Beale codes reflect a coding system that allows researchers to subordinate aggregated county level data under the metropolitan and non-metropolitan labels. ''Beale codes'' go beyond the metro-non-metro dichotomy by classifying counties in the US under a nine-class codification scheme. A dichotomic approach to the urban and rural divide was avoided by according importance to the degree of urbanization and metropolitan proximity of non-metropolitan counties. Table 7 below depicts Beale classification for the United States.

**Table 7**. Economic Research Service ''Beale codes''

| Code | Description |
|---|---|
| Metro counties: | |
| 1 | Counties in metro areas of 1 million population or more |
| 2 | Counties in metro areas of 250,000 to 1 million population |
| 3 | Counties in metro areas of fewer than 250,000 population |
| Nonmetro counties: | |
| 4 | Urban population of 20,000 or more, adjacent to a metro area |
| 5 | Urban population of 20,000 or more, not adjacent to a metro area |
| 6 | Urban population of 2,500 to 19,999, adjacent to a metro area |
| 7 | Urban population of 2,500 to 19,999, not adjacent to a metro area |
| 8 | Completely rural or less than 2,500 urban population, adjacent to a metro area |
| 9 | Completely rural or less than 2,500 urban population, not adjacent to a metro area |

The classification is built around the notion of population size and remoteness, two notions that are arguably the principal concepts upon which a sound definition of rurality should be based on. The classification depicted in the table above embodies well the concepts splitting metropolitan areas into three degree of urbanization, and rural areas into six degree of rurality.

This coding scheme has been widely use in social research addressing the urban-rural nexus. ''Beale codes'' continue to be maintained by the U.S. Department of Agriculture (USDA) Economic Research Service (ERS) and are updated every 10 years. In Canada, the Micro-Economic Analysis division at Statistics Canada maintained a similar ''Beale code'' classification for the country from 1976 to 1996 under the agriculture and rural working paper series. While Statistics Canada has discontinued the ''Definitions of 'rural'' program, its archived content offers the required material to re-construct the Beale codes through 2011.

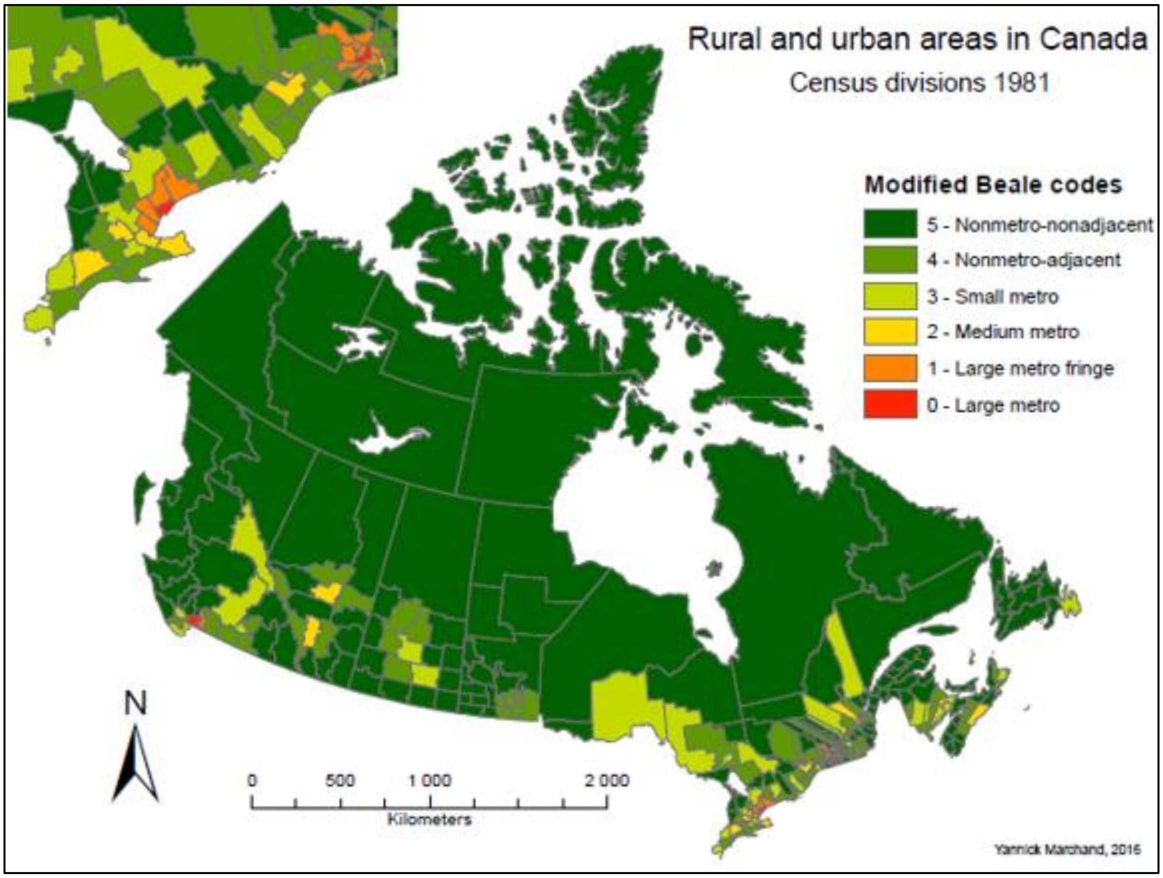
Prior to computations, Statistics Canada acknowledges that Canada and the United States have very different population densities across their territory. The demographic notion upon which Beale's coding system is built had to be adapted to the less populous Canadian territory. Henceforth, the revised codes developed by Statistics Canada present a six codes scheme as shown in Table 8 below.

**Table 8**. Statistics Canada ‘’ Beale codes’’

| Code | | Name | Description |
|---|---|---|---|
| Metro counties: | 0 | Large metro | Central and most populous census division of a CMA with a population greater than 1 million |
| | 1 | Large metro fringe | Remaining census division(s) within or partially within a CMA with a population greater than 1 million |
| | 2 | Medium metro | Census division(s) containing, within, or partially within a CMA with a population between 250,000 and 999,999 |
| | 3 | Small metro | Census division(s) containing, within, or partially within a CMA with a population between 50,000 and 249,999 |
| Nonmetro counties: | 4 | Nonmetro-adjacent | Census divisions that share a boundary with a CMA/CA and the CMA/CA has to have a population greater than 50,000 |
| | 5 | Nonmetro-adjacent | Census divisions that do not share a boundary with a CMA/CA that has a population greater than 50,000 |

Geographically, the closest equivalent of the U.S. county is found in the Canadian census divisions (CD). However, in contrast to the U.S. ‘Beale’ codes, where county population sizes are the core of the coding, the population of the Canadian census divisions is not taken into consideration here. Instead, the census metropolitan agglomeration (CMA) and census agglomeration (CA) populations serve as the defining units to classify census divisions. In other words, a CD’s relationship to a CMA or CA defines its urban-rural status. As explained in the table above census divisions either contained, partially contained, share or do not share a boundary with a CMA/CA. Figure 1 below maps the ‘’revised Beale codes’’ for 1981, with an inset map of Montreal-Windsor corridor in the top-left corner.

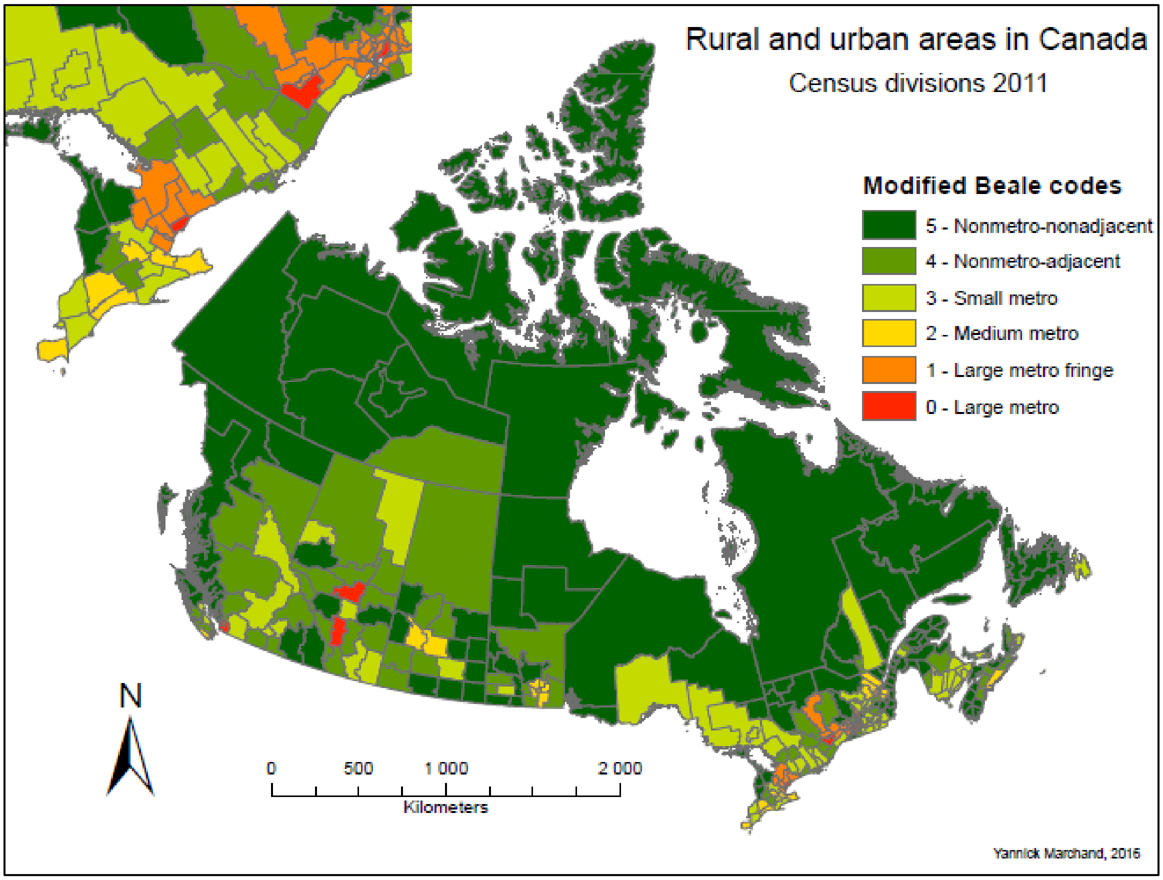
**Figure 1.** Rural and urban areas in Canada, census divisions 1981




In the above map, rural and urban areas (based on CDs) in 1981 are simply projected using Statistics Canada's Beale classification. The color scheme depicts rural areas in darker shades of green whereas urban areas are depicted in light green, yellow or red. Thus, Canada looks overwhelmingly rural in 1981. Note also that Calgary and Edmonton are still considered medium metropolitan areas back then, and so is Ottawa. The metropolitan fringes of Toronto and Montreal are also confined to their immediate surrounding downtown areas.

While one would expect significant changes in 30 years, I mentioned the program was discontinued and the last available ''revised Beale codes'' are for 1996. Using the 'Population and Dwelling count highlight tables' diffused publicly by Statistics Canada, it is possible to retrieve the CMA and CA populations for 2011 and rework a modern day 'Beale' code classification. Figure 2 maps the ''revised Beale codes'' for 2011, as compiled by the author.

**Figure 2**. Rural and urban areas in Canada, census divisions 2011




From 1981 to 2011, we see that the map is generally 'lighting up'. From west to east, we see the appearance of four small metropolitan regions in Alberta in their wake they created nonmetro-adjacent rural areas up and into the North-West Territories. In those 30 years, Winnipeg has also emerged as a medium metropolitan area that also lights up the Manitoba province. The inset map in the top left corner also shows a much more orange Quebec-Windsor corridor, with Ottawa now classified as a large metro area. Although there is a visually marked urbanisation of Canada over this 30-year period, Table 9 below shows the steady pace of this evolution through time.

**Table 9**. Summary of USDA ''Beale codes'' revised for Canada, 1981-2011.

| Code | 1981 | 1986 | 1991 | 1996 | 2001 | 2006 | 2011 | Congruent |
|---|---|---|---|---|---|---|---|---|
| **Urban** | | | | | | | | |
| 0 Large metro | 3 | 3 | 3 | 4 | 4 | 6 | 6 | 6 |
| 1 Large metro fringe | 15 | 14 | 18 | 24 | 26 | 26 | 27 | 27 |
| 2 Medium metro | 12 | 13 | 16 | 14 | 23 | 21 | 24 | 24 |
| 3 Small metro | 33 | 41 | 41 | 43 | 54 | 57 | 60 | 60 |
| Total # of urban CDs: | 63 | 71 | 78 | 85 | 107 | 110 | 117 | 117 |
| **Rural** | | | | | | | | |
| 4 Nonmetro-adjacent | 74 | 70 | 77 | 72 | 61 | 63 | 61 | 60 |
| 5 Nonmetro-Nonadjacent | 129 | 125 | 135 | 131 | 121 | 116 | 115 | 107 |
| Total # of rural CDs: | 203 | 195 | 212 | 203 | 182 | 179 | 176 | 167 |
| Total # of CDs: | 266 | 266 | 290 | 288 | 289 | 288 | 293 | 284 |

Note:Congruent, 2011, 2006 and 2001 code were created by author from population measures. Statistics Canada

The table lines up the 'Beale' codes by census years. From 1981 to 2011, each urban code (0 to 3) manages to double in size. Note that not all of this growth came from urbanizing rural areas, as new census divisions are created along the years. Nonetheless, the difference in the total number of rural CDs between 1981 and 2011 does reveal that at least 27 rural areas urbanized during that span. Even so, a majority of Canadian regions remain rural as non-metropolitan census divisions out numbered metropolitan ones, 176 to 117, in 2011.

The final column shows the harmonized number of census divisions based on the consistent geography that was presented in section 2.1.3. The harmonized 'Beale' codes were created from the 2011 CMA and CA populations (which explains why the number of CDs in urban code match the 2011 number). The rural numbers are smaller due to those CDs in the northern part of the country which had to be absorbed in more aggregate CDs when building the concordant set of census divisions presented in section 2.1.3. The final harmonized column numbers are 117 urban areas and 167 rural areas. Those are the numbers upon which Chapters 3 and 4 and rely on whenever an urban and rural component is brought to light in the analysis.

## 2.4 Conclusion

To sum up, the purpose of Chapter 2 was to go over the methodology for building a balanced panel dataset with seven time periods and 284 cross-sectional units. In developing a concordant geography and time consistent industrial classification, as well as constructing a set of sound independent variables, I am now confident in moving forward and exploring the regional patterns of income inequality in Canada.

# CHAPTER 3: DESCRIPTIVE STATISTICS

## Introduction

Inequality as a subject offers a multitude of ways it can be analysed through descriptive statistics. Analysis based on wages, income, assets, wealth and the likes, between and within countries, at static or dynamic points in time are all within the scope of inequality studies. The OECD's (2011) *Divided we stand* report and *Capital in the 21 century* (Piketty, 2014) are two major recent works on inequality conveying their arguments through an array of different graphs and summary tables.

In this chapter, I present a first set of descriptive statistics with the objective of identifying broad patterns of income inequality in Canada over a 30-year period across different sub-national geographies. Exploring those patterns within various type of spatial units allows us to begin answering part of the research questions introduced at the beginning of this thesis.

The chapter starts with an overview of patterns of income inequality at the national level. I then look at differences in the trajectories of inequality along the urban and rural divide as defined by 'Beale' codes before proceeding to the spatial analysis of census divisions, where I identify the presence of spatial autocorrelation in inequality. The chapter ends with a brief analysis of the descriptive statistics of independent variables used later on in Chapters 4 and 5 to explore how different socio-economic facets of Canada's regions have changed in 30 years.

## 3.1 Measuring income inequality

### 3.1.1 At the national level

Figures 3 and 4 (below) show the 30-year evolution of the Gini coefficient and Theil index of total income (blue line) and wages and salaries (red line).

**Figure 3**. Gini coefficients for Canada, 1981-2011

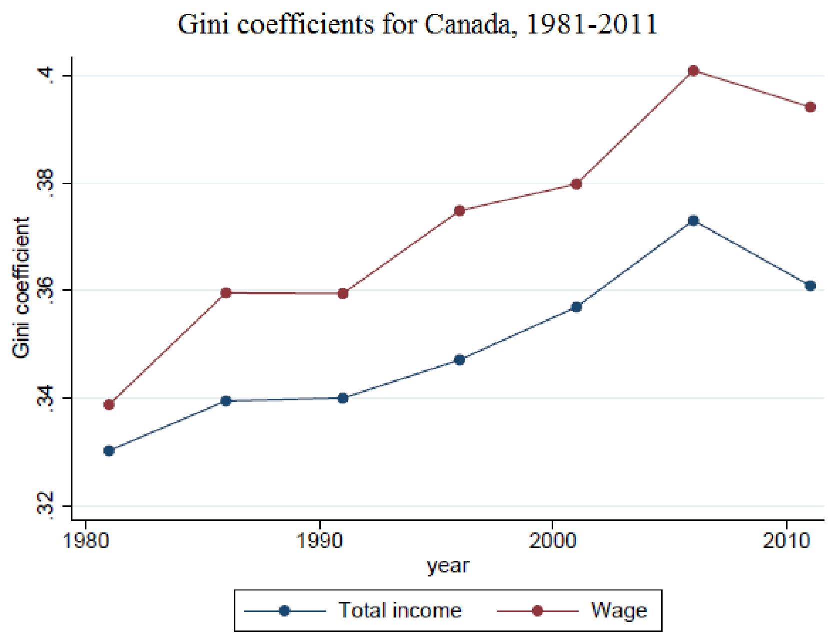
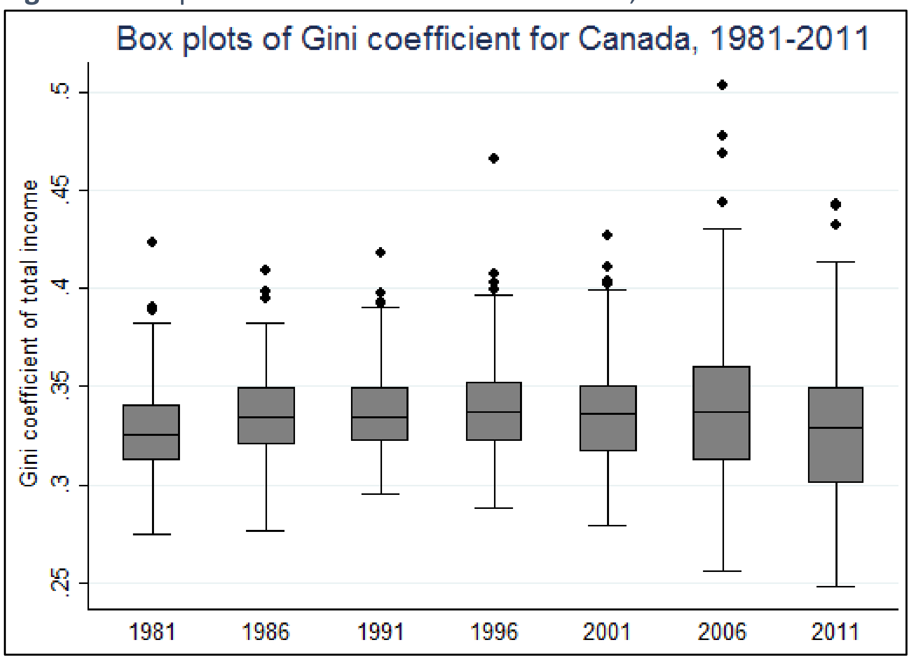



**Figure 4**. Theil indexes for Canada, 1981-2011


Following both Gini coefficients and Theil indexes of wages and total income, we see clearly that Canadian inequality experienced a net increase from 1981 to 2011. The Gini coefficients of total income and wages increased between all census years except for the 2006-2011 periods. Similarly, the Theil index of total income increased steadily in every period until it peaked, like the Gini, in 2006 before decreasing in 2011. Assuming that the 2006-2011 decreases in both metrics are in part due to the compression of incomes following the 2008 economic crisis, the above trends aligns the literature pointing to increasing inequality in Canada since the 1980s (Frenette, Green, and Milligan, 2009).

The "dip" in inequality observed post-2006 is indeed of interest as we ask how the nature of income inequality in Canada has evolved in 30 years. The 2008 recession, the most far-reaching economic crises of modern history, took a toll on top earners (Veall, 2012) which partially explains why inequality decreased. Figure 5 further investigates the evolution of Gini coefficients across regions from 1981 to 2011 using box plots.

**Figure 5**. Box plots of Gini coefficient for Canada, 1981-2011




Up until 2006, the median level of inequality, as indicated by the line splitting the middle ''box'', moves up on the scale as inequality is increasing in Canada (this also tracks the trend depicted in Figures 3 and 4). More interesting, however, is what happens to the range of inequality values across regions. The middle ''box'', which represents the inter-quartile range and contains 50% of the census divisions, essentially gets larger over time. Likewise, the whiskers are getting larger. Both of these findings suggest that there is more variation in the level of inequality between census divisions. After the 2008 crisis, inequality decreased on average, but the variation between census divisions remains high. While at first glance, Figures 3 and 4 suggest that inequality is declining during this period of time in Canada, there exist important sub-national variations in regional levels of inequality. Moreover, these regional variations are growing over time.

### 3.1.2 Along the urban-rural divide

Given the presence of important sub-national variations, patterns of income inequality in Canada can be better understood through a regional analysis. Table 10 breaks down these patterns according to urban and rural areas. Along with national figures, Table 10 reports a simple binary sub-national division that shows the 1981 and 2011 Gini coefficients and Theil indexes at the urban and rural scales, along with the variation that occurred during the 30–year period.

**Table 10.** Descriptive statistics for inequality measures, Canada, 1981 and 2011

| Mean | National | | | Urban | | | Rural | | |
|---|---|---|---|---|---|---|---|---|---|
| (standard deviation) | 1981 | 2011 | %Δ | 1981 | 2011 | %Δ | 1981 | 2011 | %Δ |
| **Total income inequality measures** | | | | | | | | | |
| Gini coefficient | 0,330 | 0,361 | 9% | 0,331 | 0,367 | 10% | 0,329 | 0,330 | 0% |
| | 0,019 | 0,044 | | 0,018 | 0,044 | | 0,024 | 0,030 | |
| Theil index (GE=1) | 0,199 | 0,273 | 27% | 0,201 | 0,287 | 30% | 0,191 | 0,202 | 5% |
| | 0,028 | 0,096 | | 0,027 | 0,097 | | 0,030 | 0,052 | |
| **Wage inequality measures** | | | | | | | | | |
| Gini coefficient | 0,339 | 0,394 | 14% | 0,336 | 0,397 | 15% | 0,348 | 0,382 | 9% |
| | 0,019 | 0,036 | | 0,016 | 0,037 | | 0,027 | 0,028 | |
| Theil index (GE=1) | 0,203 | 0,306 | 34% | 0,202 | 0,315 | 36% | 0,208 | 0,259 | 20% |
| | 0,024 | 0,086 | | 0,022 | 0,088 | | 0,031 | 0,052 | |

I begin by noting that the increase in inequality is primarily an urban phenomenon. In Table 10, we see that the Gini coefficient of total income rose by 10% in urban areas between 1981 and 2011 whereas it remained essentially the same in rural areas. In 1981, the level of inequality as measured by the Gini coefficient of total income was similar for both urban and rural spaces. Thirty years later, urban regions now have much higher levels of inequality in comparison to rural areas (with Gini coefficients of 0,367 and 0,330, respectively). That these Gini coefficients increased at different rates starting in 1981 supports the findings previously reported by Breau (2015) and are of interest to the subsequent analysis. The standard deviation for urban areas compared to rural areas in 2011 indicates that there is significantly more variation between levels of inequality in urban census divisions than between rural census divisions. The analysis of the Gini coefficient of total income thus reveals growing differences in the level of income inequality of urban and rural spaces.

To make sure the above results are robust to changes in the definition of inequality, I also calculated the Theil index of total income. The Theil index essentially reaffirms the pattern of increasing sub-national variations in inequality. Likewise, urban areas drove most of this increase with a 30% change over time compared to an increase of only 5% in rural areas. One interesting question this raises is how patterns of inequality may vary within this broad spectrum of urban spaces

As for the wage-based inequality measures presented in Table 10, these tend to reveal higher levels of inequality. I also note a more pronounced increase in inequality between 1981 and 2011 for the country as a whole as well as between urban and rural regions. Wage inequality metrics tends to be higher than those of total income since the latter includes some redistributive features and mechanisms (e.g., Employment Insurance, old age security, etc.). The Theil index reaffirms that income inequalities have been growing faster in urban areas compared to rural areas from 1981 onward. But what stands out of the bottom part of the Table 10 is significantly higher Gini coefficients and Theil indexes in rural parts of the country compared to urban areas for 1981. Rural areas had in 1981 higher levels of wage inequality than urban areas, but somehow the situation reversed itself over 30 years. I will investigate this finding in Chapter 4 when looking at the factors contributing to regional levels of inequality.

From the preliminary results presented above, I know that inequality has increased more prominently in urban areas over the last 30 years. However, inequality used to be higher in rural areas (in terms of wage inequality). I also know that the variation in levels of inequality between urban census divisions has been growing and in rural areas as well, albeit at a slower pace. The latter is a sign that urban and rural areas are not homogenous places, as discussed in Chapter 2.

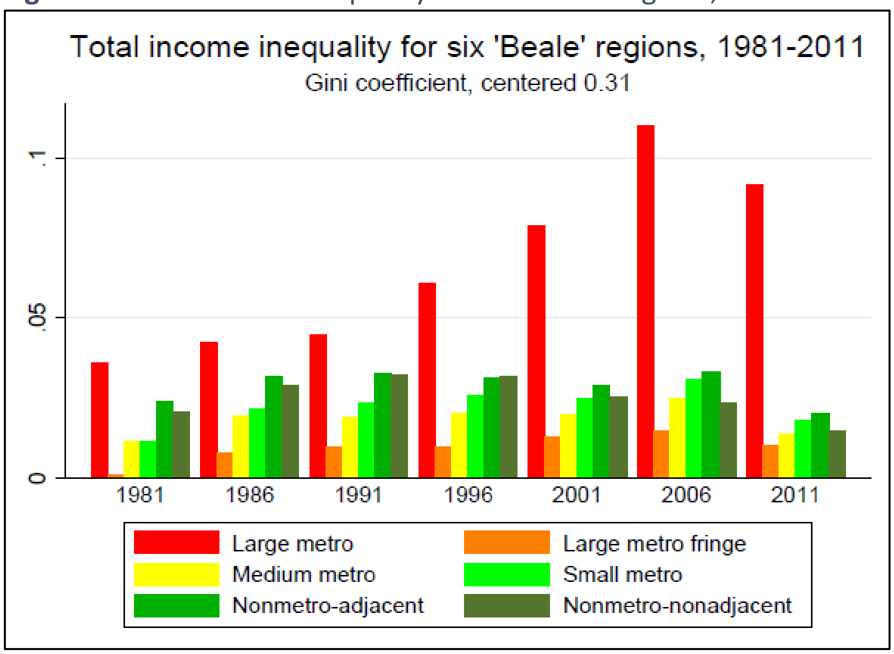
The following sub-section explores these geographical patterns of income inequality in more detail using the 'Beale' code classification developed earlier.

### 3.1.3 Looking at 'Beale' regions

Recall from our earlier discussion in Chapter 2 that the 'Beale' code classification is based on the idea that rural areas are not homogeneous and that they differ in part based on their proximity to cities. Likewise, the size of their population further characterizes differences in urban areas.

Figure 6, below, shows the Gini coefficients of total income for the six 'Beale' codes in each census year. A cursory look at the figure reveals clearly that large metropolitan areas in Canada have been in primary drivers of inequality since 1981. This creates an important spatial contrast (or divide, to use the words of Breau 2015) as neighboring large metropolitan fringe areas have the lowest Gini coefficients of total income in every year studied. Otherwise, medium and small metropolitan areas position themselves somewhere between large metropolitan areas and their fringe. Interestingly, smaller cities are generally more unequal than medium cities.

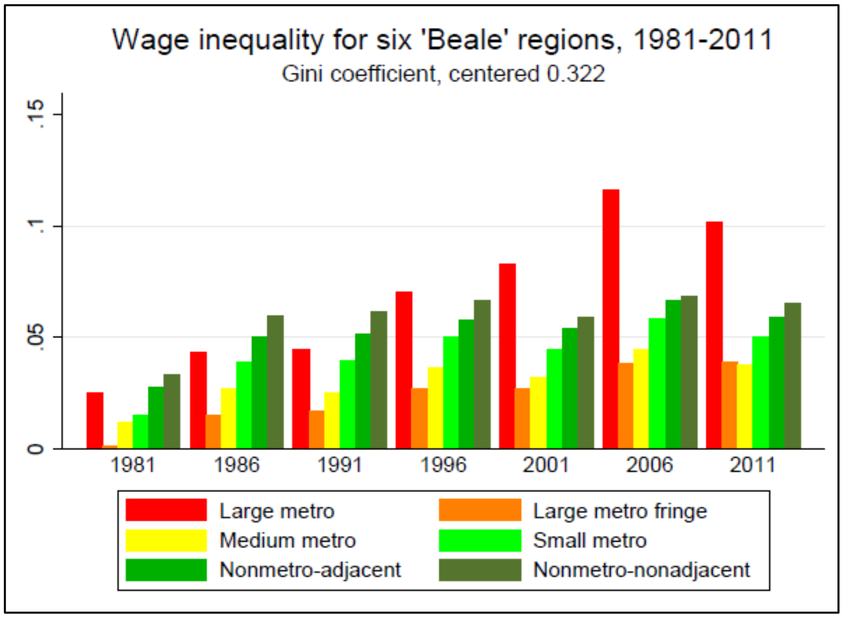
**Figure 6**. Total income inequality for six 'Beale' regions, 1981-2011




The more rural areas, shown in the last two columns of every year, tend to rank second in terms of levels of inequality after large metropolitan areas. Remote rural areas are more equal than city-adjacent countryside areas. Nonetheless, as the figure makes clear, the large metropolitan areas are the main contributors to why the Gini coefficients previously reported showed higher and faster increasing levels of inequality

Figure 7 (below) replicates the analysis using wages instead of the total income variable.

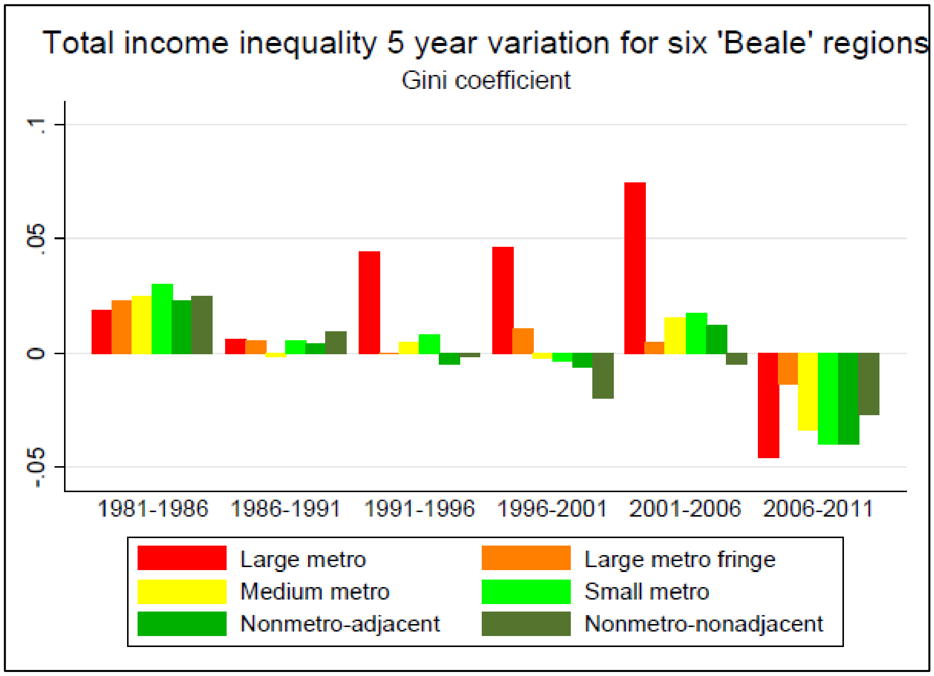
**Figure 7**. Wage inequality for six 'Beale' regions, 1981-2011




There are interesting nuances to note here. When using wages as the income definition, large metropolitan areas only become the most unequal areas of Canada in 1996. Indeed, rural areas have the highest levels of inequality for 1981, 1986 and 1991. Also, it appears that the state of inequality in large metropolitan fringe areas is slowly changing as well. By 2011, most of what makes up suburbia became more unequal than medium metropolitan areas, a finding also reached by Moos (2015).

Lastly, at the 'Beale' regional level, we can also look at the inter-census period variations to better understand the previous finding that inequality in urban areas grew while rural inequality held steady or decreased. Figure 8 shows the variation in Gini coefficient of total income for inter-census periods.

**Figure 8**. Total income inequality 5 year variation for six 'Beale' regions



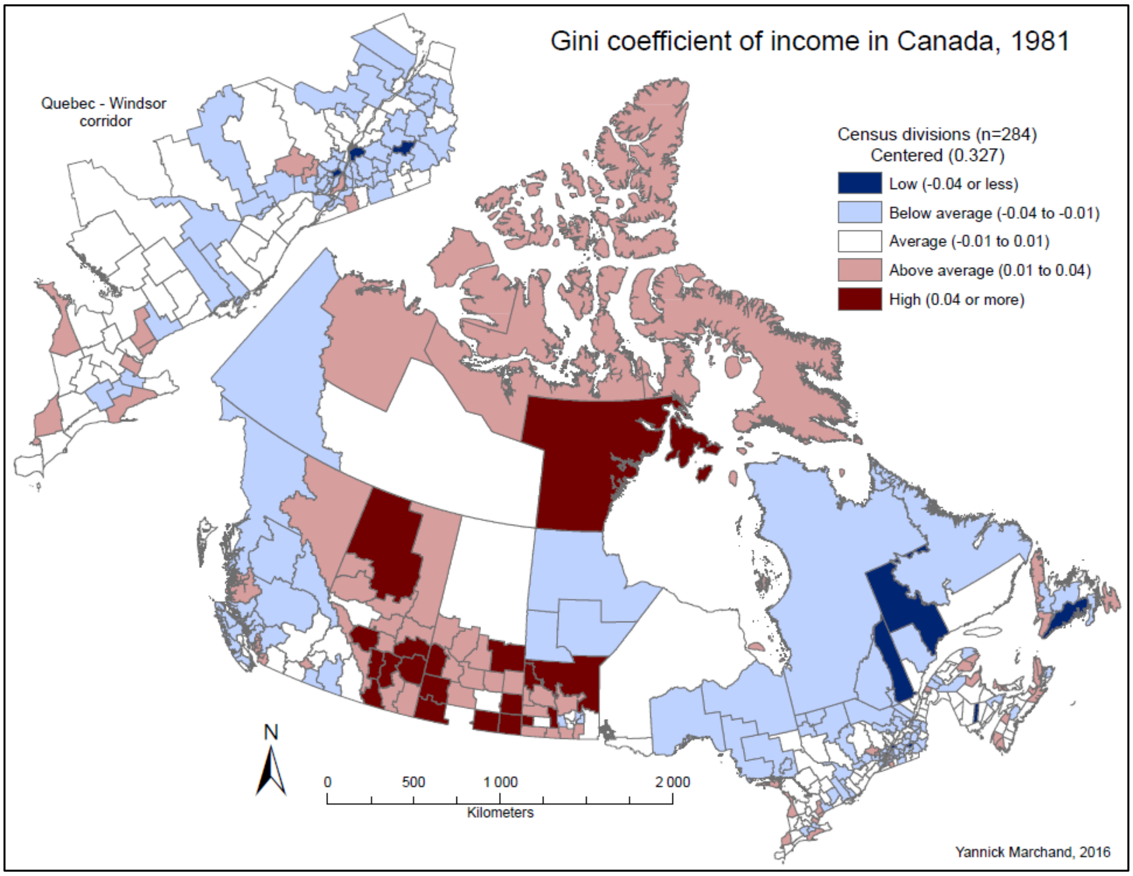

In the years of 1981 to 1991, income inequality grew between 0,02 and 0,035 points in every region. Starting in 1991, inequality in all metropolitan groups grows more rapidly than nonmetropolitan groups, and more rapidly than Canada as a whole, especially in the large metropolitan areas. Indeed, for large metropolitan areas, levels of inequality skyrocketed until 2006. From 2006 to 2011, levels of inequality decreased essentially across all Beale code regions.

Overall, the spatial dimensions of inequality in Canada are characterized by several variations across the urban-rural divide. In the following section, I examine in more detail the sub-national variations of inequality in Canada.

## 3.6 Spatial analysis of income inequality at the census divisions level

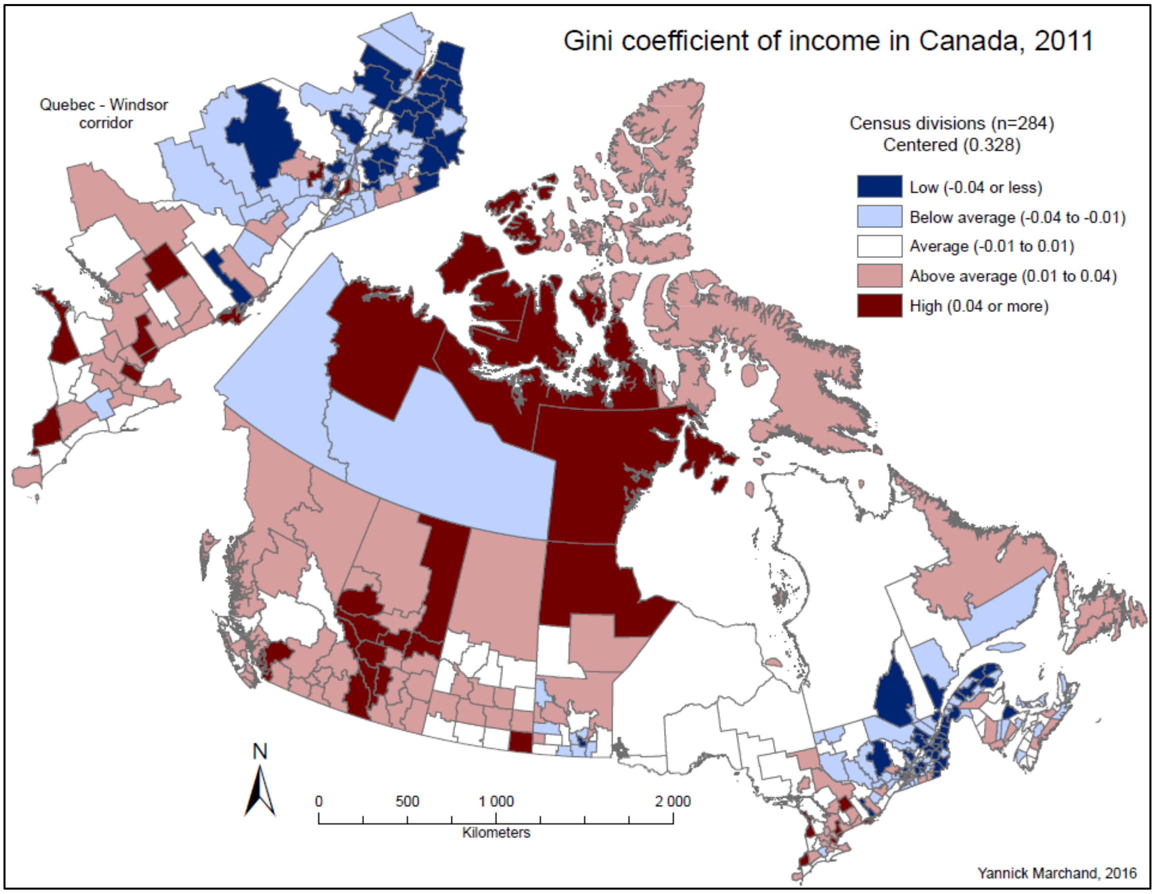
Through descriptive statistics, I have examined Canadian income inequality at the national, urban-rural and 'Beale' code regional level. Figure 9 serves as the basis for the next logical step in my analysis and maps values of the Gini coefficient of total income at the census division level in 1981.

**Figure 9**. Gini coefficient of income in Canada, 1981



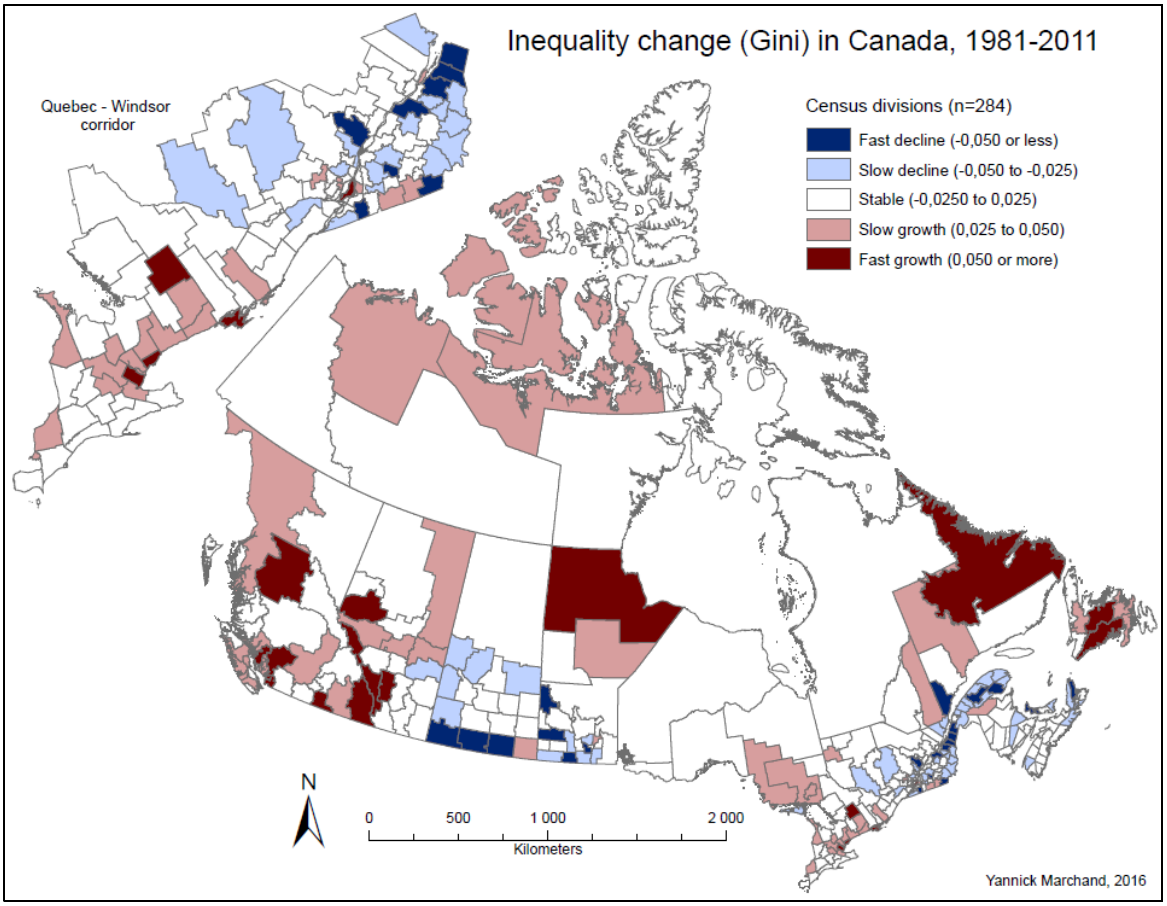

Here, red shaded areas represent census divisions with highly unequal income distributions. In contrast, blue shaded areas are those census divisions with lower levels of inequality. The darker the shading, the higher or lower the level of inequality depending on the color (red or blue). In 1981, the urban-rural divide discussed earlier is not so evident. That said, what does stand out from the map is an east-west divide in patterns of inequality. Western provinces are in general depicted by more census divisions shaded in red (higher inequality) whereas the eastern provinces, from Ontario eastward, are on majority shaded in blue (lower inequality). In previously reported findings, I noted that sub-national variations in levels of inequality increased during the 1981 to 2011 period. Figure 10, which is developed for the 2011 Gini coefficient of total income validates those findings.

**Figure 10**. Gini coefficient of income in Canada, 2011




The 2011 map (Figure 10) counts a higher number of dark shaded areas, both red and blue, meaning that national sub-variations in inequality have indeed increased. The east and west divide is again visible although pockets of high inequality are also now evident in the east. In the west, higher levels of inequality seem to have spread to British-Columbia. In the east, southern Ontario and Newfoundland are now showing higher inequality. The urban-rural divide is also characterised by increasingly unequal urban areas which isdriven by large metropolitan areas such as Toronto, Montreal, Ottawa, Calgary and Vancouver (all in dark shades of red). To simplify the analysis of the variation in inequality, I turn to Figure 11 that maps the change in Gini coefficients of total income between 1981 and 2011.

**Figure 11**. Inequality change (Gini) in Canada, 1981-2011




In this map, the color scheme is broken down to distinguish census divisions that experienced a fast decline in inequality between 1981 and 2011 with those where inequality rose sharply. Again, we see clearly the changes in the regions of Newfoundland, southern Ontario as well as in parts of Manitoba and a sharp decline in inequality in southern Saskatchewan. Adding to these changes are the Vancouver and Estrie region in Quebec, where I can denote further pockets (or clusters) of groups of census divisions that faced similar changes over the last 30 years. The term pocket is used intentionally here as I next turn to the analysis of the spatial dependence of income inequality across the Canadian landscape

### 3.6.1 Spatial autocorrelation

The above spatial analysis of a first set of maps gave us reason to believe that patterns of inequality across the Canadian landscape are positively spatially autocorrelated. Recall that spatial autocorrelation is present if the level of inequality in one census division is related to levels of inequality in neighboring census divisions. The concept is best embodied in Tobler's first law of geography: everything is related to everything else, but near things are more related than distant things (Tobler, 1970)

The Moran's I statistic can be used to detect global spatial autocorrelation. It is conveniently bound between -1 and 1 and can be interpreted like a correlation coefficient. A value of -1 indicates a pattern of perfect dispersion whereas a value of 1 indicates perfect clustering. In statistics, the presence of spatial autocorrelation leads to a violation of the assumption of independence, rendering statistical inferences invalid. Mathematically, the Moran's I statistics can be specified as follows:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

where N is the number of census divisions, X is the inequality metric for regions *i* and *j*, X bar is the mean value of inequality across all regions, and $w_{i,j}$ is an element of a spatial weight matrix (more on this below).

In spatial analysis, the nature of the relation between objects *i* and *j* is captured through the spatial weights matrix. Depending on the type of matrix adopted, the matrix records the following information for every spatial object: the number of neighbors, the distance between those neighbours, the lengths of borders, etc. The selection of a weight matrix depends on the researcher's assumptions and the spatial interaction in question. In my case, I preferred a contiguity based matrix over a distance matrix given the Canadian regional landscape which is characterized by large variations in the size of census divisions.

Building on the information presented above, Table 11 lists the Moran's I coefficients for the Gini values of total income and the Theil index based on wages. The spatial weights matrix used is a first order queen's contiguity matrix (sides plus vertices).

**Table 11**. Moran's I of inequality level in Canada, 1981 to 2011

| Year | Gini coefficient of total income | Gini coefficient of wage | Theil index of total income | Theil index of wage |
|---|---|---|---|---|
| 1981 | 0,531 | 0,345 | 0,502 | 0,337 |
| 1986 | 0,509 | 0,45 | 0,472 | 0,425 |
| 1991 | 0,487 | 0,495 | 0,316 | 0,429 |
| 1996 | 0,491 | 0,513 | 0,346 | 0,444 |
| 2001 | 0,481 | 0,523 | 0,323 | 0,363 |
| 2006 | 0,59 | 0,567 | 0,393 | 0,376 |
| 2011 | 0,538 | 0,491 | 0,349 | 0,349 |

Note: all statistics are significant at 0,05 level.

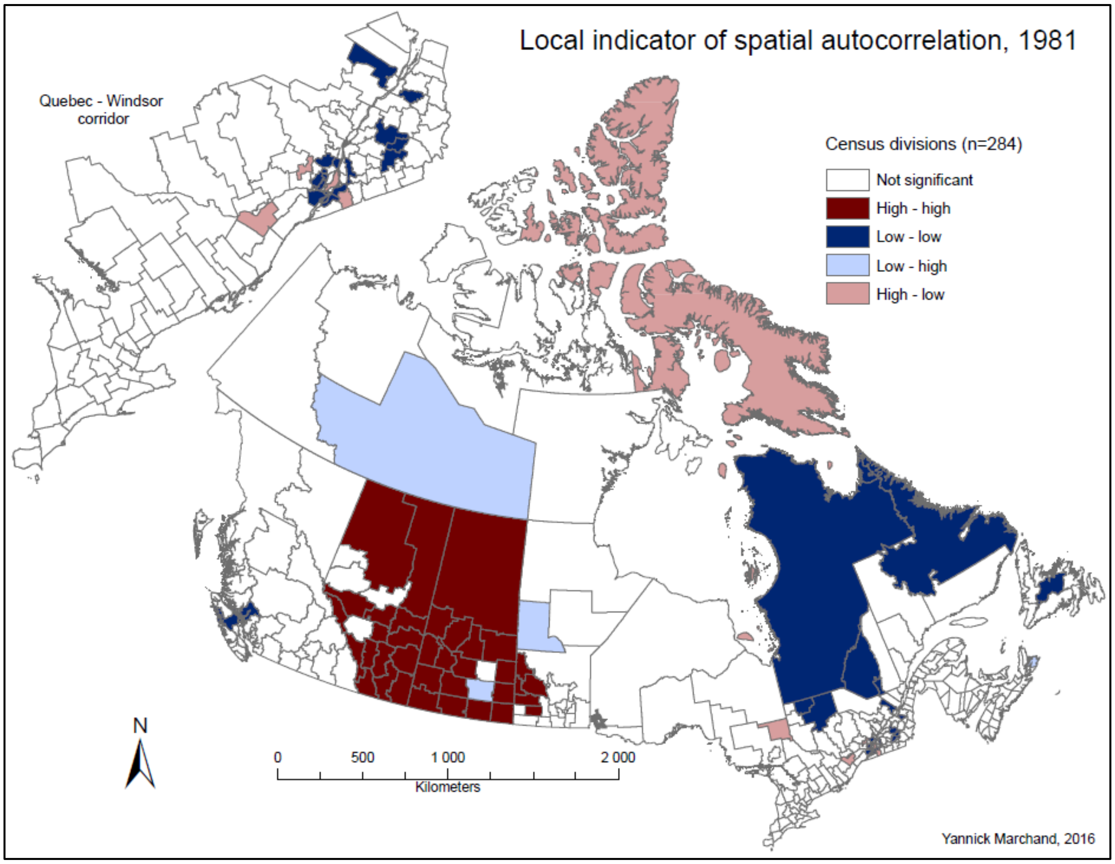
As Table 11 shows, income inequality in Canada is positively spatially autocorrelated, more so in 2011 than in 1981. These observations are robust to more than one definition of inequality. There is no clear trend as to how the concentration of inequality across regions increased to reach higher levels in 2011. In fact, the Theil index of wage showed an all-time concentration level in 1996. Similarly, spatial autocorrelation as per the Gini coefficient decreased for 15 years, before alternating between ups and downs every five years. Nonetheless, throughout the period of study, census divisions with high (low) inequality level have been a permanent feature of the Canadian landscape.

Despite allowing us to identify the presence of spatial autocorrelation, the Moran's I can not shed light on the location of those clusters within the country. To do so, I turn to Local Indicators of Spatial Association (LISAs) to further investigate the location of inequality clusters across the Canadian landscape.

### 3.6.2 LISAs

Anselin (1995) defined a LISA as a statistic that for each observation of interest gives us an indication of the extent of significant spatial clustering of similar values around that observation. It can be viewed as a local Moran's I given that the sum of LISAs for all observations is equal to the global Moran's I (presented above). Figure 12 maps the census divisions classified by their corresponding LISAs value for the year 1981.

**Figure 12**. Local indicator of spatial autocorrelation, 1981




Note that census divisions are color coded according to the type of spatial autocorrelation among Gini coefficients. Not significant LISA values are left blank. Otherwise, dark red CDs represent high-high clusters where census divisions with high levels of inequality are surrounded by census divisions with similarly high levels of inequality. Dark blue CDs represent low-low regions or those clusters with more equal census divisions. Light shaded regions represent spatial outliers where census divisions in light blue have low levels of inequality while their neighboring census divisions have high levels of inequality (or vice versa). I am most interested in positive spatial autocorrelation, namely the dark shaded region over Alberta and Saskatchewan and the low inequality clusters in Quebec, Newfoundland and British Columbia. The presence of statistically significant clusters of inequality supports my previous findings for regional patterns of inequality in 1981.

**Figure 13**. Local indicator of spatial autocorrelation, 2011

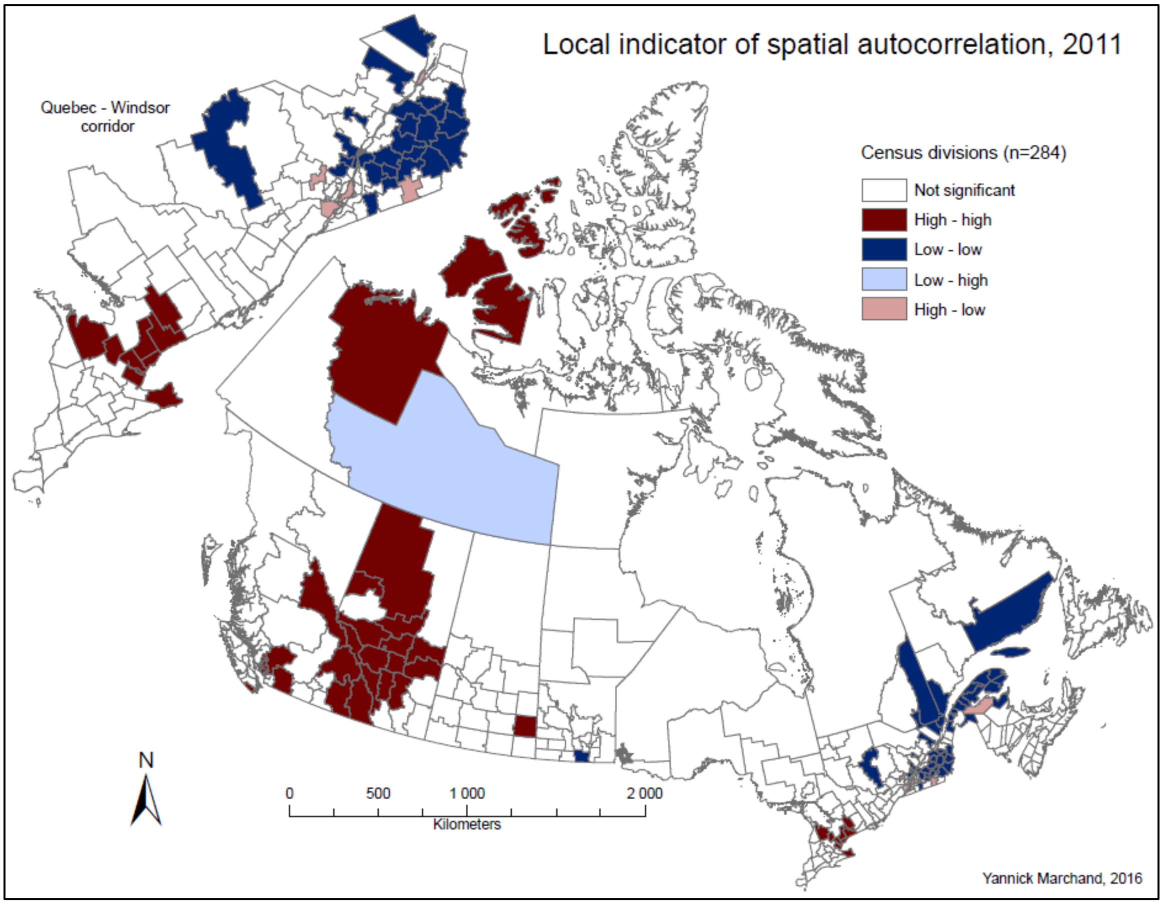



Figure 13 presents the census divisions LISA values for 2011. Several interesting changes over time are worth noting. First, we find the pocket of high inequality has moved west towards British Columbia and left Saskatchewan. Looking east, the cluster of more equal census divisions in the Quebec province, south of Montreal has more than doubled in size from 1981 to 2011. And while Newfoundland no longer appears as a cluster of income equality, the greater Toronto region has become a significant cluster of high inequality. Overall, regional patterns of inequality have evolved in 30 years, but clusters still remain an important aspect of income inequality in Canada.

Throughout Chapter 3, I have explored and identified some of the key defining trends of income inequality across regions in Canada over the 1981 to 2011 period. I found that those trends differ along an urban and rural divide, and even more so along refined rural and urban areas. In fact, important variations, patterns and clusters are found at the census divisions level itself. In Chapter 4, my attention will shift to identifying the factors associated with these patterns of inequality across census divisions using various spatial regression models. Before moving on to this analysis, the following sub-section provides a short overview of the independent variables introduced in Chapter 1, namely does who I identified in the introduction as potential determinants of regional inequality in Canada.

### 3.2 Descriptive statistics for independent variables

Table 12 provides a snapshot of the mean values of the independent variables used in the analysis for 1981 and 2011, and the 30-year variation for the national, urban and rural levels. The goal behind presenting these numbers is to look at notable differences in urban and rural economies that could help explain the findings above. This will inform the discussion for the following set of regression model results.

***Table 12***. *Descriptive statistics of potential contributors to regional inequality level*

| | | National | | | Urban | | | Rural | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **1981** | **2011** | **%Δ** | **1981** | **2011** | **%Δ** | **1981** | **2011** | **%Δ** |
| | **Labour market and economic variables** | | | | | | | | | |
| | Median wage | 7144,18 | 45853,37 | 84% | 7520,71 | 49124,45 | 85% | 6880,37 | 43561,65 | 84% |
| | Unemployment rate | 6,7% | 6,5% | -2% | 5,4% | 4,7% | -16% | 7,5% | 7,8% | 4% |
| | Part-time worker (%) | 12,8% | 13,2% | 3% | 12,0% | 12,4% | 3% | 13,4% | 13,8% | 3% |
| | Self-employed worker (%) | 3,8% | 6,3% | 40% | 2,9% | 5,8% | 50% | 4,4% | 6,7% | 34% |
| | **Industrial variables** | | | | | | | | | |
| | *Secondary sector (%)* | 26,3% | 18,5% | -42% | 29,3% | 19,5% | -51% | 24,1% | 17,9% | -35% |
| | Manufacturing (%) | 19,0% | 11,1% | -71% | 22,2% | 12,2% | -82% | 16,7% | 10,4% | -61% |
| | Resource-intensive (RII) (%) | 9,4% | 5,0% | -87% | 8,0% | 4,0% | -98% | 10,3% | 5,7% | -81% |
| industries | Labour-intensive (LII) (%) | 3,5% | 1,6% | -112% | 5,1% | 2,0% | -150% | 2,3% | 1,4% | -72% |
| | Scale-intensive (SII) (%) | 3,7% | 2,4% | -58% | 5,4% | 3,2% | -71% | 2,6% | 1,8% | -42% |
| | Differentiated goods (DG) (%) | 1,6% | 1,2% | -33% | 2,3% | 1,5% | -53% | 1,0% | 0,9% | -10% |
| | Science-based (SBI) (%) | 0,7% | 0,8% | 15% | 1,4% | 1,4% | 3% | 0,2% | 0,4% | 43% |
| | Construction (%) | 7,3% | 7,4% | 2% | 7,1% | 7,3% | 2% | 7,4% | 7,5% | 1% |
| | *Tertiary sector (%)* | 23,1% | 21,4% | -8% | 24,0% | 22,4% | -7% | 22,4% | 20,7% | -8% |
| | Transportation (%) | 5,7% | 4,5% | -27% | 5,6% | 4,3% | -30% | 5,8% | 4,7% | -24% |
| | Communication (%) | 2,1% | 2,2% | 4% | 2,3% | 2,7% | 12% | 1,9% | 1,8% | -5% |
| | Utilities (%) | 1,5% | 1,3% | -21% | 1,5% | 1,1% | -30% | 1,6% | 1,4% | -16% |
| | Retail (%) | 9,6% | 10,2% | 5% | 9,8% | 10,1% | 2% | 9,5% | 10,3% | 7% |
| | Wholesale (%) | 4,0% | 3,3% | -24% | 4,7% | 4,2% | -12% | 3,6% | 2,6% | -37% |
| | *Quaternary sector (%)* | 40,7% | 52,3% | 22% | 41,6% | 54,1% | 23% | 40,2% | 51,0% | 21% |
| | FIRE (%) | 3,8% | 4,0% | 5% | 4,6% | 4,9% | 6% | 3,3% | 3,4% | 5% |
| | Public administration (%) | 9,0% | 9,2% | 2% | 9,6% | 9,3% | -3% | 8,6% | 9,0% | 5% |
| | Education and health (%) | 17,8% | 21,8% | 18% | 16,9% | 21,2% | 20% | 18,5% | 22,3% | 17% |
| | Leisure and food (%) | 4,9% | 6,0% | 19% | 4,5% | 5,5% | 19% | 5,1% | 6,4% | 19% |
| | KIBS (%) | 1,4% | 3,9% | 65% | 1,8% | 5,2% | 65% | 1,0% | 3,0% | 65% |
| | Business services (%) | 3,8% | 7,3% | 48% | 4,2% | 8,0% | 48% | 3,5% | 6,8% | 48% |
| | **Socio-demographic variables** | | | | | | | | | |
| | Female participation rate | 35,4% | 47,8% | 26% | 36,3% | 48,0% | 25% | 34,9% | 47,6% | 27% |
| | Visible minorities (%) | 4,0% | 3,6% | -10% | 2,3% | 6,2% | 63% | 5,1% | 1,8% | -189% |
| | Education ratio | 51,0% | 31,7% | -61% | 48,1% | 32,0% | -50% | 53,1% | 31,5% | -69% |
| | Young (%) | 25,1% | 17,2% | -46% | 23,8% | 16,9% | -41% | 26,0% | 17,4% | -50% |
| | Senior (%) | 9,2% | 15,8% | 41% | 8,9% | 14,8% | 40% | 9,5% | 16,5% | 42% |
| | **Institutional variables** | | | | | | | | | |
| | Minimun wage | 1,80 | 11,62 | 84% | 1,79 | 11,64 | 85% | 1,81 | 11,60 | 84% |
| | Transfers to persons (per capita) | 600,67 | 6504,67 | 91% | 591,25 | 6371,71 | 91% | 607,28 | 6597,83 | 91% |
| | Unionization rate (%) | 32,7% | 33,5% | 3% | 32,7% | 33,1% | 1% | 32,6% | 33,8% | 4% |

I begin by looking at labour market and economic variables where we note that the median wages (the variable I use as a proxy for economic development) is higher in urban areas then in rural areas. Unemployment rates also tend to be higher in rural areas. At the national level, around 13% of the workforce consists of part-time workers with no difference along the urban rural divide. On the other hand, self-employed workers are more prevalent in rural areas, but urban areas have seen the number of such workers increase by 50% in the last 30 years.

Moving to the industrial variables, the numbers provide clear evidence of the infamous manufacturing decline of the 1980s and 90s. This deindustrialization phenomenon affected both rural and urban areas equally. Within manufacturing itself, the resource-intensive and labour-intensive sectors are the most impacted. In 2011, the percentage of workers in manufacturing industries is similar in both urban and rural areas, 12,2% and 10,4% respectively. Resource-industries tend to be higher in rural areas, while science-based industries operate almost exclusively in the cities. The tertiary sector between 1981 and 2011 is punctuated by employment declines in both rural areas and urban areas. Only the retail sector is increasing in the former, while the communications sector is growing in the latter. Similarly, urban areas suffered a 3% decrease in public administration employment share, to the expense of rural areas where it grew by 5% in 30 years. Finally, the knowledge intensive business services (KIBS) soared by 65% nationally between 1981 and 2011, an increase that is surprisingly common to both urban and rural areas.

On the socio-demographic front, the female participation rate in the workforce grew by 25% in 30 years, reaching 47.8% in 2011. During the same period, the number of visible minorities has increased in urban areas, but decreased in rural areas. Most notably, I find that the variation in educational attainment of workers dropped 61% from 1981 to 2011. In other words, the proportion of high school drop-outs to workers with a bachelor degree or more has decreased, more so in rural areas than urban areas. Lastly, the country is growing older as the number of individuals aged 15 or less declined by 46% while the proportion of seniors (65 years old or more), jumped by 41%. This is a trend observed in both urban and rural areas.

Finally, the bottom section of Table 12 shows the institutional variables which reveal similar trends along the urban and rural divide. Nationally, minimum wages increased from 1,80$ in 1981 to 11,62$ in 2011. The amount of transfers to persons (per capita) also went up by 91%. Relatively little changed in terms of union contracts over time. All three of these variables should, in theory, contribute to lowering levels of income inequality. This is a hypothesis, among others, that am ready to put to test in the following chapter.

# CHAPTER 4: MODELLING THE CAUSES OF REGIONAL INEQUALITY IN CANADA

## Introduction

In Chapter 3, I described how sub-national levels of inequality in Canada varied from 1981 to 2011 played out. However, I did not explain the mechanisms behind the changes and variations in income inequality across Canada. One key objective of this project is to identify some of the potential contributors to regional inequality and identify whether or nor there are significant differences in these mechanisms between urban and rural areas. Thus, to try and explain the changes in regional income distributions of income over the period of study, in this chapter I develop a spatial panel regression modeling approach.

Spatial panels have gained in popularity in the 2000s as they offer researchers extended modelling possibilities (Elhorst, 2012). The spatial panel regression model operates on data containing time series observations for several spatial units. As its name indicates, the models account for both spatial and time effect. Spatial regression models are used to correct for spatial autocorrelation while the panel dimension captures the information contained in the time series dimension.

As per our discussion in Chapter 3, I know that spatial autocorrelation is present in my data. Therefore, I am at risk of producing biased estimates as spatial autocorrelation violates the assumption of the independence of observations. To get around this problem, I use the spregxt.ado command in Stata in order to generate the spatial panel model results. This new suites of programs was developed by Shehata (2016) and provides the basis for all the regression models estimated and presented below.

I begin this chapter by presenting the regression model specification and a first set of results for a  spatial panel error model. I then look at a second model with more detailed industry variables. In a third section, I compute a Chow test which reveals a spatial structural break in the data that is addressed by re-estimating the models separately for urban and rural regions. I conclude with an argument suggesting that different factors contribute to inequality in rural and urban areas.

## 4.1 The modelling framework

Based on the theoretical framework presented in Chapter 1, the first model I estimate is a maximum likelihood based fixed effects spatial error model specified as:

$$y = (L_T \otimes I_N)\mu + X\beta + u \qquad \text{Eq. (4.1)}$$

$$u = \rho\,(I_T \otimes W_N)u + \varepsilon,$$

where $\rho$ is the spatial autocorrelation coefficient, $W_N$ a non-stochastic spatial weights matrix, $L_T$ a column vector of ones of dimension T, $I_N$ an identity matrix of N x N, and $\varepsilon$ the error term. The spatial weights matrix I use is a rook spatial matrix of 1st order[5]. A spatial error model is used to handle the spatial dependence due to the omitted variables or errors in measurement through the error term. The Gini coefficient of total income is my dependent variable (Y) and X is my vector of control variables as defined earlier.

Table 13 shows the results for regression model with all variables of interest in column 1 and the more detailed industrial variables presented in column 2.

Table 13 Fixed effects spatial error model results

| | Dep. Variable = Gini coefficient of total income | | | |
|---|---|---|---|---|
| | (1) All factors | | (2) with detailed industries | |
| Intercept | 0,222 | *** | 0,238 | *** |
| | | | | |
| **Labour market and economic variables** | | | | |
| Median wage | 0,000 | *** | 0,000 | *** |
| (Median wage)$^2$ | -0,000 | ** | -0,000 | * |
| Unemployment rate | 0,107 | *** | 0,088 | *** |
| Part-time worker (%) | 0,159 | *** | 0,197 | *** |
| Self-employ worker (%) | 0,136 | *** | 0,038 | |
| | | | | |
| **Industrial variables** | | | | |
| *Secondary sector (%)* | -0,040 | *** | | |
| Resource-intensive industries | | | -0,080 | *** |
| Labour-intensive industries | | | -0,036 | ** |

[5] I previously used a queen spatial matrix of 1st order in Chapter 3, both rook and queen produced similar results.

| | | | | |
|---|---|---|---|---|
| Scale-intensive industries | | | -0,093 | *** |
| Differentiated goods | | | -0,039 | |
| Science-based industries | | | -0,069 | |
| *Tertiary sector (%)* | 0,050 | *** | | |
| Transportation (%) | | | -0,079 | ** |
| Communication (%) | | | -0,020 | |
| Utilities (%) | | | 0,135 | *** |
| Retail (%) | | | -0,006 | |
| Wholesale (%) | | | -0,026 | |
| *Quaternary sector (%)* | 0,023 | ** | | |
| FIRE (%) | | | 0,057 | |
| Public administration (%) | | | -0,079 | *** |
| Education and health (%) | | | -0,011 | |
| Leisure and food (%) | | | 0,120 | *** |
| KIBS (%) | | | 0,240 | *** |
| | | | | |
| **Socio-demographic variables** | | | | |
| Female participation rate | 0,047 | ** | 0,050 | ** |
| Visible minorities (%) | 0,092 | *** | 0,071 | *** |
| Education ratio | 0,093 | *** | 0,082 | *** |
| Young (%) | 0,059 | ** | 0,145 | *** |
| Senior (%) | 0,107 | *** | 0,115 | *** |
| | | | | |
| **Institutional variables** | | | | |
| Minimun wage | -0,006 | *** | -0,007 | *** |
| Transfers to persons (per capita) | -0,000 | * | 0,000 | |
| Unionization rate (%) | -0,119 | *** | -0,123 | *** |
| | | | | |
| **Spatial and time components** | | | | |
| Rho | 0,376 | *** | 0,361 | *** |
| Sigma | 0,018 | *** | 0,017 | *** |

*Note:* *, ** and *** denote statistical significance at the .10, .05 and .01 levels, respectively.
For the industrial variables, the primary sector acts as the base category for the industrial variables, which was necessary to avoid multicollinearity problems.

A first look at column (1) reveals that labour market and economic variables are significant in explaining varying levels of regional inequality in Canada. I find that the higher the level of economic development in one region (as proxied by the median wage) positively correlates to higher levels of income inequality. That relation tends to decline as the level of economic development reaches a higher state as demonstrated by the negative value of the median wage squared.

I also find evidence confirming the hypothesis of a more precarious employment sector leading to higher level of income inequality across the regions. Here, the regional unemployment rate, the percentage of part-time workers and self-employed workers are all positively correlated with higher levels of inequality.

In terms of regional industrial structures, my findings align those of Breau (2015) where regions with higher percentages of workers in the secondary sector tend to have a more equal distribution of incomes. Those are the jobs in the manufacturing industry as well as the construction sectors. Compared to Breau (2015), whose analysis focuses on the shorter 1996 to 2006 timespan, I find that the above relationship has also been consistent from 1981 to 1996 and and up to 2011. I also push the analysis further by testing whether specific branches of the manufacturing sector are more significant than others in creating more equal distributions of income. In column (2), I present results for the manufacturing sector split in five sub-sectors (as described in Chapter 2). Regression results suggest that although all manufacturing jobs contributes to lower level of inequality, jobs in resource, labour and scale intensive industries are the main sectors leading to lower levels of inequality.

On the other hand, regions with larger tertiary and quaternary sectors tend to experience higher levels of inequality. Because these sectors are large and highly diversified in the type of jobs they represent, I again turn to column (2) of the Table for a more fine grained interpretation of the influence of a specific job sector on the distribution of income across regions. Jobs in the transportation, communications, retail and wholesale sectors all correlate negatively with inequality. However, all but the transportation employment sector, are not significant to the 0,10-confidence level. Therefore, the overall trend of a tertiary sector contributing to high inequality level comes from the strongly significant relationship between the employment share in utilities and the Gini coefficient of total income. Similarly, in the quaternary sector jobs in public administration, education and health sectors dampened the level of inequality. However, with only the public administration jobs being statistically significant, the quaternary sector as an whole contributes to higher levels of inequality, especially through jobs in the leisure, food and knowledge-intensive business sectors.

Turning to the socio-demographic variables, I find that regions with higher percentages of female participation in the labor force show higher levels of income inequality. Similarly, the percentage of visible minorities has the same effect. In both cases, the results concur with the literature in highlighting the important work that remains to be done to understand diverging patterns of income inequality between men and women and visible minorities. That these relationships hold from 1981 to 2011 is of interest.

The polarization of educational attainment within populations is another contributor to high levels of inequality (see also Glaeser et al. 2009). Here, the coefficient estimate is positive and statistically significant on the education ratio (which you will recall, reports the number of high-school drop-outs plus those with college degrees over all other workers).

In the same vein, the percentage of individuals aged less than 15 years old and older than 65 years old are significant contributors to levels of inequality across regions. This result is in line with the dependency ratio argument in the international economic development literature (see Chechhi 2004) which refers to the increasing stress placed on the active workforce the larger the ratio gets.

Looking at institutional variables, higher minimum wages and higher unionization rates tend to reduce inequality level. The amount of transfers to persons also dampens inequality, but this result is less statistically significant and it is only in one of the two models presented. Overall, results for institutional variables are as expected: higher minimum wages contribute to more equal societies as do higher unionization rates.

The last two variables presented in Table 13 show the spatial and time components of the regression models. Remember that rho is meant to account for spatial autocorrelation in the error term where a positive and significant value for rho means that spatial autocorrelation is a factor in the regression analysis and rho corrects for it. Sigma is the unobserved time invariant individual effect which is also positive and significant meaning that I control for shocks in different time periods.

## 4.2 A structural break

The divide between rural and urban areas is a central theme of the current project. Because of the spatial nature of data and specific questions raised in terms of urban and rural differences in inequality levels I ought to validate if a structural break between urban and rural regions is present in my data.

Structural stability tests are traditionally use in time series analysis to validate that coefficients vary randomly and are not based on events in specific time periods. From a statistical point of view, structural instability causes the results to be prone to forecasting errors and are considered unreliable more generally. However, in this project, I do not test to see whether there is a structural break in the time components; rather, I do so in the spatial components. From a methodological standpoint, and given the results of chapter 3 that shed light on different trends of income inequality between urban and rural spaces, it is therefore crucial to test for a spatial structural break (more so than for the periodization of 1981 to 2011). In order to test for a structural break between rural and urban spaces in Canada, we use a simple Chow test (Chow, 1960) as defined in Chapter 2.

If significant, the Chow test forces us to run separate regression models for urban and rural regions. The equation (Eq. 4.1) remains the same, though the number of observations will change.

The formula of the Chow test is as followed:

$$F = \frac{RSS_c - (RSS_1 + RSS_2)/k}{RSS_1 + RSS_2 / n - 2k}$$

$$RSS_c = combined\_RSS$$

$$RSS_1 = rural\_RSS$$

$$RSS_2 = urban\_RSS$$

where RSS stands for residual sum of squares, *n* the number of observations and *k* the number of parameters. The test statistic follows the *F* distribution with k and $N_1 + N_2 - 2k$ degrees of freedom. The null hypothesis is structural stability, if I reject the null hypothesis, it means that I have a structural break between rural and urban regions.

The calculations associated with the Chow test (which can be seen in appendix A2) show that a structural break is indeed present in the data. Accordingly, Table 14 presents the regression model estimates for both the urban and the rural spaces separately. I run a fixed effects spatial error model for the urban census divisions and one for the rural census divisions. The spatial weights matrix is a nearest neighbor matrix with three neighbors. A rook or queen matrix could not be use, as I create many 'islands' by dividing the geography into two distinct spaces, rural and urban spaces.

**Table 14**. Fixed effects spatial error model results for urban and rural regions

| | Gini coefficient of total income | | | |
|---|---|---|---|---|
| | (1) Urban | | (2) Rural | |
| Intercept | 0,344 | *** | 0,239 | *** |
| | | | | |
| **Labour market and economic variables** | | | | |
| Median wage | 0,000 | *** | 0,000 | ** |
| (Median wage)² | 0,000 | ** | 0,000 | |
| Unemployment rate | -0,009 | | 0,063 | *** |
| Part-time worker (%) | 0,209 | *** | 0,181 | *** |
| Self-employ worker (%) | 0,030 | | 0,071 | ** |
| | | | | |
| **Industrial variables** | | | | |
| *Secondary sector (%)* | | | | |
| Resource-intensive industries | -0,104 | *** | -0,067 | *** |
| Labour-intensive industries | -0,053 | * | -0,054 | ** |
| Scale-intensive industries | -0,168 | *** | -0,053 | ** |
| Differentiated goods | -0,087 | ** | -0,071 | |
| Science-based industries | -0,055 | | -0,226 | ** |
| *Tertiary sector (%)* | | | | |
| Transportation (%) | -0,199 | *** | -0,098 | ** |
| Communication (%) | -0,021 | | -0,042 | |
| Utilities (%) | -0,007 | | 0,149 | *** |
| Retail (%) | -0,170 | *** | 0,090 | ** |
| Wholesale (%) | 0,007 | | -0,025 | |
| *Quaternary sector (%)* | | | | |
| FIRE (%) | -0,102 | ** | 0,082 | |
| Public administration (%) | -0,159 | *** | -0,022 | |
| Education and health (%) | -0,057 | ** | -0,038 | ** |
| Leisure and food (%) | 0,158 | *** | 0,098 | *** |
| KIBS (%) | 0,381 | *** | 0,077 | * |
| | | | | |
| **Socio-demographic variables** | | | | |
| Female participation rate | 0,002 | | 0,066 | ** |
| Visible minorities (%) | 0,079 | *** | 0,004 | |
| Education ratio | 0,058 | *** | 0,033 | ** |
| Young (%) | 0,018 | | 0,218 | *** |
| Senior (%) | 0,169 | *** | 0,105 | *** |
| | | | | |
| **Institutional variables** | | | | |
| Minimun wage | -0,010 | *** | -0,005 | *** |

| | | | | |
|---|---|---|---|---|
| Transfers to persons (per capita) | 0,000 | | 0,000 | |
| Unionization rate (%) | -0,103 | *** | -0,125 | *** |
| | | | | |
| **Spatial components** | | | | |
| Rho | 0,247 | *** | 0,355 | *** |
| Sigma | 0,014 | *** | 0,017 | *** |

*Note:* *, ** and *** denote statistical significance at the .10, .05 and .01 levels, respectively.

Both models are interesting in that they allow us to identify factors that contribute either positively or negatively to income inequality in Canadian regions. It is especially interesting to compare the effect of independent variables between the two types of urban-rural spaces.

I begin with the labour market and economic variables.  Median wages contribute to rise inequality in both rural and urban environments. Similarly, the model results suggest that unemployment rates are particularly important in explaining inequality in rural areas as they are not significant in urban areas. On the other hand, the percentage of part time worker is an important contributor to higher levels of inequality in all Canadian regions. The last labour market variables highlight yet another difference between urban and rural areas as the percentage of self-employed is a positive and significant contributor of inequality in rural areas, but not significant in urban areas.

Turning to industrial variables, I find more evidence that the inequality in income distributions of Canadian regions is shaped differently depending on whether the region is urban or rural. Resource-, labour- and scale-intensive industries all contribute to lowering inequality regardless of whether or not the region is urban or rural. However, differentiated goods industries in cities lowers inequality while having no effect in rural areas. On the other hand, the presence of science-based industries in rural areas positively contributes to more equal distribution of incomes. On the contrary, looking at the tertiary sector, I find that utilities in rural areas are positively related to inequality. Interestingly, the opposite can be said of urban areas; while non-significant, the utilities sector there reduces inequality.

Looking at socio-demographic variables, I find similar conclusions to those reported above. First, the higher the percentage of women in the labor force, the higher levels of inequality in rural areas whereas this finding is not significant in urban areas. The opposite can be said of the percentage of visible minorities, where I find that cities with high level of visible minorities tend to be more unequal than others. The education ratio, as reported in earlier models (and as in Breau 2015), is again positively correlated with high levels of inequality regardless of the type of region. The higher the percentage of individuals aged 15 or less in a rural region suggests that inequality is higher there than other regions while it is not significantly affecting urban regions. On the other hand, the proportion of senior citizens is significantly correlated to income inequality in both rural and urban regions.

There is no difference in the conclusions reached for the institutional variables between models (1) and (2). This is not surprising given that the variables are provincial level contextual variables. The coefficients do change, but their significance and influence on inequality does not. A final note: in both models, I see that spatial autocorrelation was a factor and given a significant rho means that I at least correct for some of it.

## 4.3 Conclusion

The main goal of this chapter was to identify the determinants of income inequality across Canadian regions. I found that the best model to identify those factors was a spatial error panel model, given the time dimension observed (from 1981 to 2011) and the spatial nature of the data. Because the variables under study accounted for some of the socio-economic changes that took place in Canada over the last 30 years and because the data are built from the long form censuses and the 2011 NHS, I felt that a time structural break could not bias the results. The same could not be said of the spatial characteristics of the Canadian landscape in terms of inequality. Chapter 3 had previously identified differences in income inequality trends along the urban and rural divide. Such differences indeed lead to the positive testing of a structural break between the two spaces. I concluded Chapter 4 by presenting separate regression models for rural and urban areas. The Gini coefficient of total income was the dependant variable of choice and was regressed against a variety on independent variables identified as possible determinants in the literature. My results lead to important differences in the impact of variables across the urban-rural spectrum.

To conclude this thesis, I will be looking at one of the potential consequences of inequality level on a region by looking at how it affects economic growth there.

# CHAPTER 5: HOW DOES INEQUALITY AFFECT GROWTH? EVIDENCE FROM A PANEL OF CANADIAN REGIONS

## Introduction

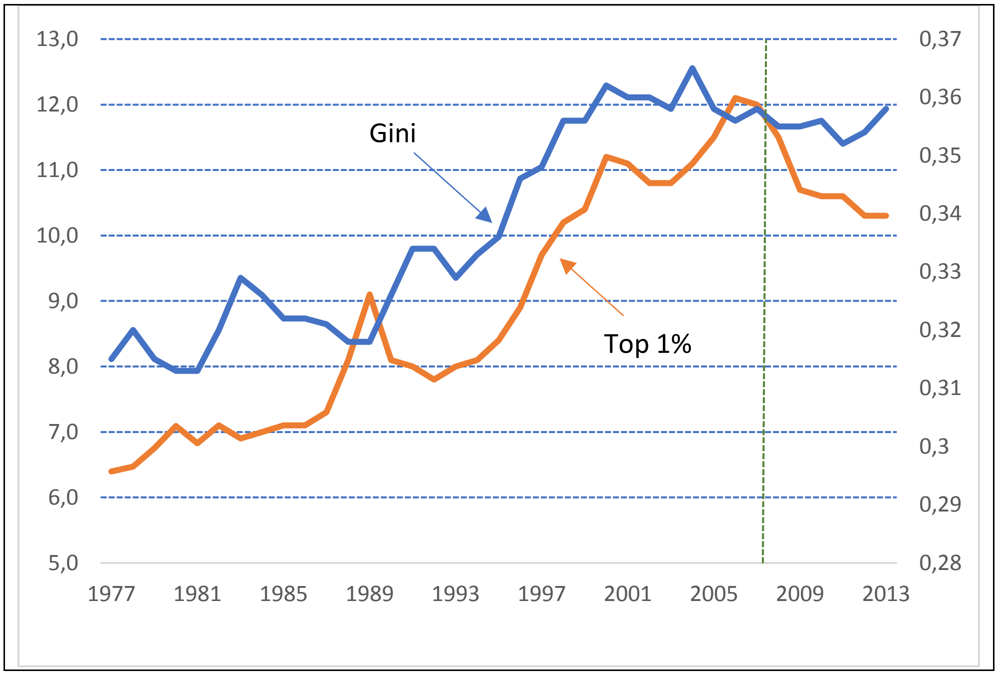
While the previous chapter focused on understanding the causes of inequality, in this chapter I examine the consequences of inequality on regional economic growth. There is a long history of studying regional disparities in Canada[6]. The general consensus among scholars is that the income gap between regions declined from the late 1950s to the mid-1980s, at which point the convergence process lost steam and became more 'episodic' with alternating periods of both convergence and divergence (see Brown and Macdonald 2015; Breau and Saillant 2016). The empirical evidence also suggests that regional income disparities remain comparatively high in Canada. They are about 50 percent higher than the average observed across US states (Coulombe 1999) and remain among the top three highest across OECD countries (OECD 2014).

While I would not expect economic disparities between regions to necessarily disappear entirely (Polèse 2014), results from the previous chapter suggest that inter-regional income inequality has been accompanied by increasing social inequality within regions (see also Breau 2015). This is part of a broader movement towards rising inequality observed in several OECD countries (see OECD 2011) which has led to a resurgence of interest in understanding distributional dynamics among economists and regional scientists (Stiglitz 2012, Piketty 2014, Atkinson 2015, Cavanaugh and Breau 2017).

In Chapter 3, I saw that inequality increased by close to 10% over the 1981 to 2011 period. Figure 14 suggest that overall levels of inequality have increased by almost 15 percent from the late 1970s to 2013 and points to an even more pronounced increase in the growth in the concentration of income among the top 1 percent of the population (almost double what it was 30 years ago). While the trajectory of inequality peaked just before the Great Recession of 2008, levels of inequality in Canada remain at historically high levels. This raises concerns about the impact of inequality on society in general and questions related to the potential impacts of higher inequality on the economic performance of regions in particular.

[6] Savoie (2017) provides a nice overview of the history of regional economic development in Canada.

**Figure 14**. Evolution of income inequality in Canada, from the late 1970s onwards



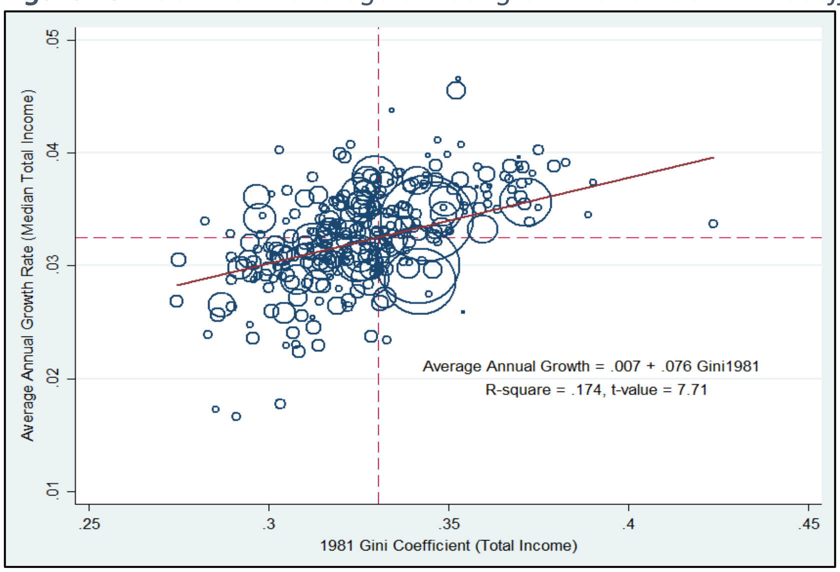

Source: CANSIM Tables 206-0031 and 202-0705.

The goal of this paper is to examine the relationship between inequality and growth using the novel panel dataset of regional income distribution measures developed earlier for the period 1981 to 2011. At first glance (see Figure 15), this relationship appears to be positive whereby regions with higher levels of inequality in 1981 subsequently experience faster average annual growth rates. Yet, with less than 20 percent of the overall variation in regional growth rates during this 30-year period explained by the initial level of inequality, the robustness of those results needs to be ascertained through the inclusion of other factors accounting for economic growth patterns across regions. Empirical studies also show that the length of the study period, the time window examined and the types of regions included tend to yield different results on the direction and strength of the relationship between inequality and growth (Partridge 1997, 2005, 2007; Panizza 2002, Frank 2009).

In order to examine the impact of those factors on the relationship between regional economic growth and inequality in Canada I ask the following questions: Does the inclusion of other factors accounting for inequality alter the effects of inequality on growth? Are the effects of inequality on growth persistent only over long periods of time or do the effects vary over the short-/medium-term horizon? Does the inequality/growth relationship vary between urban and rural regions?

As I saw in Chapter 3, the geography of income inequality varies considerably across the country. We have shown, for instance, that there is an apparent east-west divide where regions in the eastern parts of the country generally have lower levels of inequality compared to their western counterparts. The second observation derived from my earlier analysis is a strong urban-rural divide that exists within regions, with urban regions generally showing much higher levels of inequality. Thus, a further question to be examined here is just how important are differences between urban and rural regions in terms of influencing the mechanisms that shape the inequality/growth connection?

***Figure 15****. 1981-2011 average annual growth and 1981 Gini coefficient*




Using different cross-sectional models, I find that levels of inequality are positively related to regional economic growth in Canada over the long-run. However, the short-/medium-term responses are different. Results from fixed effects models point to a significant negative relationship between inequality and subsequent growth. I also find evidence of significant differences in outcomes between urban and rural regions.

The rest of the chapter is organized as follows. In the next section, I briefly review the literature examining the inequality/growth relationship at the (i) cross-country and (ii) sub-national levels. Section 5.2 then outlines my empirical approach and the data used in the analysis. Section 5.3 presents the estimation results while section 5.4 provides a further set of sensitivity analyses to test the robustness of my findings. Section 5.5 concludes the chapter with a few final thoughts.

## 5.1. A brief review of the literature

Ever since the seminal papers of Kuznets (1955) and Kaldor (1957) more than 60 years ago, economists have long been interested in the relationship between economic growth and income inequality. On the empirical front, much of the research examining whether or not there is a trade-off between growth and equity was first carried out at the macro-economic level using cross-country growth regression models typified by the work of Perotti (1993), Alesina and Rodrik (1994) and Persson and Tabellini (1994). In a much cited review paper, Benabou (1996) concluded that the overall consensus of these cross-country studies was that initially high levels of inequality were detrimental to the future economic growth of countries.

More recent macro-economic studies have challenged this consensus on several grounds (e.g., Forbes 2000, Panizza 2002). First, the estimates of several studies finding evidence of a negative effect of inequality on growth are not robust to more elaborate model specifications with additional control variables. Second, measurement error and the lack of consistent and comparable data across countries can lead to either a positive or negative bias on the impact of inequality on growth. Finally, omitted variable bias is also a possible source of important and unpredictable bias.

In an attempt to address some of the above econometric issues, regional scientists have entered the fray arguing that sub-national level data may provide a better platform to investigate the growth-equity relationship because of the consistency of the data collected by national statistical agencies. Within this body of work, much of which has been carried out in the US, there are generally two classes of modeling approaches that are adopted: ordinary least squares (OLS) growth regressions (the standard approach implemented in the cross-country literature) and panel techniques (mainly fixed effects models). Whereas the former approach is preferred when considering the long-term effects of levels of inequality on future economic growth, the latter is considered more appropriate over the short- and medium term when considering how changes in a region's level of inequality may effect changes in its growth performance (Forbes 2000).

Using state-level data from 1960 to 1980, Partridge (1997) was one of the first to investigate the growth-equity trade-off across US regions. Results from his OLS regressions suggest that states with higher levels of income inequality at the beginning of the period (as measured by the Gini coefficient) subsequently experienced greater growth. This finding of a positive relationship between inequality and long-term growth also holds from parsimonious to more complex model specifications.

In reassessing the relationship by using a similar dataset that spanned back to 1940, Panizza (2002) did not find any evidence of a positive correlation between the Gini index and growth across US states. In fact, results from fixed effects and GMMs estimations provide some evidence of a negative relationship between inequality and growth although these results are not robust. Indeed, this is arguably the most important conclusion to be drawn from Panizza's (2002) work: empirical evidence in support of either a positive or negative inequality-growth relationship is highly sensitive to small changes in the data (i.e., how the period of study is defined) and the econometric specification adopted.

In a follow-up study based on an updated panel of state-level data, Partridge (2005) tried to reconcile both long- and short-term perspectives only to acknowledge that minor differences in methodological approaches could indeed lead to mixed empirical results. Like Forbes (2000), he argued that standard OLS approaches focusing on cross-sectional differences across space better reflected the nature of the long-term effects of inequality on growth whereas modeling approaches concentrating on the time-series variation (within regions) were better suited for understanding the short-run effects of inequality on growth. His own estimates again confirmed the positive relationship between inequality and growth over the long-run while providing more ambiguous findings on the short-run dynamics of the relationship. Similarly, Frank (2009) finds that the long-run relationship between inequality and growth is positive in nature and mainly driven by the growing concentration of top-end incomes.

Rupasingha et al. (2002) and Fallah and Partridge (2007) have also examined the relationship across US counties. While the results from both studies point to varying outcomes, one novelty of the Fallah and Partridge (2007) paper is the identification of (i) a positive and significant inequality-growth link in predominantly metropolitan counties vs. (ii) a negative and significant relationship in non-metropolitan counties. Initial conditions are thus very important: even within a state, the central hypothesis of a positive inequality-growth linkage depends largely on whether or not a region is considered urban or rural. Geography matters, in other words, because of differences in the operation of economic incentives, agglomeration economies and the degree/type of social interaction. Given the results reported earlier in Chapters 3 and 4, this kind of urban-rural divide will also be important to investigate in the present chapter's analysis.

To the best of my knowledge, Dahlby and Ferede (2013) are the only ones to have applied the econometric framework developed in previous studies to study the income distribution/growth response within the Canadian context. They do so at the provincial level using real GPD per capita over 5-year growth periods from 1977 to 2006, along with Gini coefficients and the usual 'conditional' variables found as controls on the right hand side of the model (see below for more details). In contrast to US state-level studies, they find only weak evidence of a positive relationship between initial levels of income inequality and subsequent provincial economic growth, the significance of which disappears when further controls are added to the model. Such a finding, however, may not be surprising considering the rather limited potential for cross-sectional variation across provinces (*n* = 10).

In this chapter, I revisit the inequality-growth relationship using the panel dataset of Canadian regions for the period 1981 to 2011. I do so using a variety of methodological approaches to test the robustness of the relationship and identify differences in terms of the long- and short-run effects of inequality on growth.

## 5.2 Model specifications and data

I begin by estimating a baseline cross-sectional growth model that is specified as follows:

$$AAG(Y_{i2011,1981}) = \alpha + \beta INEQ_{i1981} + \gamma Y_{i1981} + \delta CONT_{i1981} + \theta IND_{i1981} + \varepsilon_i .$$

Eq. (5.1)

Here, the dependent variable represents region *i*'s average annual growth rate of median total income (*Y*) between 1981 and 2011. All variables are based on information from the micro-data files from the long-form Censuses of 1981 to 2006 and the 2011 National Household Survey (NHS). It is important to remind the reader that while the 1981 to 2006 Censuses were mandatory (with response rates hovering in the 90% range), the 2011 NHS was conducted on a voluntary basis which resulted in a lower response rate (69%). Though this raises a number of potential data quality issues for the 2011 sample (see, for instance, the discussion in Chapter 2 and Rheault et al. 2015, Smith 2015), with more than 6.7 million individual-level observations the NHS remains the single largest source of data for regional analysis in the country[7].

For the purposes of my analysis, two income concepts are used throughout. The first is total income which includes wages and salaries, old age pensions, investment income and various forms of government income support programs. The second will focus only on wages and salaries, which refers to gross wages before various deductions (e.g., income taxes, employment insurance, etc.). As mentioned above, growth is defined by looking at changes in a region's median (or average) total income (or wages and salaries). All income figures are deflated using the Consumer Price Index (for provinces) expressed in $2002.

On the left hand side of Eq. (5.1), the independent variables are all measured at the beginning of each respective growth period in order to minimize the potential for endogeneity problems (this is standard practice in the convergence literature; see, for instance, Panizza 2002 and Partridge 2005). Regional income inequality ($INEQ_{i1981}$) is measured using three different indicators. The Gini coefficient, which as mentioned earlier is the most widely used measure of inequality, will be my primary metric. To test the robustness of the inequality-growth relationship, I also supplement the Gini coefficient with two measures of general entropy: the Theil index and half the squared CV (GE2). Whereas both the Gini coefficient and the Theil index tend to be more sensitive to transfers in the middle part of the income distribution, the GE(2) is more sensitive to changes at the higher end of the distribution. $Y_{i1981}$ is the log of region *i*'s median total income (as a proxy for a region's initial level of economic development) and

[7] The models estimated in the paper were also re-estimated using 2006 (instead of 2011) as the end-year for the different growth episodes (see next section) examined. By and large, results for these models were qualitatively similar.

$CONT_{i1981}$ is a vector of control variables reflecting different socio-demographic characteristics. Among these are variables controlling for the stock of human capital (the percentage of the population with less than a high school degree and the percentage with a bachelor's degree or more), the percentage of female workers, recent immigrants and the age structure of regions (i.e., the percentage young (< 16 years of age) and senior (65+)). I also include a region's unemployment rate (to control for general economic conditions) and the log of its total population (as a coarse proxy for agglomeration effects). Finally, $IND_{i1981}$ controls for differences in the industrial composition of regions[8] and $\varepsilon_i$ is the error term.

While Eq. (5.1) is estimated by standard OLS and focuses on the long-term effects of the initial level of inequality on growth, a second model (following Forbes, 2000) investigates the relationship by focusing on short-/medium-term changes using a fixed effects model specified as:

$$AAG\left(Y_{it,t-1}\right) = \beta INEQ_{it-1} + \gamma Y_{it-1} + \delta CONT_{it-1} + \theta IND_{it-1} + \alpha_i + \eta_t + \varepsilon_{it},$$

Eq. (5.2)

where $AAG(Y_{it,t-1})$ represents the annual average growth rate of median total income from period $t-1$ to $t$ (over 10-year growth cycles), $\alpha_i$ denotes region $i$'s fixed effect, $\eta_t$ is a decade-period dummy and $\varepsilon_{it}$ is the error term. All other variables are defined as in Eq. (5.1). From my perspective, the key difference is in the interpretation of $\beta$. Whereas in Eq. (5.1), $\beta$ reflects the relationship between a region's initial level of inequality and its growth over time, in Eq. (5.2) $\beta$ is interpreted as a measure of the correlation between changes in inequality over time and changes in growth within a given region (Forbes 2000; Panizza 2002).

[8] We have 15 industry-level variables measuring the percentage of the workforce employed in a given industry. These industries are agriculture, mining, manufacturing, construction, transportation and warehousing, utilities, wholesale trade, retail trade, information and cultural services, finance and insurance, knowledge intensive business services, management services, education and health, arts and entertainment, and public administration.

## 5.3 Estimation and results

### 5.3.1. Long-run effects

Table 15 reports the first set of empirical results for the cross-sectional growth model specified in Eq. (5.1). Column 1 shows the descriptive statistics (mean and standard deviation) for each independent variable based on its initial values at the beginning of the growth period (1981). The weighted OLS results are presented in column 2 and not surprisingly, given the pattern from Figure 15, I find that the estimate for the Gini coefficient is positive and significant. In other words, regions with higher initial levels of income inequality do subsequently experience faster economic growth over the long-run (from 1981 to 2011). This is broadly consistent with the long-run impacts of inequality on growth across US states reported by Partridge (1997, 2005). Coefficient estimates for the other independent variables are also generally as expected. The coefficient for the level of economic development is negative and significant, suggesting that poorer regions have grown more rapidly than richer regions which is consistent with the catch-up effect described in the convergence literature (see, for instance, Breau and Saillant 2016). Regions with higher shares of highly educated (bachelor's degree or more) and female workers also experienced faster average annual growth rates.

**Table 15**. Cross-sectional regressions, 1981 to 2011

| | Mean (SD) | Weighted OLS | Spatial OLS§ | Weighted OLS | |
|---|---|---|---|---|---|
| | | | | Rural | Urban |
| $Gini_{1981}$ | .330 | .049** | .030** | .061 | .052** |
| | (.019) | (.012) | (.011) | (.032) | (.014) |
| Ln(median income)$_{1981}$ | 9.83 | -.015** | -.013** | -.019** | -.016** |
| | (.123) | (.004) | (.002) | (.004) | (.003) |
| % less than high school$_{1981}$ | .362 | .010 | -.001 | .004 | .013 |
| | (.070) | (.006) | (.004) | (.009) | (.011) |
| % bachelor's degree+$_{1981}$ | .128 | .020** | .010 | .007 | .027* |
| | (.044) | (.007) | (.009) | (.012) | (.012) |
| % female workers$_{1981}$ | .382 | .028** | .019** | .018 | .019 |
| | (.039) | (.010) | (.006) | (.012) | (.019) |
| % recent immigrants$_{1981}$ | .018 | -.036 | .002 | .049 | -.010 |
| | (.016) | (.043) | (.031) | (.044) | (.029) |
| % young (aged ≤ 16)$_{1981}$ | .227 | .016 | .008 | -.003 | .027 |
| | (.035) | (.016) | (.008) | (.030) | (.016) |
| % senior( aged ≥ 65)$_{1981}$ | .091 | -.010 | -.011 | -.014 | -.003 |
| | (.027) | (.013) | (.009) | (.025) | (.009) |
| Unemployment rate$_{1981}$ | .048 | -.033** | -.020** | -.019* | -.037** |

| | | | | | |
|---|---|---|---|---|---|
| | (.029) | (.007) | (.006) | (.009) | (.010) |
| Ln(total population)$_{1981}$ | 12.3 | -.001* | -.001 | .001 | -.001 |
| | (1.48) | (.001) | (.001) | (.001) | (.001) |
| Industry mix shares | | Y | Y | Y | Y |
| *rho* | | | .402** | | |
| | | | (.056) | | |
| Constant | | .076 | .075 | .219* | -.008 |
| | | (.085) | (.065) | (.100) | (.146) |
| No. of obs. | 284 | 284 | 284 | 167 | 117 |
| *R*-square | | .717 | .723 | .749 | .792 |

Notes: Heteroskedasticity robust standard errors are presented in parentheses. * indicates significance at the .10 level and ** at the .05 level. Based on the Lagrange Multiplier test, a spatial lag model was estimated.

Column 3 presents the estimates obtained from a spatial lag model. As suggested by the pattern observed in some of the maps in Chapter 3, both the average annual growth rate and Gini coefficient variables are highly clustered across the country (with Moran's I values of 0.552 and 0.486, respectively) which means the estimates from the previous OLS model could be biased and inconsistent (Rupasingha et al. 2002). Based on the analysis of a connectivity histogram, a K-6 nearest neighbour spatial weights matrix was used for estimation purposes (results from the Lagrange Multiplier test also point to the preference for a spatial lag model). The key result here is that after accounting for spatial variation, the estimate for the Gini coefficient remains positive and significant. Most of the other results are also consistent with those presented in column 2[9].

Following Fallah and Partridge (2007), I allow for the possibility that the inequality-growth transmission linkages vary between urban and rural areas. In the Canadian context, earlier work by MacLachlan and Sawada (1997) and Bolton and Breau (2012) suggests that the levels (and growth rates) of inequality are higher in metropolitan settings than elsewhere. The descriptive statistics presented in Chapter 3 also support this finding. To explore the possible divergence in urban-rural results, the last two columns of Table 15 show regression estimates

[9] Note that while I explicitly control for spatial dependence in this model, all other OLS and FE regression models presented throughout the paper are estimated using Stata's cluster function. This allows us to assume that the residuals may be correlated across certain geographic clusters (which are defined as five aggregate regions: the Atlantic provinces, Quebec, Ontario, Prairie provinces and British Columbia) but uncorrelated across of the clusters. It is an indirect way of controlling for the possibility of spatial autocorrelation in the models.

separately for urban and rural census divisions. The urban/rural classification is based on a revised and updated definition of Beale codes in Canada developed in Chapter 2 (see also Table 8 in Chapter 2). The results here confirm the importance of urbanization effects: whereas the regression estimates for the Gini coefficients are both positive in columns 4 and 5, it is only significant in the case of urban regions. In other words, it is in metropolitan areas where the subsequent growth effects of higher levels of inequality are most felt over the long-term. This could be related to urban agglomeration economies, i.e., the greater efficiency provided by the proximity of specialized production and labor activities which can also lead to greater wage differentials and the attraction of more highly skilled workers (for which the coefficient estimate is also positive and significant at the .10 level).

In Table 16, I present the results of the pooled OLS estimates of Eq. (5.1) where I have divided the 1981 to 2011 period into three 10-year growth episodes (1981 to 1991, 1991 to 2001 and 2001 to 2011) and recalculated the average annual growth rate of median total income for each of those period. In addition to the explanatory variables specified in Eq. (5.1), I also add decade dummies in the pooled model to control for possible aggregate shocks in specific time periods. In the overall model (column 1), results for the Gini coefficient again point to a positive and significant inequality-growth relationship over the 10-year periods. The pooled OLS estimations in columns 2 and 3 also confirm that the equity/growth trade-off stems primarily from urban regions. One interesting observation here is that population aging, over time, appears to have a negative impact on the long-term growth responses of regions (see also Breau and Saillant 2016).

**Table 16**. Pooled cross-sectional models, 1981 to 2011

| | Weighted OLS | Weighted OLS | |
|---|---|---|---|
| | | Rural | Urban |
| Gini | .071** | .015 | .115** |
| | (.017) | (.040) | (.041) |
| Ln(median income) | -.030** | -.043** | -.028* |
| | (.010) | (.011) | (.011) |
| % less than high school | .003 | -.020 | .022 |
| | (.011) | (.015) | (.024) |
| % bachelor's degree+ | -.064 | -.059** | -.096 |
| | (.030) | (.014) | (.053) |
| % female workers | .091** | .041 | .086* |
| | (.032) | (.023) | (.037) |
| % recent immigrants | -.088 | .078 | -.078 |

| | | | |
|---|---|---|---|
| | (.046) | (.115) | (.041) |
| % young (aged ≤ 16) | -.006 | .039 | -.009 |
| | (.033) | (.023) | (.043) |
| % senior( aged ≥ 65) | -.072** | -.029* | -.089* |
| | (.023) | (.012) | (.037) |
| Unemployment rate | -.033* | -.030* | -.029 |
| | (.013) | (.013) | (.020) |
| Ln(total population) | .001 | .001 | .001 |
| | (.001) | (.002) | (.001) |
| Industry mix shares | Y | Y | Y |
| Decade dummies | Y | Y | Y |
| Constant | -.837** | -.213 | -1.67** |
| | (.220) | (.144) | (.474) |
| No. of obs. | 852 | 501 | 351 |
| *R*-square | .846 | .776 | .884 |

Notes: Heteroskedasticity robust standard errors are presented in parentheses.

* indicates significance at the .10 level and ** at the .05 level.

In sum, results from cross-sectional models reveal that over the long-run, regions with initially higher levels of inequality do subsequently experience greater growth. Furthermore, this positive inequality-growth relationship appears to be driven predominantly by Canada's metropolitan regions[10].

### 5.3.2 Short-/Medium-run effects

In this section, I switch my focus to the fixed effects estimation of Eq. (5.2). Since I use only 10-year panels for this model, the coefficient estimates on the Gini coefficient reflect how changes in inequality may impact changes in growth over the short- to medium-term horizon. The interpretation of results is thus slightly different. Of course, one of the advantages of a fixed effect model is that it also controls for a region's unobserved time-invariant characteristics.

[10] The results presented here are for the fully specified models. Acknowledging the possibility that including so many control variables may introduce multicollinearity problems, I also re-estimated more parsimonious versions of the models. The main finding of a positive inequality-growth link over the long-run is robust to these specifications.

**Table 17**. Fixed-effects regression models, 10-year growth cycles

| | FE | FE | |
|---|---|---|---|
| | | Rural | Urban |
| Gini | -.067* | -.019 | -.078** |
| | (.029) | (.023) | (.020) |
| Ln(median income) | .117** | .123** | .121** |
| | (.007) | (.014) | (.004) |
| % less than high school | .079** | .058** | .101** |
| | (.011) | (.022) | (.027) |
| % bachelor's degree+ | .015 | .008 | -.017 |
| | (.031) | (.033) | (.038) |
| % female workers | -.007 | -.001 | -.033 |
| | (.045) | (.040) | (.061) |
| % recent immigrants | .094** | .313 | .064* |
| | (.029) | (.211) | (.026) |
| % young (aged ≤ 16) | -.128 | .021 | -.151 |
| | (.068) | (.041) | (.082) |
| % senior( aged ≥ 65) | .035 | .189** | .019 |
| | (.040) | (.047) | (.031) |
| Unemployment rate | -.015 | .019 | -.065 |
| | (.029) | (.009) | (.040) |
| Ln(total population) | -.003 | -.009 | .001 |
| | (.007) | (.010) | (.006) |
| Industry mix shares | Y | Y | Y |
| Decade dummies | Y | Y | Y |
| Constant | -.590** | -.791** | .037 |
| | (.124) | (.173) | (.811) |
| No. of obs. | 852 | 501 | 351 |
| No. of groups | 284 | 167 | 117 |
| *R*-square | .395 | .361 | .539 |

Notes: Heteroskedasticity robust standard errors are presented in parentheses.

* indicates significance at the .10 level and ** at the .05 level.

The results here are quite different than those reported earlier (see Tables 17 vs. 16). In the global model, I find that changes in the Gini coefficient have a negative though weakly significant (at the .10 level) effect on regional growth profiles. Such a finding is consistent with the work of Panizza (2002) and Partridge (2005) for US states. And again, by re-estimating the model separately for rural vs. urban regions, I find that metropolitan areas are driving this result.

As an interesting aside, the coefficient estimate for the percentage of immigrants is positive and significant suggesting that regions with higher immigrant shares benefit from higher economic growth over time. This is consistent with recent work by Kemeny and Cooke (2017) in the US that finds that metropolitan areas with a greater range of immigrant diversity and more inclusive institutions will see higher productivity levels.

## 5.4. Sensitivity analysis

As mentioned earlier, one of the key findings of the empirical literature on the equity/growth trade-off is that regression results can be very sensitive to minor changes in model specifications (see, in particular, Panizza 2002). In this section, I test the robustness of my findings by re-estimating Eq. (5.2) in a variety of different ways to test whether or not the negative short-/medium-run effects of inequality on regional growth described above are robust.

I begin by re-estimating the model using different measures of inequality. In addition to the Gini coefficient, which I have used throughout my models, I include the Theil index and the GE(2) as alternate indicators of income inequality. In both cases, I see that changing the measure of inequality does not affect the main result of a negative growth/equity trade-off (see Table 18). That said, given the sharp increase in the concentration of top incomes in Canada over the last few decades, I were surprised to see the coefficient estimate on the GE(2) being much smaller in magnitude and only significant at the .10 level[11].

**Table 18**. Sensitivity analysis

| | Coef. on inequality | Standard error | Regions | Obs. | Growth period | Estimation method |
|---|---|---|---|---|---|---|
| Inequality indicators | | | | | | |
| Gini coefficient | -.080** | (.029) | 284 | 852 | 1981-2011 | FE |
| Theil index | -.032** | (.010) | 284 | 852 | 1981-2011 | FE |
| Half squared CV (GE2) | -.002* | (.001) | 284 | 852 | 1981-2011 | FE |
| Income concept | | | | | | |
| Median total income | -.080** | (.029) | 284 | 852 | 1981-2011 | FE |
| Median wages | -.145** | (.047) | 284 | 852 | 1981-2011 | FE |
| Average total income | -.149** | (.026) | 284 | 852 | 1981-2011 | FE |
| Average wages | -.174** | (.027) | 284 | 852 | 1981-2011 | FE |
| Income groups | | | | | | |

[11] Such a finding is likely related to the fact there are much smaller numbers of top end income earners in certain regions which causes complexities when the population weights are used in Stata to estimate sampling variances (on this note, see STB-48, 1999).

| | | | | | | |
|---|---|---|---|---|---|---|
| < $15,500 | -.029 | (.048) | 75 | 275 | 1981-2011 | FE |
| $15,500 to $19,500 | -.092** | (.024) | 125 | 375 | 1981-2011 | FE |
| > $19,500 | -.134** | (.049) | 84 | 252 | 1981-2011 | FE |
| Beale category | | | | | | |
| Beale 0 | -.216* | (.076) | 6 | 18 | 1981-2011 | FE |
| Beale 1 | -.089* | (.009) | 27 | 81 | 1981-2011 | FE |
| Beale 2 | -.148** | (.044) | 24 | 72 | 1981-2011 | FE |
| Beale 3 | -.028 | (.065) | 60 | 180 | 1981-2011 | FE |
| Beale 4 | -.053 | (.044) | 60 | 180 | 1981-2011 | FE |
| Beale 5 | .017 | (.058) | 107 | 321 | 1981-2011 | FE |
| Arellano-Bond GMM | -.069** | (.016) | 284 | 1420 | 1986-2011 | A&B |

*Notes*: FE: fixed-effects, A&B: Arellano-Bond. Heteroskedasticity robust standard errors are presented in parentheses. * indicates significance at the .10 level and ** at the .05 level.

In addition to using different indicators of inequality, I also re-estimated the model using different income concepts. Whereas median income is considered the preferred proxy for growth (Partridge and Weinstein 2013), I also looked at average total income and average wages. In all cases, the relationship between inequality and regional growth remains negative and significant.

Another possibility is that the short-/medium-term impact of inequality on growth depends on a region's level of economic development. To test this, I divide regions into three separate income categories based on 1981 figures (measured in $2002) and re-estimate Eq. (5.2) for each group. Interestingly, the negative and significant relationship holds for all but the lowest income category. This is perhaps not surprising given my earlier findings that urbanization effects are important in predicting the strength of the relationship, especially since 84% of regions in the lower income category are defined as rural.

This finding also led us to re-estimate the model across different Beale code categories. As expected, evidence of the negative short-/medium-run effects of inequality on growth is found in both large and medium sized metropolitan CDs, though the impact is largest in the latter (e.g., typified by regions such as Halifax, Quebec, Waterloo, Hamilton, Saskatoon-Battleford and Victoria).

Lastly, the bottom row of Table 18 presents the results from a general methods of moments (GMM) approach (Arellano and Bond 1991). Though I have mainly focused on the FE approach to examine the short-/medium-run effects of inequality on growth, it is possible that including a lag of the endogenous variable in Eq. (5.2) may introduce bias in the estimation. The advantage of the GMM approach is that it first-differences the variables in order to eliminate the region-specific effects and allow for the use of lagged variables as instruments (Forbes 2000). In applying the GMM estimation, I used shorter 5-year panels to ensure a larger number of periods. Again, the finding of a negative short-run impact of inequality on regional growth responses holds true.

## 5.5 Conclusion

This chapter examines the relationship between income inequality and growth across Canadian census divisions. In doing so, I find that the long-run and short-/medium-run dynamics of the inequality-growth relationship are similar in Canada as those observed across US regions. Over the long-run, regions with initially high levels of inequality are found to experience greater subsequent growth. In contrast, short-/medium-run changes in both economic development and inequality are negatively correlated with each other. And in both the long- and short-/medium-run cases, I find significant differences in outcomes based on whether a region is urban or rural.

Like most cross-sectional analyses, this analysis is exploratory. Although the relationship between inequality and economic growth is robust, I cannot identify the causal channels that explain why inequality results in lower short-term growth and higher long-term growth. Establishing and examining those channels at the sub-national scale would be an important area of future research.

Hence, while these results provide new insights into the dynamics of the inequality-growth relationship across Canadian regions, I recognize that I am only beginning to scratch the surface of these complex linkages. A particularly fruitful avenue for future research would be to explore newly developed methodologies emphasizing the potential for non-linearities in the equity-growth trade-off. As Grigoli and Robles (2017) point out, most of the relevant literature has so far assumed that the relationship is best represented by a linear specification. Their own empirical evidence suggest that there may be a ‘tipping point’ (see also Weinstein and Partridge 2013) beyond which the relationship can change.

# Chapter 6 Conclusion

This thesis set out to explore the spatial patterns of income inequality in Canada by addressing the following research questions: (1) What are the patterns of income inequality across Canadian regions? (2) Do these patterns differ along an urban and rural divide? (3) What are the factors associated with increasing or decreasing regional levels of inequality? (4) Do these factors have the same effects on rural and urban regions? (5) What are the effects of income inequality on regional economic growth?

In order to answer these questions, I constructed a dataset with census divisions as the primary unit of observation. The data was drawn from the 20% long form samples of the census of population for 1981, 1986, 1991, 1996, 2001, and 2006, pooled together with the National Household Survey of 2011. While I defended the choice of pooling the two surveys into one time series dataset, it is important to note that both surveys offered the range of earnings variables and geographical detail necessary for this study. Likewise, special attention was given to the development of a consistent geography over the 1981 to 2011 period of study. And given that this thesis was interested in exploring income inequality patterns along the rural and urban lines, I chose to define the two areas using the Beale code classification.

The thesis makes several contributions to the literature. First, while I document a net increase in income inequality at the national level from 1981 to 2011, I also show interesting spatial differences in the trajectories of inequality. For instance, whereas in 1981, urban and rural areas had similar levels of inequality, the distribution of incomes remained largely stable in rural areas while it became very unequal in large urban areas. Moreover, when I examined these trends in more detail using a re-constructed Beale code classification, I find that large metropolitan areas are the hot beds of inequality. Inequality in large metropolitan fringe areas, as well as small and medium cities, has also steadily increased from 1986 onwards. In contrast, both rural adjacent and non-adjacent to metropolitan areas did not experience same changes in the distribution of income affecting the rest of the country.

I also examined the spatial clustering of inequality across Canadian regions and found significant clusters of high and low levels of inequality. These clusters of inequality (or equality) evolved (or changed their spatial morphology) from 1981 to 2011. The 2011 observations points to a growing number of regions with low levels of inequality centered in the province of Quebec, while multiple regions are facing higher levels of inequality in Alberta and British Columbia. Based on these findings, I also set out to explore the potential contributors to differences in income inequality across Canadian regions.

To do so, I turned to spatial panel regression models. Here, I find that the precariousness of labour markets (as proxied by the unemployment rate and the proportion of self-employed and part-time workers) is positively link to higher levels of inequality. Likewise, higher proportions of workers in the utilities, leisure and food, and knowledge intensive business sectors are linked to higher levels of inequality. In contrast, higher shares of workers in the manufacturing sector as well as in the transportation and public administration sectors typically lead to lower levels of inequality. Polarization in the educational attainment of individuals leads to higher inequality, so does the female participation rate, the percentage of visible minorities and the proportion of young and seniors in a region. As for the institutional variables, higher minimum wages and higher unionization rates is found to lower inequality. The regression model was also tested for a spatial structural break and re-estimated separately for urban and rural regions. Estimates from this model suggest that the determinants of inequality vary across both types of regions.

Finally, I asked whether income inequality in one region could hurt the region's economic growth profile. Here, the findings point to a positive relationship between inequality and economic growth although the relationship does not hold in the short-/medium-term. Also, the effect of inequality level on economic growth varies considerably between rural and urban areas. By answering the above questions, I have just begun to scratch the surface of the potential implications of inequality across Canadian regions. Much more research is needed in this area to understand the rapidly changing landscape of inequality in Canada and its potential policy implications.

## References

Alesina, A., and D. Rodrik (1994) Distributive Politics and Economic-Growth, *Quarterly Journal of Economics*, 109(2): 465-490.

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

Atkinson, A. B. (1975) The economics of inequality. Oxford, UK: Clarendon.

Atkinson, A. B. (2008). The changing distribution of earnings in OECD countries. Oxford: Oxford University Press.

Atkinson, A. B. (2015) *Inequality: What can be done?* Cambridge, MA: Harvard University Press.

Bartik, T. J. (1994) The Effects of Metropolitan Job Growth on the Size Distribution of Family Income. Journal of Regional Science 34, 483-502 .

Bartik, T. J. (1996) The Distributional Effects of Local Labor Demand and Industrial Mix: Estimates Using Individual Panel Data. Journal of Urban Economics 40, 150-178.

Baum-Snow, N., and R. Pavan (2012) Inequality and City Size Working Paper. Brown University, 34 pages.

Beale C.L., and Johnson K.M. The Identification of Recreational counties in Nonmetropolitan Areas of the USA. Population Research and Policy Review. 1997;17:37–53

Blank, Rebecca M., et David Card. (1993) Poverty, Income Distribution and Growth: Are They Still Correlated? Brookings Papers on Economic Activity 2, 285-325.

Blau, D. F. et L. M. Kahn (1996) International differences in male wage inequality: Institutions versus Market Forces, Journal of Political Economy, vol. 104(4).

Bolton, K., and Breau, S. 2012. Growing unequal? Changes in the distribution of earnings across Canadian cities. Urban Studies 49:1377–96.

Borjas, G. J., Freeman, R. and Katz L.. (1992) On the Labor Market Effects of Immigration and Trade." In George J. Borjas (ed.) Immigration and the Work Force. Chicago: University of Chicago.

Bound, J. et G. Johnson (1989) Changes in the structure of wages in the 1980s : An evaluation of alternative explanations, American Economic Review, vol. 82, 371-392

Bound, J. et Johnson, G. (1992) Changes in the structure of wages in the 1980s: an evaluation of alternative explanations, American Economic Review, 82(3), pp. 371–392.

Bourne, L. (1993) Close Together and Worlds Apart - an Analysis of Changes in the Ecology of Income in Canadian Cities. Urban Studies, 30 (8): 1293-1317.

Breau, S. (2007) Income inequality across Canadian provinces in an era of globalization: explaining recent trends. The Canadian Geographer, 51 (1): 72-90.

Breau, S., Kogler, D. F., & Bolton, K. C. (October 01, 2014). On the Relationship between Innovation and Wage Inequality: New Evidence from Canadian Cities. *Economic Geography, 90,* 4, 351-373.

Breau, S. (2015) Rising inequality in Canada: A regional perspective, *Applied Geography*, 61: 58-69.

Breau, S. and R. Saillant (2016) Regional income disparities in Canada: Exploring the geographical dimensions of an old problem. *Regional Studies, Regional Science*, 3(1): 464-482.

Brown, M., and R. Macdonald (2015) *Provincial convergence and divergence in Canada, 1926 to 2011*. Economic Analysis (ES) Research Paper Series, Catalogue No. 11F0027M-No. 96, Ottawa, ON: Statistics Canada.

Brzozowski, M., Gervais, M., Klein, P., & Suzuki, M. (2010). Consumption, income, and wealth inequality in Canada, Review of Economic Dynamics, 13, 1, 52-75.

Cavanaugh, A. and S. Breau (2017) Locating geographies of inequality: publication trends across OECD countries. *Regional Studies*, DOI: 10.1080/00343404.2017.1371292.

Chakravorty, S. (1996) Urban Inequality Revisited - the Determinants of Income Distribution in 18 US Metropolitan Areas. Urban Affairs Review 31: 759-777.

Chakravorty, S. (2006). Fragments of inequality: Social, spatial, and evolutionary analyses of income distribution. New York, N.Y: Routledge.

Chen, W. H., Myles, J., and Picot, G. (2012) Why Have Poorer Neighbourhoods Stagnated Economically While the Richer Have Flourished? Neighbourhood Income Inequality in Canadian Cities Urban Studies, 49, 877-896.

Chow, Gregory C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". *Econometrica*. **28** (3): 591–605

Cliff, A.D., and J.K. Ord. (1969) The problem of spatial autocorrelation. In London Papers in Regional Science, A. J. Scott (ed.), Pion, London, 25–55.

Cloke, P. J., Marsden, T., & Mooney, P. H. (2006). Handbook of rural studies. London: SAGE.

Cloutier, N. R. (1997) Metropolitan income inequality during the 1980s: The impact of urban development, industrial mix, and family structure, Journal or Regional Science, vol. 37, 459-478.

Cornia, G. A. (1999) Liberalization, Globalization and Income Distribution, World Institute for Development Economics Research.

Coulombe, S. (1999) *Economic growth and provincial disparity: A new view of an old Canadian problem* (p.36). Toronto, ON: C.D. Howe Institute.

Cowell, F. A. (2011). *Measuring inequality*. Oxford: Oxford University Press.

Dahlby, B., and E. Ferede (2013) Income inequality, redistribution and economic growth. The School of Public Policy, University of Calgary, Research Paper Series, 6(25): 42 pages.

Erksoy, S. (1994). The effects of higher unemployment on the distribution of income in Canada e 1981e1987. Canadian Public Policy-Analyse de Politiques, 20, 318-328.

Fallah, B. N., and M. Partridge (2007) The elusive inequality-economic growth relationship: are there differences between cities and the countryside?, *Annals of Regional Science*, 41(2): 375-400.

Fichtenbaum, R., & Shahidi, H. (1988). Truncation bias and the measurement of income inequality. Journal of Business & Economic Statistics, 6, 335-337.

Finnie, R (2001) Unequal Inequality: The Distribution of Individuals' Earnings by Province. Canadian Journal of Regional Science, 24: 265-88.

Forbes, K. J. (2000) A reassessment of the relationship between inequality and growth, *American Economic Review*, 90(4): 869-887.

Fortin N. M. and T. Lemieux, (2000) Income Redistribution in Canada: Minimum Wages versus Other Policy Instruments, Public Policies in a Labour Market in Transition, 211-247.

Frank, M. W. (2009) Inequality and Growth in the United States: Evidence from a New State-Level Panel of Income Inequality Measures, *Economic Inquiry*, 47(1): 55-68.

Frenette, M., D. A. Green, and K. Milligan. (2007) The tale of the tails: Canadian income inequality in the 1980s and 1990s. Canadian Journal of Economics, 40:734-764.

Frenette, M., Green, D.A., and Milligan, K. (2009). Taxes, transfers, and Canadian income inequality. *Canadian Public Policy - Analyse de Politiques, 35*(4), 389 - 411.

Freshwater, D. (2001) Delusions of Grandeur: The search for a vibrant rural America, TVA rural studies, 01-07.

Fuguitt, G. V., Brown, D. L., Beale, C. L., Pfeffer, M. J., & National Committee for Research on the 1980 Census. (1989). Rural and small town America. New York: Russell Sage Foundation.

Gilbert, A. (2007) Inequality and why it matters, Geography Compass, 1(3), pp. 422–447.

Glaeser, E. L., Resseger, M. and Tobio, K. (2009) Inequality in cities, Journal of Regional Science, 49(4), pp. 617–646.

Goodchild, M. F., L. Anselin, and U. Deichmann (1993) A Framework for the Areal Interpolation of Socioeconomic Data, *Environment and Planning A*, 25(3): 383-397.

Gray, D., J. A. Mills, and S. Zandvakili (2004) An analysis of differential provincial income inequality trends in Canada. Research on Economic Inequality, 12: 443-461.

Green, D. and J Kesselman (2006) Dimensions of inequality in Canada. Vancouver: UBC Press.

Grigoli, F. and A. Robles (2017) Inequality Overhang. IMF Working Paper WP/17/76. IMF: Washington DC.

Harrison, B., & Bluestone, B. (1988). *The great u-turn: Corporate restructuring and the polarizing of America*. New York: Basic Books.

Hamnett, C., 2003. Unequal City: London in the Global Arena. Routledge, London.

Heisz, A. (2007) Income inequality and redistribution in Canada: 1976–2004. Ottawa: Statistics Canada. Cat. no. 11F00119MIE-No.298.

Heisz, A. and L. McLeod (2004) Low-income in Census Metropolitan Areas, 1980-2000. Statistics Canada: Analytical Paper.

Irwin, M., (2007) Territories of inequality: an essay on the measurement and analysis of inequality in grounded place settings. In: Lobao, L., Hooks, G., Tickamyer, A. (Eds.), The Sociology of Spatial Inequality. SUNY Press, Albany, NY, pp. 85–109.

Jean, B., & Chaire Desjardins en développement des petites collectivités. (1997).Territoires d'avenir: Pour une sociologie de la ruralité. Sainte-Foy [Que.: Presses de l'Université du Québec.

Juhn, C., Murphy, K., et Pierce, B. (1993) Wage inequality and the rise in returns to skill, Journal of Political Economy, 101, 410-442

Kaldor, N. (1957) A Model of Economic-Growth, *The Economic Journal*, 67(268): 586-624.

Karoly, Lynn A., et Jacob A. Klerman.(1994) Demographics, Sectoral Change, and Changing Relative Wages: A Regional Approach. Rand Research Working Paper 94-20.

Katz, L. F., et Murphy, K. M. (1992) Changes in relative wages, 1963–1987—supply-and-demand factors, Quarterly Journal of Economics, 107, 35–78

Kemeny, T., and A. Cooke (2017) Urban Immigrant Diversity and Inclusive Institutions, *Economic Geography*, 93(3): 267-291.

Korpi, M. (2008) Does size of local labour markets affect wage inequality? A rank-size rule of income distribution. Journal of Economic Geography, 8 (2): 211-237.

Kuznets, S. (1955) Economic growth and income inequality. *American Economic Review, 45*, 1–28.

Levernier, W., Partridge, M. D. et Rickman, D. S. (1998) Differences in Metropolitan and Nonmetropolitan U.S. Family Income Inequality: A Cross-County Comparison. Journal of Urban Economics, September, 44(2), pp. 272-90

Levernier, W., Partridge, M. D. et Rickman, D. S. (1998) Metropolitan-Nonmetropolitan Distinctions in the Determinants of Regional Family Income. Review of Regional Studies, Winter, 28(3), pp. 83-106.

Levy, F. et Murnane, R. J. (1992) United States earnings levels and earnings inequality: a review of recent trends and proposed explanations, Journal of Economic Literature, 30(3), pp. 1333–1381.

Lobao, L., Hooks, G., (2007) Advancing the sociology of spatial inequality: spaces, places, and the subnational scale. In: Lobao, L., Hooks, G., Tickamyer, A. (Eds.), The Sociology of Spatial Inequality. SUNY Press, Albany, NY, pp. 29–61

Lobao, L., Rulli, J., Brown, L., 1999. Macro-level theory and local-level inequality: industrial structure, institutional arrangements and the political economy of redistribution, 1970 to 1990. Annals of the Association of American Geographers 89, 571–601.

Lydall, H. (1979) A Theory of Income Distribution. Oxford: Oxford University Press.

MacLachlan, I., and R. Sawada (1997) Measures of income inequality and social polarization in Canadian Metropolitan Areas, *Canadian Geographer*, 41(4): 377-397.

MacPhail, F. (March 01, 2000). What caused earnings inequality to increase in Canada during the 1980s?. *Cambridge Journal of Economics, 24,* 2, 153-175

Madden, J. F. (2000) Changes in Income Inequality within US Metropolitan Areas. Kalamazoo, MI: W. E. Upjohn Institute.

Marcouiller, D. W. (January 01, 1997). Toward Integrative Tourism Planning in Rural America. *Journal of Planning Literature, 11,* 3, 337-357

Marcouiller, D. W., Kim, K.-K., & Deller, S. C. (January 01, 2004). Natural amenities, tourism and income distribution. *Annals of Tourism Research, 31,* 4, 1031-1050.

Martin, R. (2001) Editorial: the geographer as social critic: getting indignant about income inequality, Transactions of the Institute of British Geographers, 26(3), pp. 267–272.

Martin, D., D. Dorling, and R. Mitchell (2002) Linking censuses through time: problems and solutions, *Area*, 34(1): 82-91.

McLaughlin, D., 2002. Changing income inequality in nonmetropolitan counties, 1980 to 1990. Rural Sociology 67, 512–533.

Miller, K. K. et B. A. Weber (2004) How do Persistent Poverty Dynamics and Demographics Vary across the Rural-Urban Continuum?

Mitchell, A. and Soroka, L. A. (1993) Income distributions and city size: has the structure changed?, Canadian Journal of Regional Science, 16(1), pp. 39–52.

Mok, D. (November 01, 2009). Cohort Effects, Incomes, and Homeownership Status Among Four Cohorts of Canadian Immigrants. *The Professional Geographer, 61,* 4, 527-546.

Moore, E. G., & Pacey, M. A. (2003). Changing income inequality and immigration in Canada, 1980-1995. Canadian Public Policy, 29(1), 33-52.

Murphy, K. M. et Welch, F. (1992) The structure of wages, The Quaterly Journal of Economics, 107, 285-326

Myles, J. (2003). Where Have All the Sociologists Gone? Explaining Economic Inequality. Canadian Journal of Sociology, 28, 551-560.

Myles, J. et al. (2000) Neighbourhood inequality in Canadian cities. Ottawa: Statistics Canada.

Nielsen, F., & Alderson, A. S. (1997). The kuznets curve and the great U-turn: income inequality in US countries, 1970 to 1990. American Sociological Review, 62, 12-33.

OECD. (2011) Divided we stand: why inequality keeps rising. Paris, France: OECD Publishing

OECD. (2014) *Regional Outlook*. Paris, France: OECD Publishing

Osberg, L. (1981) Economic Inequality in Canada. Toronto: Butterworths.

Panizza, U. (2002) Income inequality and economic growth: Evidence from American data, *Journal of Economic Growth*, 7(1): 25-41.

Partridge, M. D. Rickman, D. S. et Levernier, W. (1996) Trends in U.S. Income Inequality: Evidence from a Panel of States. Quarterly Review of Economics and Finance, Spring , 36(1), pp. 17-37.

Partridge, M. D. (1997) Is inequality harmful for growth? Comment, *American Economic Review*, 87(5): 1019-1032.

Partridge, M. D. (2005) Does income distribution affect US state economic growth?, *Journal of Regional Science*, 45(2): 363-394.

Partridge, M., Rickman, D., 2005. High poverty nonmetropolitan counties in America: can economic development help? International Regional Science Review 28, 415–440.

Partridge, M.D. and A.L.Weinstein (2013) Rising inequality in an era of austerity: The case of the US. European Planning Studies 21: 388-410

Pendakur, K., & Pendakur, R. (2002). Colour my world: have earnings gaps for Canadian-born ethnic minorities changed over time? Canadian Public Policy, 28, 489-512.

Pendakur, K., & Pendakur, R. (2011). Aboriginal income disparity in Canada. Canadian Public Policy, 37, 61-83.

Persky, J. et Tam, M. (1994) On the Persistent Structure of Metropolitan Income Inequality: 1900-1980. Review of Regional Studies, 24(3), pp. 211-27.

Perotti, R. (1993) Political Equilibrium, Income-Distribution, and Growth, *Review of Economic Studies*, 60(4): 755-776.

Persson, T., and G. Tabellini (1994) Is Inequality Harmful for Growth, *American Economic Review*, 84(3): 600-621.

Peters, D. J. (2014) American income inequality across economic and geographic space, 1970-2010, Social Science Researh, 42, 1490-1504.

Peters, D., 2012. Understanding micro and meso scale income inequality in the Midwestern United States, 1979–2009. Rural Sociology 77, 171–202.

Picot, W. G., & Statistics Canada. (2008). *Immigrant economic and social outcomes in Canada: Research and data development at Statistics Canada*. Ottawa: Statistics Canada.

Piketty, T. (2014) Capital in the twenty-first century. Cambridge, MA: Belknap Press of Harvard University Press.

Polèse, M. (2014) À propos de l'évolution des inégalités régionales: Un modèle simple avec un regard sur l'Europe et l'Amérique du Nord. In S. Breau & R. Saillant (eds.), *« Nouvelles perspectives en développement regional: Essais en l'honneur de Donald J. Savoie »* (pp. 25-53). Québec, QC: Presses de l'Université du Québec.

Puderer, H. (2008) *Defining and measuring metropolitan areas: A comparison between Canada and the United States*. Geography Working Paper Series, Catalogue No. 92F01138MIE-No. 2008002. Ottawa, ON: Statistics Canada.

Rodriguez-Pose, A., & Tselios, V. (2009). Mapping regional personal income distribution in Western Europe: income per capita and inequality. Czech Journal of Economics and Finance, 59(1), 41-70.

Rupasingha, A., S. J. Goetz, and D. Freshwater (2002) Social and institutional factors as determinants of economic growth: Evidence from the United States counties, *Papers in Regional Science*, 81(2): 139-155.

Rycroft, R. S. (2013). The economics of inequality, poverty, and discrimination in the 21st century. Santa Barbara, Calif: Praeger.Breau, 2014

Saez, E., (2006) Income and wealth concentration in a historical and international perspective. In: Auerbach, A., Card, D., Quigley, J. (Eds.), Public Policy and the Income Distribution. Russell Sage Foundation, New York, pp. 221–258.

Savoie, D. (2017) *Looking for Bootstraps – Economic Development in the Maritimes*, Halifax, NS: Nimbus Publishing.

Scott, A. J. (2012). A world in emergence: Notes toward a resynthesis of urban-economic geography for the 21st century. Urban Geography.vol. 32, 845-870

Smith, W.R. (2015) The 2011 National Household Survey- The complete statistical story. Available at: https://www.statcan.gc.ca/eng/blog-blogue/cs-sc/2011NHSstory. Accessed July 18 2016.

Soroka, L. A. (1999) Male–female urban income distributions in Canada: the service sector in a dependency model, Urban Studies, 36(3), pp. 563–574.

STATA Technical Bulletin-48, March 1999.

Statistics Canada. (2016) The surge of women in the workforce. Available at: https://www.statcan.gc.ca/pub/11-630-x/11-630-x2015009-eng.htm. Accessed 24 October 2016.

Statistics Canada. (2015) NHS: Data quality. Available at: https://www.statcan.gc.ca/nhs-enm/2011/ref/about-apropos/nhs-enm_r005-eng.cfm. Accessed 3 April 2018.

Stiglitz, J. (2012) *The price of inequality*. NY, NY: W.W. Norton & Co.

Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46 (Supplement): 234-240.

Topel, R. H. (1994) Wage inequality and regional labor market performances in the United States. University of Michigan Press.

United States Department of Agriculture, Economic Research Service. (2016) Rural-Urban continuum codes. Available at: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/. Accessed October 14 2016.

Ward, N. et D. L. Brown (2009) Placing the rural in regional development, Regional Studies, vol. 43, 1237-1244

Wood, Adrian (1994), 'North-South Trade, Employment and Inequality', Oxford: Clarendon Press.

World Economic Forum (2015). Deepening income inequality. Available at: http://reports.weforum.org/outlook-global-agenda-2015/top-10-trends-of-2015/1-deepening-income-inequality/. Accessed 7 April 2015.

# Appendix

### A1. Minor incongruences in Census divisions rebuilt

As explained in Chapter 2, the census divisions used in this study are custom made from census subdivisions to ensure time consistency in their boundaries. However, they were 98 minor incongruences reported during the rebuilt. Table A1 lists those minor incongruences with second column stating the Census divisions identification key and third column lists the census divisions with which it is shares a incongruence. The last columns state the year with which this incongruence appears. Given that 2011 boundary are the based year. A ‘one’ in the 1981 column indicate a CD x shares a minor incongruence with another CD y. Those incongruences occur due to boundaries change that did not emerges at the census subdivision level. Therefore, they could not be traced back smaller unit in time. Although they account for a total of 98, the minor incongruences are judge not to affect to results of this study.

**Table A1. Minor incongruences in Census divisions rebuilt**

| | CD | with CD(s) | in census year(s) | | | | | |
|---|---|---|---|---|---|---|---|---|
| Province | has a minor incongruence | | 2006 | 2001 | 1996 | 1991 | 1986 | 1981 |
| MB | 4621 | 4622 | | | | | | 1 |
| NB | 1314 | 1315 | | | | 1 | 1 | 1 |
| SK | 4712 | 4713 | | | 1 | 1 | 1 | 1 |
| AB | 4803 | 4815 | | | | | 1 | 1 |
| AB | 4811 | 4814 | | | | | 1 | 1 |
| AB | 4814 | 4818 | | | | 1 | 1 | 1 |
| AB | 4817 | 4816 | | | | | 1 | 1 |
| AB | 4812 | 4813 | | | | 1 | 1 | 1 |
| BC | 5909 | 5915 | | | | 1 | 1 | 1 |
| BC | 5919 | 5923 | | | | | | 1 |
| ON | 3548 | 3552 | | | 1 | 1 | 1 | 1 |
| ON | 3559 | 3560 | | | 1 | 1 | 1 | 1 |
| ON | 3552 | 3553 | | 1 | 1 | 1 | 1 | 1 |
| ON | 3551 | 3552 | | | | 1 | 1 | 1 |
| QC | 2402 | 2403 | | | | | | 1 |
| QC | 2402 | 2405 | | | | | | 1 |
| QC | 2403 | 2404 | | | | | | 1 |
| QC | 2404 | 2408 | | | | | | 1 |
| QC | 2404 | 2405 | | | | | | 1 |
| QC | 2405 | 2406 | | | | | | 1 |
| QC | 2406 | 2407 | | | | | | 1 |
| QC | 2406 | 2409 | | | | | | 1 |
| QC | 2407 | 2408 | | | | | | 1 |
| QC | 2407 | 2409 | | | | | | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| QC | 2409 | 2410 | | | | | 1 |
| QC | 2410 | 2413 | | | | | 1 |
| QC | 2412 | 2413 | | | | | 1 |
| QC | 2412 | 2414 | | | | | 1 |
| QC | 2413 | 2414 | | | | | 1 |
| QC | 2415 | 2416 | | | | | 1 |
| QC | 2416 | 2493 | | | | | 1 |
| QC | 2417 | 2418 | | | | | 1 |
| QC | 2421 | 2490 | | | | | 1 |
| QC | 2421 | 2493 | | | | | 1 |
| QC | 2422 | 2434 | | | | | 1 |
| QC | 2422 | 2490 | | | | | 1 |
| QC | 2425 | 2433 | | | | 1 | 1 |
| QC | 2432 | 2433 | 1 | 1 | 1 | 1 | 1 |
| QC | 2432 | 2439 | | 1 | 1 | 1 | 1 |
| QC | 2434 | 2435 | | | | | 1 |
| QC | 2434 | 2490 | | | | | 1 |
| QC | 2435 | 2436 | | | | | 1 |
| QC | 2435 | 2451 | | | | | 1 |
| QC | 2435 | 2462 | | | | | 1 |
| QC | 2435 | 2490 | | | | | 1 |
| QC | 2436 | 2451 | | | | | 1 |
| QC | 2440 | 2441 | | 1 | 1 | 1 | 1 |
| QC | 2441 | 2443 | | | 1 | 1 | 1 |
| QC | 2441 | 2444 | | | 1 | 1 | 1 |
| QC | 2442 | 2445 | 1 | 1 | 1 | 1 | 1 |
| QC | 2442 | 2448 | | | | | 1 |
| QC | 2443 | 2445 | 1 | 1 | 1 | 1 | 1 |
| QC | 2443 | 2444 | | 1 | 1 | 1 | 1 |
| QC | 2444 | 2445 | | 1 | 1 | 1 | 1 |
| QC | 2446 | 2447 | | | | | 1 |
| QC | 2448 | 2454 | | | 1 | 1 | 1 |
| QC | 2451 | 2462 | | | | | 1 |
| QC | 2452 | 2462 | 1 | 1 | 1 | 1 | 1 |
| QC | 2462 | 2463 | | | 1 | 1 | 1 |
| QC | 2462 | 2478 | | | | | 1 |
| QC | 2462 | 2479 | | | | | 1 |
| QC | 2462 | 2490 | | | | | 1 |
| QC | 2475 | 2477 | | | 1 | 1 | 1 |
| QC | 2478 | 2479 | | | | | 1 |
| QC | 2478 | 2480 | | 1 | 1 | 1 | 1 |
| QC | 2479 | 2480 | | | | | 1 |
| QC | 2479 | 2483 | | | | | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| QC | 2479 | 2490 | | | | | | 1 |
| QC | 2481 | 2482 | | | | | 1 | 1 |
| QC | 2483 | 2484 | | | | | | 1 |
| QC | 2483 | 2489 | | | | | | 1 |
| QC | 2483 | 2490 | | | | | | 1 |
| QC | 2484 | 2489 | | | | | | 1 |
| QC | 2485 | 2486 | | | | | | 1 |
| QC | 2485 | 2489 | | | | | | 1 |
| QC | 2486 | 2487 | | | | | | 1 |
| QC | 2486 | 2488 | | | | | | 1 |
| QC | 2486 | 2489 | | | | | | 1 |
| QC | 2487 | 2488 | | | | | | 1 |
| QC | 2488 | 2489 | | | | | | 1 |
| QC | 2489 | 2490 | | | | | | 1 |
| QC | 2490 | 2491 | | | | | | 1 |
| QC | 2490 | 2493 | | | | | | 1 |
| QC | 2490 | 2499 | | | | | | 1 |
| QC | 2491 | 2499 | | | | | 1 | 1 |
| QC | 2495 | 2496 | | | | | | 1 |
| QC | 2496 | 2497 | | | | | | 1 |
| QC | 2497 | 2498 | | | | | | 1 |
| QC | 2497 | 2499 | | | | | | 1 |
| Territories | 6304 | 6305 | | | | | | **1** |
| Territories | 6304 | 6307 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6304 | 6308 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6305 | 6306 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6305 | 6308 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6306 | 6307 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6306 | 6308 | 1 | 1 | 1 | 1 | 1 | **1** |
| Territories | 6307 | 6308 | 1 | 1 | 1 | 1 | 1 | **1** |

**A2. Structural break between rural and urban spaces**

| Chow test | |
|---|---|
| rss_combined_spaces | 0,719 |
| rss_rural | 0,439 |
| rss_urban | 0,216 |
| where rss = Residual sum of squares from regression output | |
| | |
| k*2 | 34 |
| k | 17 |
| N_rural (units of observation) | 1169 |
| N_urban (units of observation) | 819 |
| | |
| Chow test results: | 11,21852693 |
| | |
| Critical value: | |
| F(17, 1954) | 1,51 |

The critical value for F(3, 1954) = 1,5(5%).
As 11,21 > 1,5, we reject the null hypothesis of structural stability. We thus conclude that there is a structural break in this model which suggests that it be re-estimated as 2 sub-samples.