Modelling Canadian House Price Indices

François-Michel Boire (250994202), Western University, STATS9859A Term Project

Fall 2017

Contents

1. Introduction

The role economists traditionally assign to the banking sector is to assist in the financing of entrepreneurial initiatives. However, during the second half of the 20th century, the share of domestic banking credit allocated to the business sector plummeted in favour of mortgage loans. In turn, the risk of a financial crisis related to house price fluctuations appears increasingly daunting as mortgages further leverage private banks (Jordà, Schularick, and Taylor 2016b).

Jordà, Schularik, and Taylor (2016a) examined the composition of historical domestic credit of 17 industrialized countries with new long-term international aggregate credit data going back to 1870. They found that in the second half of the 20th century, advanced economies experienced an exceptional surge in non-financial credit (business and private sectors) relative to GDP, nearly doubling between 1980 and 2009. Moreover, aggregate household loans relative to GDP attained 68% in 2013, after averaging about 20% during the first half of the century.

In hindsight of the global financial chaos of the Great Recession, economists and policymakers are increasingly aware of housing risk. That said, Canadian house prices did not collapse during the Great Recession, mainly due to stricter mortgage regulations and a less active secondary market for mortgage-backed securities. This raises the question: What are the economic drivers of Canadian housing returns?

In this paper, I propose a prediction model for fluctuations in the house price indices of 11 Canadian cities. To do so, I considered a panel of monthly observations from May 2002 to December 2016 (175 observations per city). Fluctuations in the price indices are modelled as the response to macroeconomic fundamentals and consumer sentiment proxys. The data panel includes observations at the country level (e.g. national interest rates), the provincial level (e.g. wages), and at the metropolitan level (e.g. unemployment).

In Section 2 I address the leading trends in house prices and provide sources for the data series used to construct the data panel. I loosely classify real estate market determinants in two main categories: *fundamental* and *public sentiment* covariates. I then propose a general setup for the model. Section 3 presents coefficient results and variable selection procedures. I also investigate 4 model candidates with ANOVAs and out-of-sample performances. Section 4 concludes.

2. Housing Market Determinants & Data

This section reviews the key factors affecting the housing market in order to construct a sensible regression model. I also provide details about data sources and time series transformations where applicable.

Although macro-financial trends do have a direct impact on access to home ownership, there is growing evidence that psychological factors play an important role in explaining house price movements. Indeed, the overwhelming majority of housing market participants are amateurs in terms of financial or economic speculation. Indeed, surveys have shown that home purchases are motivated by fallacies about the housing market (Case and Shiller 2003). One of the public's most common fallacy is extrapolative expectations of returns, so recent house price growth creates optimism about the future. The problem is that house prices can vary substantially, even when the intrinsic value of housing services is unchanged. The irrational public speculation can contribute to creating a housing bubble. This real estate pricing problem is examined in the book *Animal Spirits* (Akerlof and Shiller 2010). The model will therefore include both fundamental (macroeconomic) factors and consumer sentiment proxys.

The Teranet Composite 11 House Price Index

The model's response variable is monthly percent changes in the Composite 11 House Price Index (HPI) series, available on the Teranet and National Bank of Canada House Price Index website (link). I computed percent changes in the data before applying the seasonal adjustement using the `seasonal` package. I then considered monthly observations dating back to April 2002 for eleven Canadian cities: Halifax (NS), Quebec City (QC), Montréal (QC), Ottawa (ON), Toronto (ON), Hamilton (ON), Winnipeg (MB), Edmonton (AB), Calgary (AB), Vancouver (BC), and Victoria (BC). In the plot below, Loess curves are colour-matched according to provinces. One can see that both Albertan cities follow a similar pattern during the second half of the 2000s. The British Columbian cities also share a common downture in the late 2000s. One key element of the series below is that episodes of extreme housing returns appear to be clustered, which calls for the addition of an autoregressive component in the model.

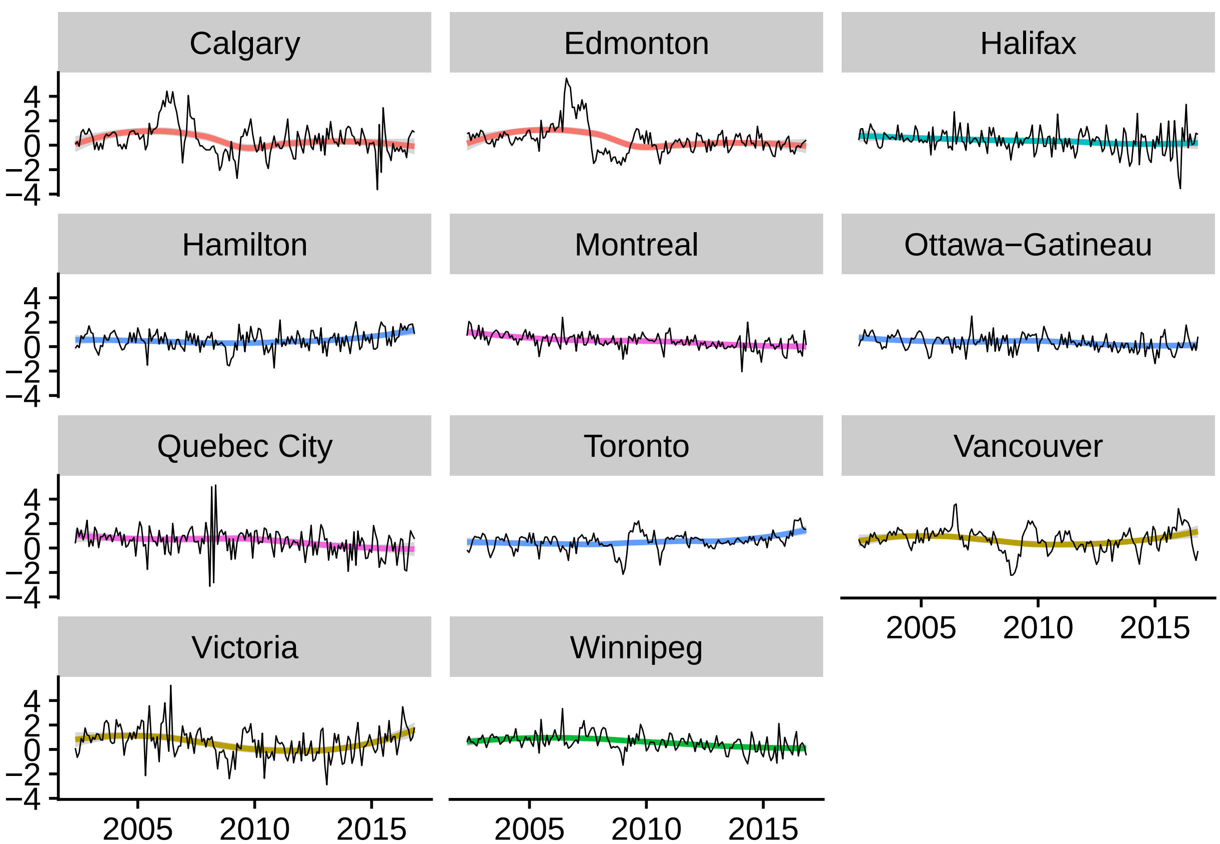


Figure 1: Historical series for percent changes in the HPI with Loess curves

In Figure 1 and 2, one can see that the distribution of house price changes is similar across provinces in terms of mean, but observations display more volatility in provinces like Alberta and British

Columbia. But in Figure 3, it is clear that house price variations have different autocorrelation structures from city to city.

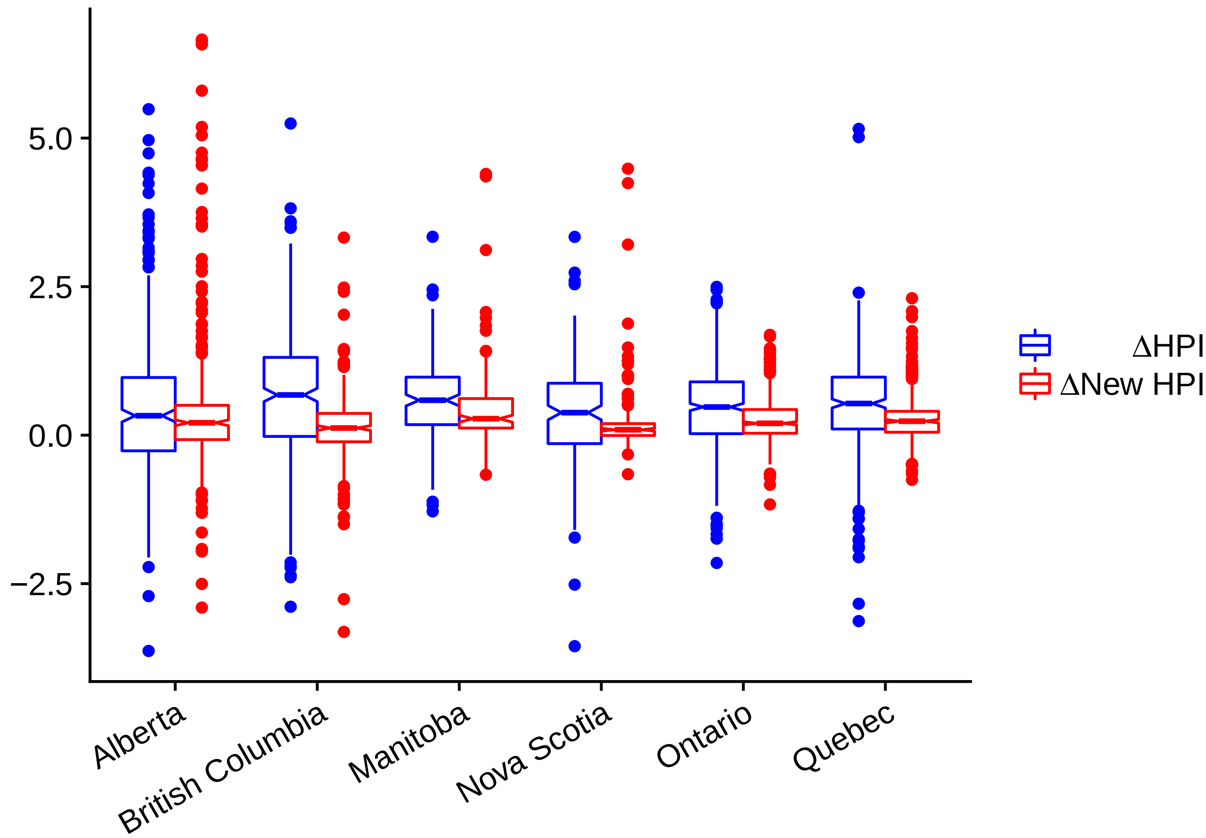


Figure 2: Boxplots for percent changes in the HPI and in the New Housing Price Index

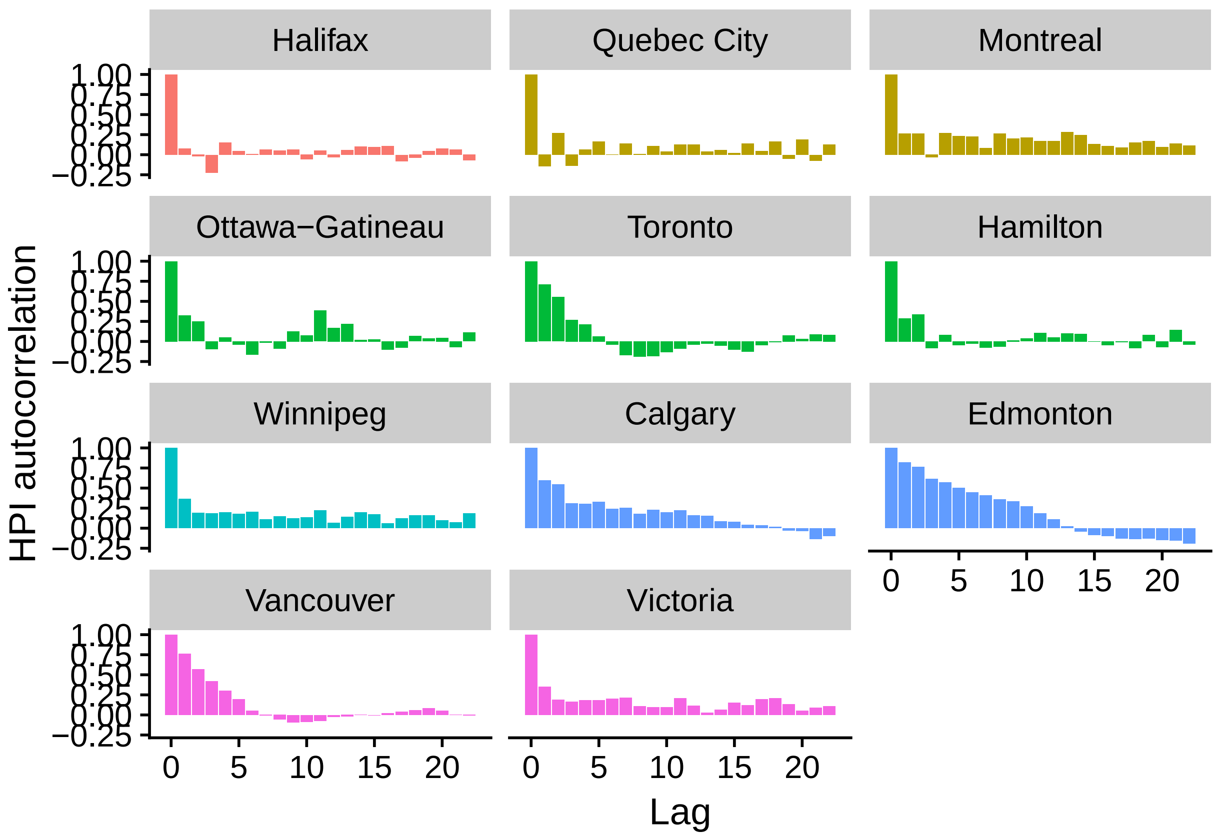


Figure 3: Autocorrelation structures for percent changes in the HPI

Public Sentiment

In the model, the presence of an auto-regressive component will reflect the market participants' extrapolative short-term expectations. Notice that monthly percent changes in city HPIs display strong autocorrelation in cities such as Toronto, Calgary, Edmonton, and Vancouver. Other cities have weak month-to-month autocorrelation, like Halifax, Montreal, and Québec City. This warrants a model specification testing procedure to assess whether coefficients should be homogeneous across cities, discussed in detail in Section 3.

In order to measure consumer sentiment in our model, I followed the work of Pavlidis et al. (2016) by generating the Backward Supremum Augmented Dickey-Fuller (BSADF) unit-root test statistic. The BSADF statistic can proxy consumer excitement in the housing market by tracking the "explosiveness" of a univariate time-series in real-time. The BSADF statistic is constructed as follows:

Consider the Dickey-Fuller equations:

$$\Delta y_t = a_{r_1,r_2} + \beta_{r_1,r_2} + \sum_{j=1}^{k} \psi_{r_1,r_2}^{(j)} \Delta y_{t-j} + \varepsilon_t,$$

where Δy_t is the first difference in the univariate time series y_t, k denotes the number of auto-regressive lags in the model, and ε_t is an iid, normally distributed error term with standard deviation σ_{r_1,r_2}. The interval $[r_1, r_2]$ designates the portion of the sample used to calculate the Augmented Dickey-Fuller (ADF) unit-root test statistic, so with a sample with periods ranging from 0 to T, the test statistic $ADF_{r_1=n/T}^{r_2=m/T}$ is based on a subset of periods ranging from n to m, inclusively $(n, m \in \{0, ..., T\} : n < m)$. The ADF test statistic is defined as:

$$ADF_{r_1=n/T}^{r_2=m/T} = \frac{\hat{\beta}_{r_1,r_2}}{s_{\hat{\beta}_{r_1,r_2}}}.$$

The Supremum ADF (SADF) (Phillips and Yu 2011) is then defined as:

$$SADF(r_0) = \sup_{r_2 \in [r_0,1]} ADF_{r_1=0}^{r_2}$$

One can see that the SADF is calculated recursively with an expanding sample of periods with a minimum window size of r_0, while fixing the starting period r_1 to 0. The SADF is suited to detect a single period of unit-root behaviour in the sample (Phillips, Shi, and Yu 2015). In order to track explosiveness over a time series the BSADF test statistic is recursively calculated over a rolling estimation window:

$$BSADF_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2},$$

where r_0 denotes the minimal window size. This procedure can produce real-time exuberance levels by setting r_2 as the current period (the t-th period corresponds to $r_2 = t/T$), and letting the start of our estimation period r_1 vary from the beginning of our sample (0) to $r_2 - r_0$. The sequence of BSADF statistics for each city's HPI were generated as such using the code made available by the Lancaster University Management School (link). I retrieved monthly estimates of consumer excitement for each city. Note that the Federal Reserve Bank of Dallas maintains these "exuberance" statistics as a part of its US housing database (link).

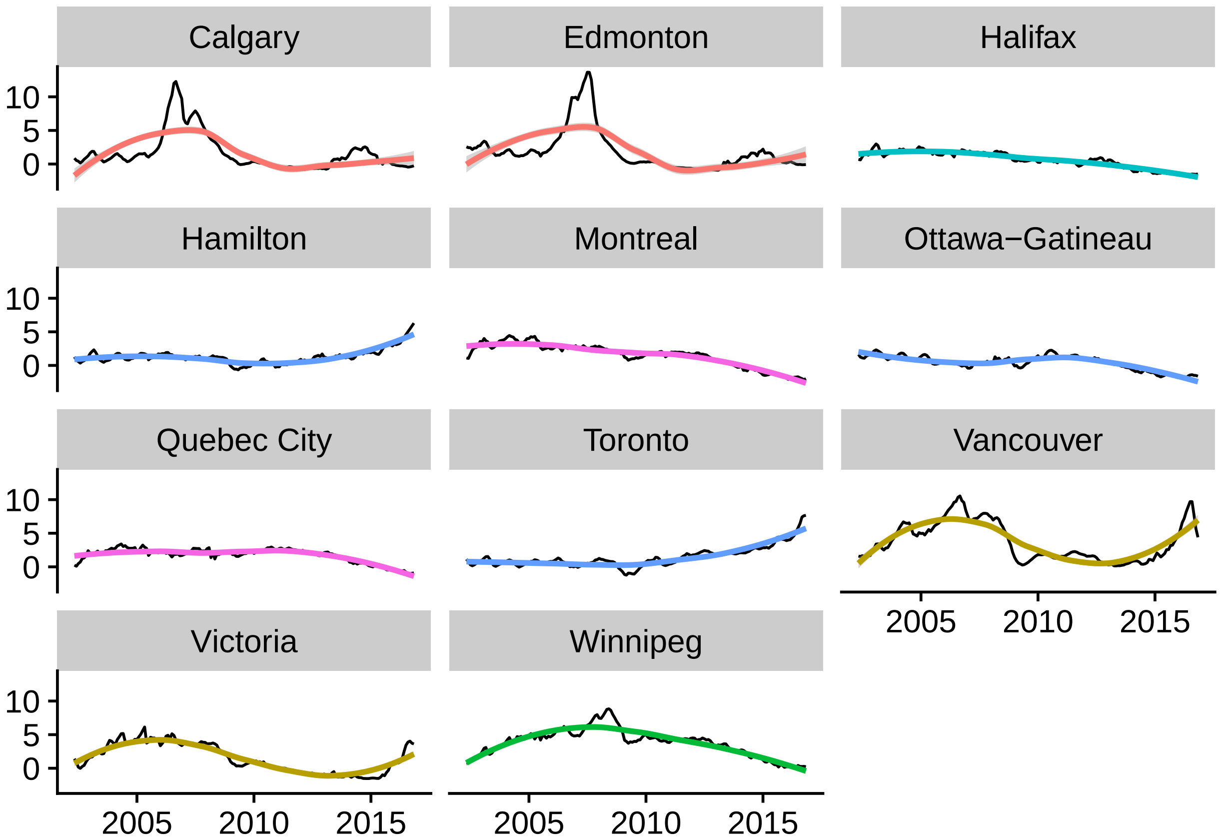


Figure 4: Series of BSADF test statistics with Loess curves

In Figure 4, monthly BSADF series are depicted along with Loess curves. Again, the latter are colour-matched according to provinces, and one can once again observe characteristic trends for Alberta, British Columbia, and Québec.

Demand Factors

The model takes account of the state of the local labour market with metropolitan unemployment rates. I used seasonally adjusted unemployment rates from the Labour Force Survey (link). Here, rising unemployment rates may stir doubt regarding real estate growth, so one can expect unemployment to be negatively correlated with regional house prices.

I also used seasonally adjusted city-specific population series from the Labour Force Survey (link) to compute monthly percent changes in population. Indeed, by taking into account changes in population size allows to control for aggregate housing demand.

Wages are also an evident demand factor to monitor the consumers' purchasing power, so I included monthly percent changes in seasonally adjusted, total provincial wages and salaries. The data used to construct our series come from Statistics Canada's estimates of labour income (link).

I also included rental vacancy rates from the Canada Mortgage and Housing Corporation (CMHC) Rental Market Survey data (link) in the model to take into accound shifts in preferences for homeownership or rentals. Here, I converted yearly rental vacancy rates to monthly observations using staircase interpolation. Homes and rentals substantially differ in terms of real estate investment strategies because a house is a large investment with some real estate risk, whereas rentals cannot generate capital gains but have the benefit of being virtually devoid of risk. In this regard, the two types of housing services may be viewed as economic substitutes. Provided the public views rentals and homes as respectively safe and risky investment strategies, low rental vacancy could indicate a flight to safety as home prices are expected to enter a downfall. Conversely, high rental vacancy rates could indicate a preference for home ownership in times of growth. On the other hand, high rental vacancy rates can be symptomatic of low demand, so the expected sign of the coefficient is ambiguous.

Since housing growth can be associated with a surge in productivity (Kahn 2009), I expect changes in real GDP to have a positive relationship with housing growth. As seen in Rünstler and Vlekke (2016), GDP and housing cycles are markedly synchronized across the 5 largest European economies and the US. Hence, there is potentially a similar relation with Canadian GDP. I first computed yearly changes in provincial GDP and then converted the series to monthly data using staircase interpolation. The original time series comes from Statistics Canada's expenditure-based real GDP series (link).

I added the national bank rates and average 5-year mortgage rates in the data panel to control for changes in credit accessibility and market expectations. The bank rate is the rate at which the Bank of Canada lend to private banks. A low bank rate stimulates growth, whereas a high bank rate aims to slow down inflation. It is important to recognize that consumer interest rates closely follow bank rates. The mortgage rate can be interpreted as a measure of risk that is specific to the residential mortgage credit. The data was retrieved from the Bank of Canada's data base (link).

Supply Factors

The number of residential building permits is a sensible candidate to measure the supply of residential construction. I included seasonally adjusted, monthly percent changes in the number of building permits reported in Statistics Canada's data base (link). This variable is included to take into account short-term sale expectations of real estate contractors.

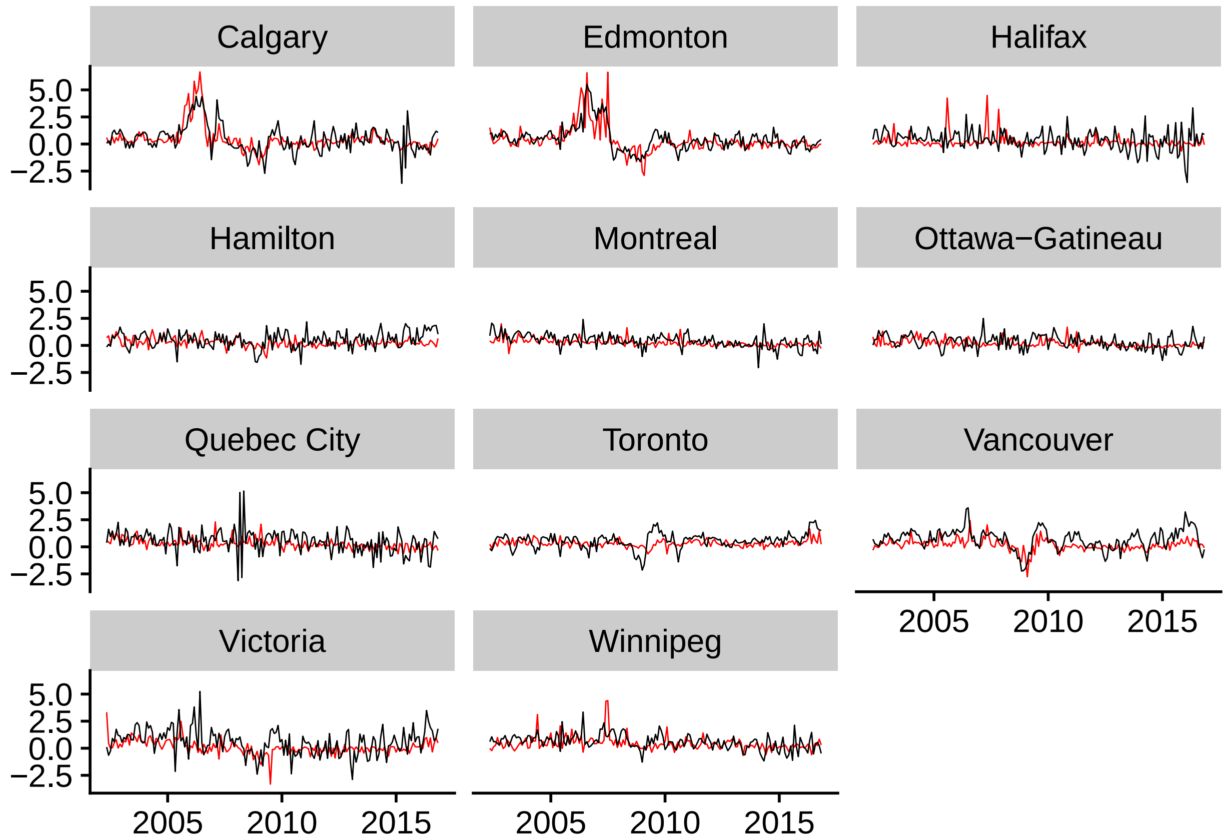


Figure 5: Historical series for percent changes in the HPI (black) and the New House Price Index (red)

I also included seasonally adjusted, monthly percent changes in the New Housing Price Index to measure changes in contractor's selling prices of new residential houses. City-specific series are available at Statistics Canada (link) and are plotted in red along with percent changes in the HPI (black) in Figure 5. Here, we can see that the percent changes in the New Housing Price Index tends to follow that of the Teranet HPI in cities like Calgary, Edmonton, and Vancouver, but not as much in other cities. This suggests that the response of housing returns to recent new housing returns varies from one region to the other.

It is widely accepted in the housing finance literature that market participants are highly influenced by recent price changes in the housing market. Indeed, the general public (including building contractors) tends to extrapolate recent housing trends when forming expectations of future housing returns (Case and Shiller 2003). This has the effect of spreading optimism, and building contractors adjust their prices accordingly. So, I expect recent changes in the New Housing Price Index to be positively correlated with percent changes in the HPI.

Financial Stability

Another covariate one might want to consider is the proportion of bank loans that are mortgage loans. I constructed this covariate by computing the ratio of Total Mortgage Loans to Total Loans for chartered banks using the Bank of Canada's data (link). This covariate can proxy the leverage of mortgage loans on the Canadian banking system.

Indeed, mortgage credit run-ups are commonly observed prior to housing market downturns in developed economies (Jordà, Schularick, and Taylor 2016b). Although Canada has never experienced a housing crash, history tells us that a market prone to public speculation and irrational excitement are prime suspects for market bubbles (Akerlof and Shiller 2010). That said, since mortgages take up a lion's share of chartered banks' balance sheets, it is fair to say that the flawed public speculation is a crucial systemic risk factor in developed economies' banking systems.

Model Setup

Because the model's purpose to predict monthly variations of house prices, I fitted HPI variations to observations from the previous month. I also considered fixed effects for provinces. The full model is formulated as follows:

$$\begin{aligned}\Delta HPI_{j,t} = \alpha_j + \beta_1 \Delta HPI_{j,t-1} + \beta_2 BSADF_{j,t-1} \\ + \beta_3 \Delta Permits_{j,t-1} + \beta_4 Unemployment_{j,t-1} \\ + \beta_5 \Delta Wages_{j,t-1} + \beta_6 \Delta Population_{j,t-1} + \beta_7 \Delta NewHPI_{j,t-1} \\ + \beta_8 Vacancy_{j,t-1} + \beta_9 \left(\frac{MortgageLoans}{TotalLoans}\right)_{t-1} \\ + \beta_{10} MRate_{t-1} + \beta_{11} BRate_{t-1} + \beta_{12} \Delta GDP_{j,t-1},\end{aligned}$$

where Δ is the percent change operator, j denotes one of the 11 cities, and α_j is the intercept for the j-th city. So, the model can either include 11 intercepts (one for each city), 8 intercepts (one for each province), or a single intercept, common to all entities (cities) in the data panel.

3. Results & Diagnostics

In this section, I examine multicollinearity in the data, and use iterative model building to highlight the driving factors of the Canadian housing market. In light of the different city-specific autocorrelation functions illustrated in Figure 2, it is appropriate to investigate the out-of-sample performance of a model with autoregressive slope parameters that differ with respect to provinces.

Multicollinearity

First, I examined the correlation matrix for a preliminary assessment of multicolinearity in the data. It is not uncommon to observe correlation between covariates with economic data. In Figure 6, it is not surprising to observe a high correlation between mortgage rates and bank rates or between percent changes in the HPI and percent changes in the New Housing Price Index.

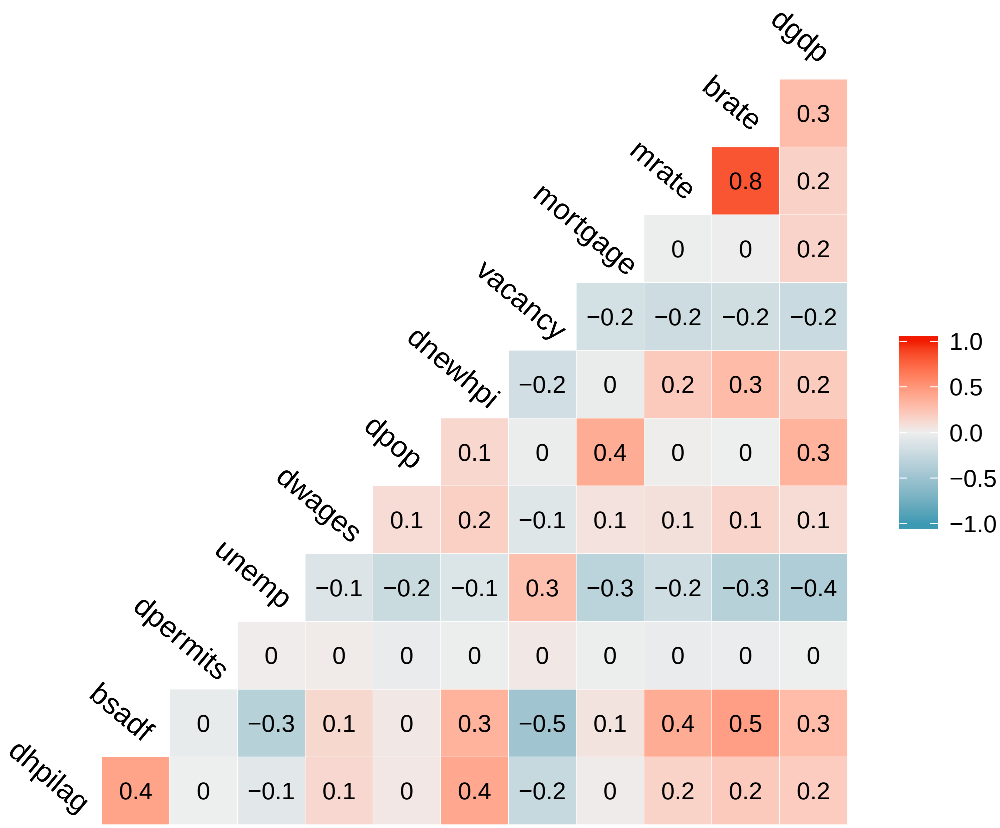


Figure 6: Correlation matrix

Although pairwise correlations can help in detecting multicollinearity, it is a somewhat limited tool because multicollinearity may arise from a linear correlation between more than 2 covariates. The Variance Inflation Factor (VIF) solves this problem by providing a measure of the impact of multicollinearity on the slope estimates' variance. My covariates' VIFs are illustrated in Figure 7. Although mortgage rates and bank rates appear to be the main source of multicollinearity, their VIFs are relatively low, suggesting that there is no issue with multicollinearity in my data.

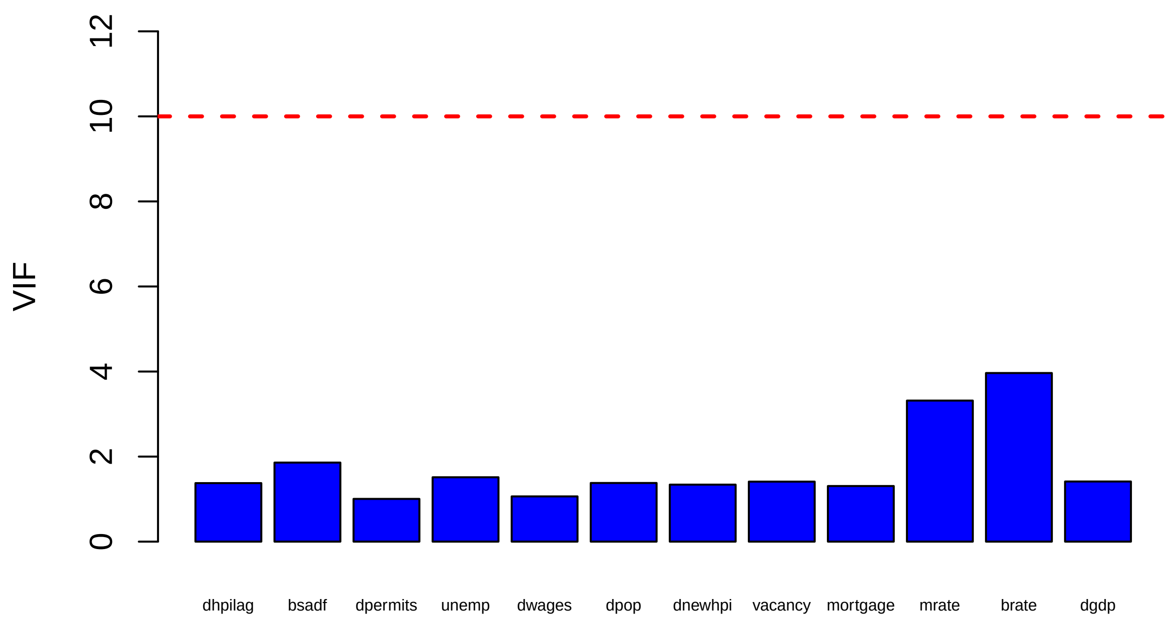


Figure 7: Variance Inflation Factors

Variable Selection

My goal here is to reduce my model to a more parsimonious form. In order to do so, I used stagewise variable selection procedures on a traning data set composed of the 120 first months (out of 175) in the data. The remaining 55 months are used for out-of-sample performance tests. In Table 1, summaries for the models selected by the AIC and the BIC criteria are presented.

Table 1: Models Selected by Stepwise Regression

	AIC	BIC
	(1)	(2)
Constant	0.525**	0.595**
	(0.247)	(0.245)
ΔHPI	0.319***	0.322***
	(0.027)	(0.027)
Unemployment	0.056***	0.054***
	(0.017)	(0.017)
ΔWages	0.065*	
	(0.033)	
ΔNew HPI	0.234***	0.243***
	(0.032)	(0.031)
Vacancy	−0.087***	−0.086***
	(0.019)	(0.019)
5-year Mortg. Rate	−0.125***	−0.136***
	(0.046)	(0.046)
Bank Rate	0.087***	0.094***
	(0.028)	(0.028)
ΔGDP	0.022***	0.023***
	(0.005)	(0.005)
Observations	1,320	1,320
R^2	0.294	0.292
Adjusted R^2	0.289	0.288
Residual Std. Error	0.784 (df = 1311)	0.785 (df = 1312)
F Statistic	68.146*** (df = 8; 1311)	77.170*** (df = 7; 1312)

Note: *p<0.1; **p<0.05; ***p<0.01

Using the AIC criterion, the preferred model considers the following covariates: percent changes in wages, mortgage rates, bank rates, unemployment, percent changes in the New Housing Price Index, and lagged percent changes in the HPI. Using the BIC criterion, the preferred model is the same with the exception that percent changes in wages are not included in the model. Note that the selected models do not include city-specific or province-specific fixed effects (intercepts), nor do they include the consumer sentiment proxy. Rather, all cities are assigned a common slope. I chose to further investigate the more parsimonious model selected by the BIC criterion, hereafter referred to as the simple model.

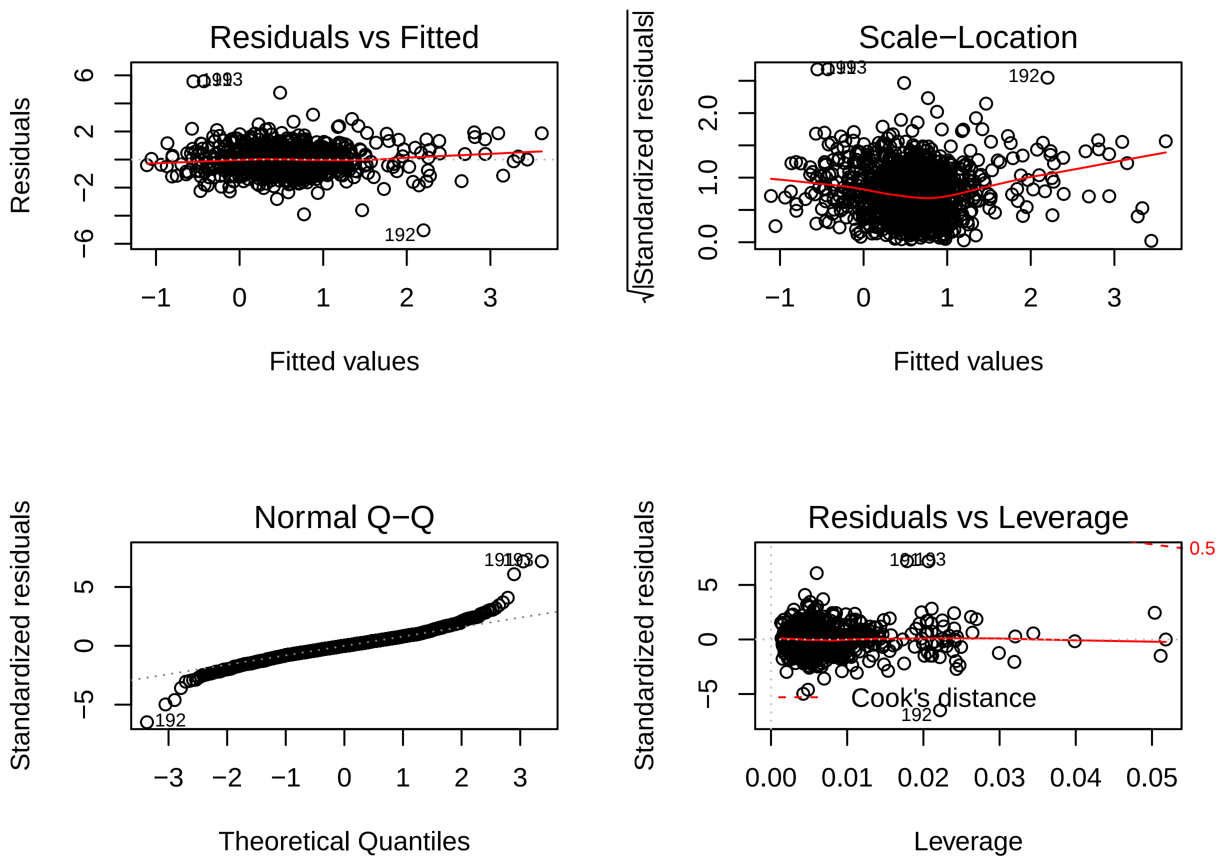


Figure 8: Diagnostic plots for the simple model

From the diagnostic plots in Figure 8, the *Residuals vs Fitted* and the *Scale-Location* plots do not indicate heteroscedasticty. Moreover, all the observations have a Cook's distance below 0.5 and the QQ-plot indicates a symetric, fat-tailed distribution of residuals. The Jarque-Bera test for normality (included in the `tseries` package) reports strong evidence that the residuals are not normally distributed with a p-value of 0.

Testing for Different Slopes

In Figure 9, autocorrelation appears to persist in the residuals, especially for Toronto, Edmonton, and Vancouver. Moreover, Halifax and Québec residuals have substantial first-lag autocorrelations below 0. Since the persistence in house price changes appear to have different structures from one province to another, forcing a common slope parameter for the ΔHPI and ΔNew HPI variables may result in a poor fit for cities with characteristic autocorrelation structures. Perhaps some provinces have strong autocorrelation of house price while others do not. In turn, I considered more flexible models that allow for ΔHPI and ΔNew HPI coefficients to vary across provinces.

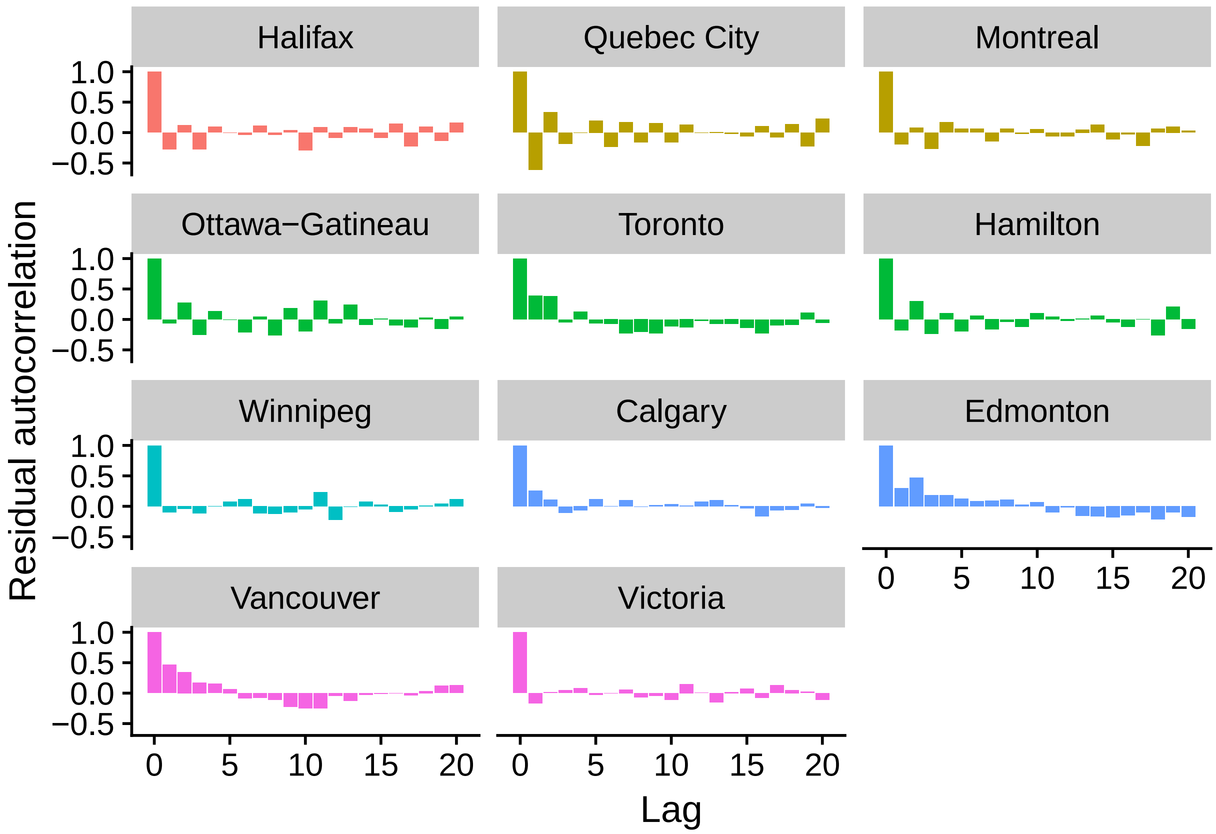


Figure 9: Autocorrelation structures for residuals

In Tables 3 and 4 are reported ANOVA test statistics for 2 sets of nested model: one with different ΔHPI slopes and one with different ΔNew HPI slopes. Table 3 shows ANOVA tests results for the simple model (1), the model with province-specific ΔHPI slopes (2), and the model with both province-specific ΔHPI and ΔNew HPI slopes (3). Similarly, Table 4 shows results for the same procedure with the simple model (1), the model with province-specific ΔNew HPI slopes (2), and the model with both province-specific ΔHPI and ΔNew HPI slopes (3). Here, the ANOVAs test the models against one another in the order mentionned. In both cases, the use of a full set of province-specific covariates appears to be the preferred model specification.

Table 2: ANOVA test for equality of ΔHPI slopes

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1312	808.48				
2	1307	727.64	5	80.84	29.41	0.0000
3	1302	715.80	5	11.84	4.31	0.0007

Table 3: ANOVA test for equality of ΔNew HPI slopes

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1312	808.48				
2	1307	782.86	5	25.62	9.32	0.0000
3	1302	715.80	5	67.06	24.40	0.0000

Indeed, the simple model is rejected in favour of the model with different ΔHPI slopes with a p-value of 0% and 0% for the model with different ΔNew HPI slopes. Then, the models with different slopes are compared to a more complex model, where both the ΔHPI and ΔNew HPI coefficient vary across cities. The model with different ΔHPI slopes is rejected in favour of the full model with a p-value of 0.1%, and likewise for the model with different ΔNew HPI with a p-value of 0%.

One can notice that the ΔHPI slope coefficient in the simple model is 0.32, but when the model allows for different slope coefficients, Alberta's coefficient jumps to 0.6, whereas Nova Scotia and Québec's coefficient drop to 0.1 and -0.1. The other provinces coefficient remain stable. Also, the ΔNew HPI slope coefficient in the simple model is 0.24, but Manitoba, Nova Scotia, and Ontario slopes are now close to 0.

Also, note that from the simple model to the model with flexible ΔHPI slope, the adjusted R^2 goes from 0.288 to 0.357, a somewhat sizeable gain. The full model also appears to be a significant improvement from the simple model, but the ΔNew HPI slope estimates are somewhat unstable from a model to another, so I chose the model with different ΔHPI slopes and a common ΔNew HPI slope.

In Figure 10, the different-slope model effectively reduces the residual autocorreation observed in the simple model for Edmonton, but not quite so for Toronto and Vancouver. The different-slope model also attenuates the increasing slope in the *Location-Scale* graph in Figure 11.

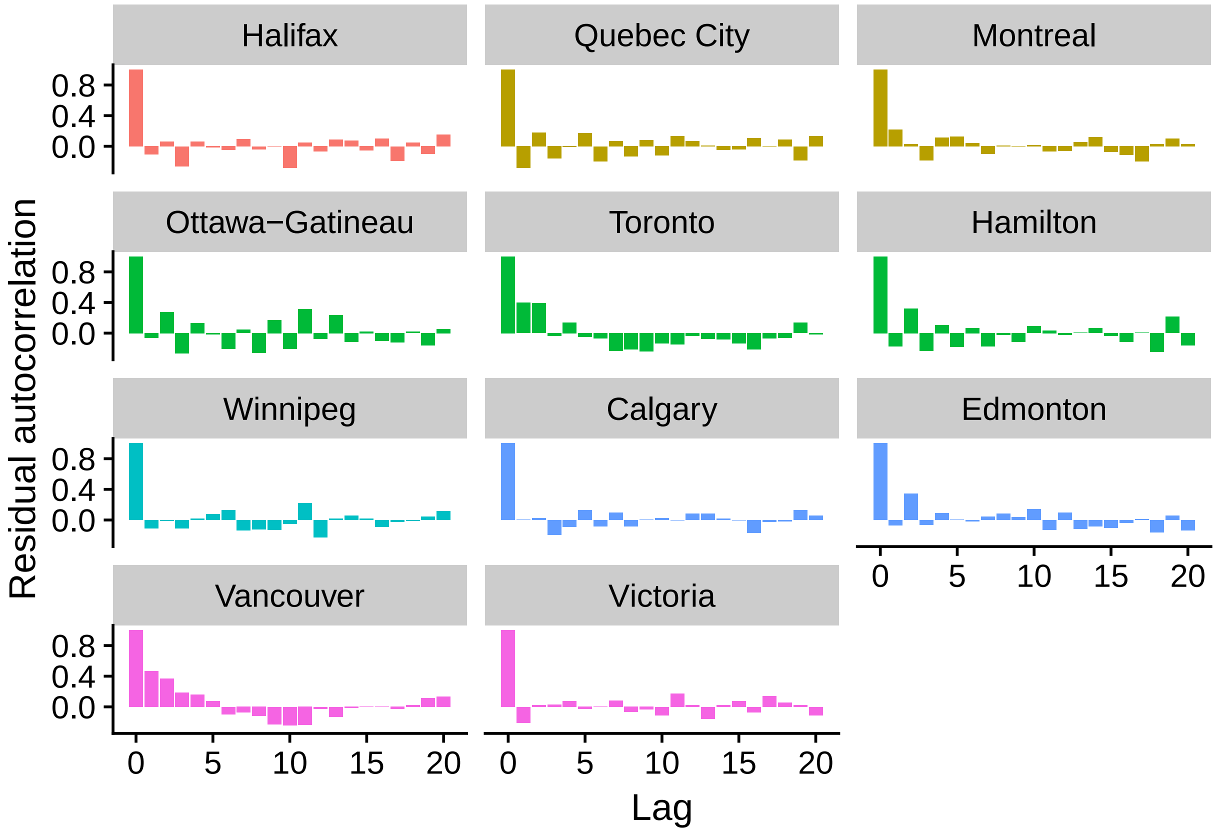


Figure 10: Autocorrelation structures for residuals for the model with province-specific slope parameters

Table 4: Model Summaries

	Common Slopes	Different ΔHPI Slopes	Different ΔNew HPI Slopes	Different Slopes for Both
	(1)	(2)	(3)	(4)
Constant	0.595**	0.372	0.448*	0.363
	(0.245)	(0.235)	(0.245)	(0.236)
Unemployment	0.054***	0.099***	0.075***	0.096***
	(0.017)	(0.017)	(0.018)	(0.017)
Vacancy	−0.086***	−0.123***	−0.089***	−0.108***
	(0.019)	(0.019)	(0.020)	(0.020)
5-year Mortg. Rate	−0.136***	−0.118***	−0.121***	−0.117***
	(0.046)	(0.044)	(0.046)	(0.044)
Bank Rate	0.094***	0.108***	0.103***	0.110***
	(0.028)	(0.026)	(0.027)	(0.026)
ΔGDP	0.023***	0.014***	0.019***	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)
ΔHPI	0.322***		0.292***	
	(0.027)		(0.027)	
ΔNew HPI	0.243***	0.136***		
	(0.031)	(0.032)		
AB ΔHPI		0.613***		0.574***
		(0.042)		(0.049)
BC ΔHPI		0.362***		0.328***
		(0.042)		(0.048)
MB ΔHPI		0.334***		0.412***
		(0.074)		(0.087)
NS ΔHPI		0.104		0.118
		(0.082)		(0.083)
ON ΔHPI		0.316***		0.369***
		(0.053)		(0.055)
QC ΔHPI		−0.124**		−0.177***
		(0.050)		(0.054)
AB ΔNew HPI			0.372***	0.187***
			(0.039)	(0.047)
BC ΔNew HPI			0.307***	0.262***
			(0.075)	(0.080)
MB ΔNew HPI			0.073	−0.012
			(0.082)	(0.093)
NS ΔNew HPI			−0.040	0.005
			(0.096)	(0.094)
ON ΔNew HPI			−0.060	−0.142
			(0.093)	(0.093)
QC ΔNew HPI			−0.032	0.340***
			(0.100)	(0.104)
Observations	1,320	1,320	1,320	1,320
R^2	0.292	0.362	0.314	0.373
Adjusted R^2	0.288	0.357	0.308	0.365
Residual Std. Error	0.785 (df = 1312)	0.746 (df = 1307)	0.774 (df = 1307)	0.741 (df = 1302)
F Statistic	77.170*** (df = 7; 1312)	61.927*** (df = 12; 1307)	49.877*** (df = 12; 1307)	45.533*** (df = 17; 1302)

Note: *p<0.1; **p<0.05; ***p<0.01

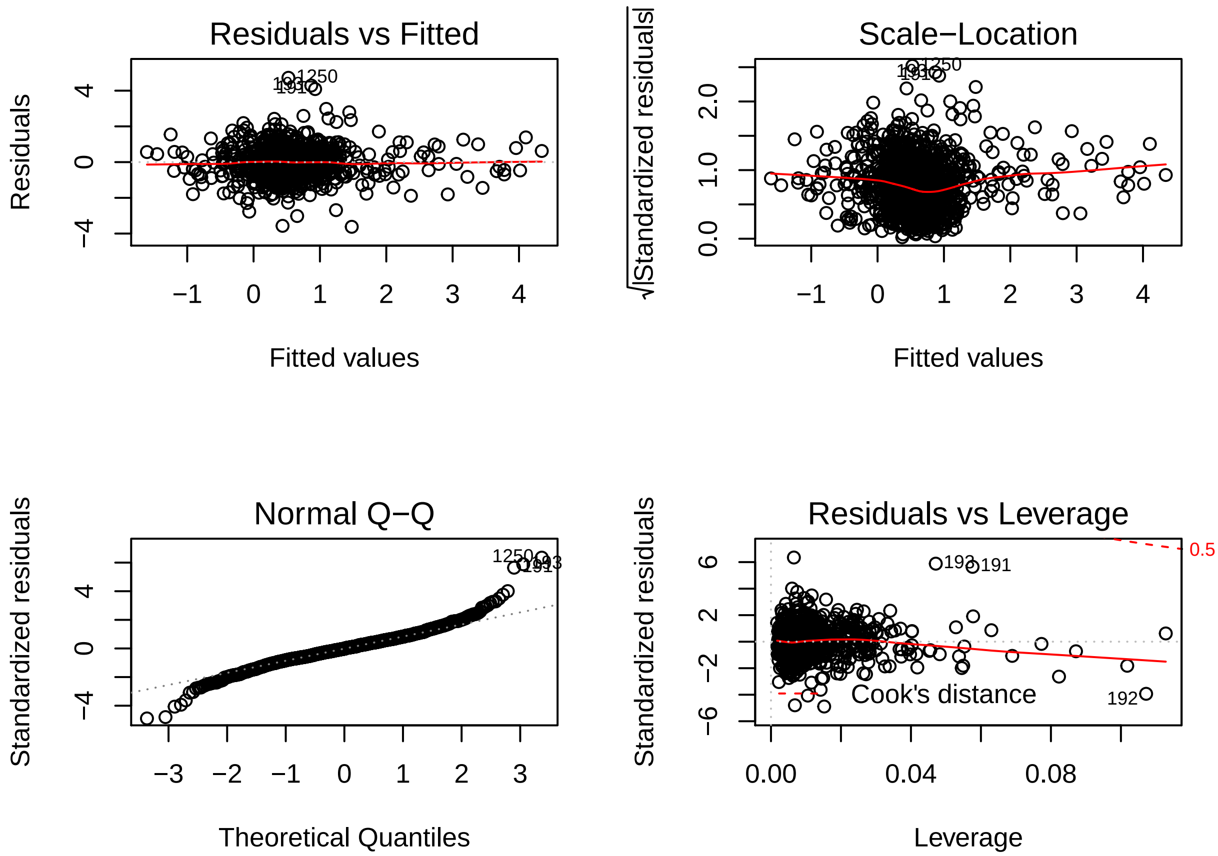


Figure 11: Diagnostic plots for the model with different slopes

Out-of-sample Fit

I measured the improvement made by the different-slope model in reducing residual autocorrelation by computing Durbin-Watson tests for each city's first lag. In Figure 12, a heatmap is produced using the funtion `pheatmap` from the `pheatmap` package.

A Durbin-Watson (DW) test statistic between 2 and 4 indicates negative serial correlation, and a DW test statistic between 0 and 2 indicates positive serial correlation. Here, the different-slope model deals better with serial correlation than the simple model in the training data. In the test data however, the two seem to have similar performance. Indeed, the simple model fails to tackle residual autocorrelation for Toronto, Calgary and Vancouver in the test data. The different-slope model shows residual autocorrelation in Montréal, Toronto, Calgary, and Vancouver.

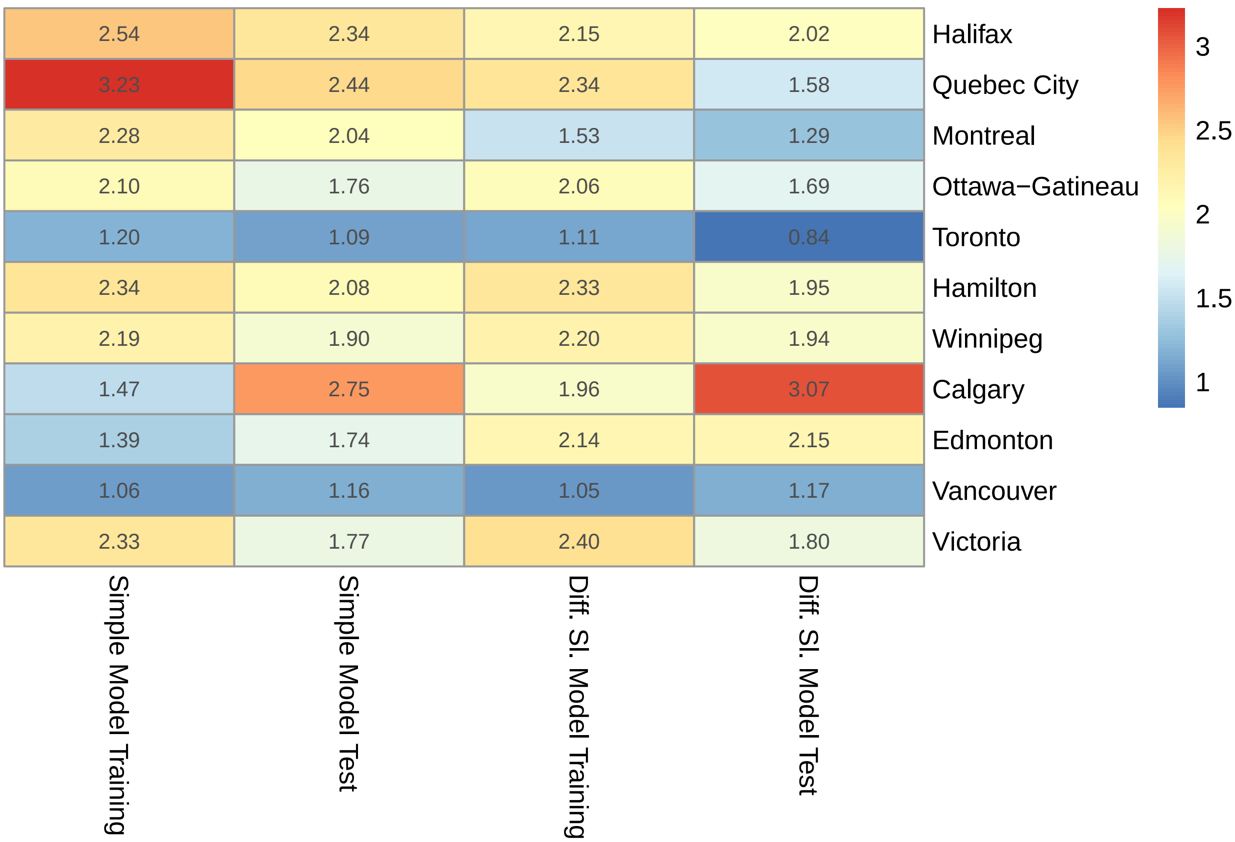


Figure 12: Durbin Watson Test Statistics

One can also observe that the simple model performs better in terms of RMSE in the test data, even though the different-slope model performs better on the training data (as expected). This is an example of the over-training dilemma: the prediction error of a model increases with its complexity. In this case, the simple model can be considered as the the *High Bias/Low Variance* model (see Figure 13), which is the model I finally chose.

Table 5: RMSE

Simple Model Training	Simple Model Test	Diff. Sl. Model Training	Diff. Sl. Model Test
0.783	0.863	0.742	0.880

In Figure 13, the historical data (black) is plotted along with the fits of the simple model (blue) and the different-slope model (red), along with 95% prediction interval (dashed lines). In Figure 14, there is no visible pattern in the Residual Dependence plots, so error terms appear homoscedastic, although autocorrelation is present for some cities.

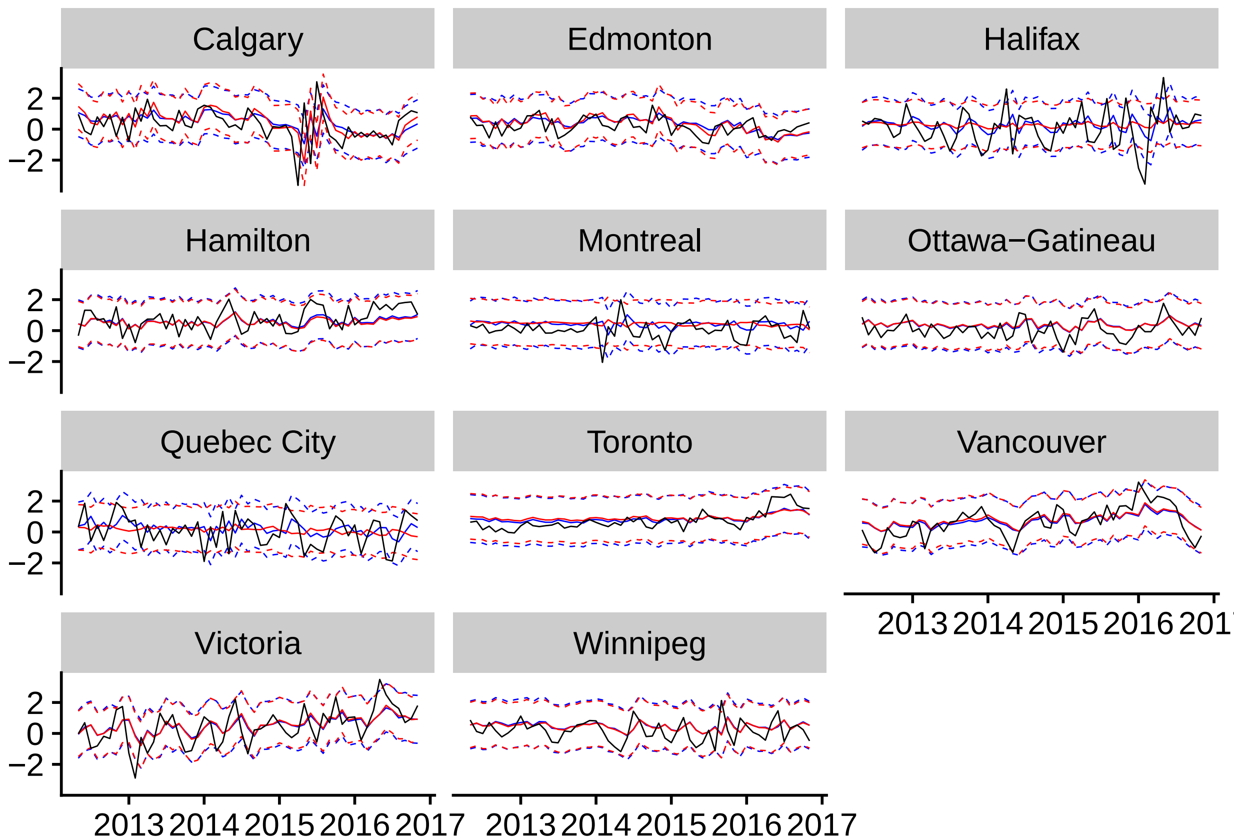
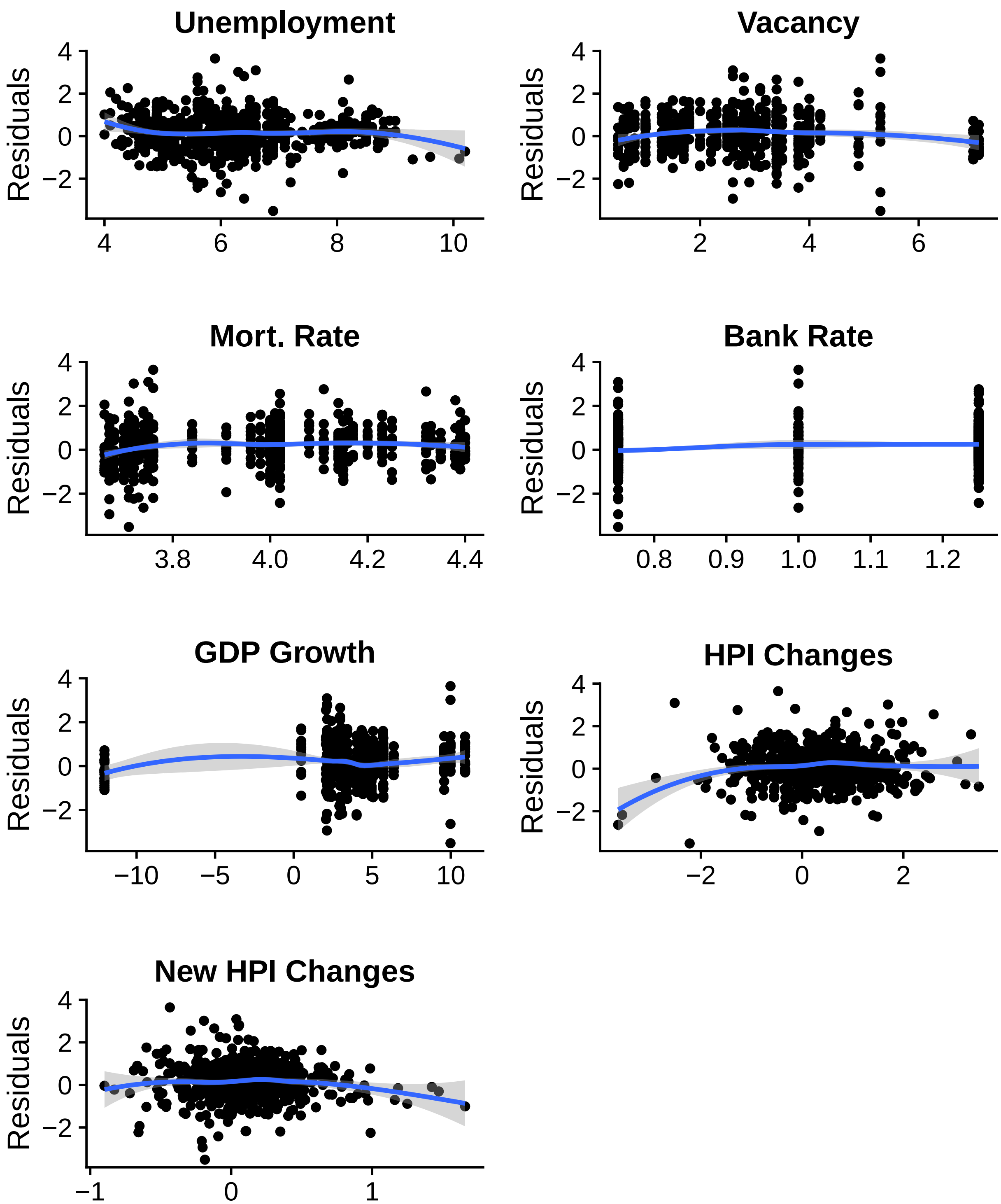


Figure 13: Test data fitted values with 95% prediction intervals


Unemployment
Residuals
4 2 0 −2
4 6 8 10
Vacancy
Residuals
4 2 0 −2
2 4 6
Mort. Rate
Residuals
4 2 0 −2
3.8 4.0 4.2 4.4
Bank Rate
Residuals
4 2 0 −2
0.8 0.9 1.0 1.1 1.2
GDP Growth
Residuals
4 2 0 −2
−10 −5 0 5 10
HPI Changes
Residuals
4 2 0 −2
−2 0 2
New HPI Changes
Residuals
4 2 0 −2
−1 0 1

4. Conclusion

To conclude, I proposed a Canadian house price model with fundamental variables and consumer sentiment variable. Using stepwise regressions I found that the most important predictors of house prices are mortgage rates, bank rates, unemployment, percent changes in the New Housing Price Index, and lagged percent changes in the HPI, which account for 29% of the variance. Since some provinces had characteristic autocorrelation structures, I investigated whether different slopes for the autoregressive component could help my model reflect province-specific persistence in prices. Residual autocorrelation tests seem to indicate that the different-slope model reduces autocorrelation in the training data, but does not eliminate it. In the test data, it appears the simple model performs better out of sample, with a lower RMSE and comparable residual autocorrelation. So, I chose the simple model as the preferred regression.

In the future, a modelling approach that one might want to consider to correctly account for non-identically distributed residuals is using *clusters*. Data clustering is a useful regression tool widely used in econometrics when the distribution of the error terms in a panel regression depends on the entity they belong to. Indeed, the residuals' volatility may vary from an entity to another. Even more interestingly, this technique allows the user to estimate correlations *across* the panel's entities. In a house price model, cross-sectional dependence is a crucial aspect to consider. Indeed, copula-based model specification tests for US house prices at the state level show strong evidence of cross-sectional dependence (Zimmer 2012). So, it is reasonable to assume that Canadian house prices are susceptible to display a similar dependence structure.

5. References

Akerlof, George A, and Robert J Shiller. 2010. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* Book. Princeton, New Jersey: Princeton University Press Princeton. doi:10.1002/hrm.20337.

Case, Karl E, and Robert J Shiller. 2003. "Is There a Bubble in the Housing Market?" Journal Article. *Brookings Papers on Economic Activity* 2003 (2): 299–342. doi:10.1353/eca.2004.0004.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2016a. "Macrofinancial History and the New Business Cycle Facts." Book Section. In *NBER Macroeconomics Annual 2016, Volume 31*. University of Chicago Press. doi:10.3386/w22743.

———. 2016b. "The Great Mortgaging: Housing Finance, Crises and Business Cycles." Journal Article. *Economic Policy* 31 (85): 107–52. doi:10.3386/w20501.

Kahn, James A. 2009. "Productivity Swings and Housing Prices." Journal Article. *Current Issues in Economics and Finance* 15 (July 3). doi:10.2139/ssrn.1434114.

Pavlidis, Efthymios, Alisa Yusupova, Ivan Paya, David Peel, Enrique Martínez-García, Adrienne Mack, and Valerie Grossman. 2016. "Episodes of Exuberance in Housing Markets: In Search of the Smoking Gun." Journal Article. *The Journal of Real Estate Finance and Economics* 53 (4): 419–49. doi:10.1007/s11146-015-9531-2.

Phillips, Peter CB, and Jun Yu. 2011. "Dating the Timeline of Financial Bubbles During the Subprime Crisis." Journal Article. *Quantitative Economics* 2 (3): 455–91. doi:10.3982/QE82.

Phillips, Peter CB, Shuping Shi, and Jun Yu. 2015. "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500." Journal Article. *International Economic Review* 56 (4): 1043–78. doi:10.1111/iere.12132.

Rünstler, Gerhard, and Marente Vlekke. 2016. "Business, Housing and Credit Cycles." Working Paper. European Central Bank. doi:10.1111/j.1468-0289.2010.00551_31.x.

Zimmer, David M. 2012. "The Role of Copulas in the Housing Crisis." Journal Article. *Review of Economics and Statistics* 94 (2): 607–20. doi:10.1162/REST_a_00172.