How Does Immigration Impact Canada's Housing Prices – Evidence from a Spatial Perspective

4/10/2024

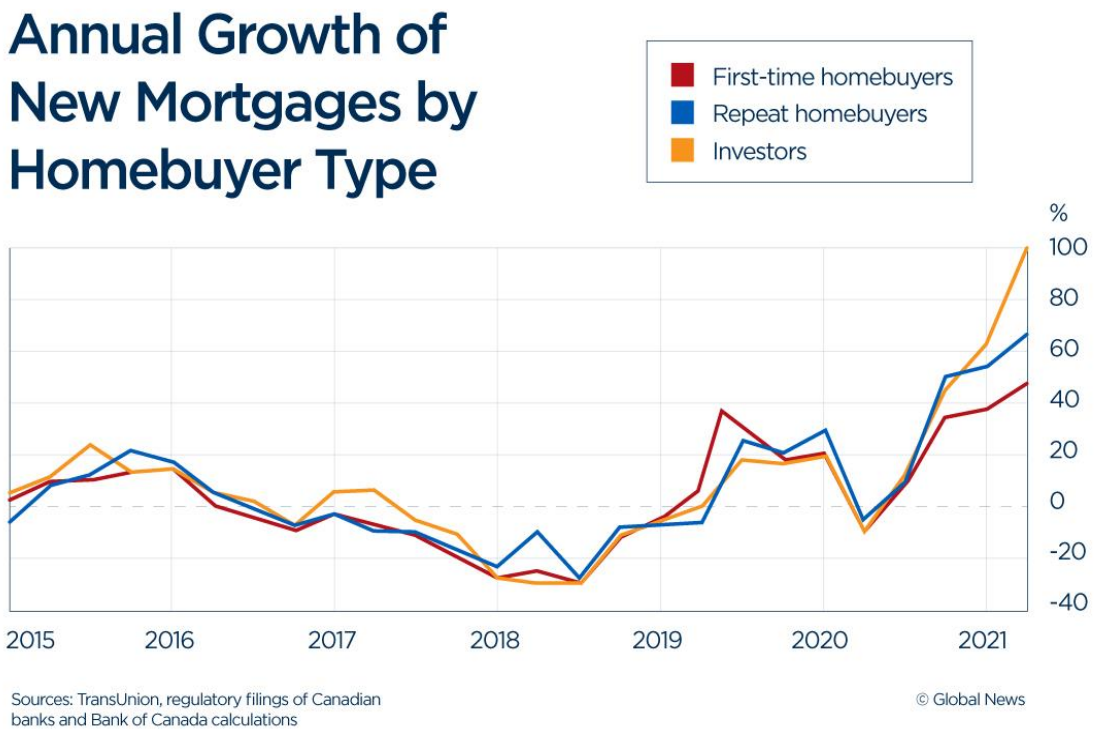
Yuxin Gong,
Department of Economics

UNIVERSITY OF WATERLOO

CONTENTS

1. Background & Motivation

- Current situation: Canada's housing market faces a challenge of unaffordability due to high demand. People are taking up more loans than ever to afford a home.
- Meanwhile, Canada welcomes high numbers of immigrants (aiming for 500,000 annually by 2025), fueling up population growth as well as the potential demand for housing.



Source: https://globalnews.ca/news/8416870/real-estate-investors-canada-housing-market/

1. Background & Motivation

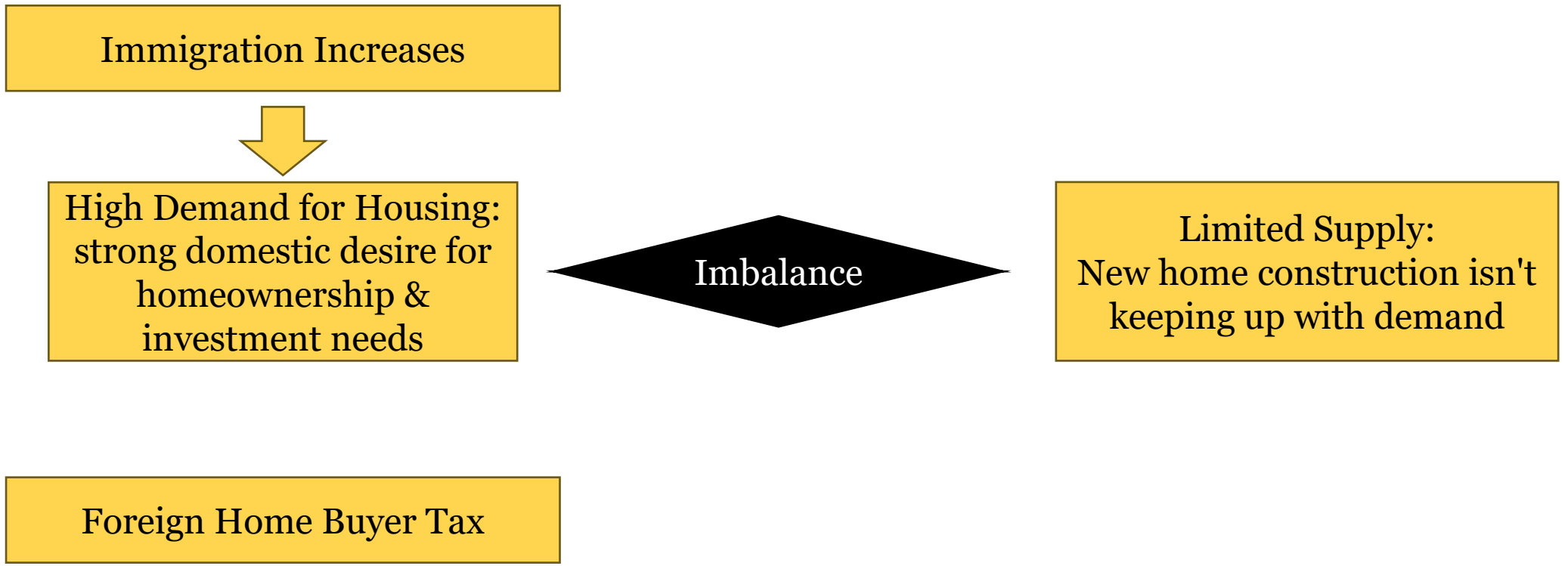
- Goal: This project aims to study if, over the years, immigration impacts Canada's housing prices through a spatial econometrics model, and if so, how?
- Current outline: The project aims to

  ① use a special dynamic panel data model to consider the spatial and temporal lag in the relationship between immigration and housing price;

  ② take a closer look at how the introduction of Foreign Home Buyer tax in Ontario (Ontario's Non-Resident Speculation Tax (NRST)) to see if the tax policy changes the impact.

2. Economic Theory & Hypotheses

- Economic Theory (Rational): Demand-supply relationship in housing market



2. Economic Theory & Hypotheses

- Hypotheses:

(1) Immigration in Canada's provinces has a positive effect on Canada's housing price, and there also exists a positive spillover effect.

  - To be tested in the main regression – the spatial dynamic panel data regression

(2) The introduction of the NRST mitigates this positive effect in Ontario.

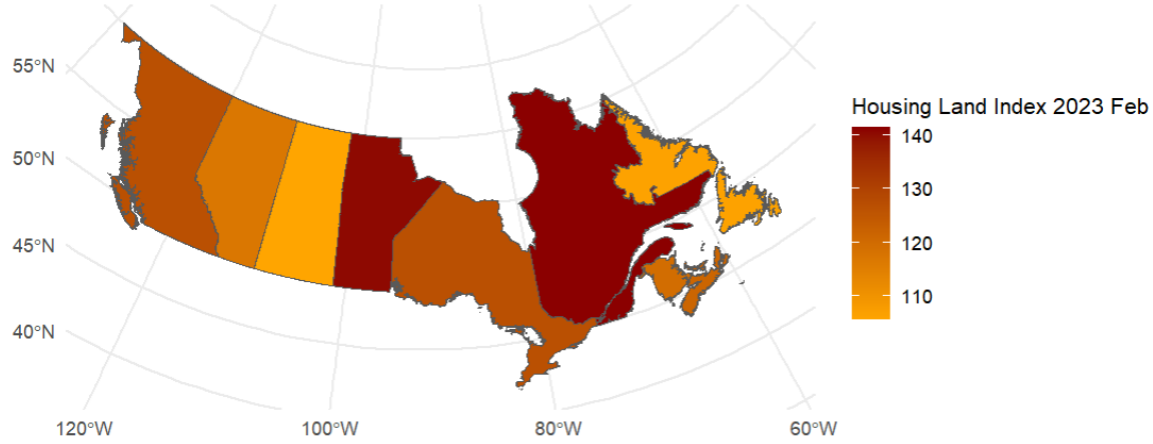
  - To be tested in the follow-up regression – pooled OLS regression

3. Econometric Model

- Why Use a Spatial Econometrics Model?

Many types of data are spatially and temporally autocorrelated

- Political elections
- Contaminant transfer
- Disease spread
- Housing market
- Weather



3. Econometric Model

- A Spatial Dynamic Panel Data Model (SPDM)

$$y = \alpha lag(y) + \rho Wy + X\beta + WX\theta + \lambda Wu + \epsilon$$

y: dependent variable (n x 1)

Lag(y): temporal lag of the dependent variable (n x 1)

X: Explanatory variable (n x k)

W: Spatial weights matrix (n x n), measures the spatial correlation between two individuals

ρ: Spatial autoregressive coefficient (n x n), measures the endogenous interaction effects.

λ: Spatial autocorrelation coefficient (n x n), measures the interaction effects among the error terms.

β: Coefficient for x (k x 1), measures the effect of x on y in the same individual as well as that of x on y of a different individual (exogenous interaction effects).

3. Econometric Model

- Spatial Autoregressive Model (SAR)

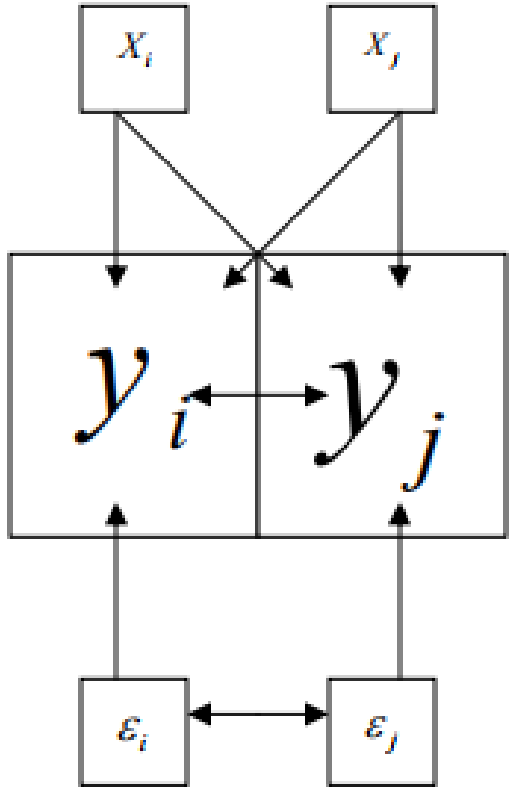
$$y = \rho W y + X\beta + WX\theta + \epsilon$$

- Spatial Error Model (SEM)

$$y = X\beta + WX\theta + \lambda W u + \epsilon$$

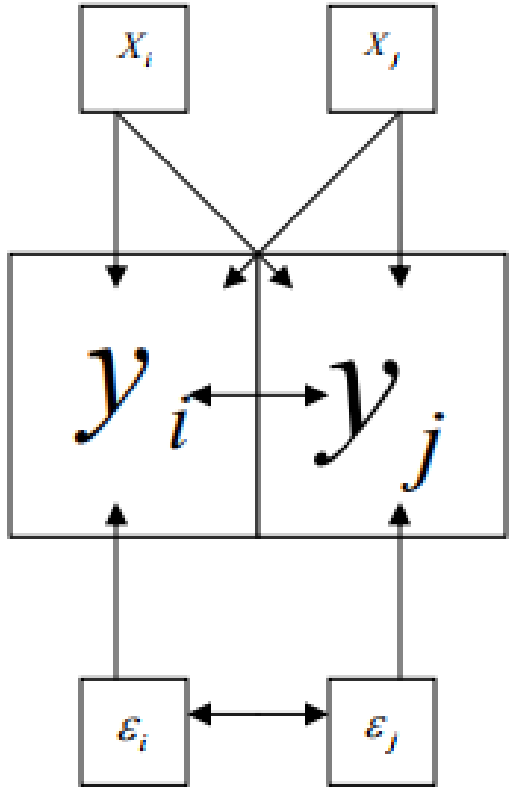
- Spatial Durbin Model (SDM)

$$y = \rho W y + X\beta + WX\theta + \lambda W u + \epsilon$$



3. Econometric Model

- Estimation Methods for Spatial Durbin Models
  - (1) Approach based on maximum likelihood (ML) or quasi maximum likelihood (QML)
  - (2) Approach based on instrumental variables or generalized method of moments (IV/GMM)
  - (3) Approach based on the Bayesian Markov Chain Monte Carlo (MCMC) approach
- Potential problem: Anselin et al. (2008) argue that the parameters of the dynamic spatial panel data model could not be identified.



4. Data & Variables

- Dataset used in this project are from 1910 samples: 10 provinces over 191 periods (monthly from 2018/Feb to 2023/Dec). Data are formed into a balanced panel dataset. All data are sourced from Statistics Canada published datasets.
- Data Source:

Variable Name	Data source
Housing price index	https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501
immigration level	https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1410008201
income level	https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3610020501
inflation: CPI	https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810000413
unemployment rate	https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1410001701
population density	population(15 and above): https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1410008201
	area: https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=9810000101
Map shapefile	https://open.canada.ca/data/en/dataset/a883eb14-0c0e-45c4-b8c4-b54c4a819edb

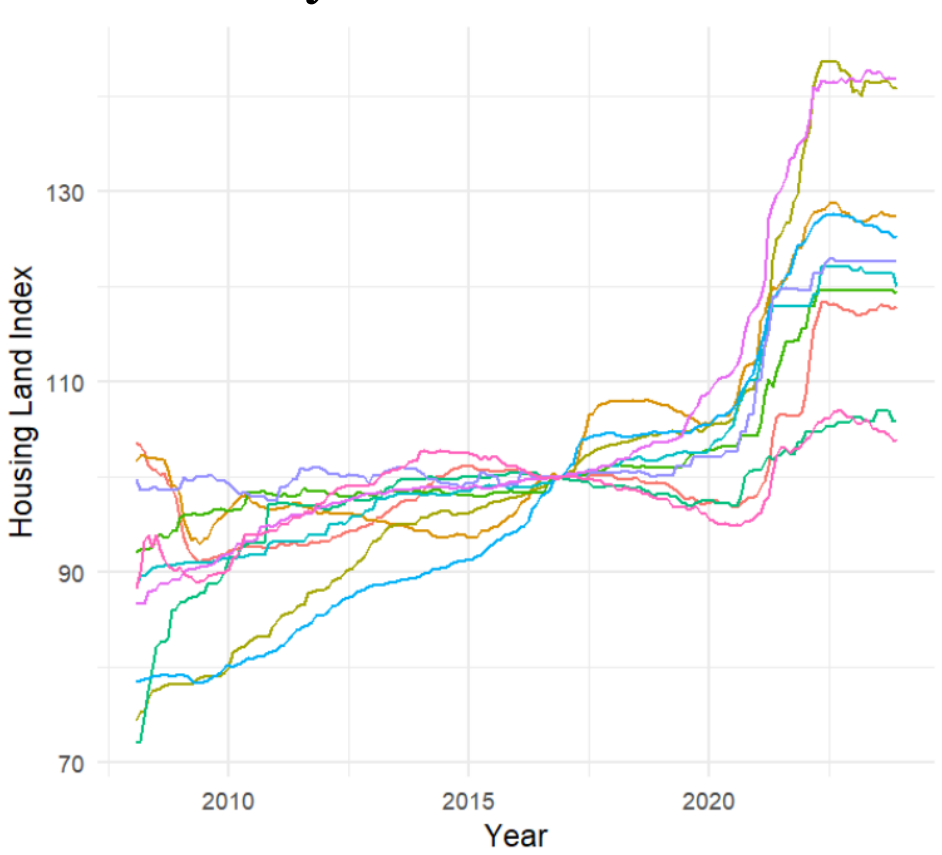
4. Data & Variables

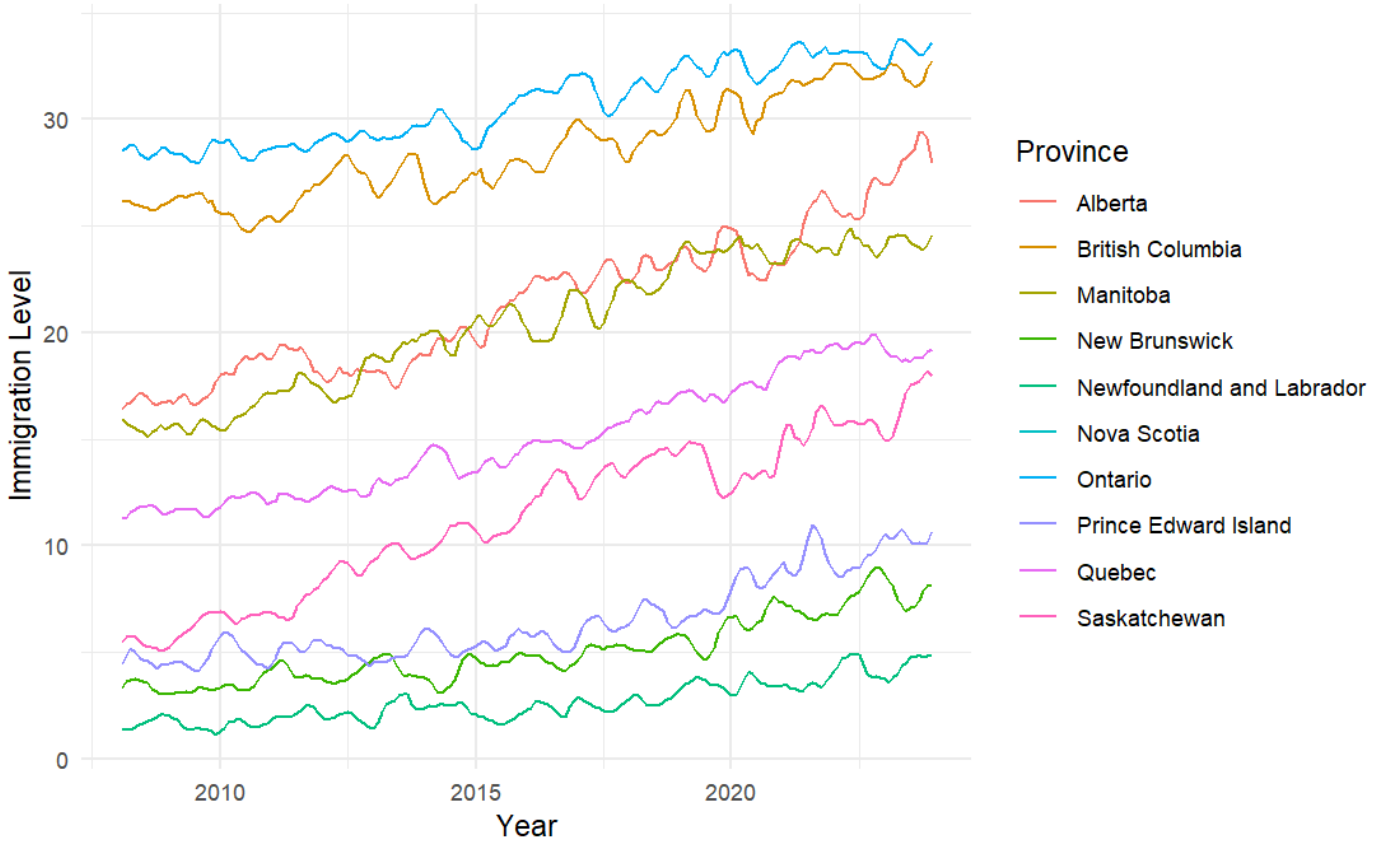
- Below are the variable table

Variable Name	Description	Construction
housing_land_index	The dependent variable, measures the level of new housing and land prices.	New housing price index from Stats Can (2016/Dec = 100)
immigrant_ratio_employmnt	The independent variable, measures the level of immigration among employed population in Canada.	Ratio of immigrants among all employed population (aged 15 and over)
population_density	Control 1, measures the population density in one province at a certain period	Population/Land area (Provincial land area is from 2021 Census)
unemployment	Control 2, measures the unemployment rate in one province at a certain period	Unemployment rate from Stats Can
wage	Control 3, the average wage in one province at a certain period	Total distributed wage/employed population
COVID	Control 4, measures if it is a post-COVID year	2020 - 2023 = 1 2008 - 2019 = 0
after_tax	Control 5, measures if it is a post-NRST year	2018 - 2023 = 1 2008 - 2017 = 0

4. Data & Variables

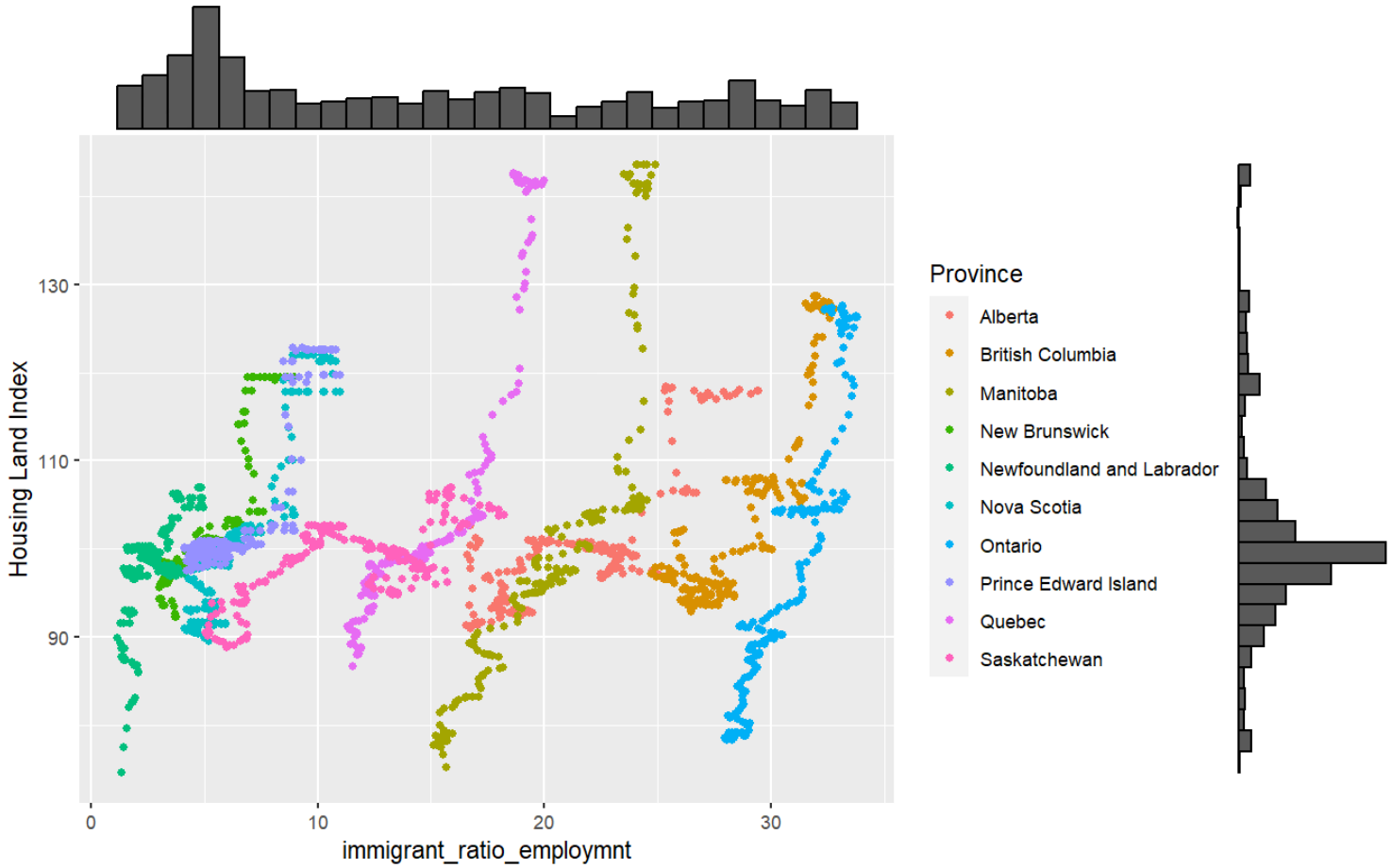
- Preliminary Variable check – Housing Price Index by Time





4. Data & Variables

- Preliminary Variable check – Housing Price Index & Immigration



5. Pre-regression Examinations

- Stationarity of y: ADF test and Levin–Lin–Chu test

Augmented Dickey-Fuller Test
Data: province$housing_index_log
Dickey-Fuller = -6.3034, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary

Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and Trend)
data: housing_index_log ~ trend
z = -1.5797, p-value = 0.05708
alternative hypothesis: stationarity

- ADF test for time series; Levin-Lin-Chu test for panel data.
- In both tests, **we reject the null and accept the alternative, concluding stationary.**
- Now that we have tested that the housing index variable is stationary, it is safe to include AR elements in my econometric model.

5. Pre-regression Examinations

- Spatial Effects: Moran's I test

- $I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$

- N is the number of spatial units indexed by i and j
- x is the variable of interest; $\bar{x}$ is the mean of x
- w is a matrix of spatial weights
- W is the sum of all w_{ij}

- Moran's I test measures the spatial autocorrelation for continuous data.
- Define a neighbor list and weight matrix: contiguous neighbor or k-nearest neighbor.
- Queen contiguity weights or distance-band weights (KNN) (Here I use KNN neighbor and weights).
- Calculate Moran' I statistics and conduct a hypothesis test.

5. Pre-regression Examinations

- Spatial Effects: Moran's I test

```
        Global Moran I for regression residuals

data:
model: lm(formula = housing_land_index ~ immigrant_ratio_employmnt, data =
province_panel_small_data)
weights: weights_2023_6mon

Moran I statistic standard deviate = 14.474, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Observed Moran I      Expectation         Variance
     0.946092120     -0.016801696      0.004425427
```

Strong positive spatial autocorrelation

- In general,
  - I ~ 1 means strong positive autocorrelation
  - ~ -1 means strong negative autocorrelation
  - ~ 0 means no autocorrelation
- Null hypothesis: I is (approximately) zero
- Alternative hypothesis: I is greater than zero

5. Pre-regression Examinations

- Collinearity: Check Variance Inflation Factor (VIF)

The variance inflation factor of an explanatory variable in a model is a function of how collinear that variable is with the over explanatory variables in the model are. The higher the number, the more collinear and the most the variance estimates of the slopes are being inflated by including that variable.

	lag(housing_land_index)	immigrant_ratio_employmnt	population_density	unemployment	wage	CPI
Pooled OLS	2.650889	2.009668	1.435508	1.533092	3.633255	4.686434
Random Effects	3.741416	1.559336	1.176622	1.144370	5.322900	6.021596

5. Pre-regression Examinations

- Random effects: LM test and Hausman test

(1) Breusch–Pagan Lagrange Multiplier Test: Pooled OLS vs Random Effects

```
        Lagrange Multiplier Test - (Breusch-Pagan)

data:  housing_land_index ~ lag(housing_land_index) + immigrant_ratio_employmnt +  ...
chisq = 27.328, df = 1, p-value = 1.717e-07
alternative hypothesis: significant effects
```

Random Effects Panel Model

5. Pre-regression Examinations

- Random effects: LM test and Hausman test

(2) Hausman Test: Fixed Effects vs Random Effects

```
        Hausman Test

data:  housing_land_index ~ lag(housing_land_index) + immigrant_ratio_employmnt +  ...
chisq = 59.039, df = 6, p-value = 7.053e-11
alternative hypothesis: one model is inconsistent
```

Random Effects Panel Model

6. Preliminary Regression Results

- Random effects Panel Data Model

```
Effects:
                    var   std.dev share
idiosyncratic 7.515e-05 8.669e-03  0.85
individual    1.328e-05 3.644e-03  0.15
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6753  0.6753  0.6753  0.7225  0.8299  0.8299

Residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.021464 -0.006158 -0.000488 -0.000231  0.003623  0.057069

Coefficients:
                           Estimate  Std. Error  z-value  Pr(>|z|)
(Intercept)               3.7550e+00  7.7710e-03 483.2041 < 2.2e-16 ***
lag(housing_land_index)    8.0768e-03  6.4183e-05 125.8410 < 2.2e-16 ***
immigrant_ratio_employmnt  3.9592e-04  1.2876e-04   3.0748  0.002107 **
population_density         7.8610e-04  1.8374e-04   4.2782 1.884e-05 ***
unemployment              -4.8087e-04  2.0903e-04  -2.3005  0.021421 *
wage                       7.0365e-06  1.5572e-06   4.5188 6.220e-06 ***
COVID                      1.6786e-02  1.7965e-03   9.3436 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    79.908
Residual Sum of Squares: 0.050618
R-Squared:      0.99937
Adj. R-Squared: 0.99937
Chisq: 44120.1 on 6 DF, p-value: < 2.22e-16
```

- Significant and positive as it is, the effect of immigration is small.
- R_Squared reached as high as 99%.

6. Preliminary Regression Results

- Random effects Panel Data Model

```
Effects:
                  var  std.dev share
idiosyncratic 0.001628 0.040343 0.172
individual    0.007837 0.088528 0.828
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.9344  0.9344  0.9344  0.9444  0.9670  0.9670

Residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.164075 -0.023769  0.001246 -0.000864  0.024901  0.132697

Coefficients:
                           Estimate  Std. Error  z-value  Pr(>|z|)
(Intercept)              4.1391e+00  4.5617e-02  90.7347 < 2.2e-16 ***
immigrant_ratio_employmnt  6.3167e-03  1.1837e-03   5.3363 9.485e-08 ***
population_density         8.1841e-03  3.3464e-03   2.4457   0.01446 *
unemployment              -1.3210e-02  8.2086e-04 -16.0926 < 2.2e-16 ***
wage                       1.0235e-04  7.2672e-06  14.0840 < 2.2e-16 ***
COVID                      5.6623e-02  9.2719e-03   6.1069 1.016e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.698
Residual Sum of Squares: 1.0128
R-Squared:      0.84941
Adj. R-Squared: 0.84819
Chisq: 1394.46 on 5 DF, p-value: < 2.22e-16
```

- Significant and positive as it is, the effect of immigration is small.
- R_Squared is decent.

6. Preliminary Regression Results

- Random effects Panel Data Model

  - Significant and positive as it is, the effect of immigration is small.
  - R_Squared is decent.
  - The research of Akbari & Aydede (2011) has similar results: their result indicates a statistically significant but small effect of immigration on prices of privately owned dwellings in Canada.
  - Their explanation: An out-migration of the native born from the areas, where new immigrants settle, or an increased supply of housing due to expectations of higher demand in those areas may have caused this result.

6. Preliminary Regression Results

- Spatial Dynamic Panel Data Model
- blmpSDPD() function from SDPDmod library in R.
- Possible reasons that caused issues during estimation:

(1) Computational cost: W matrix is an n x n matrix, in panel data weight matrix creation, the software actually treats each observation as a different location, even if there are only 10 unique map datapoints. This prevents me from creating the full weight matrix

(2) Data cleaning issue: even for smaller data set whose weight matrix I am able to get, errors occur when running the regression. According to posts online, this may be due to the unmatching dimension of weight matrix and my panel data.

(3) Further exploration needs to be done for more accurate solutions.

6. Preliminary Regression Results

- Pooled OLS

```
Call:
lm(formula = housing_index_log ~ immigrant_ratio_employmnt +
    population_density + unemployment + wage + CPI, data = ontario_before2017)

Residuals:
      Min        1Q    Median        3Q       Max
-0.018741 -0.009048 -0.000871  0.005890  0.032930

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.9550522  0.0965210  20.255  < 2e-16 ***
immigrant_ratio_employmnt 0.0072553  0.0018205   3.985  0.00012 ***
population_density        0.2043982  0.0189430  10.790  < 2e-16 ***
unemployment             -0.0078190  0.0012554  -6.228 8.29e-09 ***
wage                     -0.0000106  0.0000187  -0.567  0.57180
CPI                      -0.0009271  0.0011663  -0.795  0.42834
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01125 on 113 degrees of freedom
Multiple R-squared:  0.9837,    Adjusted R-squared:  0.983
F-statistic:  1364 on 5 and 113 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = housing_index_log ~ immigrant_ratio_employmnt +
    population_density + unemployment + wage + CPI, data = ontario_after2017)

Residuals:
      Min        1Q    Median        3Q       Max
-0.042496 -0.021131 -0.000113  0.017236  0.069710

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              4.234e+00  2.936e-01  14.420  < 2e-16 ***
immigrant_ratio_employmnt 2.338e-02  6.819e-03   3.429 0.001048 **
population_density       -1.610e-01  4.102e-02  -3.926 0.000209 ***
unemployment              1.424e-03  2.128e-03   0.669 0.505707
wage                      1.943e-04  3.385e-05   5.740  2.6e-07 ***
CPI                       6.795e-03  1.677e-03   4.051 0.000137 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02507 on 66 degrees of freedom
Multiple R-squared:  0.9198,    Adjusted R-squared:  0.9137
F-statistic: 151.4 on 5 and 66 DF,  p-value: < 2.2e-16
```

6. Preliminary Regression Results

- AR(1) Model

```
Call:
lm(formula = housing_index_log ~ lag(housing_index_log) + immigrant_ratio_employmnt +
    population_density + unemployment + wage + CPI, data = ontario_before2017)

Residuals:
       Min         1Q     Median         3Q        Max
-3.420e-16 -6.040e-18  9.100e-19  7.260e-18  8.188e-17

Coefficients:
                          Estimate Std. Error    t value Pr(>|t|)
(Intercept)             -1.954e-15  6.491e-16 -3.010e+00  0.00323 **
lag(housing_index_log)   1.000e+00  2.940e-16  3.401e+15  < 2e-16 ***
immigrant_ratio_employmnt  9.986e-19  6.077e-18  1.640e-01  0.86977
population_density       1.322e-16  8.456e-17  1.563e+00  0.11980
unemployment            -1.728e-19  4.547e-18 -3.800e-02  0.96975
wage                    -1.094e-19  5.851e-20 -1.870e+00  0.06404 .
CPI                      2.956e-18  3.655e-18  8.090e-01  0.42034
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.515e-17 on 112 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 1.183e+32 on 6 and 112 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = housing_index_log ~ lag(housing_index_log) + immigrant_ratio_employmnt +
    population_density + unemployment + wage + CPI, data = ontario_after2017)

Residuals:
       Min         1Q     Median         3Q        Max
-3.593e-17 -4.939e-18  1.376e-18  5.079e-18  1.151e-17

Coefficients:
                          Estimate Std. Error    t value Pr(>|t|)
(Intercept)             0.000e+00  1.863e-16  0.000e+00    1.000
lag(housing_index_log)   1.000e+00  3.833e-17  2.609e+16   <2e-16 ***
immigrant_ratio_employmnt -2.454e-20  2.305e-18 -1.100e-02    0.992
population_density       1.656e-17  1.419e-17  1.167e+00    0.247
unemployment            -4.452e-19  6.650e-19 -6.690e-01    0.506
wage                    -5.433e-21  1.291e-20 -4.210e-01    0.675
CPI                     -3.542e-19  5.836e-19 -6.070e-01    0.546
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.807e-18 on 65 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 1.415e+33 on 6 and 65 DF,  p-value: < 2.2e-16
```

7. Critiques, Limitations, and Next Steps

(1) Core issue: the estimation of the spatial effects.

(2) Rigorousness of the panel data analysis: a lot of tests and models are from cross-sectional data, including Spatial Durbin model which was initially designed for cross-sectional data.

(3) Regarding the different effects before and after the NRST: there is a difference, but is this difference significant? (Possible solution: simply add a dummy? add interaction? DID? conduct regime analysis?)

(4) Further interpretation of the results is needed: so far, I have only managed to get some preliminary results from pure data analysis, but have not yet explained them very well based on their economic meaning and the context.

Bibliography

1. Akbari, A. H., & Aydede, Y. (2012). Effects of immigration on house prices in Canada. *Applied Economics*, *44*(13), 1645–1658. https://doi.org/10.1080/00036846.2010.548788
2. Head, A., & Lloyd-Ellis, H. (2016). Has Canadian house price growth been excessive? *The Canadian Journal of Economics / Revue Canadienne d'Economique*, *49*(4), 1367–1400. http://www.jstor.org/stable/24915782
3. Mussa, A., Nwaogu, U. G., & Pozo, S. (2017). Immigration and housing: A spatial econometric analysis. *Journal of Housing Economics*, *35*, 13–25. https://doi.org/10.1016/j.jhe.2017.01.002
4. Rherrad, I., Mokengoy, M., & Kuate Fotue, L. (2019). Is the Canadian housing market "really" exuberant? Evidence from Vancouver, Toronto and Montreal. *Applied Economics Letters*, *26*(19), 1597–1602. https://doi.org/10.1080/13504851.2019.1588941
5. HIEBERT, D. (2009). Newcomers in the Canadian housing market: a longitudinal study, 2001-2005. *The Canadian Geographer / Le Géographe Canadien*, *53*(3), 268–287. https://doi.org/10.1111/j.1541-0064.2009.00263.x
6. Barbu, T.C, Vuţă, M., Străchinaru, A.I. and Cioacă, S.I. (2017). An Assessment of the Immigration Impact on the International Housing Price. *Amfiteatru Economic*, 19(46), 682-695
7. Chow, S.-C., Cunado, J., Gupta, R., & Wong, W.-K. (2017). Causal relationships between economic policy uncertainty and housing market returns in China and India: evidence from linear and nonlinear panel and time series models. *Studies in Nonlinear Dynamics & Econometrics*, *22*(2). https://doi.org/10.1515/snde-2016-0121
8. Hossain, B., & Latif, E. (2009). Determinants of housing price volatility in Canada: a dynamic analysis. *Applied Economics*, *41*(27), 3521–3531. https://doi.org/10.1080/00036840701522861
9. J. Paul Elhorst. (2009). Spatial Panel Data Models. *Springer EBooks*, 377–407. https://doi.org/10.1007/978-3-642-03647-7_19
10. *Spatial Regression*. (n.d.). Www.emilyburchfield.org. https://www.emilyburchfield.org/courses/gsa/spatial_regression_lab


UNIVERSITY OF
WATERLOO