Detecting Temporal Anomalies with Pseudo Age Groups: Homeownership in Canada, 1981 to 2016

Yue Yuan, PhD student
School of Public Policy and Administration, Xi`an Jiaotong University,
Xi`an, Shaanxi, 710049, China
Email: yuanyuepink@hotmail.com

Jiaxin Gu, PhD student
Department of Sociology, The University of British Columbia,
Vancouver, BC, V6T 1Z1, Canada
Email: gujiaxinsoci@gmail.com

Dr. Xin Guo, Assistant Professor
Department of Applied Mathematics, The Hong Kong Polytechnic University,
Hong Kong, China
and Senior Lecturer
School of Mathematics and Physics, The University of Queensland,
Brisbane, QLD, 4072, Australia
Email: xin.guo@uq.edu.au

Dr. Yushu Zhu, Assistant Professor
Urban Studies Program and School of Public Policy, Simon Fraser University,
Vancouver, BC, V6B 5K3, Canada
Email: yushu_zhu@sfu.ca

Dr. Qiang Fu, Associate Professor
Department of Sociology, The University of British Columbia,
Vancouver, BC, V6T 1Z1, Canada
Email: qiang.fu@ubc.ca

Data availability statement: The data that support the findings of this study are openly available in Statistics Canada at https://www12.statcan.gc.ca/census-recensement/pc-eng.cfm.

Acknowledgement: Jiaxin Gu and Yue Yuan contributed equally to this work and share first authorship. The authors gratefully acknowledge financial support from an Insight Grant from the Social Sciences Humanities Research Council of Canada (#435-2021-0720, **PI**: Qiang Fu). The work described in this paper is also supported partially by the Research Grants Council of Hong Kong [Project No. PolyU 15304917]. Please direct all correspondence to Qiang Fu, Department of Sociology, The University of British Columbia, Vancouver, BC, V6T 1Z1, Canada, Email: qiang.fu@ubc.ca.

Abstract

Methodological advances in demographic research, especially age-period-cohort (APC) analysis, primarily focus on developing new models, yet often fail to consider practical concerns in empirical analysis. We propose a mixed approach that integrates multiple data imputation and structural change analysis in time series so that scholars can 1) construct pseudo age groups based on more coarsely grouped age data; and 2) identify temporal anomalies. This approach is illustrated using multiple waves of Canadian Population Census data (1981-2016). We construct pseudo age groups based on more coarse age information available in the Census data, and identify a local anomaly in the temporal trajectory of homeownership in Canada's less populous provinces and territories. These findings are assessed and validated in comparison with results from more populous Canadian provinces. This research broadens the methodological repertoire for demographers, geographers, and social scientists in general, and extends the literature on homeownership in an understudied area.

Keywords: age-period-cohort analysis, structural change analysis, multiple imputation, homeownership, pseudo data

1. Introduction

Demographers, geographers, and other social scientists have long used age-period-cohort (APC) analysis as a major tool to study the contribution of age, period, and cohort effects on social events, behaviors, and attitudes. The age effect assesses the variation brought about by the chronological aging process and the corresponding changes of social status and roles; the period effect represents the changes due to the social forces that equally affect individuals across all age groups and birth cohorts; and the cohort effect occurs when individuals who share membership in the group (such as those born in the same year) collectively experience social and historical changes. For example, Reither, Hauser, and Yang (2009) conduct APC analysis to study subjective well-being across the life course in the United States using the General Social Survey. Employing the hierarchical age-period-cohort (HAPC) logistic model, Fu (2016) assesses the temporal patterns of homeownership in China amid housing reforms. In a more recent study, Gu et al. (2020) separate the age, period, and cohort effects of adolescent marijuana use in the United States from 1991 to 2018.

As Mason et al. (1973) point out, since the 1950s the identification issue of separating the age, period, and cohort effects due to the perfect collinearity of the three temporal indicators ($age = survey\ year - birth\ cohort$) has become a major methodological challenge in various disciplines, including demography, epidemiology, and sociology. In recent years there has been a rapid development of models and methods for APC analysis, which features the introduction of intrinsic estimators, cross-classified models, and entropy-based approaches to deal with the identification problem in APC analysis (Fosse & Winship, 2019; Land et al., 2016; Reither et al., 2015). Yet methodological advances in this field primarily focus on the development of new models and fail to consider two practical concerns in empirical analysis.

First, the grouping or categorization of data retrieved from official statistics does not always align with the needs of empirical analysis. Temporal data retrieved from official census statistics are often in more coarse age groups, which makes it difficult, and sometimes impossible, to specify one's exact birth cohort. In cases where the interval of survey periods is not in accordance with the interval of age groups, birth cohorts of respondents interviewed in successive survey periods overlap with each other. For example, consider the specification of the birth cohorts of respondents from two successive surveys conducted in 2006 and 2011. If only the 10-year age intervals (e.g., 10–19 years old, 20–29 years old, 30–39 years old) of survey respondents are available, those who were 30–39 years old in the 2006 survey correspond to the birth cohorts of 1967–1976, while those who were 30–39 years old in the 2011 survey correspond to the birth cohorts of 1972–1981. Given that the birth cohorts of the same 10-year age group in the two surveys overlap, we cannot divide the repeated cross-sectional data into mutually exclusive birth cohorts for subsequent age-period-cohort analysis. In other words, neither a sequence of 1937–1946, 1947–1956, 1967–1976, 1977–1986, and so forth, nor a sequence of 1932–1941, 1942–1951, 1952–1961, 1962–1971, 1972–1981, 1982–1991, and so forth, is appropriate to categorize the birth cohorts of all respondents in the two successive surveys.

Second, existing methods for APC analysis only test whether a specific temporal effect is significantly different from zero or the average age, period, or cohort effect. Such hypothesis testing cannot detect local anomalies. For example, although the coefficient and standard error associated with a particular time period suggest whether the effect of this period is statistically different from zero, we cannot tell if it is different from the effects of neighboring time periods. However, practitioners and policymakers who rely heavily on census or survey data in their decision making are often concerned about these local anomalies so that they can track the immediate impacts of policy or social changes.

By incorporating data imputation and structural-change analysis into APC analysis, we propose a mixed approach for handling these two practical concerns, namely, the construction of pseudo age groups and the detection of temporal anomalies. It should be noted that one specific study may correspond to either one or both of these two practical concerns. If both are present in a temporal analysis, the construction of pseudo age groups should clearly precede the detection of temporal anomalies. In particular, the multiple data imputation addresses the first concern in which coarse age groups prevent the specification of mutually exclusive birth cohorts. We construct mutually exclusive birth cohorts based on pseudo five-year age intervals created by multiple data imputation. Structural-change analysis is carried out to detect local anomalies: it treats period effects obtained from APC analysis as a time series and detects whether and where structural changes may emerge. We next illustrate this mixed approach using hierarchical (cross-classified) models for APC analysis (Fu & Land, 2017; Yang & Land, 2006). It is worth noting that our method is flexible and can be applied to other models for temporal analysis with coarse age or cohort groups.

2. Homeownership in Canada

Scholars have long investigated how inequalities in homeownership in Canada are explained by individual or household-level factors such as immigrant status (Constant, Roberts, & Zimmermann, 2009; Edmonston, 2005), race and ethnicity (Haan, 2007), sexual orientation (Dilmaghani & Dean, 2020), socioeconomic status (Harris, 1984), and household income (Haan, 2008). However, these studies often treat the nation as a homogenous entity. As Zhu, Fu, and Ren (2014) point out, scholars cannot gain a holistic understanding of inequalities in homeownership unless spatial disparities are seriously considered. Likewise, Chauvet (2016) suggests that regional differences resulting from sample selection bias, demographic composition, nonresponse rates, and calibration strategies may also shape our understanding of housing issues.

Although all levels of government in Canada are involved in some aspect of making housing policy, constitutional jurisdiction and regulations of housing in Canada ultimately fall to the provincial and municipal governments (Dalton, 2009; Hulchanski, 2003). This governance structure means that provincial-level housing policies and regional conditions may shape housing inequalities in Canada (Hulchanski, 2003). To date, a large body of research on housing tenure has paid particular attention to the influx of immigrants, housing affordability, and racial/ethnic disparities in Canada's metropolitan areas (Grigoryeva & Lay, 2019; Haan, 2008; Li, 1998; Moore & Skaburskis, 2004; Skaburskis, 1996). Nevertheless, the existing literature overlooks the less populous provinces and territories (hereafter simplified to "less populous provinces") in Eastern Canada, the Prairies, and Northern Canada. Overall, these less populous provinces make up about 64.3% of Canada's total geographical area, yet only contribute about 25.3% to Canada's population. In contrast, the remaining three provinces (Ontario, Quebec, and British Columbia), where Canada's major global cities of Toronto, Montreal, and Vancouver are located, are home to almost three-quarters of Canadians. Important questions remain as to whether homeownership in the less populous provinces shares a temporal pattern with homeownership in the well-studied, more populous provinces.

To date, there appear to be only two empirical studies on APC analysis of homeownership and housing needs in Canada. Using eight waves of Canadian Census data from 1971 to 2006, Hou (2012) finds that the homeownership rate among Canadians generally increases with age and reaches its peak among the elderly. Specifically, the homeownership rate increases drastically among adults younger than 40. The increase slows down, levels off around age 75, and eventually declines afterwards. This study also investigated cohort differences in homeownership and observed an increase across birth cohorts since the 1970s. In a related study, Li and Shan (2020) also use Canadian Census data from 2001 to 2016 to examine temporal patterns of housing needs by age groups, time periods, and birth cohorts. Their results suggest that young and senior Canadians have higher housing needs, while middle-aged individuals have lower housing needs. Meanwhile, the study does not identify salient period or cohort patterns of housing needs.

Over the last few decades, the impact of shifting housing policies on the period effects of homeownership in Canada warrants special attention. The neoliberal housing policies implemented by the federal government during fiscal austerity have greatly facilitated the financialization of housing in Canada since the late 1990s (Kalman-Lamb, 2017; Walks, 2013). The supply of social housing in Canada has been sharply reduced and an increasing number of low-income Canadians have to rent private housing (Kalman-Lamb, 2017). Meanwhile, since more affordable rental housing units have been gradually replaced in the gentrification process across Canadian cities, many middle-income tenants have had to rely on the mortgage market to achieve homeownership (Walks, Hawes & Simone, 2021). In particular, the implementation of new government-insured mortgage programs such as the introduction of the Canada Mortgage Bonds in 2001 have greatly promoted housing financialization and housing investment (Kalman-Lamb, 2017; Walks & Clifford, 2015).

Given the focus of the existing literature on national-level analyses or more populous provinces in Canada, important questions remain about how the trajectories of homeownership in the less and more populous provinces coevolve over time, and how the neoliberal housing policy, together with social and economic changes in Canada, may lead to a fundamental shift in homeownership. In an effort to answer these questions and gain a better understanding of housing inequalities in Canada, we specifically investigate homeownership in the less populous provinces and compare their age, period, and cohort effects of homeownership with these in more populous provinces.

3. Data

Our study draws on datasets from Public Use of Microdata Files (PUMF), which consist of eight waves of the Canadian Population Census. The eight censuses were conducted every five years from 1981 to 2016. In each wave a dataset contained a 10% sample of a corresponding census. We selected respondents whose residency at the time of the censuses was in Eastern Canada (New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island); the Prairies (Alberta, Manitoba, and Saskatchewan); or Northern Canada (Northwest Territories, Nunavut, and Yukon). To assess our research findings based on these less populous provinces, we also applied the same analytical procedure to analyze homeownership in the more populous provinces (Ontario, Quebec, and British Columbia).

The dependent variable is a binary variable indicating whether some member of a household owns the dwelling at the time of a census, where owners are coded as 1 and renters are coded as 0. The three key variables in our APC analysis are age, birth cohort, and survey period. Although censuses have collected the exact age information of respondents, the census datasets only provide information about coarse age groups (five- or 10-year groups in different waves) instead of the exact ages in each calendar year. We used multiple imputation to construct (pseudo) age groups with five-year intervals, ranging from 20–24 years old to 75 years old and above. The method is discussed in detail below. Because the Canadian Census is conducted on a quinquennial basis, the resulting birth cohorts are five-year groups ranging from 1906 and before to 1992 and after. We combined individuals born in 1987–1991 with those born in 1992 and after, due to there being fewer observations in the most recent birth cohorts. We had 12 five-year age groups, eight five-year survey periods, and 18 five-year birth cohorts in total.

Similar to previous studies on housing tenure in the Canadian context (Balakrishnan & Wu, 1992; Moore & Skaburskis, 2004; Okkola & Brunelle, 2018; Skaburskis, 1996), our analysis draws on the sociodemographic characteristics of primary household maintainers. The definition of a primary household maintainer from the Canadian Census refers to the individual in a household who is responsible for shelter costs for mortgage payments, property taxes, strata fees, utility fees, and rent (Statistics Canada 2011). We included three binary variables to account for the influence of the sociodemographic backgrounds of the primary household maintainers. These are: male (versus female), married or in common-law relationships (versus unmarried, divorced, or other marital status), and high school diploma and below (versus college and above). In addition, we included a categorical variable denoting the three regions of Eastern Canada, the Prairies, and Northern Canada.

Table 1 provides a summary of the sociodemographic characteristics of the primary household maintainers in the less populous provinces. The total number of observations in our sample is 41,649, comprised of 4,985 in 1981, 5,407 in 1986, 5,183 in 1991, 5,136 in 1996, 5,341 in 2001, 5,105 in 2006, 5,125 in 2011, and 5,367 in 2016. The overall proportion of homeownership is 70%. Most primary household maintainers (68%) were male, and about 60% were married or in common-law relationships. More than half of the individuals had completed college. These three regions account for 31% (Eastern Canada), 67% (Prairies), and two percent (Northern Canada) of the sample. Specifically, individuals in earlier waves had lower levels of homeownership than those in more recent waves. The proportion of male primary household maintainers shows a downward trend throughout the eight waves. With respect to marital status, the proportion of these who were married or in common-law relationships gradually decreases from 1981 to 2006, but shows an upward trend in the 2011 and 2016 waves, which is possibly linked with changes in the census response categories for marital status. Since the 2011 Census, Statistics Canada has added a category of living in a common-law relationship to the response categories for marital status, but this category was not included prior to 2011 (Statistics Canada, 2011). Less-educated individuals constitute a smaller share of the sample in more recent waves, which is related to the gradual expansion of higher education in Canada (Gyourko & Linneman, 1996). The population distribution in these regions is relatively stable over the years of study, although the number of residents in the Prairies increased faster than the number of residents in the other two regions. Furthermore, we note that the distribution of five-year age groups across survey waves is in line with our expectation, which suggests the validity of constructing pseudo age groups: the shifting age structure reveals that the Canadian population in these regions has been getting older in recent years.

Table 2 shows homeownership rates across different sociodemographic groups and their temporal variations. For the overall sample, the level of homeownership sharply increases until individuals reach their 40s, and then increases more gradually until people reach their 60s. The homeownership rate levels off around 60 years old and then declines. The age pattern of homeownership across different waves is similar to that of the overall population. Those who were born between the 1920s and 1940s appear to achieve a higher level of homeownership. The peak level of homeownership is 78.9% for the 1927–31 cohorts, 80.3% for the 1932–36 cohorts, 81.2% for the 1937–41 cohorts, and 79.9% for the 1942–46 cohorts. Males' levels of homeownership are generally higher than females' in all waves, yet females' levels of homeownership show more fluctuation over the period of study. As expected, married individuals show higher levels of homeownership than their non-married counterparts; this conclusion remains robust across different waves. The educational disparities in homeownership vary from 1981 to 2016. Better-educated individuals show a lower level of homeownership from 1981 to 1991 but an opposite pattern is observed after the 1996 wave. The crossover in educational disparities possibly depends on the evolution of Canadian housing policy. From the 1950s to 1970s, different levels of Canadian governments played an active role in providing affordable housing to marginalized populations, including low-income households and Indigenous families (Carter, 1997). As a result, even less-educated primary household maintainers showed a high level of homeownership up until the 1980s. Since the late 1980s, the Canadian federal government has advocated housing privatization, retreated from social housing programs, and allowed provinces to cancel existing projects related to social housing (Walks, 2016). Due to this policy change, most Canadian households began to achieve homeownership through home purchases in the market. As expected, better-educated individuals with more financial resources tended to achieve homeownership in more recent years. For primary household maintainers from the more populous provinces (Ontario, Quebec, and British Columbia), their sociodemographic characteristics and homeownership rates across sociodemographic groups are shown in Appendix A and Appendix B.

[Table 1 and Table 2 about here]

4. Methods

4.1 Imputing key variable(s) using multiple imputation of missing data

In the PUMF data, the age information of primary household maintainers differs in various survey years. Age groups segmented by five-year intervals are not available in the 1991, 1996, and 2001 waves, while they are available in the 2006 and 2011 waves for individuals younger than 55. Instead, only age groups by 10-year intervals are available across all eight waves of the census data. As discussed above, these coarse 10-year age groups can produce overlapping birth cohorts when we assign individuals to their birth cohorts with the equation: $birth\ cohort = survey\ year - age$.

We adapted the multiple-imputation method to address this issue. As a general and flexible method widely adopted by scholars to impute missing values, multiple imputation has several important advantages over other alternative imputation methods (Berglund & Heeringa, 2014; Buuren, 2018; McCleary, 2002). First, multiple imputation is flexible and able to handle various missing scenarios. Second, this method processes all available data and maximizes the statistical power. Third, estimates from multiple imputation are unbiased and readily interpretable. Specifically, we first constructed pseudo age groups with five-year intervals. For primary household maintainers included in the 1991, 1996, and 2001 waves and for those older than 55 in the 2006 and 2011 waves, we treated the five-year age group information of these individuals as missing. For other individuals whose five-year age groups were available, we treated their age information as observed. We then conducted multiple missing data imputation on a pseudo five-year age group using the Fully Conditional Specification (FCS) in SAS (Berglund, 2015). The typical procedure for multiple imputation on missing data follows a three-step procedure (Berglund, 2015; Bodner, 2008; Graham, Olchowski, & Gilreath, 2007; White, Royston, & Wood, 2011): the imputation phase (phase 1) generates n copies of the datasets and fills in each copy with an independent set of estimates for the missing values of the target variable(s); the analysis phase (phase 2) analyzes each imputed dataset using standard procedures, and produces n sets of estimates and standard errors of regression coefficients; the pooling phase (phase 3) combines results from the n sets into one pooled set based on Rubin's rule (1987) for statistical inference.

To take advantage of information embedded in existing (more coarse) age groups and insights from multiple imputation, we adapt the imputation phase (phase 1) to construct pseudo age groups in APC analysis. When more coarse age groups are available, it should be noted that age information is not completely missing. In other words, a researcher knows that one's age falls into a more coarse group (say, 20–29 years old) but is uncertain about whether one's age falls into the first half (20–24 years old) or second half (25–29 years old) of that coarse age group. When multiple imputation consisting of three (imputation, analysis, and pooling) phases is applied to the overall dataset, we cannot guarantee that one's imputed age *only* falls into either of the two target five-year age groups. In fact, the resulting five-year age group from multiple imputation (e.g., 60–64 years old) based on the entire sample can be outside the known 10-year group of 20–29 years old. To address this issue, we divided the overall dataset into subsamples according to the available (more coarse) age groups, applied only the first imputation phase to each subsample, used binary variables to indicate target (more precise) age groups, and then constructed pseudo age groups of the overall dataset for APC analysis. Another advantage of constructing pseudo age groups is that researchers can further validate pseudo age groups by comparing their descriptive statistics with those in survey waves where the information about more precise age groups is available.

With regard to the six 10-year age groups (i.e., 15–24, 25–34, 35–44, 45–54, 55–64, and 65–74), we first generated six binary variables denoting whether a pseudo age group falls into the first half (e.g., 15–19) or second half (e.g., 20–24) of a 10-year interval (e.g., 15–24). For survey waves and individuals with information available for five-year age groups, we coded and treated these values as observed. For those without such information (i.e., individuals interviewed in the 1991, 1996, and 2001 waves, and those older than 55 in the 2006 and 2011 waves), we treated values of these six binary variables as missing in the imputation phase.

To perform the imputation phase for the six binary variables with missing values, we next divided the data into seven subsamples based on their 10-year age group (15–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75 and above). Except for the last age group (75 and above) in the APC analysis, we repeated the imputation procedure for the first six subsamples. To impute the missing values in these six binary variables of five-year age groups, we followed the general rule of including as many variables as possible (Rubin, 1996; Schafer & Graham, 2002) and used a total of 20 variables at the individual and household levels (census metropolitan area, province or territory of residence, number of children aged 0 to 17 in the household, number of persons aged 65 and over in the household, household size, household type, marital status, sex, place of birth, year of immigration, race/ethnic origin, language preference, educational attainment, mobility status, occupation, total household income, number of employment income recipients in the household, number of total income recipients in the household, the major source of household income, and income status with low-income cutoffs) in the census dataset as covariates for imputation. The imputation was implemented using SAS PROC MI (Berglund, 2015; Bodner, 2008; Buuren, 2018; Graham, Olchowski, & Gilreath, 2007; White, Royston, & Wood, 2011).

We illustrate the ideas of imputation methods under different scenarios as follows (Berglund, 2015; Bodner, 2008; Graham, Olchowski, & Gilreath, 2007; White, Royston, & Wood, 2011). For data with a monotone missing pattern, it is possible to reorder the variables in a way that if one observation is missing on variable Y_p, this variable is also missing on all remaining variables Y_q, where $q > p$. Under the assumption of a monotone missing pattern, we can use the following steps to impute missing values based on a multivariate model with its parameter θ and an (optional) completely observed covariate matrix X.

- Step 1: Draw a new regression model from the posterior distribution of parameters for imputing the missing values: $\dot{\theta}_1 \sim P(\theta_1 \mid Y_1^{obs}, X)$;
- Step 2: Impute the missing values for the first variable $\dot{Y}_1$ based on the new regression model: $\dot{Y}_1 \sim P(Y_1 \mid \dot{\theta}_1, X)$;
- Step 3: Update the missing part of the variable Y_1 with the imputed values: $Y_1 = (Y_1^{obs}, \dot{Y}_1)$.

Given the monotone missing pattern, we can repeat these steps for subsequent variables with missing values. The variable-by-variable procedure also means that these updated variables with missing values imputed are readily available as covariates for subsequent parameter inference. We have the following procedure to impute the missing values for the remaining variables.

- Step 1A: Draw a new regression model from the posterior distribution of parameters for imputing the missing values: $\dot{\theta}_p \sim P(\theta_p \mid Y_1, \cdots, Y_{p-1}, Y_p^{obs}, X)$;
- Step 2A: Impute the missing value for the variable $\dot{Y}_p$ based on the new regression model: $\dot{Y}_p \sim P(Y_p \mid \dot{\theta}_p, Y_1, \cdots, Y_{p-1}, X)$;
- Step 3A: Update the missing part of the variable Y_p with the imputed values: $Y_p = (Y_p^{obs}, \dot{Y}_p)$.

The inference for regression parameters (Step 1 and Step 1A) is a critical issue for implementing this algorithm. For each of the six binary variables with missing values, we fitted the following logistic regression based on Y_1^{obs} and Z (Buuren, 2018; Rubin, 1987):

$$\log\frac{g}{1-g}=Z\beta=\beta_0+\beta_1 Z_1+\cdots+\beta_j Z_j,$$

where $g=P(Y_1=1\mid Z_1,Z_2,\cdots,Z_j)$. We obtained parameter estimates $\hat{\beta}$ and its covariance matrix $\hat{\Sigma}$ from the fitted logistic model, and sampled parameters $\dot{\beta}$ from the posterior predictive distribution:

$$\dot{\beta}\sim N(\hat{\beta},\hat{\Sigma}).$$

We then predicted the latent probability of $Y_1=1$:

$$g_1=\frac{\exp(Z\dot{\beta})}{1+\exp(Z\dot{\beta})}.$$

Finally, we draw a random uniform variable $t\sim U(0,1)$ and impute $\dot{Y}_1^{(0)}$ as:

$$\dot{Y}_1^{(0)}=\begin{cases}0 & if\ t\geq g_1\\ 1 & if\ t<g_1\end{cases}.$$

In some imputation steps (e.g., Step 2A), we can put together the covariate matrix X as well as the previously imputed variables (e.g., $Y_1,\cdots,Y_{p-1}$) as the covariate matrix Z.

For data with arbitrary missing patterns, we update the original filled-in values with predicted values from each iteration until the convergence criterion is reached (Rubin, 1987). More specifically, we implement the following steps at iteration $s+1$:

Step 1.1: $\dot{\theta}_1^{(s+1)}\sim P(\theta_1\mid Y_1^{obs},Y_2^{(s)},\cdots,Y_p^{(s)},X)$;

Step 1.2: $\dot{Y}_1^{(s+1)}\sim P(Y_1\mid\dot{\theta}_1^{(s+1)},Y_2^{(s+1)},\cdots,Y_p^{(s+1)},X)$;

Step 1.3: $Y_1^{(s+1)}=(Y_1^{obs},\dot{Y}_1^{(s+1)})$;

Step 2.1: $\dot{\theta}_2^{(s+1)}\sim P(\theta_2\mid Y_1^{(s+1)},Y_2^{obs},Y_3^{(s)},\cdots,Y_p^{(s)},X)$;

Step 2.2: $\dot{Y}_2^{(s+1)} \sim P(Y_2 \mid \dot{\theta}_2^{(s+1)}, Y_1^{(s+1)}, Y_3^{(s+1)}, \cdots, Y_p^{(s+1)}, X)$;

Step 2.3: $Y_2^{(s+1)} = (Y_2^{obs}, \dot{Y}_2^{(s+1)})$;

…

Step p.1: $\dot{\theta}_p^{(s+1)} \sim P(\theta_p \mid Y_1^{(s+1)}, Y_2^{(s+1)}, Y_3^{(s+1)}, \cdots, Y_p^{obs}, X)$;

Step p.2: $\dot{Y}_p^{(s+1)} \sim P(Y_p \mid \dot{\theta}_p^{(s+1)}, Y_1^{(s+1)}, Y_2^{(s+1)}, \cdots, Y_{p-1}^{(s+1)}, X)$;

Step p.3: $Y_p^{(s+1)} = (Y_p^{obs}, \dot{Y}_p^{(s+1)})$.

The steps above are usually referred to as the Fully Conditional Specification method (Berglund, 2015; Buuren, 2018). Following the general rule that 3–5 rounds of imputations are enough to produce valid results (Rubin 1987; Schafer and Olsen 1998), we carried out the imputation algorithm five times and subsequently obtained five datasets for the six binary variables, denoting whether a respondent was in the first or second half of a 10-year age interval. Next, we manually selected all observations with a binary variable imputed as 1 in at least three out of the five datasets, and coded these as 1, and 0 otherwise. Again, observations coded as 1 belong to the first half (five-year age group) of the corresponding 10-year age group, while those coded as 0 belong to the second half (five-year age group) of the corresponding 10-year age group. After data imputation we generated five-year pseudo age groups with 12 categories (24 and below, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and 75 and above) based on observed and imputed missing values. Since the imputation phase corresponds to a stochastic process governed by parametric models, scholars may obtain slightly different pseudo age groups after these procedures, and it is recommended to compare the distribution of pseudo age groups with that of observed age groups where data are available. If the resulting pseudo age groups were less desirable, researchers could include more variables for imputing age variables or perform more rounds of imputations (Graham, Olchowski, & Gilreath, 2007; Schafer & Graham, 2002).

4.2 Hierarchical Age-Period-Cohort (HAPC) model

We employed hierarchical age-period-cohort (HAPC) logistic regression models to analyze homeownership in the rural and less populous areas in Canada. Using best linear unbiased predictors (BLUPs), HAPC models treat age as an individual-level fixed effect, while treating period and cohort effects as population-level, cross-classified random effects. The hierarchical modeling not only circumvents the collinearity problem of temporal effects, but also allows social scientists to consider individual-level variations along with population-level clusters such as survey periods and birth cohorts. All HAPC models were estimated using SAS PROC GLIMMIX (Littell et al., 2006).

We define the individual-level within-group model as:

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\beta_{0jk}+\sum_{p=1}^{11}\beta_p a_{pi}+\sum_{q=1}^{4}\beta_q X_q \ , \qquad (1)$$

where p_{ijk} denotes the probability of homeownership of the i^{th} individual born in the k^{th} cohort interviewed in the j^{th} survey year; a_p denotes all 11 age groups (with the 35–39 age group as the reference group); and X_q denotes all the control variables (sex, marital status, educational attainment, and region).

We define the population-level between-group model as:

$$\beta_{0jk}=\gamma_0+p_{0j}+c_{0k} \ , \qquad (2)$$

where γ_0 denotes the mean averaged over eight survey years and 18 birth cohorts when age and control variables at the individual level are held as zero; p_{0j} is the residual random effect of period j averaged over 18 birth cohorts; and c_{0k} is the residual random effect of cohort k averaged over eight survey years. We further calculated predicted probabilities based on estimates from the HAPC model described above using the equation $p = \frac{odds\ ratio}{1 + odds\ ratio}$. Covariates were held at their sample means when the predicted probabilities were calculated.

4.3 Estimating structure breaks in time series

To identify local anomalies in homeownership trends, we used structural change analysis proposed by Bai and Perron (1998) to detect potential structural breaks in time series. Different from previous methods that focus on a single structural change, Bai and Perron (1998) used a least-squares method to identify one or multiple structural changes. The Bai-Perron model tests the null hypothesis of having no structural change in time series against the alternative hypothesis of having at least one breakpoint. To determine the number of structural breaks, the method carries out a sequence of tests of n breaks against n+1 breaks for n=1, 2, 3, ..., until the null hypothesis can not be rejected. A dynamic programming algorithm that minimizes the sum of squared residuals of recursive least squares is used to determine the location of the breaks.

The following equation represents a linear regression with n breaks and the corresponding $n+1$ segments in the time series:

$$y_i = x_i^{\mathrm{T}} \beta_j + \varepsilon_i \quad (i = i_{j-1} + 1, \ldots, i_j,\ j = 1, \ldots, n+1) \ , \tag{3}$$

where y_i denotes the predicted probability of homeownership at time i, x_i and β_j denote covariates and their coefficients, j denotes the segment index, and ε_i is the residual at time i.

5. Results

The predicted age, period, and cohort effects of homeownership based on HPAC analysis, net of other effects, are shown in Figure 1 to Figure 3. For individuals living in the less populous provinces (portrayed by the solid lines), Figure 1 shows that the predicted probabilities of homeownership across age groups continue to increase until age 55. After age 55, the age trajectory in homeownership levels off and declines somewhat. Figure 2 shows the period effects across eight survey waves. In early waves the level of homeownership remains stable and then reaches its peak in the year 2006, which is then followed by a sharp decrease in the most recent waves. It is possible that this peak in 2006 is related to policy changes in housing marketization in those provinces prior to 2006 (Kalman-Lamb, 2017). However, the positive effect of policy shifts in homeownership is later compensated for by rising housing prices (Walks, 2014). The cohort effects, as depicted in Figure 3, fluctuate moderately across the 18 birth cohorts and tend to increase in the most recent cohorts. The temporal patterns of homeownership in the more populous provinces (represented by the dashed lines) are similar to those in the less populous provinces, with two notable exceptions. First, the levels of homeownership in the more populous provinces are generally lower than in the less populous provinces, and this pattern holds across age groups, time periods, and birth cohorts. Second, members of the younger generation born after the early 1980s were able to achieve a higher level of homeownership than other birth cohorts in the less populous provinces. Yet, the younger generation's advantage in achieving homeownership is not observed in the more populous provinces. It should be noted that, since we divide the entire Canadian population into two parts (i.e., the more and less populous provinces), the age, period, or cohort effect of homeownership at the national level (dotted lines in Figure 1 to Figure 3) is the arithmetic average of their corresponding temporal effects in these two parts of the country and follow a similar pattern to that found in the more and less populous provinces.

Table 3 shows the results from the HAPC analysis of homeownership in the less populous provinces (results based on the more populous provinces are presented in Appendix C). All age effects are statistically significant at the 0.001 level. Males and married individuals are positively associated with the likelihood of homeownership, while the better educated are also more likely to be homeowners. People living in Eastern Canada and the Prairies are more likely to attain homeownership than are people in the Northern provinces. Due to a secular decline in period effects, the 2006 wave shows the highest level of homeownership and the 2016 wave shows the lowest level of homeownership. Moreover, only period effects in these two waves are statistically significant. The variance components associated with both period and cohort effects are significant at the 0.001 level, suggesting the presence of both temporal effects.

[Figure 1, Figure 2, Figure 3 and Table 3 about here]

Based on results from the HAPC analysis, we next applied structural change analysis to the predicted period effects of homeownership in the less and more populous provinces in Canada (Figure 4.1 and Figure 4.2), and identified one and only one structural break in both regions: the year 2001 (95% CI [1996, 2006]). It is also noteworthy to mention that homeownership at the national level has the same period-effect structural break as that in the less and more populous provinces. As suggested by Walks and Clifford (2015), one possible explanation for this structural break is a shift toward neoliberal housing policies and housing financialization in Canada around the early 2000s. A further comparison of results obtained from APC analysis and structural change analysis suggests that the two methods follow different principles and do not necessarily produce similar conclusions. Statistical tests in APC analysis tell whether a (random) temporal effect is significantly different from zero, and tend to identify the zenith (the year 2006) and nadir (the year 2016 based on the less populous provinces; the years 1981 and 1986 for the more populous provinces) of a period trajectory. In contrast, structural change analysis tells us whether one time period deviates from an existing temporal pattern and represents, from an econometric perspective, a shock to the system, such as the Canadian policy shift towards neoliberal housing and housing financialization in the early 2000s. Therefore, the subsequent increases or decreases in temporal effects, as revealed by statistical testing in HAPC analysis, can be viewed as an outcome of the present shock identified by structural change analysis. The latter provides a valuable tool for scholars and policymakers to explore and identify the timing of major policy, social, and economic changes, especially in an era of urban transformation (Fu & Lin, 2013).

[Figures 4.1 and 4.2 about here]

6. Conclusion

People in modern societies widely treat homeownership as one of the most important family assets, an indicator of economic well-being, and a symbol of personal success (Rohe & Stegman, 1999; Rosenbaum, 1996; Zhu, Breitung, & Li, 2012). Canadians are no exception to these general beliefs (Kalman-Lamb, 2017; Walks, 2016). However, empirical analysis of homeownership in Canada tends to focus on the country's more populous provinces or metropolitan areas. Given this imbalance in the research focus, studies on how homeownership evolves over time in the less populous regions in Canada are warranted. Our empirical analysis partially fills this gap in the literature. Using data retrieved from the Canadian Population Censuses between 1981 and 2016, we analyzed the overall trend and different temporal patterns of homeownership in Eastern Canada, the Prairies, and Northern Canada. Our results generally show a higher level of homeownership among middle-aged and elderly Canadian adults, despite a slight decrease in homeownership among those aged 70 and older. We also find that both period and cohort effects are significant. Most importantly, there is a mismatch between the significant period effects identified by HAPC analysis and structural change analysis. The former is associated with a secular decline in homeownership from 2006 to 2016, while the latter is possibly attributable to policy changes around the early 2000s.

The similarities and differences in temporal patterns between the more and less populous regions in Canada greatly advance our understanding of homeownership in the country, especially in the less populous provinces. First, the levels of age, period, or cohort effects in the more populous provinces are consistently lower than in the less populous provinces. Moreover, the younger generation (cohorts born in the early 1980s) in the less populous provinces can achieve a higher rate of homeownership than other birth cohorts, but this conclusion does not apply to residents in the more populous provinces. These findings point to the issue of housing (un)affordability in Canada's metropolitan areas and global cities (Bunting, Walks, & Filion, 2004; Moore & Skaburskis, 2004). Second, despite some sporadic changes and the overall differences in levels, homeownership in the less and more populous regions largely share the same trend across time periods, birth cohorts, and age groups. These similarities in temporal patterns are possibly related to the absence of institutional barriers for interprovincial migration within Canada (Chan

& Zhang, 1999; Fu & Ren, 2010; Maas, 2013). Third, the structural change analysis identifies the year 2001 as the only structural break for both the more and less populous provinces. Despite the focus on how the Canadian housing market is shaped by the influx of migrants, the rise of global cities in Canada, and regional disparities, this finding suggests that the federal shift towards a neoliberal housing policy in the early 2000s remains a fundamental factor shaping homeownership regardless of one's place of residence in Canada.

We synthesized statistical methods and proposed a mixed approach to practical concerns in APC analysis. By linking HAPC analysis with multiple imputation and structural change analysis, this mixed method can simultaneously deal with more-coarse temporal groups and identify local anomalies in temporal patterns. Because census or other de-identified datasets available for scholarly use often do not provide more precise age groups in every survey wave, but instead provide more coarse age groups, the resulting overlapping birth cohorts prevent a further investigation of separate age, period, and cohort effects. We thus addressed the overlapping birth cohorts through a construction of pseudo age groups using multiple imputation. Our procedures described here also provide a feasible solution to a wide range of research problems which could not be properly addressed without more precise age information (Baron et al., 2012; Crost & Guerrero, 2012; Crowcroft et al., 2003; Gorstein et al., 1994; Park, Park, & Choi, 2019). Furthermore, we also applied structural change analysis and detected any structural break as a local anomaly of temporal (period) effects. This analysis of homeownership in less and more populous provinces in Canada demonstrates the use and validity of this mixed method in a comparative way. The synthesis of methods for APC analysis and computational tools in other research fields not only addresses practical concerns in temporal analysis but also facilitates a valuable conversation that can inform demographic research.

REFERENCES

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*(1), 47-78.

Balakrishnan, T. R., & Wu, Z. (1992). Home ownership patterns and ethnicity in selected Canadian cities. *Canadian Journal of Sociology/Cahiers Canadiens de Sociologie, 17*(4), 389-403.

Baron, A., Rayson P., Greenwood P., Walkerdine J., & Rashid A. (2012). Children online: A survey of child language and CMC corpora. *International Journal of Corpus Linguistics*, 17(4), 443-81.

Berglund, P. A. (2015). Multiple imputation using the fully conditional specification method: A comparison of SAS®, Stata, IVEware, and R. SAS Institute. *Proceedings of the SAS Global Forum Conference*, *Paper 2081-2015*. Cary, NC: SAS Institute Inc.

Berglund, P., & Heeringa, S. G. (2014). *Multiple imputation of missing data using SAS*. Cary, NC: SAS Institute Inc.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(4), 651-675.

Bunting, T., Walks, R. A., & Filion, P. (2004). The uneven geography of housing affordability stress in Canadian metropolitan areas. *Housing Studies*, *19*(3), 361-393.

Buuren, S. V. (2018). *Flexible imputation of missing data*. Boca Raton, FL: CPC Press.

Carter, T. (1997). Current practices for procuring affordable housing: The Canadian context. *Housing Policy Debate, 8*(3), 593-631.

Chan, K. W., & Zhang, L. (1999). The hukou system and rural-urban migration in China: Processes and changes. *The China Quarterly*, (160), 818-855.

Chauvet, G. (2016). Variance estimation for the 2006 French housing survey. *Mathematical Population Studies, 23*(3), 147-163.

Constant, A. F., Roberts, R., & Zimmermann, K. F. (2009). Ethnic identity and immigrant homeownership. *Urban Studies, 46*(9), 1879-1898.

Crost, B., & Guerrero S. (2012). The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age. *Journal of Health Economics*, 31(1), 112-21.

Crowcroft, N. S., Stein C., Duclos P., & Birmingham M. (2003). How best to estimate the global burden of pertussis? *The Lancet Infectious Diseases*, 3(7), 413-18.

Dalton, T. (2009). Housing policy retrenchment: Australia and Canada compared. *Urban Studies, 46*(1), 63-91.

Dilmaghani, M., & Dean, J. (2020). Sexual orientation and homeownership in Canada. *Journal of Housing Economics, 49*, 1-18.

Edmonston, B. (2005). Who owns? Homeownership trends for immigrants in Canada. Paper presented at the Annual Meeting of the Canadian Population Society, Winnipeg, MB.

Fosse, E., & Winship, C. (2019). Analyzing age-period-cohort data: A review and critique. *Annual Review of Sociology, 45*(1), 467-492.

Fu, Q. (2016). The persistence of power despite the changing meaning of momeownership: An age-period-cohort analysis of urban housing tenure in China, 1989–2011. *Urban Studies*, *53*(6), 1225-1243.

Fu, Q., & Land, K. C. (2017). Does urbanisation matter? A temporal analysis of the socio-demographic gradient in the rising adulthood overweight epidemic in China, 1989–2009. *Population, Space and Place, 23*(1), 1-17.

Fu, Q., & Lin, N. (2013). Local state marketism: an institutional analysis of China's urban housing and land market. *Chinese Sociological Review*, *46*(1), 3-24.

Fu, Q., & Ren, Q. (2010). Educational inequality under China's rural–urban divide: The hukou system and return to education. *Environment and Planning A*, *42*(3), 592-610.

Gorstein, J., Sullivan K., Yip R., De Onis M., Trowbridge F., Fajans P., & Clugston G. (1994). Issues in the assessment of nutritional status using anthropometry. *Bulletin of the World Health Organization*, *72*(2), 273.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*(3), 206-213.

Grigoryeva, I., & Lay, D. (2019). The price ripple effect in the Vancouver housing market. *Urban Geography, 40*(8), 1168-1190.

Gu, J. X., Guo, X., Veenstra, G., Zhu, Y., & Fu, Q. (2020). Adolescent marijuana use in the United States and structural breaks: An age-period-cohort analysis, 1991 to 2018. *American Journal of Epidemiology, 190*(6), 1056-1063.

Gyourko, J., & Linneman, P. (1996). Analysis of the changing influences on traditional households' ownership patterns. *Journal of Urban Economics, 39*(3), 318-341.

Haan, M. (2007). The homeownership hierarchies of Canada and the United States: The housing patterns of White and non-White immigrants of the past thirty years. *International Migration Review, 41*(2), 433-465.

Haan, M. (2008). The place of place: Location and immigrant economic well-being in Canada. *Population Research and Policy Review, 27*(6), 751-771.

Harris, R. (1984). *Class and housing tenure in modern Canada*. Toronto, ON: Centre for Urban and Community Studies, University of Toronto.

Hou, F. (2012). Homeownership over the life course of Canadians: evidence from Canadian censuses of population. *SSRN Electronic Journal* (325).

Hulchanski, J. D. (2003). What factors shape Canadian housing policy? The intergovernmental role in Canada's housing system. Paper presented at the Conference on Municipal-Federal-Provincial Relations in Canada, Queen`s University, Kingston, ON.

Kalman-Lamb, G. (2017). The financialization of housing in Canada: Intensifying contradictions of neoliberal accumulation. *Studies in Political Economy, 98*(3), 298-323.

Land, K. C., Zang, E., Fu, Q., Guo, X., Jeon, S. Y., & Reither, E. (2016). Playing with the rules and making misleading statements: A response to Luo, Hodges, Winship, and Powers. *American Journal of Sociology, 122*(3), 962-973.

Li, Peter S. (1998). *The Chinese in Canada.* Don Mills, ON: Oxford University Press.

Li, M. X., & Shan R. (2020). Housing need in Canada: age-period-cohort effects and transitions (March 23, 2020). *SSRN Electronic Journal*.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. (2006). *SAS for mixed models*. Cary, NC: SAS Institute.

Maas, W. (2013). Free movement and discrimination: Evidence from Europe, the United States, and Canada. *European Journal of Migration and Law*, *15*(1), 91-110.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review, 38*(2), 242-258.

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research*, *51*(5), 339-343.

Moore, E., & Skaburskis, A. (2004). Canada's increasing housing affordability burdens. *Housing Studies*, *19*(3), 395-413.

Okkola, S., & Brunelle, C. (2018). The changing determinants of housing affordability in oil-booming agglomerations: A quantile regression investigation from Canada, 1991–2011. *Housing Studies*, *33*(6), 902-937.

Park, J., Park S. K., & Choi Y. H. (2019). Environmental pyrethroid exposure and diabetes in US adults. *Environmental Research,* 172, 399-407.

Reither, E. N., Hauser, R. M., & Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Social Science & Medicine, 69*(10), 1439-1448.

Reither, E. N., Land, K. C., Jeon, S. Y., Powers, D. A., Masters, R. K., & Zheng, H., . . . Yang, C. Y. (2015). Clarifying hierarchical age–period–cohort models: A rejoinder to Bell and Jones. *Social Science & Medicine*, *145*, 125-128.

Rohe, W. M., & Stegman, M. A. (1999). The effects of homeownership: On the self-esteem, perceived control and life satisfaction of low-income people. *Directions in Person-Environment Research and Practice*, *60*(2), 173-184.

Rosenbaum, E. (1996). Racial/ethnic differences in home ownership and housing quality, 1991. *Social Problems, 43*(4), 403-426.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: John Wiley & Sons.

—. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473-489.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schafer, J. L., & Olsen M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*(4), 545-571.

Skaburskis, A. (1996). Race and tenure in Toronto. *Urban Studies, 33*(2), 223-252.

Statistics Canada. (2011). Guide to the Census of Population, 2011 [Internet]. Available at: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm.

Walks, A. (2013). Mapping the urban debtscape: the geography of household debt in Canadian cities. *Urban Geography*, 34(2), 153-87.

Walks, A. (2014). Canada's housing bubble story: Mortgage securitization, the state, and the Global Financial Crisis. *International Journal of Urban and Regional Research*, *38*(1), 256-284.

Walks, A. (2016). Homeownership, asset-based welfare and the neighbourhood segregation of wealth. *Housing Studies*, *31*(7), 755-784.

Walks, A., & Clifford, B. (2015). The political economy of mortgage securitization and the neoliberalization of housing policy in Canada. *Environment and Planning A: Economy and Space*, *47*(8), 1624-1642.

Walks, A., Hawes E., & Simone D. (2021). Gentrification in large Canadian cities: tenure, age, and exclusionary displacement 1991-2011. *Urban Geography*, 1-31.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399.

Yang, Y., & Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, *36*(1), 75-97.

Zhu, Y. S., Breitung, W., & Li, S. (2012). The changing meaning of neighbourhood attachment in Chinese commodity housing estates: Evidence from Guangzhou. *Urban Studies, 49*(11), 2439-2457.

Zhu, Y. S., Fu, Q., & Ren, Q. (2014). Cross-city variations in housing outcomes in postreform China: An analysis of 2005 microcensus data. *Chinese Sociological Review, 46*(3), 26-54.