Possible income misstatement on mortgage loan applications: Evidence from the Canadian housing market

Basiri, Kiana and Mahmoudi, Babak

Ryerson University Ted Rogers School of Business, Toronto, Ontario, Canada, Canada Mortgage and Housing Corporation, Ottawa, Ontario, Canada

2020

DOI: 10.1111/1540-6229.12310

ORIGINAL ARTICLE

Possible income misstatement on mortgage loan applications: Evidence from the Canadian housing market

Kiana Basiri[1] | **Babak Mahmoudi**[2]

[1]Ryerson University Ted Rogers School of Business, Toronto, Ontario, Canada

[2]Canada Mortgage and Housing Corporation, Ottawa, Ontario, Canada

Correspondence
Kiana Basiri, Ryerson University, Ted Rogers School of Business, 350 Victoria Street, Toronto, Ontario, Canada M5B 2K3.
Email: kiana.basiri@ryerson.ca

Abstract

We construct a measure of possible income misstatement (PIM) for first-time homebuyers by quantifying the gap between growth in incomes reported on mortgage applications and growth in incomes reported on tax files from 2004 to 2014 in Canada. Using a two-stage least square framework to correct for the endogenous nature of house prices and PIM, we find robust evidence that part of the observed dispersion in PIM is caused by house price variation. This suggests borrowers have greater incentive to misstate income in high-priced markets. We report evidence that markets with a tendency for income misstatement also had higher default rates.

1 | INTRODUCTION

To what extent do house prices affect borrowers' incentive to overstate their income on mortgage applications? Using Canadian data, this article measures the gap between growth in first-time homebuyers' (FTHBs') incomes reported on mortgage applications and growth in FTHBs' incomes reported on tax files from 2004 to 2014. The basic idea is that in the absence of any change in income misstatements on mortgage applications, income reported on mortgage applications and tax files rise (or decline) at the same rate over a certain period. Therefore, we expect to see little to no growth difference (*i.e.,* no expansion or contraction in the gap).

This article has three objectives: (i) to construct a measure of possible income misstatement (PIM) on mortgage applications, (ii) to test whether or not part of the observed dispersion can be associated with house prices and (iii) to document the relationship between PIM and ex post outcomes such as mortgage defaults. To achieve these objectives, we use data from the Canada Mortgage and Housing Corporation (CMHC), a federal Crown corporation that provides mortgage insurance, and we match these data at the level of forward sortation area, or FSA (the geographic area designated by the first three characters of a Canadian postal code), with tax record information from the Canada Revenue Agency (CRA) provided by the Longitudinal Administrative Databank (LAD). At the time of loan origination, the CMHC collects borrower and loan characteristics, such as mortgage interest rate, loan amount, house price, house location, terms and amortization, loan performance, income, credit score and debt service ratios. To make our sample homogeneous, we focus on FTHBs with a loan-to-value (LTV) ratio between 80% and 95%. To identify FTHBs in the LAD, we include filers who used the Home Buyers' Plan (HBP) to purchase their house. The HBP is a federal program that allows FTHBs to borrow, tax-free, from their registered retirement savings plan (RRSP) in order to make a down payment on a house.[1] We are able to construct FSA-level measures of the degree of PIM for FTHBs from 2004 to 2014 in Canada. Our measure does not suggest a notable amount of PIM for insured mortgages in the Canadian mortgage market from 2004 to 2014.

We hypothesize that borrowers have a greater incentive to misstate their income in high-priced markets. However, if we assume borrowers overstate their income to enter the housing market, then an increase in income misstatement potentially translates to higher house prices if the supply of housing is inelastic in the short term. Therefore, we are faced with an endogeneity problem. To mitigate the scope for an alternative interpretation of our evidence, we use the average rent prices in each FSA as an instrument for house prices. We observe that the rent prices are highly correlated with the house prices, while also being uncorrelated with our error term. We find that when other controls are kept constant, PIM is higher in high-priced markets. More specifically, a 1% increase in house prices is associated with a 0.12% increase in PIM. Our results also suggest that among loan characteristics (*e.g.*, total debt service [TDS] ratio, gross debt service [GDS] ratio and LTV ratio), GDS ratio is negatively associated with PIM. This is consistent with the notion that tighter borrowing constraints could lead to a higher probability of income misstatement.

Given our findings, it is important to ask whether FSAs with higher PIM experience worse loan performance. We use two definitions of default as an indicator of loan performance: (i) mortgage in arrears, when a mortgage is 90 or more days past due within five years of the loan origination date, and (ii) mortgage in claims, when a mortgage is foreclosed and a mortgage insurance claim is submitted by the lender within five years of loan origination.[2] Arrears is an early indicator of a borrower who has cash flow problems.[3] Borrower cash flow problems could be due to income shocks, such as unemployment, loss of secondary income or marital split; increases in mortgage payments caused by an increase in a variable rate; health shocks; or mismanagement of funds. We track mortgage performance data for five years for mortgages that originated between 2004 and 2009. For example, a mortgage that originated in 2005 is flagged as a mortgage in arrears if the mortgage was 90 or more days past due between 2005 and 2010. Our results suggest that a 1% increase in predicted PIM is associated with a 0.12% increase in the share of arrears.

[1]See Steele (2007) for a comprehensive analysis of this program.

[2]Rates of claims and arrears are two measures of loan performance that are widely used in the industry. See U.S. Financial Crisis Inquiry Commission (2011, p. 215). Default is defined as "90-days or more past due or in foreclosure." Some authors, for example, Mian and Sufi (2009) also use 30-day or more delinquent as definition of default.

[3]Arrears could also be an indicator of a borrower who is in the process of strategic default. Strategic default is when a borrower with negative equity is able to make regular mortgage payments, but chooses not to pay. Strategic default is correlated with significant negative equity (Bhutta, Dokko, & Shan, 2017). From 2004 to 2014, the Canadian housing market did not experience a significant house price decline and because strategic default is less likely to happen for insured mortgages that are recourse (Ghent & Kudlyak, 2011), strategic default does not represent a big share of arrears in our sample.

Our results should be viewed as a lower bound. The sample we use in this study is less associated with misrepresentation than the population. Therefore, there might be larger risks that cannot be measured in the current sample. Our study covers insured mortgages, where regulators have strong oversight on underwriting through mortgage insurers. PIM is known to be prevalent in the unregulated space, where regulators do not have enough oversight, for example, Ben-David (2011); Jiang, Nelson, & Vytlacil (2014); Griffin & Maturana (2016); Piskorski, Seru, & Witkin (2015).

We performed several robustness checks that are reported in the Online Appendix. Taken together, these results suggest that high-priced markets appear more likely to create incentives for borrowers to falsify their income in order either to enter the housing market or to obtain bigger loans. This creates a channel for increased defaults due to possible cash flow problems.

A number of papers demonstrate that misrepresentations of borrowers' incomes on mortgage applications increased significantly in the United States during the period prior to the 2007–2008 financial crisis. These papers compare income reported on Home Mortgage Disclosure Act (HMDA) data; covering the vast majority of mortgages, with alternative sources of income data. Blackburn & Vermilyea (2012) compare the incomes of a subsample of U.S. homebuyers from the American Housing Survey (AHS) with income reported on HMDA data and, because a loan-level match is not possible, they use borrower and mortgage characteristics to calculate an HMDA-estimated income for each AHS observation. Avery et al. (2012) conduct a similar exercise, but use the 2000 Census and the 2005 American Community Survey instead of the AHS. Mian and Sufi (2017) use income reported on mortgage applications from HMDA and compare it with tax-reported income. They use ZIP code–level differences in growth in income reported on mortgage applications and growth in tax-reported income of all the tax filers in a ZIP code. All of these studies found that income reported on mortgage applications was highly overstated during the housing boom. While there is widespread evidence of income misstatement in the United States, no academic study has previously examined system-wide evidence for Canada on the existence and magnitude of PIM.

Using data from a single lending institution from 2004 to 2008, Jiang et al. (2014) compared average neighborhood income reported on tax files to average income reported on mortgage applications and concluded that income exaggeration occurred on low-doc loans, resulting in elevated defaults. Ambrose, Conklin, and Yoshida (2016) follow the method outlined in Jiang et al. (2014) to measure PIM with a focus on differences in employment status. They show that the majority of adverse selection and income falsification is attributed to a specific borrower group that selected into low-doc loans. They also support the findings of Mian and Sufi (2017) that mortgages granted to borrowers who were most likely to overstate income were concentrated in lower income neighborhoods. Yavas and Zhu (2020) use a nationwide mortgage servicing data set that includes the residential mortgages serviced by the nine out of the ten largest U.S. residential mortgage servicers. They match these data with the country-level Recorder's Data (public record) to identify the underreporting of second liens. They find significant amounts of second lien misreporting in both portfolio loans and privately securitized loans in years 2005 and 2006.

Our article complements the findings of these studies in a number of dimensions. First, similar to Mian and Sufi (2017), our article compares average income reported on mortgage applications to average tax income at the FSA level (first three characters of a postal code). However, in order to create a better match between the two data sources, we focus on FTHBs who reported their home purchases on income tax files to take advantage of a federal tax program. Our measure of PIM compares growth differences in average income reported in the tax files of these FTHBs to the average income that

FTHBs reported to lenders. This reduces the concerns about gentrification present in the previous study, which included the average income of all tax filers in a ZIP code.[4]

Second, we develop a simple empirical model of default rates during 2004 and 2014, in which PIM is allowed to be a potential explanation. To our knowledge, this article is the first study that uses Canadian data consisting of the entire historical loan portfolio of the largest mortgage insurance company in Canada to measure the importance of PIM during this period. We demonstrate that as PIM increases, loan performance deteriorates. Our article also contributes to mortgage insurance practices by proposing appropriate measures to reduce risk, and it supports the need for better income verification.

The rest of the article proceeds as follows: Section 2 describes the institutional environment, in particular mortgage default insurance and fraud detection in Canada. Section 3 presents a detailed analysis of the data. Section 4 presents the empirical analysis, and Section 5 outlines some conclusions. The Online Appendix contains additional results.

2 | MORTGAGE DEFAULT INSURANCE AND FRAUD DETECTION IN CANADA

2.1 | Mortgage default insurance

Mandatory mortgage default insurance was introduced in Canada over 60 years ago to protect lenders against defaulting borrowers. When a borrower defaults on an insured mortgage, the lender typically forecloses and sells the property. The lender then files an insurance claim that includes the shortfall of the sale amount of the property to cover the outstanding balance of the mortgage and some legal and sale costs.[5]

Legislation generally requires mortgage insurance for any mortgage loan that exceeds 80% of the value of the mortgaged property (high-ratio mortgage is any loan that exceeds an 80% LTV ratio). Mortgage insurance is also available for mortgages with LTV ratios of 80% and under. In Canada, 45% of all mortgage credit is insured.[6] Mortgage insurance is provided by the CMHC, a federal Crown corporation, and two private mortgage insurers, Genworth and Canada Guarantee. The CMHC covers approximately half of the market value of flow of insured mortgages.[7] Making high-ratio mortgage access conditional on mortgage insurance allows for greater scrutiny of underwriting practices for insured mortgages.

2.2 | Residential mortgage underwriting

The OSFI requires a Federally Regulated Financial Institution (FRFI) to develop a risk appetite framework (RAF), which takes into account the FRFI's risk profile. The OSFI sets out guidelines for prudent residential mortgage underwriting and is applicable to all FRFIs. Lenders follow these underwriting

[4]One concern when comparing income of FTHBs with average income of all tax filers in a neighborhood is that the discrepancy might be a reflection of gentrification of poor neighborhoods. Relatively high-income homebuyers may be attracted to poorer neighborhoods that are more affordable. One important contribution of this article is that we try to select (to the best the data allow us) FTHBs from both sources, that is, LAD and CMHC. Therefore, we are not comparing the high-income FTHBs purchasing homes in poorer neighborhoods with the average income of the whole neighborhood.

[5]Alternatively, the lender can transfer the property to the mortgage insurer for the entire outstanding balance of the mortgage. In such a case, the mortgage insurer would sell the property.

[6]According to the Office of the Superintendent of Financial Institutions (OSFI).

[7]According to CMHC estimates.

standards for borrowers' reported income. For example, it is common practice to require self-employed borrowers to provide evidence of at least three years of income.[8] Moreover, the OSFI requires lenders to exercise rigorous due diligence in underwriting loans that are materially dependent on rental income or for borrowers relying on income from sources outside of Canada. Lenders are required to discount reported incomes that are temporarily high due to overtime wages, irregular commissions and bonuses. These guidelines create a "*structural difference*" between borrowers' income reported on mortgage applications and borrowers' income reported for income tax.

Additionally, the OSFI expects lenders to put in place programs that continuously monitor and audit the information received on mortgage loans to reduce the potential for inaccuracies and to help detect fraud or "gaming."[9] Lenders also have underwriting requirements related to fraud detection, such as income verification, screening tools and risk management frameworks. Third-party tools such as credit bureau reports or platforms specializing in detecting fraud are also available.[10]

2.3 | Why is income misrepresentation still possible?

Mortgage approval requires borrowers to disclose any sources of income they have. This includes employment income, investment income, rental property income, pension income, spousal support and self-employed income. The income verification process requires borrowers to submit a copy of their latest pay stub, a letter of employment, Canadian tax documents such as T4 income slips, T1 General tax returns and notices of assessment (NOAs), as well as documents showing other sources of income and legal agreements to support a spouse or make child support payments.

Letters of employment, pay slips and T4 slips are falsifiable; and tax returns from previous years provide a noisy proxy for present income. Lenders and mortgage insurers take steps to detect misstatement of income, including contacting a borrower's employer for income verification. However, they do not necessarily verify a borrower's income directly with the tax authority (the CRA). Over the years, mortgage insurers and lenders have detected occasional sophisticated fraud schemes, such as fake phone numbers and fake companies.

3 | DATA DESCRIPTION AND METHODOLOGY

The sample for this analysis is constructed by combining two main databases. Our primary data on mortgage contracts and borrowers is loan-level administrative data collected by the CMHC from 2004 to 2014. At the time of origination of insured mortgages, the CMHC collects loan characteristics such as interest rate, loan amount, house price, house location, terms and amortization and loan performance; and borrower characteristics such as income, credit score and debt service ratios.

Our second source of data is income and demographic variables from the CRA tax files provided by the LAD. The LAD provides detailed information about both individual and family income for people who filed an income tax return between 1982 and 2014. The LAD takes a random 20% sample of all tax filers, and individuals remain in the sample for as long as they file their taxes. Income tax returns are filed mainly in the spring following the year of reference "y." LAD includes the T1 files for income year "y," which are received from CRA in January of the year "y+2." To ensure that the sample we used

[8]Average income for the past three years is considered for mortgage applications for self-employed borrowers.

[9]Mortgage insurers have fraud detection tools and processes at the underwriting stage and at the mortgage insurance claim processing stage.

[10]Equifax's Citadel, for example, is a data sharing platform used by lenders to flag cases of fraud.

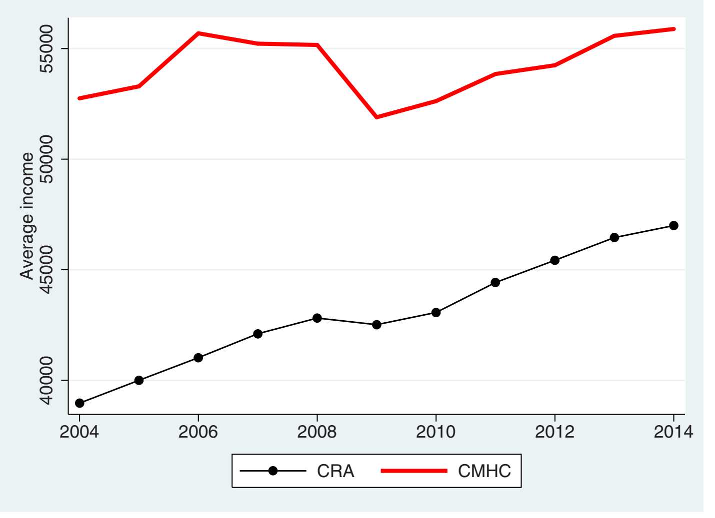

FIGURE 1 CMHC and CRA reported income of first-time homebuyers (FTHBs) [Color figure can be viewed at wileyonlinelibrary.com]

Note: FTHBs' income as reported on mortgage applications and tax files from 2004 to 2014. Both incomes are adjusted for inflation using the CPI (base 2002).

for this article is representative of the entire population, we multiply calculated aggregate variables by their corresponding weight.

To avoid mismatch errors, we compare the realized income in the same year mortgage was obtained and link that with house prices in the same year. Consider a home purchase in April 2011, borrowers declare their current income amount on their mortgage application. Lenders have to verify and substantiate borrowers' declared income. Borrowers provide the current proof of income (*e.g.,* pay stubs) and the 2010 NOAs for verifying the history of income.[11] Therefore, income recorded on CMHC data for a 2011 application is borrower's declared income in 2011. We compare CMHC income to the "actual" 2011 income reported on the tax files.

Figure 1 shows CMHC and CRA average-reported incomes for FTHBs from 2004 to 2014. The change in the gap between these two reported incomes is what we are interested in explaining. This gap could be explained by structural differences between income reported on mortgage applications and tax files.[12] It could also be due to inflated income reported on mortgage applications. We use the average rent prices at FSA in each year as an instrumental variable.[13] Our source for average rents is the Rental Market Survey (RMS), which is conducted by the CMHC every year to estimate the relative strengths in the rental market. The survey is conducted on a sample basis in all urban areas

[11]For mortgage applications early in the year, last year's NOAs may not be available. In these cases, lenders rely on NOAs from two years before the loan application and other forms of proof of income (*e.g.,* T4 slips) from last year.

[12]For example, lenders are required to discount reported incomes that are temporarily high due to overtime wages, irregular commissions and bonuses. Self-employed borrowers with less than three years of income, borrowers reporting rental income or those relying on income from sources outside of Canada are subject to rigorous due diligence. These sources of income are included in tax data.

[13]The FSA is the geographic area designated by the first three characters of a Canadian postal code. The average FSA has a radius of 7.6 km, while the median is much lower, at 2.6 km. We observed nearly 1,518 FSAs in the sample. In Online Appendix C, we report the number of FSAs in each Canadian province and territory.

with populations of 10,000 or more, and it targets only privately initiated structures with at least three rental units and which have been in the market for at least three months.[14]

To control for neighborhood characteristics, we include information from a number of sources, one of which is the 2011 National Household Survey (NHS), designed to collect social and economic data about the Canadian population. Information in the NHS is available at the FSA level. We also include FSA-level credit score information from Equifax (a credit bureau).

3.1 | Sample selection and summary statistics

We restrict our sample to contracts with homogenous terms. We exclude homeowners who were either refinancing or renewing their mortgage contract and focus only on newly issued mortgages.[15] We focus on FTHBs with LTV ratios between 80% and 95%.[16] We also drop contracts with more than one borrower on the mortgage application (no coborrower).

To identify FTHBs in the LAD, we track filers who used the HBP to purchase their house. The HBP is a federal program that allows FTHBs to borrow, tax-free, from their RRSP to make a down payment on a house. Ideally, we would have liked to compare income reported on mortgage applications with income reported on tax files at the loan level. However, because of confidentiality issues, individual-level data are not available through the LAD. As a result, we constructed an FSA-level measure of FTHBs' PIM.

Prior literature also uses U.S. ZIP code–level data to construct an aggregate measure of income misstatement. To improve our aggregate measures, we focus on FTHBs who reported their home purchases and not the average income of all tax filers in each FSA. First, we calculate average income for FTHBs in each FSA and year using CMHC data. Next, we match average income in each FSA for FTHBs using LAD data from 2004 to 2014 with CMHC data, resulting in 1,518 FSAs in each year. These 1,518 FSAs represented 92% of the population of Canada in 2011. Table 1 presents summary statistics for the sample used in our analysis.

Figure 2 plots mortgage credit and income growth at origination from 2004 to 2014 in Canada using CMHC data. The mortgage size growth is the growth in the total amount of mortgages originated for home purchases in an FSA from 2004 to 2014. Both income measures and mortgage size are annualized and adjusted for inflation.[17] To ensure outlier values were not included in our sample, we Winsorized the sample at the 1st and 99th percentiles. Figure 2 shows a rapid increase in mortgage credit during the period 2006–2009, followed by a decline during the housing bust in 2009–2011 and an upward trend since then.[18] During the 10-year window, mortgage credit grew substantially by 40%, while income grew by only 7%, in real terms. This is some indication of rising leverage among FTHBs over the past decade.

[14]See https://www03.cmhc-schl.gc.ca/hmip-pimh/en/tablemapchart/rmsmethodology for more information on this program.

[15]Allen, Clark, and Houde (2014) use similar data and show that repeat buyers have similar TDS ratios and lower LTV ratios compared to FTHBs, and that they take out, on average, larger loans to purchase larger homes.

[16]FTHBs are defined as households that do not own housing assets at the time of mortgage origination. FTHBs represent 77% of CMHC's data.

[17]We use the Consumer Price Index (CPI), excluding food, energy and the effect of indirect taxes (index, 2002 = 100) to adjust for inflation.

[18]Mortgage size includes mortgage amount plus mortgage default insurance premium.

TABLE 1 Summary statistics

Variable	Mean	*SD*	*P*(25)	*P*(50)	*P*(75)
Property value	177,634	77,129	118,848	168,900	226,153
Mortgage size	169,043	72,678	113,733	161,222	214,678
Income (CMHC)	54,196	14,085	44,863	51,829	60,901
Income (CRA)	43,074	11,823	35,555	40,462	47,263
Loan-to-value	92.9	1.32	92.3	93.1	93.8
Total debt ratio	34.6	2.56	33.2	34.8	36.2
Credit score	715	36	700	719	735
Mortgage interest rate	4.58	0.90	3.86	4.49	5.36
Bank	0.536	0.191	0.426	0.538	0.65
Credit union	0.18	0.22	0	0.09	0.3
Financial services	0.13	0.13	0.03	0.11	0.2
Other	0.15	0.14	0.04	0.13	0.22
Self-employed	0.01	0.05	0	0	0
Broker	0.12	0.13	0	0.09	0.19
Arrears	0.06	0.09	0	0.03	0.08
Claims	0.03	0.07	0	0	0.05

Note: This table presents a summary of statistics for FSAs in our sample. The sample contains 585,343 observations on FTHBs with LTV between 80% and 95% and with no coborrower on the contract from 2004 to 2014. Property value, mortgage size and income are adjusted for inflation using the CPI (base 2002) and are Winsorized at the 1st and 99th percentiles. Income (CRA) is taxable income reported to the CRA. "Other" is a category that includes proportion of mortgages originated by life insurance, monolines and trusts in our sample.

3.2 | Loan performance measures

We use two indicators of loan performance: (i) mortgage in arrears, when a mortgage is 90 or more days past due within five years of the loan origination date, and (ii) mortgage in claims, when a mortgage is foreclosed and a mortgage insurance claim is submitted by the lender within five years of loan origination. To make claims and arrears comparable for mortgages originated in different years, we use a vintage-based approach. We track mortgage performance data for mortgages originated between 2004 and 2009 for five years. For example, we flag a mortgage originated in 2005 as a mortgage in arrears if it was 90 or more days past due at any time from 2005 to 2010. Similarly, we flag the loan as defaulted if the property was foreclosed and a mortgage insurance claim was submitted at any time between 2005 and 2010. By definition, every mortgage in default was also flagged as mortgage in arrears. The arrears rate is on average 6%, and the mortgage insurance claims rate is on average 3% for loans that originated between 2004 and 2009.[19]

3.3 | PIM measure

This subsection documents the construction of our main variable: PIM. Our methods closely follow those of Mian and Sufi (2017). Specifically, we calculate the difference between growth in average income according to CRA tax data and growth in average income according to mortgage applications

[19] Arrears and claims rates capture defaults in the first five years of the life of the mortgage. The likelihood of default is significantly higher in the early years of mortgages. As a result, aggregate arrears and claims rates on the entire stock of mortgages on CMHC's book are significantly lower. For example, current aggregate arrears rate for CMHC's book is around 0.4%.

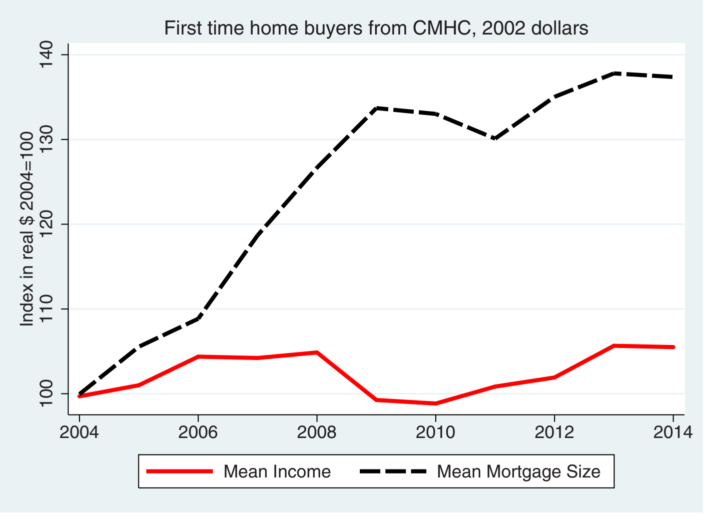

FIGURE 2 Percent change in average income and mortgage size, CMHC data [Color figure can be viewed at wileyonlinelibrary.com]

as recorded in CMHC data for FTHBs in each FSA from 2004 to 2014. In order to create a better match between the two data sources, we focus on FTHBs who reported their home purchases on income tax files to take advantage of a federal tax program, and not the average income of all tax filers in each FSA. The basic idea is that, in the absence of any income misstatement on mortgage applications, income reported on mortgage applications and income reported on tax files rise (or decline) at the same rate over a certain period. Therefore, we expect to see little to no growth difference (*i.e.,* no expansion or contraction in the gap).

We assume a linear relationship between the log of tax income and the log of income reported on mortgage applications. For each borrower j in FSA_i, we have the following equation:

$$I^{CMHC}_{j,t} = I^{CRA}_{j,t} + C_{j,t} + \psi_{j,t} + \epsilon_{j,t}, \tag{1}$$

where I^{CMHC} is the log of income reported on mortgage application, I^{CRA} is the log of tax income and C is all the structural differences between tax income and mortgage income. C includes all forms of a borrower's income that are reported as taxable income and are not reported on mortgage applications, and vice versa. ψ is the misstatement of income on mortgage applications and ϵ is the error term.[20] We calculate the average income for all FTHBs in each FSA_i at time t

$$\frac{1}{n_{i,t}} \sum_{j \in FSA_i} I^{CMHC}_{j,t} = \frac{1}{n_{i,t}} \sum_{j \in FSA_i} \left[I^{CRA}_{j,t} + C_{j,t} + \psi_{j,t} + \epsilon_{j,t} \right], \tag{2}$$

where $n_{i,t}$ is the number of FTHBs at time t. The above expression holds for $t-1$. We take the first difference, which gives the following for all FSA_i

$$\begin{aligned} \bar{I}^{CMHC}_{i,t} - \bar{I}^{CMHC}_{i,t-1} &= \left(\bar{I}^{CRA}_{i,t} - \bar{I}^{CRA}_{i,t-1} \right) + \left(\bar{C}_{i,t} - \bar{C}_{i,t-1} \right) + (\bar{\psi}_{i,t} - \bar{\psi}_{i,t-1}) \\ &+ \left(\bar{\epsilon}_{i,t} - \bar{\epsilon}_{i,t-1} \right), \ \forall \ FSA_i, \end{aligned} \tag{3}$$

[20]For example, self-employment income from less than two years is taxable, but cannot be reported on mortgage applications.

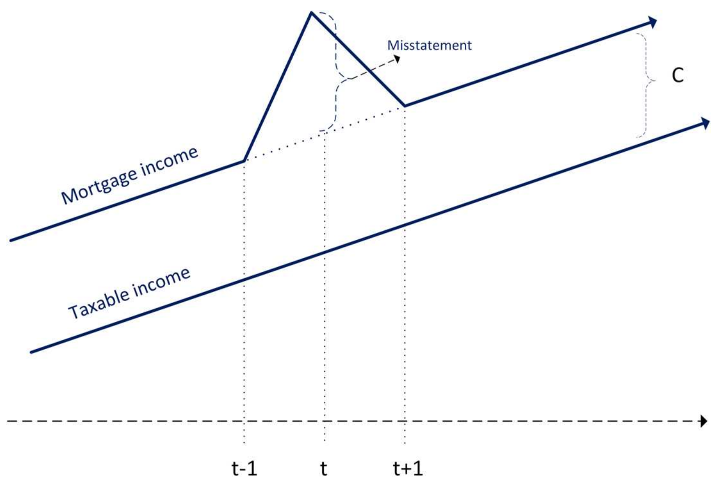

FIGURE 3 Possible income misstatement [Color figure can be viewed at wileyonlinelibrary.com]

where "$\bar{I}$" shows the average. We assume that for each FSA, the average structural differences between two types of income and the average of the error term remains steady over time: $\bar{C}_{i,t} - \bar{C}_{i,t-1} = \bar{\epsilon}_{i,t} - \bar{\epsilon}_{i,t-1} = 0$, which gives

$$\bar{\Psi}_{i,t} - \bar{\Psi}_{i,t-1} = \left(\bar{I}^{CMHC}_{i,t} - \bar{I}^{CMHC}_{i,t-1}\right) - \left(\bar{I}^{CRA}_{i,t} - \bar{I}^{CRA}_{i,t-1}\right), \ \ \forall\ FSA_i. \tag{4}$$

In Online Appendix A, we discuss in greater detail the construction of PIM. Figure 3 illustrates the structural differences between the two, income and the misstatement, over time. We define PIM at FSA i as

$$PIM_{i,t} = \Delta\bar{I}^{CMHC}_{i,t} - \Delta\bar{I}^{CRA}_{i,t}. \tag{5}$$

The Δ operator represents a one-year change in reported income. PIM is the change in log of income ratios; thus, it is unitless. Figure 4 plots PIM across FSAs in each year. PIM varies across FSAs, but it was on average not sizeable from 2004 to 2014, meaning that overall there was no significant difference in income growth between CMHC and CRA. In Online Appendix C, we also include geographical distribution of PIM across Canada. Table 2 presents summary statistics for PIM in each year. Overall, average PIM is not very different from zero in our sample period. Table 3 documents more extensive descriptive FSA statistics for different PIM levels. By decile of PIM, Table 3 compares mean and standard deviation of credit score (2004–2014, CMHC data), unemployment rate (2011, NHS data), share of home ownership (2011, NHS data), share of self employed (2004–2014, CMHC data), share of households in bottom half of the income distribution (2011, NHS data), share of immigrants (2011, NHS data) and income volatility (2004–2014, CMHC data).[21] The average PIM in the first and the 10th decile are −32.66 and 30.13, respectively. Except credit score, there is no clear trend between decile of PIM- and FSA-level characteristics. The PIM distribution in each year is what we are interested in explaining. In the next section, we test whether variation in PIM across FSAs can be explained by house price variation.

[21] Income volatility is defined as standard deviation divided by mean (the coefficient of variation), in each FSA.

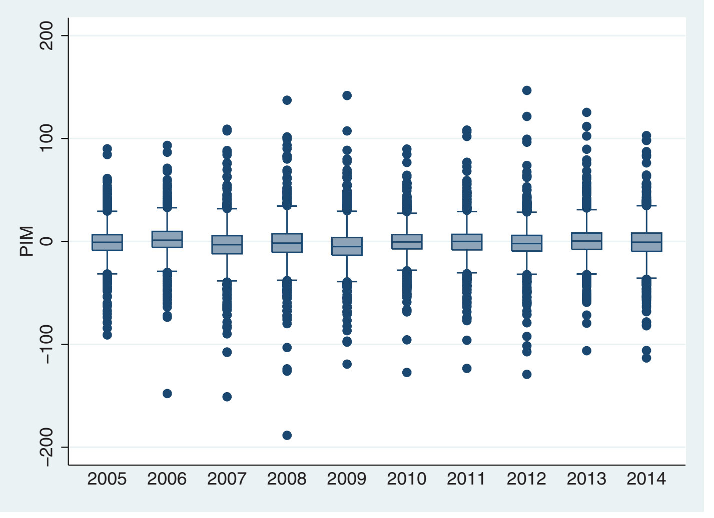

FIGURE 4 Variation of PIM across FSAs in each year [Color figure can be viewed at wileyonlinelibrary.com]
Note: The box plot indicates the lower and upper quartiles of PIM in each year. The median is represented by a line subdividing the box.

TABLE 2 Summary statistics on PIM by year

Year	Mean	*SD*	*P*(25)	*P*(50)	*P*(75)
2005	−1.12	15.8	−8.7	−.846	6.57
2006	1.77	16.2	−5.84	1.19	9.68
2007	−3.19	18.4	−12	−3.14	5.6
2008	−1.85	20.2	−10.6	−1.64	7.48
2009	−4.28	18.6	−13.5	−5	3.78
2010	−0.323	15.6	−7.25	−0.504	6.65
2011	−0.75	16.8	−8.21	−0.0514	6.85
2012	−1.68	17.6	−9.31	−2.06	5.85
2013	0.273	17	−7.78	0.408	8.16
2014	−0.682	18.5	−9.71	−0.759	8.1

Note: This table presents descriptive statistics for possible income misstatement across FSAs from 2005 to 2014.

4 | EMPIRICAL SPECIFICATIONS AND RESULTS

The results in this section are divided into two subsections. Section 4.1 documents the relationship between PIM and house prices across FSAs from 2004 to 2014. We find that, controlling for other factors, PIM is higher in high-priced markets. Section 4.2 uses arrears rates and mortgage insurance claims rates to demonstrate that, as PIM increases, loan performance deteriorates. Taken together, these results suggest that high-priced markets appear more likely to create incentives for borrowers to misstate their incomes either to enter the housing market or to obtain larger loans. This creates a channel for more defaults due to possible cash flow problems.

TABLE 3 Summary statistics for different PIM levels

Deciles	PIM Mean	PIM SD	Unemployment Mean	Unemployment SD	Home ownership Mean	Home ownership SD	Selfc-employed Mean	Selfc-employed SD
1st	−32.7	16.8	8.39	4.66	0.72	0.20	0.01	0.06
2nd	−14.7	1.99	7.98	3.43	0.70	0.19	0.02	0.06
3rd	−9.37	1.24	7.95	3.47	0.70	0.18	0.01	0.03
4th	−5.69	0.94	7.9	3.2	0.70	0.17	0.01	0.04
5th	−2.6	0.85	7.82	3.09	0.71	0.18	0.01	0.03
6th	0.33	0.86	7.89	3.23	0.70	0.18	0.01	0.03
7th	3.32	0.91	8.01	3.56	0.70	0.17	0.01	0.04
8th	7.02	1.28	8.09	3.65	0.72	0.18	0.01	0.04
9th	12.4	2.02	8.27	3.81	0.71	0.18	0.01	0.04
10th	30.1	16.7	8.53	4.8	0.72	0.2	0.02	0.08
Deciles	**Credit score Mean**	**Credit score SD**	**Immigrants Mean**	**Immigrants SD**	**Low income Mean**	**Low income SD**	**Income volatility Mean**	**Income volatility SD**
1st	718	48.1	0.17	0.16	0.49	0.16	16.8	8.65
2nd	718	30.8	0.18	0.17	0.51	0.14	10.9	5.68
3rd	718	29.5	0.17	0.16	0.50	0.14	9.72	4.67
4th	717	28.3	0.16	0.16	0.50	0.13	9.16	4.75
5th	717	25.8	0.17	0.16	0.50	0.13	8.7	4.08
6th	718	26.1	0.18	0.17	0.51	0.13	8.98	4.49
7th	717	26.7	0.18	0.17	0.51	0.13	9.04	4.13
8th	716	32	0.17	0.16	0.50	0.14	9.89	4.93
9th	714	35.5	0.16	0.16	0.51	0.14	11.1	5.4
10th	711	55.3	0.17	0.16	0.50	0.16	16.8	8.91

Note: This table presents descriptive statistics at FSA level for decile of PIM from 2004 to 2014.

4.1 | Do house prices impact PIM?

We start by analyzing the relationship between house prices and PIM. There are two empirical challenges associated with estimating a relation between house prices and PIM. First, if we consider that the supply of housing is inelastic in the short term, then increases in income misstatement potentially translate to higher house prices. This simultaneous relationship between house prices and PIM creates a potential endogeneity problem and biases an ordinary least squares (OLS) framework. Second, there may be an unobservable factor that explains increases in both PIM and house prices. We employ two-stage least squares (2SLS) and fixed effects to correct for the potential endogeneity problem of reverse causality and omitted variable bias.

For our 2SLS model, we need an instrument that is highly related to the house prices and unrelated to error term. We use the average rent price in each FSA as an instrument for house prices. The average rent prices in each FSA is positively correlated with increases in house prices, which satisfies the relevance condition of an instrumental variable. The simple correlation between our instrument (rent prices) and endogenous variable (house prices) is about 0.69 in absolute value, which is not an indication of a weak instrument. A large first-stage F-statistic also indicates that our instrument is not weak. In addition, this instrument should not have a direct causal effect on the outcome. The average rent price in each

FSA is a function of supply and demand for the rental market and is unlikely to be correlated with PIM, which satisfies the exclusion condition of an instrumental variable.[22]

We construct the instrumental variable as follows. We use the RMS, which is a survey the CMHC conducts every year to estimate the relative strengths in the rental market. The survey collects market rent levels, turnover and vacancy unit data for all sampled structures. We compute the average rent prices of all available FSAs for each year in our sample and use that as our instrument measure. Because rent information is not available for all FSAs, we lose some observations by matching the rent data to our main sample.

In addition to house prices, loan origination channel, credit score, neighborhood income level and employment status are expected to influence income misstatement. Jiang et al. (2014) point out that mortgage brokers tend to originate lower quality loans, which result in borrower information falsification, known as "liar's loans." We generate a variable that is a proxy for whether a loan was generated through a broker.[23]

Using U.S. data, Mian and Sufi (2017) show that PIM is highest in low credit score, low-income ZIP codes. We exploited this idea using the median income from the 2011 NHS and average credit score from Equifax for each FSA. FSAs are sorted according to the 2014 credit score data and 2011 median income from census data and remained in the same group throughout the sample.

Focusing on no-doc loans, Ambrose et al. (2016) show how borrower heterogeneity with respect to employment status contributed to income misrepresentation. They argue that documenting income and assets and conducting verification activities are more costly for both borrowers and lenders when a borrower is self-employed. The complexity of the process provides an avenue for some borrowers to inflate or exaggerate their income in order to qualify for larger mortgages. We exploit this idea by using borrowers' employment status as reported to the CMHC. The key difference with Ambrose et al. (2016) is our study focuses on full-doc loans. We generate a dummy variable that changes from zero to one for self-employed borrowers. We then calculate the share of self-employed borrowers in each FSA. We run the following level-log regression:

$$PIM_{i,t} = \alpha + \beta \, log\,(HP_{i,t}) + \gamma X_{i,t} + YEAR_t + PROV_i + YEAR_t \times PROV_i + \epsilon_{i,t}, \tag{6}$$

where i is FSA and t is the year, $PIM_{i,t}$ is the difference in growth in income according to mortgage applications and growth in income according to tax files in each FSA year, as explained in Section 3.3, $log\,(HP_{i,t})$ is the log of house prices reported on the mortgage applications. The vector $X_{i,t}$ includes explanatory variables, such as credit score, LTV ratio, TDS ratio, GDS ratio, median income, share of self-employed borrowers and share of loans generated through brokers.

Our regression set-up controls for possible systematic variation of PIM over time and across provinces. The time fixed effects in our model are in terms of year dummies ($YEAR_t$). We also include province fixed effects ($PROV_i$) to control for any variation in economic factors, such as socioeconomic backgrounds, labor market differences and credit supply differences at the province level. The interaction between time and province fixed effects controls for any province-wide macroeconomic

[22] We also used the average mortgage interest rate in each FSA as an instrument for house prices. Results are available in Online Appendix B. The results are robust to the choice of instrument.

[23] Our proxy for broker-generated mortgages is equal to one when a mortgage is originated through a Mortgage Finance Company (MFC) and zero otherwise. MFCs are nondepository financial institutions that underwrite and administer mortgages through noncorresponding brokers.

TABLE 4 Explaining possible income misstatement (PIM; 2004–2014)

			2SLS		
Dependent variable: PIM	**OLS**	**F.E.**	**First stage**	**Second stage**	**Second stage**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
log(HP)	21.48***	25.45***			
	(0.94)	(1.02)			
Rent prices			0.001***		
			(0.00)		
$\widehat{log(HP)}$				10.12***	12.15***
				(1.37)	(1.75)
TDS	0.03	0.18	−0.00	0.03	0.15
	(0.12)	(0.13)	(0.00)	(0.16)	(0.17)
GDS	−2.18***	−2.65***	0.05***	−1.42***	−1.80***
	(0.11)	(0.12)	(0.00)	(0.12)	(0.14)
LTV	0.77***	1.09***	−0.07***	0.01	0.04
	(0.22)	(0.25)	(0.00)	(0.25)	(0.31)
Broker	−7.21***	0.02	0.13***	−1.20	1.70
	(2.09)	(2.62)	(0.03)	(2.34)	(2.65)
Self-employed	−8.06	−3.65	0.67***	−0.77	7.81
	(6.08)	(8.03)	(0.11)	(6.02)	(7.58)
Credit score	−0.02**	−0.03**	−0.00	−0.01*	−0.03***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
log (median income)	−7.75***	−7.85***	0.17***	−3.73***	−3.69***
	(0.71)	(0.82)	(0.02)	(0.58)	(0.72)
Constant	−181.09***	−246.96***	14.98***	−38.86	−49.85
	(25.27)	(29.44)	(0.41)	(30.08)	(40.93)
Observations	10,009	10,009	10,991	10,009	10,009
Adjusted R^2/Pseudo R^2	0.13	0.18	0.78	0.10	0.15
First-stage F-statistic			1,017.03		
Prov F.E.		✓	✓		✓
Year F.E.		✓	✓		✓
Prov × Year F.E.		✓	✓		✓

Note: This table presents coefficient estimates from an OLS, fixed effect and instrumental variable regressions of PIM on log of house prices reported on the mortgage applications, total debt service ratio, gross debt service ratio, loan-to-value ratio, credit score, share of self-employed borrowers, share of loans generated through brokers and log of median income at FSA level. Average rent prices are used as an instrumental variable. Heteroskedasticity-robust standard errors adjusted for clustering at the FSA level are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

factors, including unemployment rates. Identification of β after introducing province fixed effects relies on province-specific time variation relative to the average level in a particular year. To address heteroskedasticity in $\epsilon_{i,t}$, we use cluster-robust standard errors at the FSA level.

The results are reported in Table 4. Column 1 reports results of OLS regression with borrower and loan characteristics variables. Column 2 reports results using the fixed effect model. Adding province fixed effects and year fixed effects and the interaction between the two does not affect the statistical significance of the result. Columns 3–5 use average rent prices in each FSA as an instrument for house prices. Column 3 reports the first-stage regression results and columns 4 and 5 the second-stage regression results without and with fixed effects, respectively. In all three specifications, we find that house prices are positively associated with PIM and that the magnitude of the effect is statistically significant at the 1% level. This suggests that borrowers in high-priced markets are more likely to misstate their income.

In the fixed effect specification, a 1% in $HP_{i,t}$ corresponds to a 0.25% increase in PIM. In the IV regression, the estimated coefficient on $HP_{i,t}$ is smaller, implying that a 1% increase in house prices is associated with a 0.12% increase in PIM. This is consistent with the notion of upward bias in the OLS regression due to reverse causality. The IV estimator would correct this and deliver a smaller coefficient. As expected, the coefficient of the credit score is negative and statistically significant at the 1% level. This is consistent with earlier findings that borrowers with lower credit score are more likely to misstate their income. We do not find any evidence that the broker channel plays a substantial role in variation in PIM.

We have analyzed the robustness of the results to the choice of house price measures. Instead of price level, we have used level of price to income ratio, the change in log price, lagged change in log price and the change in price to income ratio. Next, we used more fixed effects at the FSA level. Results are available in Online Appendix B. The main results are robust to all these checks.

4.2 | Does PIM deteriorate loan performance?

We now turn to documenting whether PIM is correlated with loan performance. We use arrears and claims as indicators of loan performance, as discussed in Section 3.2.

We define $D_{i,t,t',claims}$ as a variable that measures the number of mortgages that originated at year t and went to claims at year $t' \geq t$ in FSA i. Similarly, $D_{i,t,t',arrears}$ measures the number of mortgages that originated at year t and went to arrears at year $t' \geq t$ in FSA i. Five-year claims and arrears rates at FSA i, year t are defined as

$$Y_{i,t,j\in\{claims,\ arrears\}} = \frac{\sum_{t'=t}^{t+5} D_{i,t,t',j\in\{claims,\ arrears\}}}{total\ \ mortage\ \ originations_{i,t}} \times 100. \tag{7}$$

We start by asking whether FSAs with higher PIM experience worsen loan performance. We use the following regression, where the loan performance can be represented as

$$\begin{aligned} Y_{i,t,j\in\{claims,\ arrears\}} &= \alpha + \beta PIM_{i,t} + \gamma X_{i,t} + YEAR_t \\ &\quad + PROV_i + YEAR_t \times PROV_i + \epsilon_{i,t}, \end{aligned} \tag{8}$$

where $X_{i,t}$ is a vector of control variables. $YEAR_t$ and $PROV_i$ are year and province fixed effects, respectively. The set of control variables include change in log house prices between t and $t+5$, credit score, mortgage interest rate and LTV ratio. The interaction between province and year fixed effects allow us to control for any province-level time variations. We cluster standard errors at the FSA level. Our results are displayed in Table 5. The impact of PIM on arrears is reported in columns 1–3. Columns 4–6 repeat the estimate for claims. The columns differ in the PIM measures used as well as the control variables. As a robustness check, we also use "early default" as an indicator for poor loan performance. Early default is defined as loans that have gone to arrears within the first two years after origination.

TABLE 5 Relationship between possible income misstatement (PIM) and loan performance at FSA level (2004–2009)

	Arrears			Claims			Early Dedfaults		
	OLS	F.E.	F.E.	OLS	F.E.	F.E.	OLS	F.E.	F.E.
Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PIM	0.02**	0.01		0.00	−0.00		0.01*	0.0	
	(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)	
$\widehat{PIM}_{i,t}$			0.12**			0.04			0.10**
			(0.05)			(0.04)			(0.10)
Δ log(price)		1.82***	1.85***		0.92**	0.96**		2.38***	2.42***
		(0.54)	(0.54)		(0.39)	(0.39)		(0.38)	(0.38)
Credit score		−0.02***	−0.02***		−0.02***	−0.01***		−0.02***	−0.01***
		(0.01)	(0.01)		(0.00)	(0.00)		(0.00)	(0.01)
Interest rate		0.05	0.11		−0.03	0.02		−0.02	0.05
		(0.59)	(0.60)		(0.51)	(0.52)		(0.50)	(0.51)
LTV		0.72***	0.74***		0.45***	0.46***		0.38***	0.40***
		(0.15)	(0.15)		(0.13)	(0.12)		(0.14)	(0.13)
Constant	6.03***	−71.08***	−76.53***	3.54***	−42.25***	−44.57***	3.45***	−52.27***	−57.09***
	(0.13)	(18.11)	(17.12)	(0.09)	(14.19)	(13.21)	(0.10)	(15.04)	(13.86)
Observations	7,561	7,411	7,411	7,561	7,411	7,411	7,561	7,411	7,411
Adjusted R^2	0.00	0.11	0.11	0.00	0.12	0.12	0.00	0.09	0.10
Prov F.E.		✓	✓		✓	✓		✓	✓
Year F.E.		✓	✓		✓	✓		✓	✓
Prov × Year F.E.		✓	✓		✓	✓		✓	✓

Note: This table presents coefficient estimates from an OLS and fixed effect regressions of arrears, claims and early defaults on PIM. The predicted PIM is derived from a 2SLS model of PIM on log of house prices reported on mortgage applications, total debt service ratio, gross debt service ratio, loan-to-value ratio, credit score, share of self-employed borrowers, share of loans generated through brokers and log of median income at FSA level. The Δ log(price) is the change in log house prices between t and $t + 5$. The sample period is from 2004 to 2009. Heteroskedasticity robust standard errors adjusted for clustering at the FSA level are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

The $\widehat{PIM_{i,t}}$ is the predicted value of $PIM_{i,t}$ from the 2SLS regression (6). The coefficient on the $\widehat{PIM_{i,t}}$ variable suggests that a one point increase in predicted PIM is associated with a 0.12 point increase in the rate of arrears. These results are consistent with the notion that borrowers who misstate their income on mortgage applications are, on the average, more likely to default on their mortgages.

The PIM's impact through misrepresentation are observed after house price changes are controlled for. The impact is more pronounced in arrears, but no so much for claims. This can be due to the fact that once a loan is in arrears, CMHC provides lenders with a variety of policies to help lenders effectively manage loan default situations. These policies are called "default management tools"[24] and they are designed to reduce the number of defaulted loans that would eventually become insurance claims.

To further support our previous findings, we performed our analysis at the loan level. We compared *ex post* arrears and claims conditional on borrower and loan characteristics observable at loan origination as well as area characteristics. Online Appendix B describes these results. The results indicate that the risk that a borrower will default is not just a function of the borrower's characteristics, the loan terms and economic trends, but that it also depends significantly on the neighborhood in which the borrower lives.

5 | CONCLUSION

Empirical analysis of PIM on mortgage applications in Canada has been relatively rare, which limits evidence-based policy recommendations to mitigate mortgage fraud risks. The evidence presented here shows that there was not a significant amount of PIM for insured mortgages in the Canadian mortgage market from 2004 to 2014. The most important feature of the results presented here is the unique nature of the data used to construct them. The sources of data consist of the entire historical loan portfolio of the largest mortgage insurance company in Canada covering a time period from 2004 to 2014, as well as income tax data collected by the LAD, aggregated at the FSA level.

Our tests provide support for the hypothesis that part of the observed dispersion in PIM measure is caused by house price variation. In particular, we find that borrowers in high-priced markets have more incentive to misstate their income. We also demonstrate the importance of PIM on loan performance. A number of studies have documented the link between income misstatement and loan performance (see, *e.g.*, Blackburn & Vermilyea, 2012; Mian and Sufi, 2017). In line with those studies, we demonstrate that higher predicted income misstatement coincides with higher rates of arrears.

Our findings have important implications for regulators designing mortgage market policies. First, mortgage lenders have traditionally relied on credit scores to assess the risk of a borrower. However, when a borrower or broker misrepresents fundamental characteristics such as income, employment, debt or the value of property, the credit score risk assessment is not as effective. Moreover, documenting and verifying income is a costly process for both borrowers and lenders, especially when the borrower is self-employed (see Ambrose et al., 2016). The complexity of the process provides an avenue for some borrowers to inflate or exaggerate their income in order to qualify for larger mortgages. Enhancing the income verification process by, for example, adopting a more sophisticated channel through which lenders must verify a borrower's income directly with the tax authority (CRA) as part of the underwriting process would improve the efficiency of the process and help reduce delinquency and

[24]These default management tools include conversion of variable rate mortgage payment to fixed rate mortgage payment, use of the existing equity in the mortgage to manage missed payments, mortgage payment deferrals, extension of amortization period and conversion of mortgage to interest only payment.

improve risk assessment models. Second, income reported on mortgage applications, especially in high-priced markets, should be considered with caution, as a borrower's income could be misstated.

Better understanding of the mechanisms of PIM is an important field of future research. Understanding whether PIM is driven by borrowers or mortgage originators would require additional loan-level information. Extending our framework beyond 2014, when the Canadian housing market experienced an upward shift in house prices, especially in Toronto and Vancouver, would also improve our understanding of the impact of house prices on possible income misstatement.

ACKNOWLEDGMENTS

We thank Statistics Canada for providing us with the data. We greatly benefited from thoughtful comments of our anonymous referee. We are also grateful for the valuable comments and suggestions of Bert Pereboom, Guillaume Neault, Brennan Thompson, Robert Clark, Decio Coviello, Murtaza Haider, Jaclene Begley and the participants at the 2018 Applied Econometrics Workshop at Ryerson University, the 2017 CEA annual conference, the 2018 FMA annual meeting and the 2018 AREUEA National Conference. The views in this article are those of the authors and do not necessarily reflect those of CMHC. All errors are our own.

ORCID

Babak Mahmoudi https://orcid.org/0000-0003-2094-0801

REFERENCES

Allen, J., Clark, R., & Houde, J.-F. (2014). Price dispersion in mortgage markets. *The Journal of Industrial Economics*, *62*(3), 377–416.

Ambrose, B. W., Conklin, J., & Yoshida, J. (2016). Credit rationing, income exaggeration, and adverse selection in the mortgage market. *The Journal of Finance*, *71*(6), 2637–2686.

Avery, R. B., Bhutta, N., Brevoort, K. P., & Canner, G. B. (2012). The mortgage market in 2010: highlights from the data reported under the Home Mortgage Disclosure Act. *Federal Reserve Bulletin*, *98*(6), 1–46. https://www.federalreserve.gov/pubS/bulletin/2011/pdf/2010_HMDA_final.pdf

Ben-David, I. (2011). Financial constraints and inflated home prices during the real estate boom. *American Economic Journal: Applied Economics*, *3*(3), 55–87.

Bhutta, N., Dokko, J., & Shan, H. (2017). Consumer ruthlessness and mortgage default during the 2007 to 2009 housing bust. *The Journal of Finance*, *72*(6), 2433–2466.

Blackburn, M. L., & Vermilyea, T. (2012). The prevalence and impact of misstated incomes on mortgage loan applications. *Journal of Housing Economics*, *21*(2), 151–168.

Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: Evidence from US states. *The Review of Financial Studies*, *24*(9), 3139–3186.

Griffin, J. M., & Maturana, G. (2016). Who facilitated misreporting in securitized loans? *Review of Financial Studies*, *29*(2), 384–419.

Jiang, W., Nelson, A. A., & Vytlacil, E. (2014). Lian's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics*, *96*(1), 1–18.

Mian, A., & Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics*, *124*(4), 1449–1496.

Mian, A., & Sufi, A. (2017). Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. *The Review of Financial Studies*, *30*(6), 1832–1864.

Piskorski, T., Seru, A., & Witkin, J. (2015). Asset quality misrepresentation by financial intermediaries: Evidence from the RMBS market. *The Journal of Finance*, *70*(6), 2635–2678.

Steele, M. (2007). The Canadian home buyers' plan: Tax benefit, tax expenditure, and policy assessment. *CANADIAN TAX JOURNAL*, *55*(1), 1.

Yavas, A., & Zhu, S. (2020). Misreporting of second liens in portfolio mortgages an privately secularized mortgages. Working Paper.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Basiri K, Mahmoudi B. Possible income misstatement on mortgage loan applications: Evidence from the Canadian housing market. *Real Estate Econ*. 2020;1–19. https://doi.org/10.1111/reec.12310