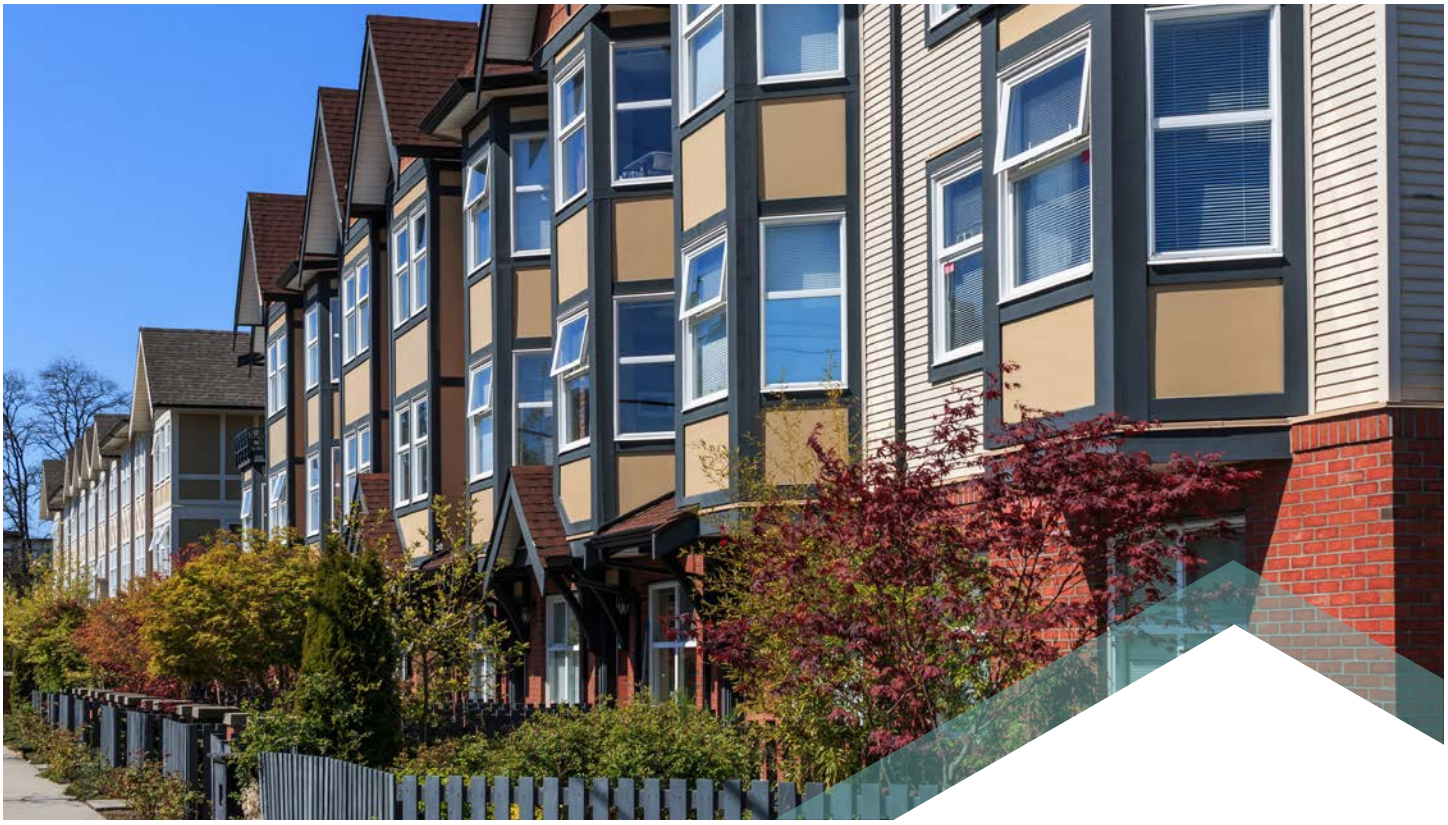


HOUSING RESEARCH REPORT

House Price and Income Inequality in Canada:
The Instrumental Variable Approach

Canada

CMHC SCHL

This study was conducted for Canada Mortgage and Housing Corporation (CMHC) and Natural Resources Canada (NRCan) under Part IX of the *National Housing Act* and the Program for Energy Research and Development (PERD). The analysis, interpretations and recommendations are those of the consultant and do not necessarily reflect the views of CMHC or NRCan.

La SCHL fera traduire le document sur demande.

Pour recevoir une copie traduite de ce document, veuillez envoyer un courriel ou remplir la partie ci-dessous et la retourner à l'adresse suivante :

Centre du savoir sur le logement
Société canadienne d'hypothèques et de logement
700, chemin Montréal
Ottawa (Ontario) K1A 0P7

Titre du rapport : _____

_____

Je demande que ce rapport soit disponible en français.

Nom : _____

*Courriel : _____

*No de téléphone : _____

Adresse : _____
rue app.

_____
ville province code postal

*Obligatoire

House Price and Income Inequality in Canada: the Instrumental Variable Approach

Duangsuda Sopchokchai* Chenggang Zhou†

CMHC CMHC

August 2020

Abstract

This study examines how residential house prices impact income inequality across Canadian CMAs and CAs. To the extent that house prices could be endogenous with respect to unobserved determinants of income, we propose a supply constraints instrument à la Saiz (2010) using GIS data from Natural Resources Canada. We find evidence that house price is indeed endogenous and a simple regression would lead to an inconsistent estimate. In our two-stage least square regressions, we find that an increase in median house price decreases income inequality measures, both when we consider Gini coefficients from total income and from residual income. This finding serves to encourage discussions among policy makers on the contradicting objectives of different policies at play, for instance, policies aimed at reducing income inequality versus policies aimed at assisting homebuyers, thereby putting an upward pressure on demand and price.

*Research Division, Canada Mortgage and Housing Corporation, 700 Montreal Road, Ottawa, ON; email: dsopchok@cmhc-schl.gc.ca.

†Research Division, Canada Mortgage and Housing Corporation, 700 Montreal Road, Ottawa, ON; email: czhou@cmhc-schl.gc.ca

Contents

1 Introduction **1**

2 Methodology **6**

2.1 Instrumental Variable: Developable Land per Capita 6

2.2 Empirical Model 9

3 Data **10**

4 Discussion **16**

5 Conclusion **19**

A Variable Definition **23**

1 Introduction

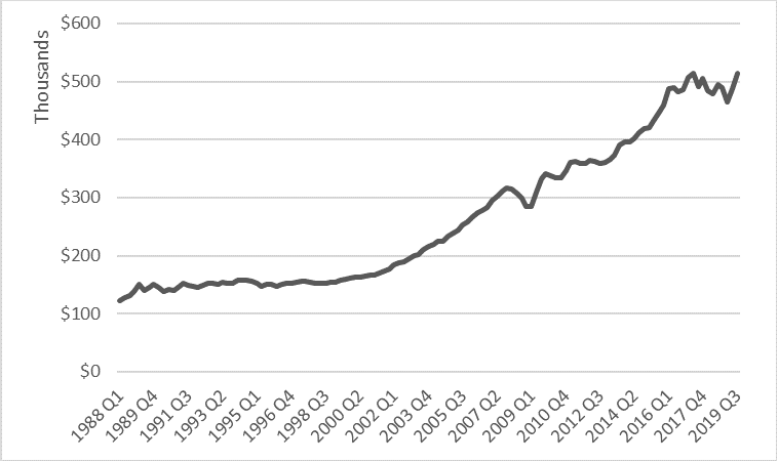
Income inequality has been increasing in Canada, especially among major cities. Between 1982 and 2010, the income of the bottom 90 % increased by a meagre two percentage points while the income of the top 10 % increased by more than 75 % (Green, Riddell, and St-Hilaire, 2017). The increase in income inequality seen in Canada is almost exclusive to major cities; for instance, in 2014, Calgary posted an increase in inequality four times higher than the national average since 1982, and Vancouver and Toronto followed closely behind with increases that are 2.5 and three times higher, respectively (Fong, 2017).

During the same period, average house price in Canada has increased significantly. Figure 1 shows the average quarterly residential price for Canada from 1988 to 2019. The average residential house price increases dramatically from around $150,000 in the late 1990s to around $500,000 in recent years. A similar trend is also observed in the rental market where the average rental price has almost doubled for a two-bedroom apartment, from $568 in 1992 to $962 in 2016 (CMHC Rental Market Survey).

It is well-documented in the income inequality literature that the long-term trend in inequality can be explained by structural changes in the labour market;[1] nevertheless, the pivotal question yet to be answered is whether, and to what extent, a housing market boom counteracts or further exacerbates income inequality across major cities in Canada. This question warrants an investigation as researchers observe how housing drives a long-term rise in income in seven large developed economies including Canada, but that the winners in this phenomenon are the ones at the top part of the

[1] See, for instance, Katz and Murphy (1992), Autor, Katz, and Krueger (1998), DiNardo, Fortin, and Lemieux (1996), and Card, Lemieux, and Riddell (2003).

Figure 1: Average Residential House Price



Data retrieved from the CREA website: https://creastats.crea.ca/en-CA/

income distribution, further propagating inequality for generations to come.[2]

There are a number of reasons why increases in house prices, and therefore limited access to housing, can lead to growing income inequality. Property is indeed an important asset that yields many income advantages, including a return on investment from increases in house prices and savings one make from rental. Low-income individuals are barred entirely from these financial benefits due to financial inaccessibility to housing.

Rising house prices also hampers the flow of unskilled labour to regions where an increase in house prices is coupled with an economic boom. This, in turn, prevents income mixing in the region, which could reduce income inequality. Furthermore, it is well known and established in the literature that the size and physical condition of a home and its location affect the socioeconomic outcomes of the children growing up in said homes. As such, the housing market can often perpetuate existing inequalities in the society.

[2]See Rognlie (2015) for instance.

Many studies examine a relationship between income inequality and rising house prices; however, the approach varies and the direction of this relationship is ambiguous. For instance, studies employing macroeconomic models to study the theoretical relationship between income distribution and the housing price distribution typically consider whether inequality affects house prices. In this strand of the literature, income inequality can affect house prices via demand mechanisms. First, when houses are considered as consumption goods, an increase in income inequality raises the number of people who are willing to pay high prices for their residence (see for instance Gyourko, Mayer, and Sinai (2013), Määttänen and Terviö (2014), and Matlack and Vigdor (2008)). Second, when houses are considered as rent-generating assets, inequality is expected to increase the absolute amount of savings, which in turn raises the total demand for houses as investment good (see for instance Nakajima (2005) and Zhang (2016)).

Another study uses a time series method to unearth the association between income inequality and house prices. Goda, Stewart, and Torres (2019) hypothesize that the co-movement of income inequality and house prices is systematic, and that the increase in inequality leads to an increase in the demand of homeownership and therefore house prices. They conducted co-integration tests for a panel of 18 OECD[3] countries from 1975 to 2010, using both Gini coefficient and its variance as measures of income inequality. They found that increasing income inequality raises real house prices in the majority of the OECD countries.

Hyun Choi and K. Green (2017) apply a micro-econometrics approach to investigate how fluctuations in house prices influenced households income inequality in the U.S.. The researchers considered not only nominal income inequality, but also

[3]Organization for Economic Co-operation and Development

inequality in the income after housing expenses (refer to as "real income inequality" in their study). Since housing expense accounts for a large proportion of household income, including such cost will alter the income distribution. Using the 2000 Census and American Community Survey for years 2005 to 2015, they found a negative relationship between the change in house prices and the change in nominal income Gini coefficients; nevertheless, the changes in the housing expenses fully offset the reduction in income inequality.

This study examines how residential house prices impact income inequality measures for Canadian census metropolitan areas (CMAs) and census agglomerations (CAs). Since housing costs account for a large proportion of household income and have long been recognized as a factor contributing to poverty and to inequality, similar to Hyun Choi and K. Green (2017), we consider both total nominal household income and residual income in calculating the Gini coefficients.[4]

To the extent that house prices could be endogenous with respect to unobserved determinants of income, an instrumental variable is essential to identify exogenous variation in house prices.[5] We develop an instrumental variable similar to the concept suggested in Saiz (2010), developable land per capita,[6] using Geographic Information Systems (GIS) data from Natural Resources Canada. We find that an increase in house price reduces income inequality but that the impact is smaller when we consider

[4]Residual income is the total household disposable income less total housing expenses. Housing expenses for homeowners include,where applicable, mortgage payments, property taxes and condominium fees, costs associated with electricity, heat, water and other municipal services. For renters, housing expenses include the rent and costs associated with electricity, heat, water and other municipal services. See Sopchokchai and Shewchuk (2020) for an application of the residual income concept to measure housing affordability problems.

[5]Hyun Choi and K. Green (2017) uses land unavailability à la Saiz (2010) as an instrumental variable.

[6]Developable land per capita is defined as land available within each boundary for housing development divided by the number of people within that boundary. The available land cannot be a body of water (for example, oceans, lakes, etc.) and the slope cannot be greater than 15 degrees (for example, mountainous land).

the residual income.

The finding that house price negatively impacts income inequality does not imply that house price escalation is a panacea for income inequality. One of the reasons we observe such a phenomenon is likely due to the fact that there is matching by individual ability and regional productivity such that an inflow of high-skilled workers to a relatively more productive region leads to a rise in income level within the region. This increase in income results in more competition for scarce land, which puts an upward pressure on house prices and hinders the migration of low-skilled workers into the region. As a consequence, there is less mixing of income levels in the area. See, for instance, Moretti (2013), Gyourko, Mayer, and Sinai (2013), and Ganong and Shoag (2017) for the literature discussing labour migration, productivity and house prices in detail.

The findings from this study also promote a discussion among policy makers on some contradictions between different policies being implemented. For instance, one objective of tax policies is to reduce income inequality, while housing policies, especially ones that generate demand, could favour escalating house prices, which in turn exaxerbate the inequalities.

This study is organized as follows. Section 2 discusses the methodology, section 2.1 expands on the instrumental variable developed for this research and section 2.2 presents the empirical model. Section 3 details the data sources. We discuss our results in detail in section 4 and conclude in section 5.

2 Methodology

2.1 Instrumental Variable: Developable Land per Capita

The key insight derived from the housing literature is that given a common positive housing demand shock, cities with a more limited housing supply will experience larger house price increases. To estimate housing supply elasticity and causal inference, an instrument with exogenous variation is necessary to avoid potential sources of endogeneity obfuscating such economic relationships. Specifically, we face the typical issue of endogeneity in testing our hypothesis. The variation of house prices across cities is endogenous with respect to unobserved determinants of local household income. A change in the income distribution might affect local demand, which in turn affect the house price. In this study, we use developable land per capita as an instrumental variable to isolate exogenous variation in house prices as our instrument variable.

Saiz (2010) measures cross-geography metropolitan-level housing market elasticities as a non-linear combination of land lost to steep slopes and water and a proxy for housing market regulations.[7] Since regulation constraints are likely endogenous, the land unavailability itself is considered as a candidate instrument variable for home prices. Indeed, many studies since have exploited this land unavailability as an instrument in the prediction of housing markets for policy purposes, to test housing market theories and to compute causal estimates through the instrumental variable approaches.

Davidoff (2016) posited that supply constraints may not be valid instruments for

[7] Saiz (2010)'s proxy for regulation constraints is conducted using historical government expenditure on preventive and regulatory activities, and the 1971 fraction of local Christians that belonged to non-traditional Protestant denominations.

home prices;[8] however, a more recent study by Lutz and Sand (2019) demonstrated that there is little evidence to support such concern through the use of innovative large-scale data techniques and comprehensive high-resolution satellite imagery. In particular, Lutz and Sand (2019) expanded the popular unavailable land proxy from Saiz (2010) by constructing new data sets to measure the percentage of undevelopable land as well as buildable land in the U.S.. They demonstrated that undevelopable land is uncorrelated with housing demand proxies, validating it as an instrument for house prices, and tested the supply side speculation theory with the buildable land data set. Lutz and Sand (2019) concluded that future research may employ the land unavailability (conversely the developable land) as an instrument in the estimation.

To derive our developable land per capita instrumental variable, we first compute the developable land area by substracting the area deemed undevelopable from the total land mass of our regions of interest. Following Saiz (2010), we consider an area to be undevelopable if it is covered by water or at a grade steeper than 15 degrees. We then calculate the developable land per capita (in square metres) by simply dividing the total developable land area by the regional population.

We compute the land slope using the elevation data from Natural Resources Canada Geo-spatial Data Extraction.[9] Geographic boundaries and water bodies (lake and river polygons) are both identified using Statistics Canada's 2016 boundary files. We also consider the protected countryside and urban river valley designations established by the *Ontario Greenbelt Plan (2017)* as undevelopable for the Greater Golden Horseshoe regions.

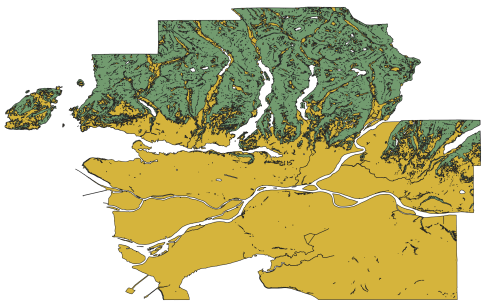
[8]The researcher uses data from Saiz (2010) to show that supply constraints are positively associated with housing unit growth and observable demand characteristics, and that their relationship with price growth and volatility falls significantly in the presence of demographic and productivity controls.

[9]Canadian Digital Elevation Model (CDEM) raster files for this study are at a resolution of 0.75 arc second along a profile in both the south-north and west-east directions.

Figure 2 depicts land mass for Vancouver, Toronto, Montréal and Calgary. The green colour represents areas with a slope greater than 15 degrees, the blue colour represents areas protected by the Greenbelt Plan, and the yellow colour represents land mass not covered by water in the CMA. From figure 2, it is clear that the area north of the Vancouver CMA is made up of mountains and the Toronto CMA has a sizable area of protected land, both of which cannot be used for residential development. Montréal and Calgary, on the other hand, have areas that are considered developable in this study.

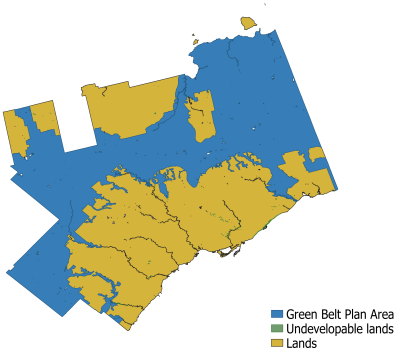
Figure 2: Land Maps

(a) Vancouver



(b) Montréal

(c) Toronto



(d) Calgary

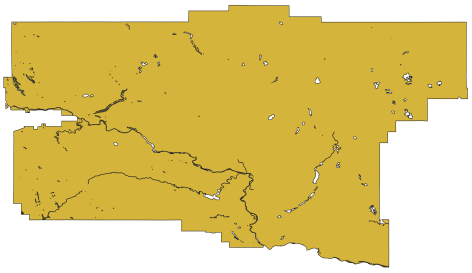
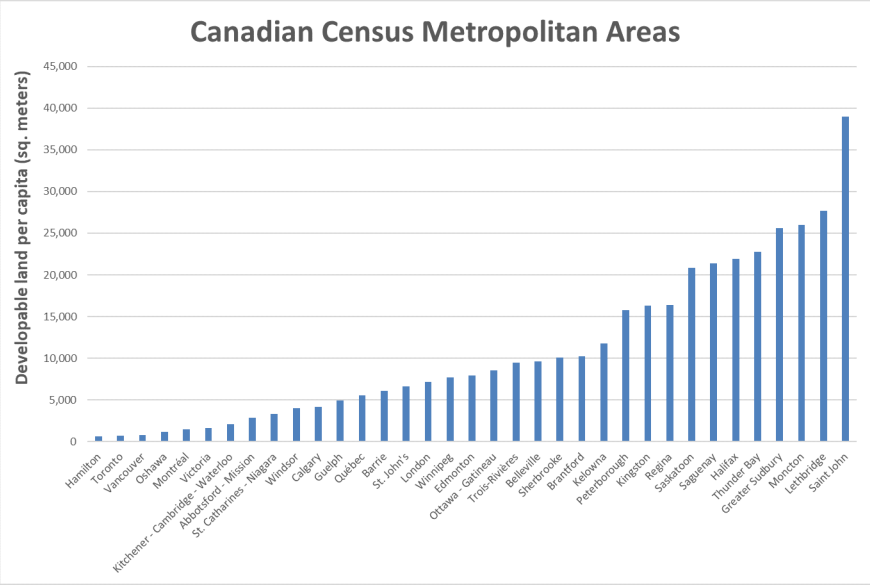


Figure 3 shows developable land per capita by CMAs. Both Toronto and Vancou-

ver, the two Canadian cities experiencing high house price growth in the last decade, have relatively small developable land per capita, at 673 and 801 m^2 per capita, respectively. Cities near the GTA, such as Hamilton, Kitchener, Cambridge and Waterloo, also experience limited developable land as a result of the Greenbelt Plan. Other major CMAs appear to have moderate developable land per capita, ranging from 4,133 m^2 for Calgary to 8,528 m^2 for Ottawa-Gatineau. Halifax, with its small population, has an extremely high value at 21,893 m^2 per capita.

Figure 3: Developable Land per Capita



2.2 Empirical Model

We estimate the following equation:

$$\text{Gini}_{kt} = \beta_0 + \beta_1 \log P_{kt} + \gamma X_{kt} + \mu_{pr} + \nu_t + \epsilon_{kt}, \tag{1}$$

where, for CMA k at time t, Gini_{kt} is the Gini coefficient, $\log P_{kt}$ is the natural log

of median sale house price, X_{kt} is a vector of demographic controls,[10] μ_{pr} and ν_t are the provincial and time fixed effect, respectively, and ϵ_{kt} is the error term.

To control for the endogeneity between house prices and unobserved determinants of income, we use the two-stage least squares (2SLS) regression with developable land per capita as instrumental variables. The variable P_{kt} in equation1 is predicted using the following first stage equation:

$$\text{Log}P_{kt} = \alpha_0 + \rho Z_{kt} + \delta X_{kt} + \xi_{kt}, \tag{2}$$

where Z_{kt} is our instrumental variable, developable land per capita, and ξ_{kt} is the error term.

3 Data

This study uses data from the 2006 and 2016 Canadian censuses and the 2011 National Household Survey (NHS). The census files provide high-quality information about people and housing units in Canada by their demographic, social and economic characteristics. Statistics Canada conducts a census every five years, and all residents of Canada are legally required to complete the census questionnaire, according to the *Statistics Act*. The Census Program consists of two parts: a short questionnaire with a basic set of questions and a long questionnaire. In 2011, the mandatory long form was replaced by a voluntary survey, that is, the NHS.

For statistical purposes, information on income, income taxes, contributions to registered savings plans and selected expenditures was extracted from respondents' personal income tax and benefits records[11] and added to their responses to the census

[10]The vector includes median age, median level of education, ratio of male head of the household, ratio of homeowners.

[11]Including the T1 income tax return, various information slips held by the Canadian Revenue

files. With this information at hand, we use the adjusted total household income and the adjusted disposable income in deriving our two key variables of interest: the Gini coefficient and the residual income (disposable income less housing expenses) Gini coefficient.[12] [13]

For another key variable of interest, the median sale house price of each CMA and CA, we use CMHC's confidential Property Sales and Assessment Database (PSAD). The PSAD is a large repository of past sales, property characteristics, and property assessment values with almost 10 million homeowner residential properties covering most of Canada.

We exclude households with multiple economic families since we cannot allocate housing expenses to each family unit living in the same household. We exclude farm-operating households from our analysis as the housing expenses pertaining to living cannot be separated from expenses in operating a farm. Household with non-positive income as well as those living in band housing are also excluded.

For variables such as age, education and immigrant status, we consider the information of the representative person of the household, specifically the person with the highest income in the household. In cases where more than one person reports equal amounts of income, we assign the older person as the representative of the household.

The total number of observations is 136 CMAs/CAs each year for a total of 408 units of observation. A full list of variables with detailed definitions can be found in

Agency (CRA) and the Canadian Child Tax Benefit (CCTB) and Goods and Services Tax (GST) credit programs.

[12] Adjusted income, both total and disposable, refers to income adjusted for economies of scale. The equivalence scale applied by Statistics Canada is the square root of the number of persons in the statistical unit. The adjustment made to income addresses the fact that individuals living together can share resources and the marginal increase in need decreases as the number of individuals sharing resources increases.

[13] Note that income-related information was not linked to the tax files in previous census years and, therefore, we cannot derive our key variable: total household disposable income.

appendix A.

Table 1 presents the summary statistics. On average, Gini coefficients of Canadian CMAs and CAs from residual income are higher than those of total income, 0.3707 for the first and 0.3191 for the latter. This implies that households in the lower part of the income distribution indeed spend a greater proportion of their income on housing expenses. The average median age of the household's representative person is about 52, with approximately 60% identifying as male, 11% being immigrant, 46% being married, and 70% being homeowners. The majority of our observations are made in Ontario, about 31%, followed by Quebec and British Columbia at about 18-19%.

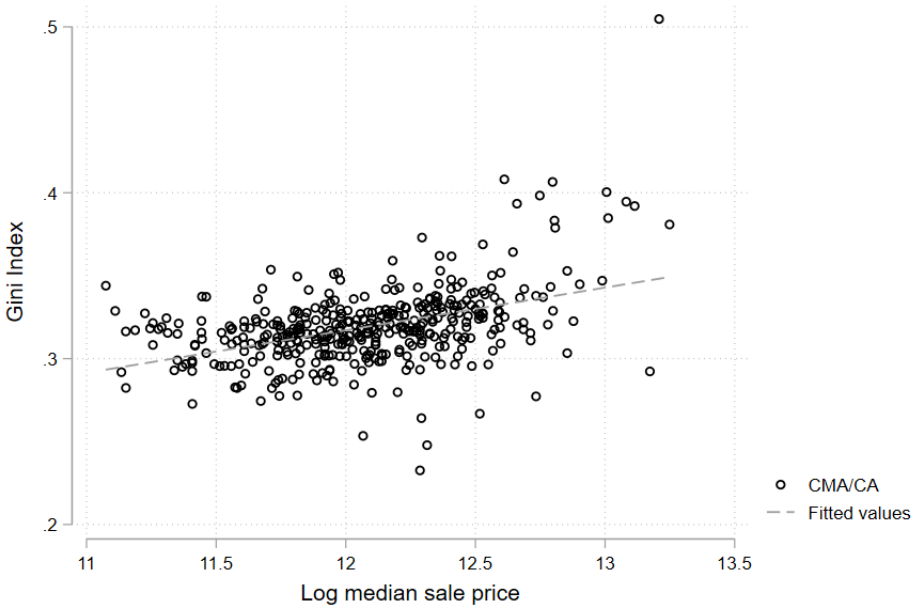
Figures 4 and 5 show scatter plots of Gini coefficients and log median sale prices using the total income and the residual income, respectively. In both cases, the correlation between the Gini coefficients and log median prices appear to be positive, though visibly less so for the residual income Gini. These positive relationships are shown to be faulty when we consider results from IV regressions, which we will discuss in detail in section 4.

Table 1: Summary Statistics

Variable	**Mean**	**Std. Dev.**
Gini–total income	0.3191	0.0012
Gini–residual income	0.3707	0.0011
Age	51.5294	0.1915
	(%)	
Male	59.93	0.0022
Immigrant	11.36	0.0045
Married	46.01	0.0042
Homeowners	70.23	0.0034
Median level of education:	(%)	
High school diploma or equivalent	15.44	0.0179
Trade certificate	45.83	0.0246
Certificate of Qualification	22.79	0.0207
College program $<$ 1 year	7.35	0.0129
College program 1-2 year	7.10	0.0127
College program $>$2 year	1.47	0.0059
Province:	(%)	
Newfoundland & Labrador	2.94	0.0083
Prince Edward Island	1.47	0.0059
Nova Scotia	3.67	0.0093
New Brunswick	5.14	0.0109
Quebec	19.11	0.0194
Ontario	30.88	0.0229
Manitoba	2.94	0.0083
Saskatchewan	5.88	0.0116
Alberta	9.55	0.0145
British Columbia	18.38	0.0191

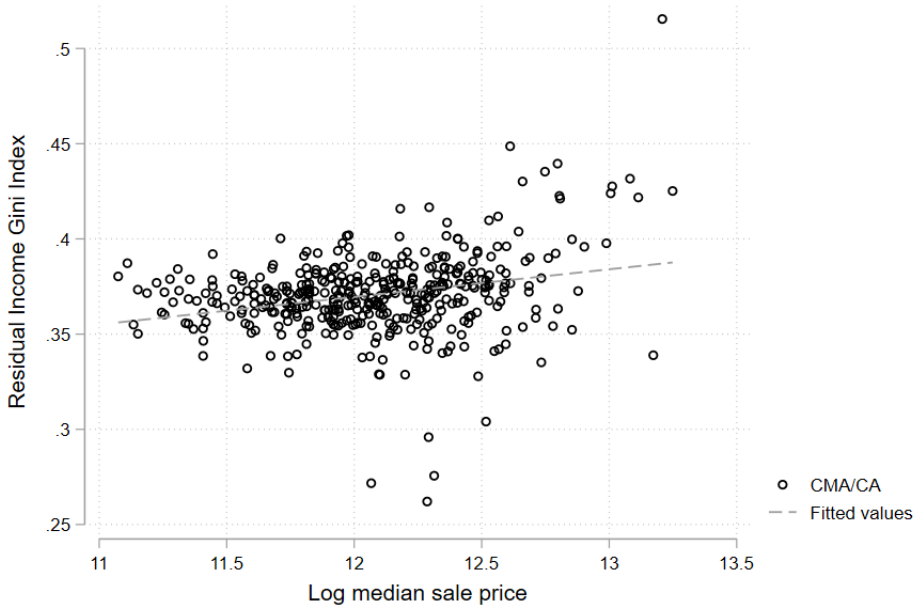
Data sources: 2006 and 2016 censuses; 2011 National Household Survey (NHS)

Figure 4: Scatter Plot: Gini Coefficients and Median Prices across CMAs and CAs



Data sources: Gini coefficients are calculated from the 2006 and 2016 census data and the 2011 NHS data; Median prices are retrieved from the PSAD.

Figure 5: Scatter Plot: Residual Income Gini Coefficients and Median Prices across CMAs and CAs



Data sources: Gini coefficients are calculated from the 2006 and 2016 census data and the 2011 NHS data; Median prices are retrieved from the PSAD.

4 Discussion

Before presenting the results from our 2SLS regressions, we present in table 2 and 3, the results from the ordinary least squares (OLS) regressions of nominal income Gini coefficients and residual income Gini coefficients, respectively. Specification 1 of both tables have one regressor, the log price, and thus the coefficients here are simply the slope presented in figures 4 and 5. These positive relationships disappear once we include other control variables, such as a set of demographic characteristics, time and provincial fixed effects. Using simple regressions to estimate the relationship between house prices and Gini coefficients would thus conclude that there is no statistical link between the two. We also find that the fit of the model improves significantly as we add the control variables to the model.

Tables 4 and 5 summarize the results from the 2SLS regressions of nominal income Gini coefficients and residual income Gini coefficients, respectively. The column *OLS* contains results from specification 4 of tables 2 and 3 for comparison purposes.

In contrast to results from simple regressions, we find a statistically significant and negative relationship between the house price and the Gini coefficient using the 2SLS approach. When we consider developable land per capita as an instrument, we find that an increase of 1% in the median sale house price leads to a reduction in the Gini coefficient by about 6 % for the nominal income and by about 4.5 % for the residual income. We perform endogeneity tests by testing the null hypothesis that price is exogenous. We strongly reject the null hypothesis, leading us to conclude that price is indeed endogenous in our model. [14]

[14] If the price variable is in fact exogenous, our estimates would still be consistent but less efficient.

Table 2: Pooled OLS Regressions of Gini Coefficients

	(1)	(2)	3)	(4)
Log price	0.026***	0.006	0.007	0.001
	(0.00)	(0.00)	(0.00)	(0.01)
Control variables:				
Demographic X	N	Y	Y	Y
Census cycles	N	N	Y	Y
Provinces	N	N	N	Y
N	408	408	408	408
Adjusted R-square	0.189	0.322	0.323	0.521

Data sources: 2006 and 2016 censuses; 2011 National Household Survey (NHS)
Note: Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$; Demographic X set includes median age, median education level, ratio of male population, ratio of immigrants and non-permanent residents, ratio of married couples, ratio of homeowners; Y = Yes, N =No.

Table 3: Pooled OLS Regressions of Residual Income Gini Coefficients

	(1)	(2)	(3)	(4)
Log price	0.014***	0.002	0.004	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Control variables:				
Demographic X	N	Y	Y	Y
Census cycles	N	N	Y	Y
Provinces	N	N	N	Y
N	408	408	408	408
Adjusted R-square	0.070	0.306	0.313	0.483

Data sources: 2006 and 2016 censuses; 2011 National Household Survey (NHS)
Note: Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$; Demographic X set includes median age, median education level, ratio of male population, ratio of immigrants and non-permanent residents, ratio of married couples, ratio of homeowners; Y = Yes, N =No.

Table 4: Pooled IV Regressions of Gini Coefficients

	OLS	IV
Log price	0.001	-0.061**
	(0.01)	(0.02)
Control variables:		
Demographic X	Y	Y
Census cycles	Y	Y
Provinces	Y	Y
N	408	408
First-stage regression summary statistics:		
Adjusted R-squared	N.A.	0.784
F-statistics	N.A.	22.50
Endogeneity tests		
Ho: Instrumented variable is exogenous:		
Robust score–chi-square	N.A.	p=0.002
Robust regression F	N.A.	p=0.000

Data sources: 2006 and 2016 censuses; 2011 National Household Survey (NHS)
Note: Robust standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$; Demographic X set includes median age, median education level, ratio of male population, ratio of immigrants and non permanent residents, ratio of married couples, ratio of homeowners; Y = Yes, N =No, N.A. = Not Applicable; See section 2.1 for a detailed discussion of how variable "developable land per capita" is developed.

Table 5: Pooled IV Regressions of Residual Income Gini Coefficients

	OLS	IV
Log price	-0.000	-0.045*
	(0.00)	(0.02)
Control variables:		
Demographic X	Y	Y
Census cycles	Y	Y
Provinces	Y	Y
N	408	408
First-stage regression summary statistics:		
Adjusted R-squared	N.A.	0.784
F-statistics	N.A.	22.50
Endogeneity test		
Ho: Instrumented variable is exogenous:		
Robust score–Chi-square	N.A.	p=0.012
Robust regression F	N.A.	p=0.004

Data sources: 2006 and 2016 censuses; 2011 National Household Survey (NHS)
Note: Robust standard errors in parentheses. $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$; Demographic X set includes median age, median education level, ratio of male population, ratio of immigrants and non-permanent residents, ratio of married couples, ratio of homeowners; Y = Yes, N =No, N.A. = Not Applicable; See section2.1 for a detailed discussion of how variable "developable land per capita" is developed.

5 Conclusion

This study examines how house prices relate to measures of income inequality in Canada. To answer this question, we follow Saiz (2010)'s popular supply constraints concept in constructing our instrumental variable to address the endogeneity issue presented in our model. Specifically, we propose *developable land per capita* using the definition of undevelopable land in Saiz (2010), that is, if the area is covered by water

or at a grade steeper than 15 degrees. We use the GIS data from Natural Resources Canada to identify sloped areas and geographic boundaries and water bodies data from Statistics Canada's 2016 files. With the use of data from the 2006 and 2016 Canadian censuses and the 2011 National Household Survey (NHS), we obtain 136 CMAs/CAs each year for a total of 408 units of observation.

We find that price is indeed endogenous in our model and, therefore, it is essential to apply an instrumental variable technique to obtain a consistent estimate of the coefficient of interest. We consider Gini coefficients calculated from both total income and residual income, and we find that an increase of 1% in median sale price reduces the Gini index by about 6 % and by 4.5 %, respectively.

The finding that house price negatively impacts income inequality is not to be taken lightly, and further study is needed to explain the observed relationship. One hypothesis that can be drawn from the labour and productivity literature is that high-skilled workers are likely to flow into high productivity areas, increasing income level in the region, while the competition for scarce land becomes fierce. This upward pressure on house prices further prevent low-skilled workers from moving into the area, resulting in less income ranges and therefore an improved income inequality measures. With the Census of Population and the NHS being cross-sectional data, we are unable to test this hypothesis here.

Nevertheless, the finding from this study serves to start a discussion among policy makers on the contradicting objectives of different policies at play. For instance, one objective of tax policies is to reduce income inequality, while housing policies, especially ones that generate demand, could favour escalating house prices, which in turn exaxerbate the inequalities.

References

Autor, D. H., L. F. Katz, and A. B. Krueger (1998). "Computing inequality: have computers changed the labor market?" *The Quarterly Journal of Economics* 113 (4), pp. 1169–1213.

Card, D., T. Lemieux, and W. C. Riddell (2003). *Labor market institutions, and the distribution of wages, 1973-1992: a semiparametric approach*. NBER Working Paper No. 9473.

Davidoff, T. (2016). "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated With Many Demand Factors". *Critical Finance Review* 5 (2), pp. 177–206.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). "Labor market institutions, and the distribution of wages, 1973-1992: a semiparametric approach". *Econometrica* 64, pp. 1001–1044.

Fong, F. (2017). *Income Inequality in Canada: The Urban Gap*. CPA Paper.

Ganong, P. and D. Shoag (2017). "Why has regional income convergence in the U.S. declined?" *Journal of Urban Economics* 102, pp. 76–90.

Goda, T., C. Stewart, and A. Torres (June 2019). "Absolute income inequality and rising house prices". *Socio-Economic Review* early online view, pp. 1–36.

Green, D. A., W. C. Riddell, and F. St-Hilaire (2017). *Income inequality in Canada: driving forces, outcomes and policy*. Institute for Research on Public Policy.

Gyourko, J., C. Mayer, and T. Sinai (2013). "Superstar cities". *American Economics Journal: Economic Policy* 5 (4), pp. 167–199.

Hyun Choi, J. and R. K. Green (2017). *House prices shock and changes in inequality across cities*. AEA Conference paper.

Katz, L. F. and K. M. Murphy (1992). "Changes in relative wages, 1963-1987: supply and demand factors". *The Quarterly Journal of Economics* 107 (1), pp. 35–78.

Lutz, C. and B. Sand (2019). *Highly disaggregated land unavailability*. Available at SSRN: https://ssrn.com/abstract=3478900.

Matlack, J. L. and J. L. Vigdor (2008). "Do rising tides lift all prices? Income inequality and housing affordability". *Journal of Housing Economics* 17 (3), pp. 212–224.

Moretti, E. (2013). "Real wage inequality". *American Economics Journal: Applied Economics* 5 (1), pp. 65–103.

Määttänen, N. and M. Terviö (2014). "Income distribution and housing prices: An assignment model approach". *Journal of Economic Theory* 151, pp. 381 –410.

Nakajima, M. (2005). *Rising earnings instability, portfolio choice and housing prices.* Mimeo, University of Illinois, Urbagna Champaign.

Rognlie, M. (2015). *Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity?* Brookings Papers on Economic Activity.

Saiz, A. (2010). "The geographic determinants of housing supply". *The Quarterly Journal of Economics* 123 (3), pp. 1253–51296.

Sopchokchai, D. and S. Shewchuk (2020). *Introducing the Housing Hardship concept.* Research Insight, Canada Mortgage and Housing Corporation.

Zhang, F. (2016). "Inequality and house prices". Job Market Paper. PhD thesis. University of Michigan, Department of Economics.

Appendix A Variable Definition

Variable	Description
Gini	Gini coefficients calculated using household's total income for each CMA. The Gini coefficients have the traditional interpretation—measuring the disposable income inequality in the CMA.
Residual Income Gini	Gini coefficients calculated using household's residual income, i.e. disposable income less shelter expenses, for each CMA.
Median sale price	Median sale price of residential properties sold in that CMA of the respective year. This variable is obtained from the PSAD.
Median age	Median age of the Major Income Earner (MJIE) of the household for each CMA.
Median level of education	Median of highest level education of the MJIE of the household for each CMA. The variable education is a categorical variable with responses: No certificate, diploma or degree; High school diploma or equivalency certificate; Certificate of Apprenticeship or Certificate of Qualification; Other trades certificate or diploma; College, CEGEP or other non-university certificate or diploma from a program of 3 months to less than 1 year; College, CEGEP or other non-university certificate or diploma from a program of 1 year to 2 years; College, CEGEP or other non-university certificate or diploma from a program of more than 2 years; University certificate or diploma below bachelor level; Bachelor's degree; University certificate or diploma above bachelor level; Degree in medicine, dentistry, veterinary medicine or optometry; Master's degree; Earned doctorate.
Immigrant ratio	Ratio of household's MJIEs who are immigrants in the CMA. Immigrant includes persons who are, or who have ever been, landed immigrants or permanent residents.
Married ratio	Ratio of households' MJIEs who are married in the CMA. 'Marital status' refers to whether or not a person is living in a common-law union as well as the legal marital status of those who are not living in a common-law union.
Male ratio	Ratio of households' MJIEs who are male.
Homeowner ratio	Ratio of households who own their dwelling in the CMA.
Unemployment rate	Unemployment rate for each CMA, calculated from an individual-level labour force status.

69676_RR 20200904-001

cmhc.ca