RPS Real Property Solutions

RPS House Price Index

Product & Methodology Information

Date: January 2024

RPS House Price Index Overview

The RPS House Price Index (HPI) is recognized as the most comprehensive source for house price trends in Canada. House price trends are provided for thousands of discreet local markets and five core property types in all Canadian markets, delivering the most robust national view of Canadian house price trends.

Current clients include federally regulated financial institutions, mortgage investment companies, mortgage brokerages, government, economists, and other financial industry participants.

The RPS HPI leverages RPS' extensive, unique residential property database comprising millions of transactions that are updated continuously from sources across Canada. RPS employs a stratified central tendency approach[1] that enables the delivery of robust house price trend information down to the neighbourhood level, across multiple property types including the high, median and low ends of these real estate markets.

The HPI dataset delivered to clients is highly structured and includes detailed, supplemental information for added utility and enhanced insights.

RPS House Price Index - Public Release

The RPS HPI is available as a complimentary monthly public release for download from the RPS website that includes median index and dollar values back to the base period of January 2005 for the Aggregate property type (weighted combination of all property types in each geography) including:

- National Index – the weighted median of the top 1,000 Census Subdivisions (CSDs) across Canada based on population. A CSD is a municipality determined by provincial / territorial legislation, or an area deemed to be equivalent to a municipality for statistical reporting purposes. There are 5,161 CSDs in Canada as of the 2021 Census. Many are sparsely populated, however the top 1,000 CSDs used in the RPS HPI measured by population account for 92% of the total population of Canada.
- National Index Published – the weighted median of the top 1,000 Census Subdivisions (CSDs) across Canada based on population. The current month's values for this index are calculated exactly as the National Index, however the historical values are not adjusted and remain as originally published.
- Weighted National Index – represents the weighted median of 13 Census Metropolitan Areas (CMAs) – Calgary, Edmonton, Halifax, Hamilton, Montréal, Ottawa-Gatineau, Québec, Regina, Saskatoon, Toronto, Vancouver, Victoria, and Winnipeg.
- CMA Indices for Calgary, Edmonton, Halifax, Hamilton, Montréal, Ottawa-Gatineau, Québec, Regina, Saskatoon, Toronto, Vancouver, Victoria, and Winnipeg.

[1] Stratified central tendency is one of four main approaches to constructing a house price index described in the *Handbook on Residential Property Price Indices* (RPPIs). The Handbook was developed in coordination with the Statistical Office of the European Union (Eurostat) under the joint responsibility of six organizations, including the International Monetary Fund (IMF).

RPS House Price Index – Enterprise Solution

The comprehensive RPS HPI Enterprise solution is delivered monthly to subscribers. This complete view of national house price trends includes data that:

- Covers: National, 10 provincial, 33 CMA, 1,000 city, and 1,500 neighbourhood geographies
- Includes: aggregate, single family detached, semi-detached, row / townhouse, condominium apartment and plex property types, where available
- Provides: median index value and high, median, low dollar values
- Includes: additional geographic hierarchy identifiers, confidence score, national and regional rankings, urban/rural indicator, neighbourhood names, and census population and dwelling count data

Access to the HPI Enterprise solution requires a subscription and is geographically configurable.

HPI Enterprise Solution Use Cases

The comprehensive HPI Enterprise solution enables mortgage lenders, mortgage insurers and other real estate focused industry organizations to:

- Gain extensive visibility into real estate markets and the related impacts to their property portfolios
- Provide improved reporting to stakeholders (executives, investors, regulators and more)
- Make smarter, more informed decisions related to risk management and identification of new business opportunities
- Perform analysis at more localized and precise geographic segmentation to distinguish differential behaviours between smaller markets and larger urban centers
- Perform analysis across property types and value segments in the same geographies to uncover trends that could impact the way they do business

The RPS HPI provides enterprises, regardless of size or national footprint, with the insights needed to evaluate markets and better manage their business.

RPS Real Property Solutions Inc. (RPS) is a Brookfield company and a leading Canadian provider of outsourced appraisal management, mortgage-related services and real estate business intelligence to financial institutions, real estate professionals and consumers. The company's expertise in real estate valuation, together with its proprietary property database and innovative technologies and services, has established RPS as the trusted source for Canadian residential real estate intelligence and analytics.

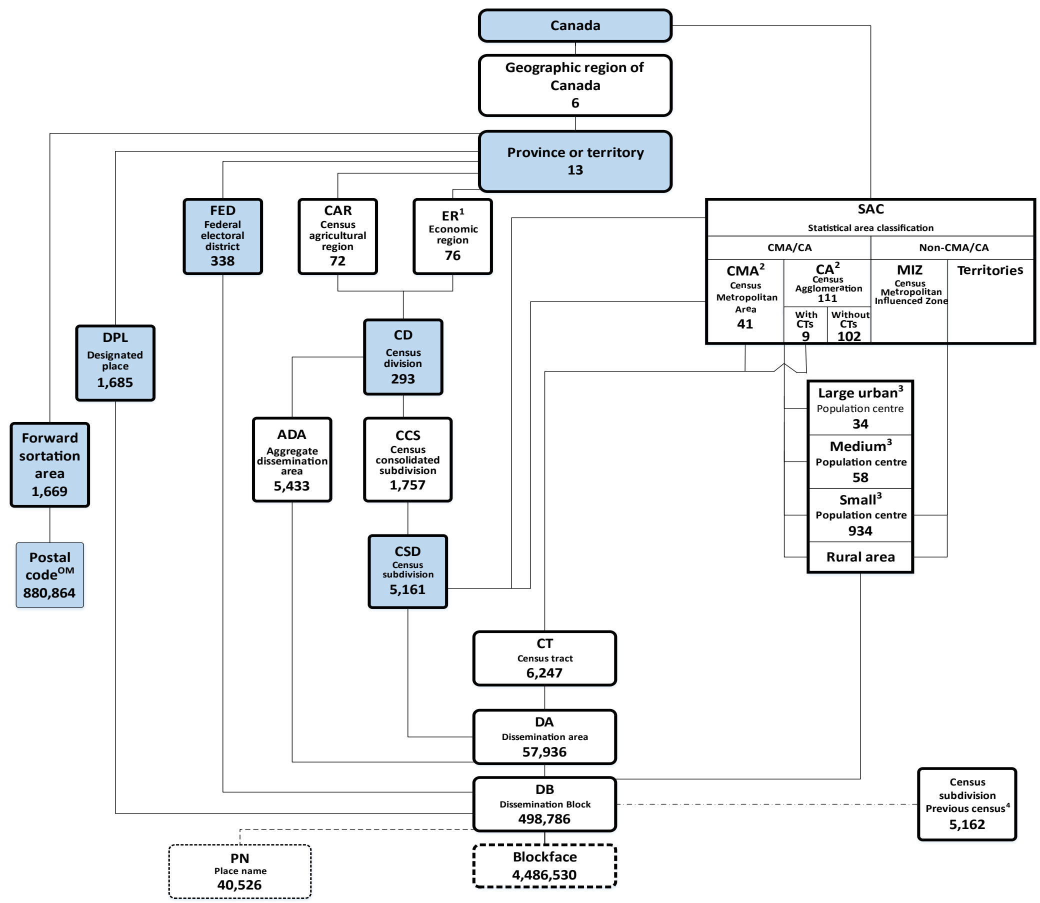
The RPS House Price Index is part of a complete suite of risk and valuation solutions, including portfolio analytics services, automated valuation models (AVMs), supplementary market reports like our homogeneity reports and house price forecasts and more.

HPI Geographic Stratification and Coverage

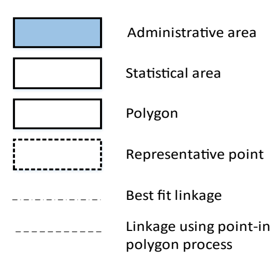
The RPS HPI uses Statistics Canada and Canada Post geographic boundary standards to stratify indices into five geographic levels: National, Provincial, Metropolitan (CMA), City (CSD), and Neighbourhood (FSA).

Figure 1 shows the Statistics Canada geographic hierarchy which provides context for the subsequent descriptions of the RPS geographic stratifications provided in the RPS HPI.

Figure 1: Statistics Canada Geographic Hierarchy



1. Economic regions (ER) are composed of complete census divisions (CD) except for one CD in Ontario.
2. Some census metropolitan areas (CMA) and census agglomerations (CA) cross provincial boundaries.
3. Previous census population centres are used in the delineation of the current census CMAs and CAs ; some population centres (POPCTR) cross provincial boundaries.
4. A best fit linkage is created between the census subdivisions (CSD) - previous census and the current census dissemination blocks (DB) to facilitate historical data retrieval.



Source: Statistics Canada, 2021 Census of Population.

National Composite Indices

The RPS HPI has three national composite indices:

1. National Index – the weighted median of all markets across Canada, including all towns and cities. This robust view accurately assesses house price trends across Canada, using Census Subdivisions (CSDs) as the base data. A CSD is a municipality determined by provincial / territorial legislation, or an area deemed to be equivalent to a municipality for statistical reporting purposes. There are 5,161 CSDs in Canada as of the 2021 Census. Many are sparsely populated, however the top 1,000 CSDs used in the HPI measured by population account for 92% of the total population of Canada.
2. National Index Published – new for the January 2023 data series, and to coincide with implementation of RPS' methodology enhancement described in Appendix A, this index is the assembly of published monthly National Index values. Historical values are not adjusted and remain as originally published.
3. Weighted National Index – the weighted median of 13 CMAs – Calgary, Edmonton, Halifax, Hamilton, Montréal, Ottawa-Gatineau, Québec, Regina, Saskatoon, Toronto, Vancouver, Victoria, Winnipeg

HPI Index Value Adjustments

RPS recalculates index values every month starting from the base period of January 2005. The historical index values may change from month to month due to:

- Model smoothing and trending – the RPS HPI model design, more fully described beginning on Page 8, refines historical index values as a function of monthly processing.
- Annual re-weighting – weighted indices are updated annually to ensure accurate contribution of underlying stratifications to the weighted index, this includes property type contributions to the Aggregate property type.
- New data sources / updated data – RPS may procure additional property transaction data or, as part of its continuous data management practices, be able to use previously deficient data in the HPI model. An example of enabling previously unusable data would be appending property type information to a property record that did not contain property type data.

These changes are monitored; mean absolute percentage change of historical index values for the previous six HPI National Index publications (202212 – 202305) for the Aggregate property type was 0.01537%.

Provincial Indices

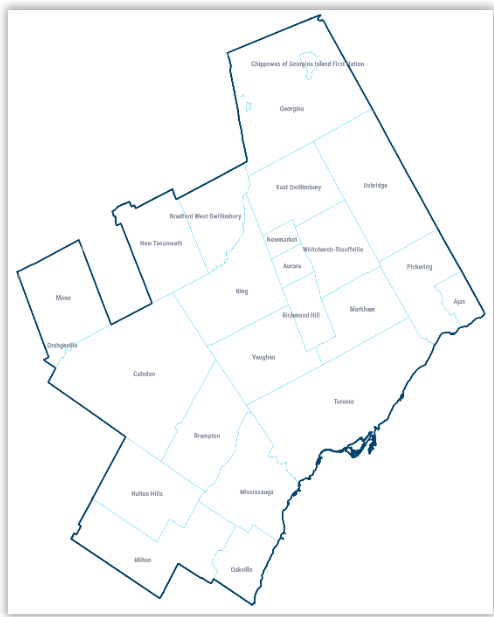
Canada is divided into ten provinces and three territories. Province and territory designations refer to second-tier political geographies in Canada. According to Statistics Canada, the occupied private dwelling counts by province, territory, and structural dwelling type as of 2021 are:

Statistics Canada 2021 Census - Occupied Private Dwellings							
Province/Territory	Total Dwellings		Type of Dwelling [2]				
	Count / National %		Single Detached	Semi-Detached	Row	Apartment	Other
Alberta	1,633,220	10.9%	994,560	98,740	127,735	365,645	46,540
British Columbia	2,041,835	13.6%	866,340	62,885	168,585	889,165	54,855
Manitoba	518,055	3.5%	343,990	18,185	19,720	125,975	10,185
New Brunswick	337,650	2.3%	228,945	13,490	9,735	70,130	15,330
Newfoundland	223,250	1.5%	161,410	8,695	10,770	40,670	1,705
Northwest Territories	15,210	0.1%	8,600	1,070	1,630	3,180	730
Nova Scotia	428,225	2.9%	272,980	21,605	11,220	106,390	16,040
Nunavut	9,925	0.1%	4,280	960	3,020	1,655	15
Ontario	5,491,200	36.7%	2,942,995	303,255	505,270	1,714,480	25,205
Prince Edward Island	64,570	0.4%	43,855	3,640	2,680	11,625	2,770
Quebec	3,749,035	25.0%	1,671,925	199,085	98,625	1,739,895	39,500
Saskatchewan	449,580	3.0%	322,070	13,675	19,865	84,130	9,840
Yukon	17,180	0.1%	10,355	1,270	1,255	2,735	1,570
Canada	**14,978,940**	**100.0%**	**7,872,305**	**746,560**	**980,110**	**5,155,670**	**224,300**

Metropolitan Indices

RPS metropolitan indices refer to a CMA. CMAs consist of one or more neighbouring municipalities codified by Statistics Canada as Census Subdivisions (CSDs) that are situated around a core. Examples of CMAs are the Greater Vancouver Area (GVA), Greater Montreal Area (GMA) and Greater Toronto Area (GTA).

The Toronto CMA, shown to the right, is comprised of the municipality of Toronto (the core) and the CSDs of Ajax, Aurora, Bradford West Gwillimbury, Brampton, Caledon, East Gwillimbury, Georgina, Halton Hills, King, Markham, Milton, Mississauga, Mono, New Tecumseth, Newmarket, Oakville, Orangeville, Pickering, Richmond Hill, Toronto, Uxbridge, Vaughan, and Whitchurch-Stouffville.



The Toronto Index Published – new for the January 2024 data series, this index is the assembly of published monthly Toronto Metropolitan (CMA) level Index values. Historical values are not adjusted and remain as originally published.

[2] Type of dwelling refers to structural characteristics / dwelling configuration. Apartment includes all stories purpose-built rental and condominium / strata, and flats in duplex. Other includes mobile homes and other single attached homes.

Per the 2021 Census, the Toronto CSD had 1.16M private dwellings and a population of 2.79M, while the Toronto CMA had 2.26M private dwellings and a population of 6.20M.

City Indices

Cities (or towns) identified in the RPS HPI are codified using Statistics Canada Census Subdivision (CSD) geographies. A CSD is a municipality determined by provincial / territorial legislation, or an area deemed to be equivalent to a municipality for statistical reporting purposes, such as a First Nations reserve.

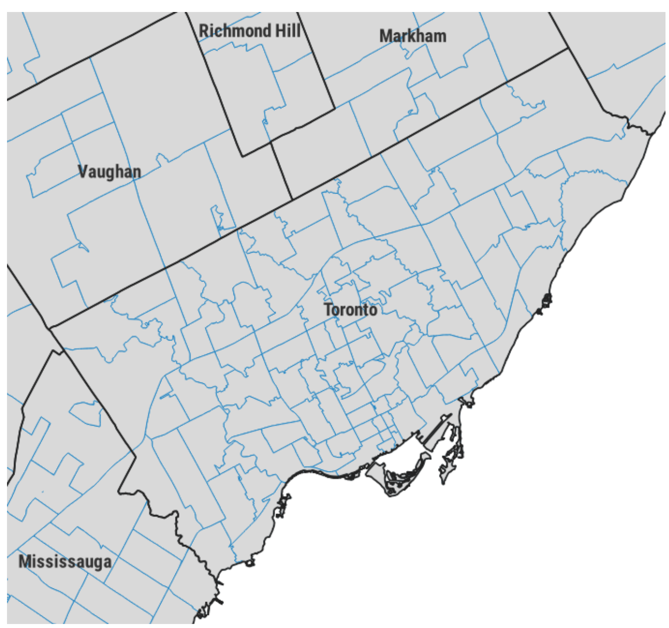
There are currently 44 different CSD types, ranging from City, Hamlet, Municipality, Parish, Town, Village, Settlement, etc.

There are 5,161 CSDs in Canada as of the 2021 Census. Many are sparsely populated, however the top 1,000 CSDs measured by population account for 92% of the total population of Canada.

Neighbourhood Indices

The RPS HPI defines a neighbourhood as a Canada Post Forward Sortation Area (FSA). An FSA is denoted by the first three characters of a postal code. There are about 1,670 FSAs in Canada. Most FSAs are populated (~1,500) while the remainder are either sparsely populated or cover non-residential areas. The RPS HPI covers the 1,500 most populous FSAs across Canada which equates to areas where over 90% of people reside.

In urban areas, FSAs tend to be bounded by major roads encompassing neighbourhoods with common econometric characteristics. The City of Toronto is currently comprised of 104 FSAs. The blue lines on the map to the right show the City of Toronto's FSAs. An FSA is large enough to provide the required monthly property transaction volumes to compute reliable house price trend information. FSAs are discreet enough however to generally represent homogeneous areas from which discernable localized home price changes can be derived to enable meaningful in-depth analyses at the neighbourhood level. Neighbourhoods do not always mirror CSD / CMA level trends. Neighbourhood indices reflect the realities of distinct local markets and equip clients to analyze local markets independently.



In rural areas where FSAs cover larger land areas, at times the corresponding city index may be a more appropriate dataset to perform analyses. The RPS HPI includes urban and rural geographic indicators to assist clients with determining how to best implement the indices.

RPS HPI Property Type Definitions

In addition to geographic stratification, the RPS HPI stratifies indices by five residential property types and an aggregate property type for each level of geography. Some property types are typically found only in certain markets – semi-detached properties are usually located in larger urban centers, and plex properties are predominantly located in Quebec. RPS HPI data is only provided for the types of properties located in each level of geography. Multi-purpose properties are not included in the HPI.

- Condominium apartments are units in a building owned by individuals. Common element parts of the property are owned jointly by all unit owners.
- Single Family Detached are houses that are not attached to another house and may be one level or multi-level.
- Semi-Detached houses are built side by side and share a common wall; usually each house’s layout is a mirror image of its attached neighbour.
- Row / Townhouse are built side by side as units sharing one or more common walls, usually in such a way that each house’s layout is very similar; tenure of these properties can be either condo or freehold.
- Plex houses are buildings with apartments that have separate entrances for multiple households, and include properties labelled duplex, triplex and maisonette. It is a property type found mainly in Quebec.
- Aggregate is the weighted combination of the property types listed above at each level of geography.

RPS HPI Model Methodology

Methodologies for house price prediction can be classified into data-driven approaches and model-driven approaches. The model-driven approach applies house price theory to inference price dynamics based on partial observation of price data. The advantage of the model-driven approach is that it can obtain generally accurate house price state estimation from fewer observations, however the performance of the model-driven approach can be poor if the applied models are not well calibrated. The data-driven approach relies on the spatial-temporal correlation of price states for which future price states can be estimated based on historical time series data.

The RPS HPI uses a Bayesian dynamic linear model (DLM) to derive property price trends over time. Bayesian theory is a branch of mathematical probability theory that uses prior knowledge and observational evidence to allow for the modelling of trends based on data that may exhibit uncertainty and volatility. By having a good estimate of the current state and the dynamics of a system, it is possible to derive assumptions about their evolution. The objective of Bayesian inference is to use prior information and new measurements to determine volatility and infer current estimates.

The RPS DLM is most closely related to autoregressive integrated moving average models, among the different data-driven machine learning methods that are widely used for house price prediction such as neural networks, support vector machine methods, nearest neighbor classification methods, and ensemble learning approaches.

The RPS stratified central tendency DLM provides a systematic approach based on Bayes' theorem for system states updating and prediction. A stratified central tendency method addresses sampling bias issues when computing property price indices by stratifying the national geography into many small geographic and property type-based strata combinations. The properties within these strata tend to have homogeneous characteristics which diminishes variance in the sample set over time. Median sales prices from monthly transactions within each strata form the basis of the index. The DLM considers system states as unknown stochastic variables to be estimated. The prior distribution of system states is quantified based on historical data. By collecting new data over time, the posterior distribution of system states can be estimated based on the Bayes' theorem. This sequential learning framework provides an adaptive learning process for handling time series price prediction.

Knowing that model parameters need to adapt with system behavior, RPS developed methodology to detect change points of system states and update system parameters to catch system behavior changes.[3] The RPS HPI change-point detection method employs the Bank Prime Rate[4] feature, was designed to monitor prediction errors and detect market change, providing feedback to adjust model prediction and reduce prediction errors. By applying the Bank Prime Rate feature to the filtering process, RPS achieved adaptive model parameters relative to the system behaviour and removed data noise from the median time series to obtain the best estimate of unknown states, considering the set of available observations at a given point in time.

RPS HPI Data & Processing

Only valid residential at-market property transactions are included in the RPS HPI model processing. Transaction data is filtered through automated ETL and machine learning processes that:

- Perform data standardization and normalization, remove any duplicate records, and test records meet minimum data element requirements
- Geocode property addresses to a minimum of interpolated street-level accuracy and assign additional census geographies
- Detect and remove property value outliers in local markets using machine learning assessments of price per property characteristic (gross living area/bedroom/bathroom/etc.) to determine if metrics are within statistical norms for the property type and local market
- Stratifies data by geography and property type and derives a monthly median value for each stratum using one, two, three, four, five and six months of data based on the transaction volume – this adaptive moving window aggregation ensures statistically valid samples to compute the strata median

[3] Xiao & Maughan, *Report: RPS HPI Enhancement – Revised Kalman Filter* - December 21, 2022 – refer to Appendix A
[4] Bank of Canada https://www.bankofcanada.ca/rates/banking-and-financial-statistics/posted-interest-rates-offered-by-chartered-banks/ CANSIM V80691311

- Computes time series trend using Kalman (Bayesian) filtering for each period in the stratified data and stock-weights to determine the aggregate trend for each local geography
- Computes price trends and aggregations at the higher-level geographies (CSD, CMA, provincial and national) based on historical transactions and most recent Statistics Canada data

RPS HPI Base Period

The base period for the RPS House Price Index is January 2005. The HPI results are updated and published monthly on or about the 10th business day.

RPS HPI Testing & Maintenance

To ensure model accuracy, RPS continually performs blind back testing on HPI results to gauge the accuracy of the index values. Using a property's historical market value at T_1 RPS applies the index for the property's local market and property type to calculate the value at $T_{CURRENT}$ and compares that to a known market value transaction:

$$\frac{PV_{HPI} - PV_{MARKET}}{PV_{MARKET}} = PE_{HPI}$$

where PV_{HPI} is the property value calculated by applying the HPI, PV_{MARKET} is the known at-market property value, and PE_{HPI} is the percent error of the property value calculated by applying the HPI. Results continue to demonstrate reliable commercial-grade accuracy in line with other automated valuation methodologies.

RPS HPI File Structure

The RPS HPI is provided to subscribed clients via SFTP in .csv format and includes the following fields:

Theme	Field Name	Description
Geographic	RowID	Unique ID for each record
	HPI_Level	Geographic level identifier (National, Provincial, CMA, CSD, FSA)
	HPI_GeoName	Unique ID of the geography being reported on
	HPI_GeoID	Neighbourhood / City name of the geography relative to the Parent_GeoName based on FSA
	Parent_GeoName	Parent geography of the HPI_GeoName
	Parent_GeoID	Parent geography ID, typically Province, CSD number
	Province	Province name
	ProvinceID	Census province ID number
	Geo_Type	Geographic type identifier (urban, rural, provincial, national)
Property	Style	Property type identifier (0_Aggregate, 1_SFD, 2_SD, 3_CondoApt, 4_Row)
HPI Data	Date	Year and month of HPI Data as YYYYMM
	HPI_LowerValue	Lower estimated value of median property value based on 80% confidence

	HPI_Value	Market value of a neighbourhood's median property
	HPI_UpperValue	Upper estimated value of median property value based on 80% confidence
	HPI_Index	Index value of property value relative to January 2005
	ConfScore	Confidence score for the last 60 months (1 = high confidence, 0 = low confidence)
Additional Metrics	Year_over_year_PctChange	Year-over-year percentage change
	National_Rank_1YR	Proprietary ranking identifying the percentile of price appreciation related to national price appreciation
	Regional_Rank_1YR	Proprietary ranking identifying the percentile of price appreciation related to parent geography price appreciation
	POP2006	2006 population count
	POP2011	2011 population count
	POPPctChange	2006-2011 percentage population change
	POD2011	2011 dwelling count
	POD2016	2016 dwelling count
	POP2016	2016 population count
	POPPctChange_2011_2016	2011-2016 percentage population change
	POP2021	2021 population count
	POD2021	2021 dwelling count
	POPPctChange_2016_2021	2016-2021 percentage population change

Appendix A

Report: RPS HPI Enhancement – Revised Kalman Filter

Kai Xiao[1], Robert Maughan[2]

Published December 21, 2022

Abstract

Prior to March 2020 residential property values had generally and steadily increased across Canada since the sub-prime mortgage crisis of 2007-2008, regardless of property type and location. The sub-prime mortgage crisis was a period of severe contraction of liquidity in global financial markets that originated in the United States.

Like the sub-prime mortgage crisis, the COVID-19 global pandemic that began in March 2020 and its related macro-economic impacts continues to cause Canadian residential property market values to fluctuate significantly. This report asserts that a primary driver of both positive and negative property value change is tightly correlated to changes in the Canadian Bank Prime Rate[3] (IR) used for mortgage lending.

The RPS Real Properties Solutions Inc. (RPS) House Price Index (HPI), that was first made commercially available in 2014, leverages the Kalman filter to smooth its estimations of housing prices to derive its index values. Since the sub-prime mortgage crisis, as property values steadily increased in an era of moderate and infrequent changes in interest rates, the application of the Kalman filter in the RPS HPI model proved to be an extremely accurate reflection of the monthly change in property values across multiple stratifications: property type, property location, and property value. RPS recognized the need to enhance the Kalman filter's sensitivity with the onset of dramatic property value changes brought on by the pandemic and other related global economic conditions.

To enhance the sensitivity of the Kalman filtering and smoothing methods employed by the RPS HPI, an IR feature was introduced into the HPI model as a linear deterministic coefficient to adjust the Kalman Gain: the weighting between state and observation matrices. Where an interest rate change is either large, or occurs in rapid succession, the Kalman Gain is augmented to increase the use of observed data to offset any risk of lower confidence in the state matrix. The benefit of the IR feature in the RPS HPI model is that by increasing the weight of observed data coincident to an interest rate change provides improved localized sensitivity at each stratification of property type and geography.

The addition of the IR feature to the RPS HPI enhances the model's ability to reflect local market sensitivity in real estate markets sensitive to the rate and/or relative size of interest rate change. The IR feature also improved the accuracy of the HPI estimate outputs. A review of the top Census Metropolitan Areas across Canada shows the mean absolute error percentage reduced by 8.46% for single family detached property type indices and by 3.69% for condominium apartment property type indices.

Keywords: HPI, Kalman Filter, Interest Rate

1. Introduction

The Kalman filter is an algorithm that provides estimates for unknown variables given a set of observed measurements over time. The Kalman filter and Kalman smoother have demonstrated their usefulness in various applications[4], such as signal processing, guidance, navigation, and control of vehicles, object tracking, robotics, and econometrics and financial technology.

The Kalman filter is a method of estimating the current state of a dynamical system, given the observations so far. The underlying model is a hidden Markov model in which everything is multivariate normal – so in particular, the hidden variables are continuous, rather than discrete. The Kalman filter is the forward algorithm, except that each step can be computed analytically due to Gaussian functions. There is also a backward algorithm that is referred to as the Kalman smoother algorithm. The smoother allows one to refine estimates of previous states, in the light of later observations. In the case of discrete-state hidden Markov models, the results of the Kalman filter and smoother can also be combined with expectation-maximization to estimate the parameters of the model.

Compared to some more advanced machine learning methods, such as Prophet[5] and the deep learning algorithms including Long-Short Term Memory and Temporal Fusion Transformer[6], which require sophisticated modelling and statistical inference, the Kalman filter produces accurate predictions and does not require intensive computational power.

The real advantage of the Kalman filter is that its outputs are generated by estimating states based on linear dynamical models (systems) in state space format. This is different from general statistical filters for time-series data, such as ones using a moving average approach, which smooth a series by consolidating and averaging the data points into longer units of time where no dynamic inference process is involved.

The parameters of the linear dynamic model are assumed to be known; by implementing the Kalman filter we seek to infer the posterior state over state space[7].

The Kalman smoother's estimate of state is retrospectively improved by using the Rauch–Tung–Striebel (RTS) smoother (backward algorithm)[8] with the RTS parameters deducted from Kalman filter. Being strongly associated with the parameters deduced from Kalman filter at the forward step, Kalman smoothers are used widely to estimate the state of a linear dynamical system from noisy measurements [9,10].

In the following sections of the report, the terms Kalman filter and Kalman smoother are jointly used since the two algorithms are sharing the same parameters. In the traditional formulation, the dynamics and output matrices are considered fixed attributes of the system; the covariance matrices of the process and sensor noise are tuned by the designer, within some limits, to obtain good performance in simulation or on the actual system. For example, it is common to use noise levels in the Kalman smoother in excess of the actual noise to obtain practical robustness[11].

In this report we take a macro-economic approach to the problem of tuning Kalman smoother. We start with the observation that (by our definition) only the output is observed. This implies that the only way to verify that a

Kalman smoother is working well is to compare the predicted outputs with those that occur on new or unseen test data – data that was not used by the Kalman smoother.

In machine learning terms, one would consider the output prediction error to be the error for minimization. We consider the noise covariance matrices, as well as the system matrices, as parameters that can be varied to obtain different estimators, in this case, different Kalman smoothers. These are varied, within limits, to obtain good test performance; this final Kalman smoother can then be checked on entirely new data.

To do this we formulate the Kalman smoothing problem, with missing observations, as a simple least squares problem, with a coefficient matrix that depends on the parameters, i.e., the system and noise covariance matrices. We show how to efficiently compute the derivative of the test error with respect to the parameters and use a simple proximal gradient method to update them to improve the test error. This method yields a Kalman smoother auto-tuning method. It uses one or more observed output sequences, and the usual prior knowledge in determining the starting system matrices as well as a description of the set over which we are allowed to vary them.

The same formulation works for tuning robust Kalman smoothers, where the process and sensor noises are assumed to have a non-Gaussian distribution, typically with fatter tails. In this case the least squares formulation of the Kalman smoother becomes a convex optimization problem, and the effect of the parameters is even less obvious, and therefore harder to tune manually. Our auto-tuning method extends immediately to such problems. In this report:

- We describe a Kalman smoother enhancement using interest rate change as a feature to adjust the weighting of state and observed matrices (Kalman gain),
- We describe how the interest rate feature provides improved, accurate local market sensitivity at all stratifications, and
- We provide the implementation of the interest rate feature to illustrate the method via numerical examples using real data.

2. Kalman Filter and Smoother

The goal in smoothing is to reconstruct or approximate missing measurements given known measurements.

Since the outputs and states are jointly Gaussian, the maximum likelihood and conditional mean estimates of the missing output values are the same and can be found as the solution to a constrained least squares problem.

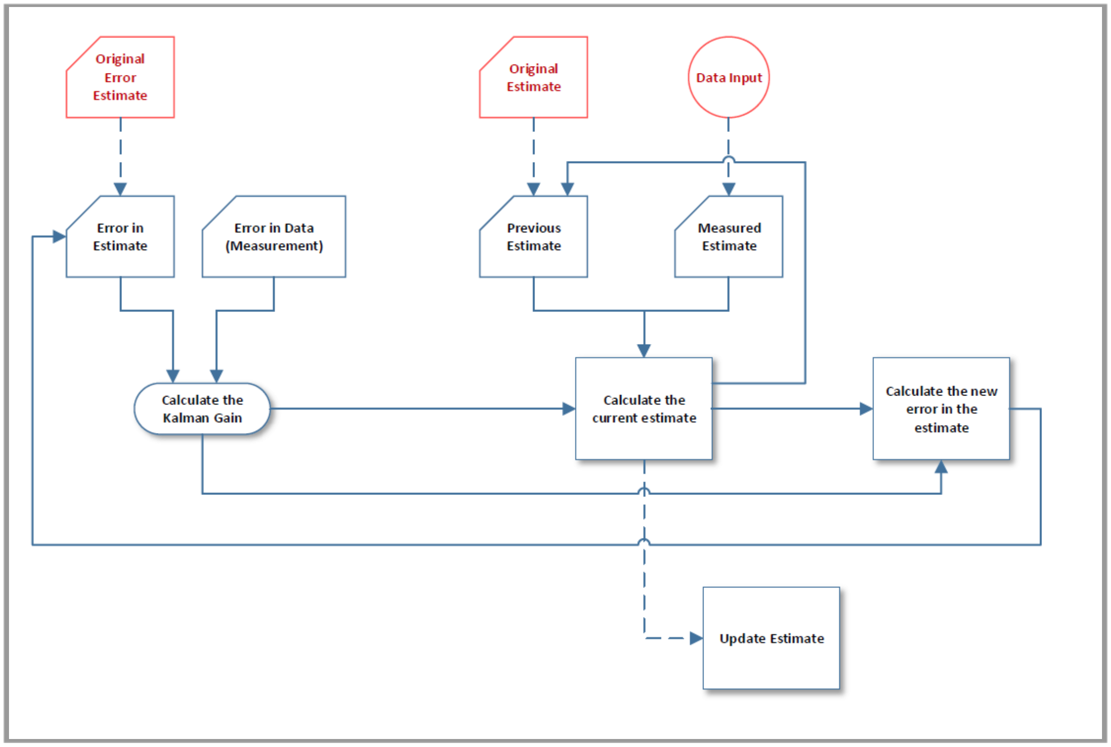


Fig. 1: Diagram of Kalman filter algorithm.

As denoted in Figure 1, three fundamental equations in the Kalman filter algorithm are used to calculate the $Kalman\ Gain$ (KG), the $Current\ Estimate$ (EST_t) and the $New\ Error\ in\ Estimate$ (E_{EST_t}).

These calculations are shown in Equation (1), (2) and (3) respectively:

$$KG = \frac{E_{EST}}{E_{EST} + E_{MEA}} \quad (1)$$

$$EST_t = EST_{t-1} + KG[MEA - EST_{t-1}] \quad (2)$$

$$E_{EST_t} = [1 - KG]E_{EST_{t-1}} \quad (3)$$

where EST denotes estimate, MEA denotes measurement, E denotes error, and KG denotes $Kalman\ Gain$.

To achieve higher computational efficiency, the Kalman filter algorithm is normally implemented in the matrix form shown in Figure 2.

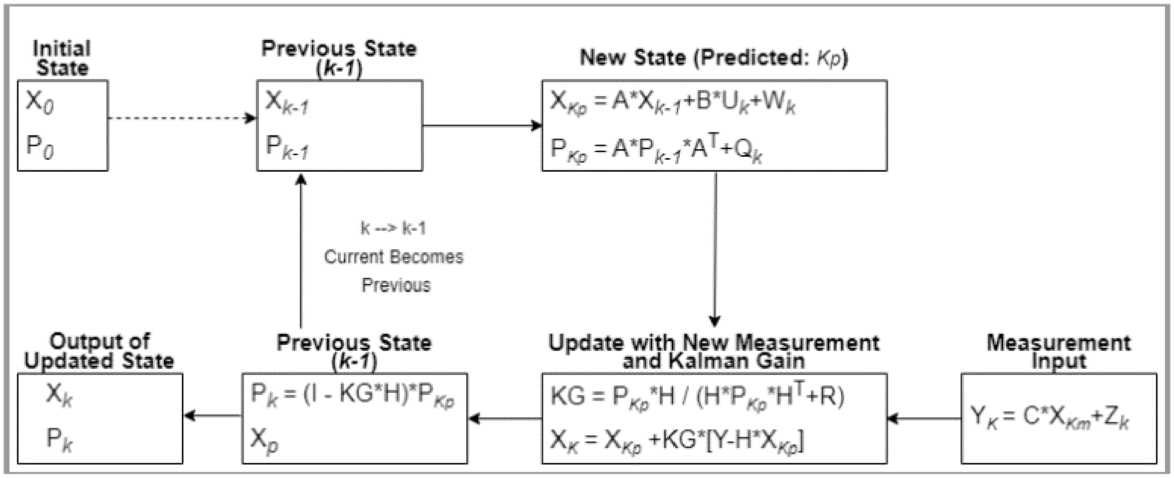


Fig. 2: Diagram of Kalman filter algorithm in matrix form.

where:

- X denotes the State Matrix
- P denotes the Process Covariance Matrix (represents error in the estimate)
- KG denotes the Kalman Gain
- R denotes the Sensor Noise for Covariance Matrix
- I denotes the Identity Matrix
- Y denotes the Measurement of the State

In the New State calculation process:

- U denotes the Control Variable Matrix
- W denotes the Predicted State Noise Matrix
- Q denotes the Process Noise Covariance Matrix

The Kalman Gain (KG) is the weight given to the measurements and current state estimate. With a high Kalman Gain, the filter places more weight on the most recent actual measurements, and thus conforms to them more responsively. With a low Kalman Gain, the filter conforms to the model predictions more closely.

3. Issues Associated with the Kalman Filter

The Kalman filter has the advantage over general statistical filters for its capacity to make interferences of data trends. To show how the Kalman filter responds to sudden trend changes, we use a sequence of 1000 random data points centered around 20 with an interval between 16 and 24 (error of observations of 4) following a normal (Gaussian) distribution. The sequence of data points (1 to 1000) is used as a time-series, and their values are plotted along the time axis as t followed by the value's sequence number (e.g., t100 refers to data point 100 in the sequence).

Figures 3(a) through 3(d) are the visualized observations and corresponding estimates of the Kalman filter (with an initial estimation of 50) on the sequence of data described above. The Kalman filter can infer the trend, and the variation of observations does not notably affect the trend.

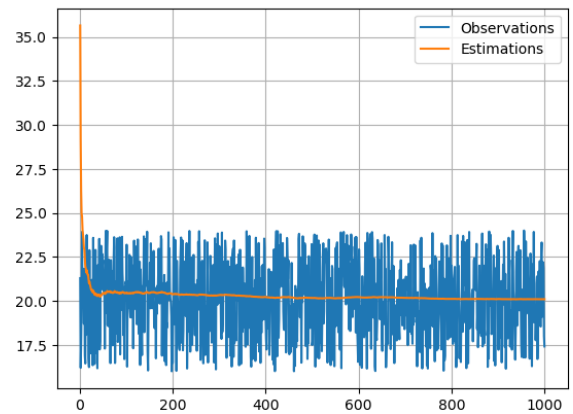


Fig. 3(a): Observed and corresponding estimations, visualized from *t*1 to *t*1000 (all data points plotted).

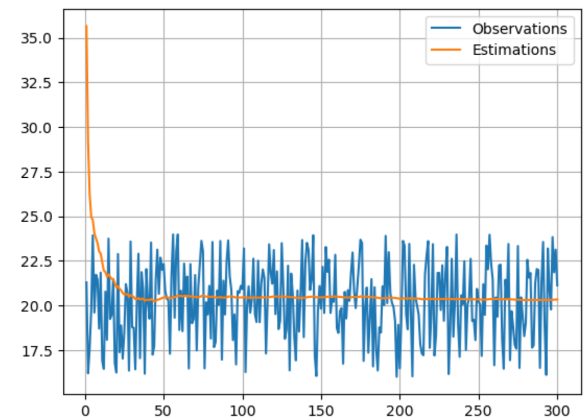


Fig. 3(b): Observed and corresponding estimations, visualized from *t*1 to *t*300 (first 300 data points plotted).

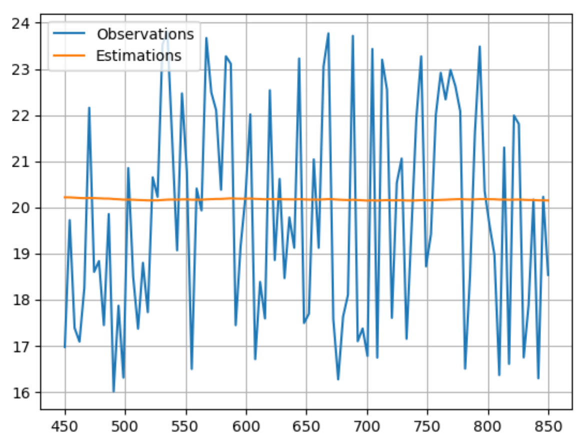


Fig. 3(c): Observed and corresponding estimations of 100 data points, visualized from *t*450 to *t*550.

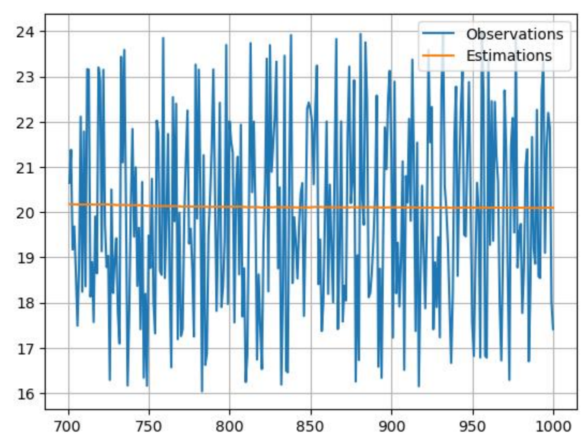


Fig. 3(d): Observed and corresponding estimations, visualized from *t*700 to *t*1000 (last 300 data points plotted).

However, the Kalman filter can be slow to respond to sudden changes in observed data as its inferred data relies on the estimated state. If the weight of the previously estimated state is high, the Kalman filter will produce estimates which significantly lag any new data that is substantially different from the calculated state estimated using previous data.

With the same data distribution as above, we multiply the 100 data points between *t*300 and *t*400 with a scalar of 1.5, thus creating a marked change in the values of the data points for this period in the series as visualized in Figure 4(a) through 4(d). With the original Kalman filter applied on the entire set of data points, the observed and estimated results show that, although the original Kalman filter is responding to the change of data distribution, the response is a smooth and slow transition across the sudden change. This demonstrates that the original Kalman filter does not respond to a sudden trend change quickly, despite creating an accurate overall estimation.

These examples demonstrate that the Kalman filter itself does not have a mechanism to adjust its responsiveness to an abrupt change in the trend of the data. While the Kalman filter works well across the entire data set and accurately estimates the trend, it works poorly to promptly react to data that quickly deviates from its range of error. This underperformance becomes problematic when an accurate estimation is needed period-over-period, especially in an application where there is no assurance that the data follows a Gaussian distribution.

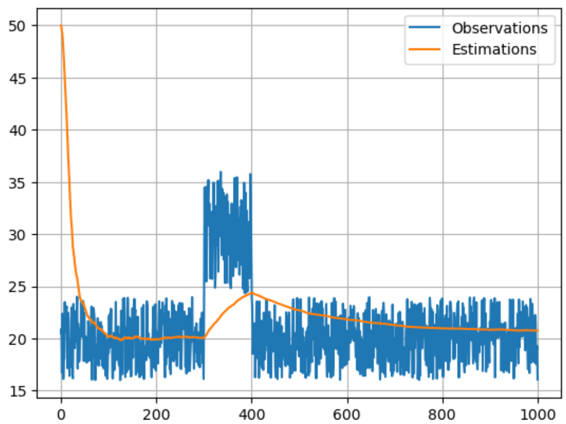


Fig. 4(a): Observed and the corresponding estimations, visualized from $t1$ to $t1000$.

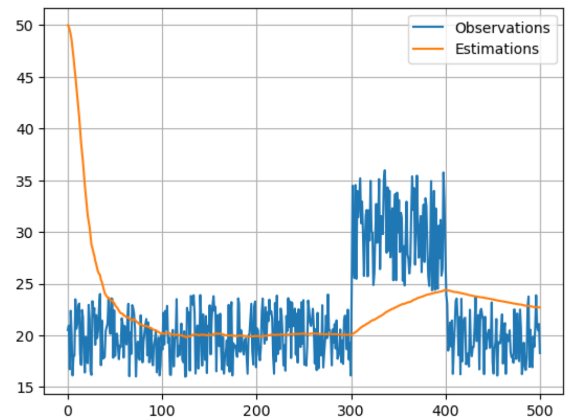


Fig. 4(b): Observed and the corresponding estimations, visualized from $t1$ to $t500$.

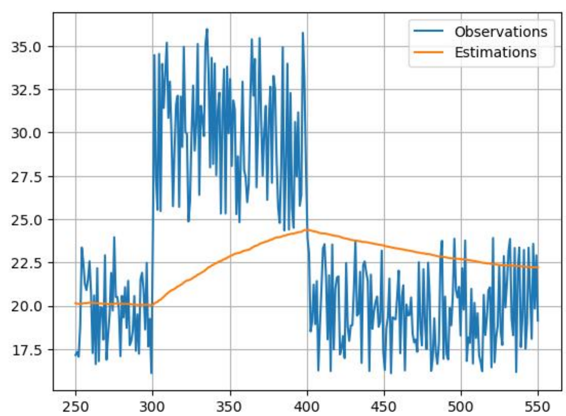


Fig. 4(c): Observed and the corresponding estimations of 300 data points, visualized from $t250$ to $t550$.

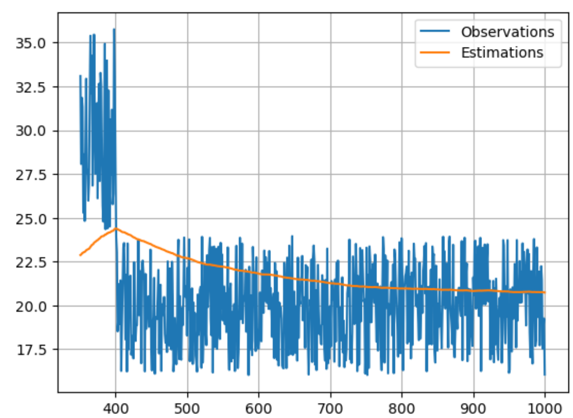


Fig. 4(d): Observed and the corresponding estimations, visualized from $t350$ to $t1000$.

4. Enhancement of the Responsiveness of Kalman Filter

As previously discussed, the Kalman Gain determines weight / ratio of estimation and observation. Therefore, it was reasonable to assume that to enhance the responsiveness of the Kalman filter where a sudden change of data distribution occurs, a scalar factor a can be introduced into the KG increasing the KG to a higher rate. The use of the scalar factor results in the Kalman filter placing more weight on the most recent measurements, and its responsiveness is increased accordingly.

We therefore modify Equation (1) to Equation (4) as follows:

$$KG = \frac{E_{EST}}{E_{EST} + E_{MEA}} \cdot \lambda \qquad (4)$$

where λ denotes a scalar factor associated with the extent of the data distribution, which is a vector (or matrix) when applied in time-series data.

Figures 5(a) through 5(d) show visualized observations and corresponding estimates of the modified Kalman filter, where weighting of observed measurements will be enhanced at times where sudden changes on distribution occur, on an array of data presented in time-series with a sudden distribution change. To test the effectiveness of this approach, we use the same data and same distribution as shown in Figure series 4, but we apply the λ scalar factor of 1.5 in Equation (4) as a factor to adjust the Kalman Gain calculation at points where sudden changes occur (starting at $t300$ and $t400$). By comparing the estimation output with the corresponding outputs shown in Figure series 4, it can be observed that the estimations produced by the modified KG calculation in Equation (4) is more responsive to sudden changes and provides more accurate estimations that are more reflective of the actual changes in the data.

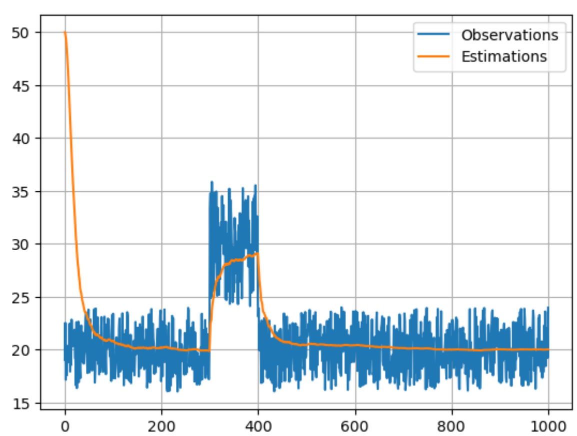


Fig. 5(a): Observed and the corresponding estimations, visualized from $t1$ to $t1000$.

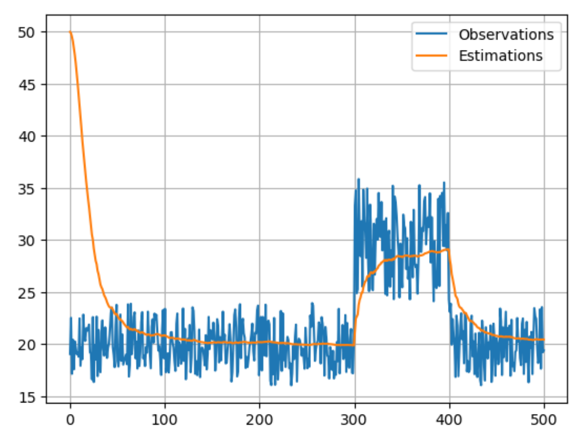


Fig. 5(b): Observed and the corresponding estimations, visualized from $t1$ to $t500$ (first 500 data points plotted).

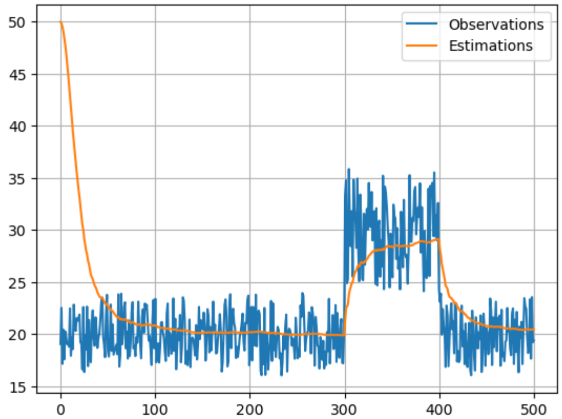


Fig. 5(c): Observed and the corresponding estimations of 300 data points, visualized from *t*250 to *t*550.

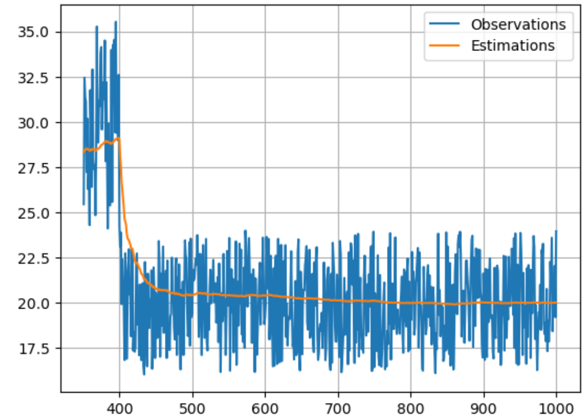


Fig. 5(d): Observed and the corresponding estimations, visualized from *t*350 to *t*1000.

5. Interest Rate as Macro-Economic Indicator

Mortgage interest rates are one of the key factors that affect the affordability of buying or maintaining residential real estate. When interest rates increase, mortgage payment increases on existing mortgages may follow, depending on the type of mortgage.

Mortgage interest rate increases also impact buyers' ability to secure mortgage funding, and for real estate investors to cover carrying costs on rental properties. These impacts affect the overall housing market which in turn affects the value of homes. Historically in general, increased interest rates lead to decreases in the value of homes, and the reverse is seen where decreased interest rates lead to increased home values.

The Bank Prime Rate published by the Bank of Canada shown in Figure 6(a) was used to create a vector, and its consecutive difference was calculated as shown in Figure 6(b).

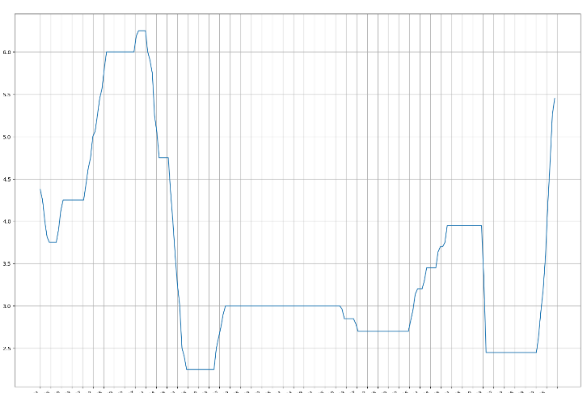


Fig. 6(a): Bank Prime Rate from Bank of Canada, for the period of 2004-01 through 2022-10.

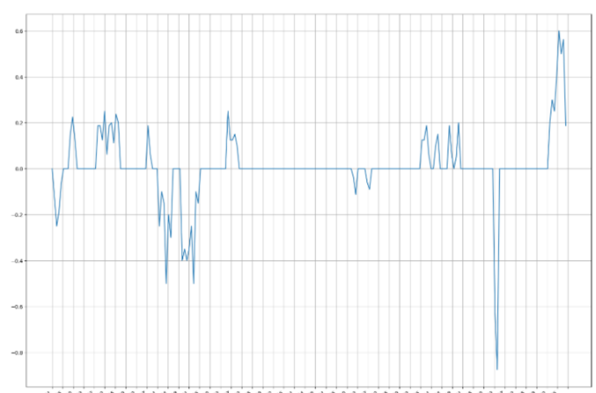


Fig. 6(b): Consecutive difference of Bank Prime Rate, for the period of 2004-01 through 2022-10.

The effect on housing pricing from interest rate change can be assumed to be a Gaussian distribution[7]. To create a vector of factors that correspond to the effect of interest rates, the consecutive change of interest rates has been used to create a probability distribution function as shown in Equation (5).

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{e^{-(x-\mu)^2}}{2\sigma^2}} \qquad (5)$$

where μ is mean of distribution and σ^2 is the variance of the distribution.

The cumulative abstract values created at each time period by the function are summed and set at the location of the vector which is associated with housing prices at each timeframe, as shown in Figure 7, as a coefficient (λ in Equation (4) in terms of a vector) in which is associated with the change in interest rate frequency and amplitude.

It can be observed that large coefficient values correspond to frequent and large interest changes. This coefficient is referred to as the IR feature.

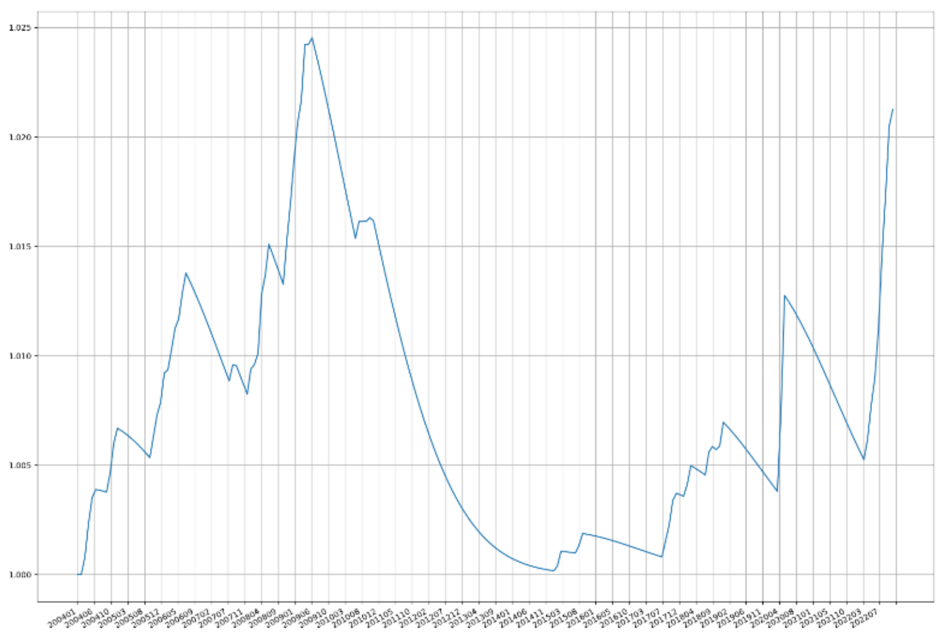


Fig. 7: Interest rate change converted to Gaussian Probability Density Function for the period of 2004-01 through 2022-10.

6. Implementation of Application of Interest Rate Change to RPS HPI

To incorporate the IR feature of interest rate change with the RPS HPI model, the vector data created by Equation (5) is added as exogenous variables, while the measured variables (actual house prices) are used as endogenous variables.

To enable the exogenous variables to impact the Kalman Gain value, the *"smooth trend"* model is selected[12]. The data used for HPI model computations are pre-processed, so to alleviate the potential issue that actual home values may not always be available, NULL endogenous values are replaced by the nearest available actual value in the time-series.

6.1 CMA-Level Performance

Testing was performed on the top ten most-populous Census Metropolitan Areas (CMAs) across Canada, according to the 2021 Statistics Canada census, at the postal geography Forward Sortation Area (FSA) stratification and by property type stratification. In each case, the application of the Interest Rate change feature improved the sensitivity of the Kalman Filter used in the RPS HPI.

The series of Figures 8 through 10 show RPS HPI smoothing results on select CMAs for Single Family Detached property types.

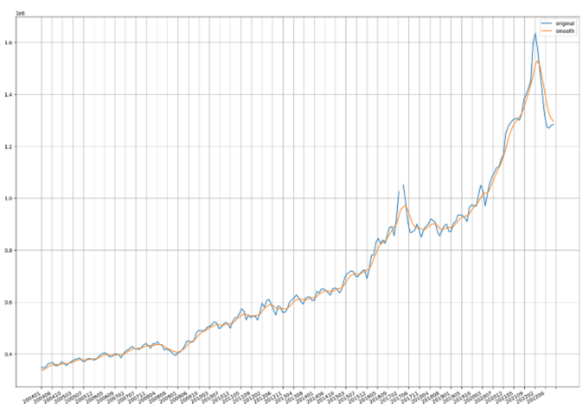


Fig. 8(a): Toronto CMA Single Family Detached property type with Original Kalman Smoother

Fig. 8(b): Toronto CMA Single Family Detached property type with Interest Rate Incorporated Kalman Smoother

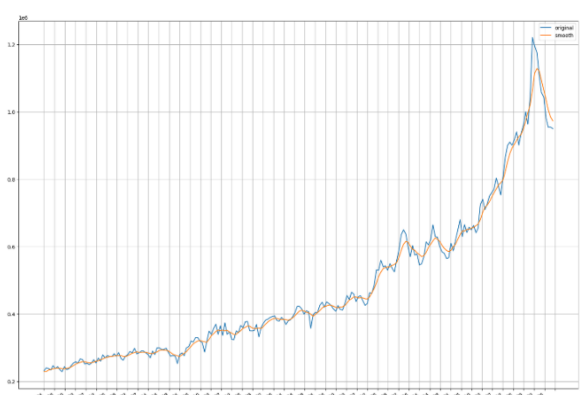


Fig. 9(a): Hamilton CMA Single Family Detached property type with Original Kalman Smoother

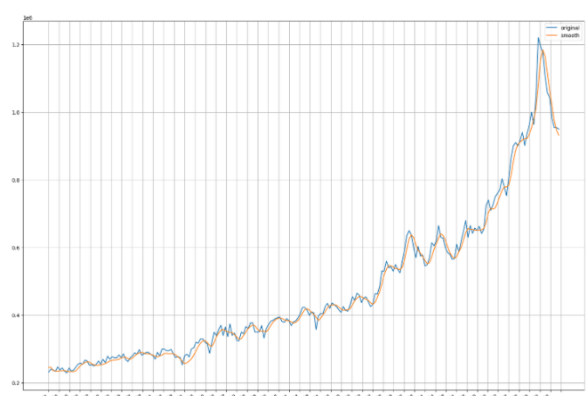


Fig. 9(b): Hamilton CMA Single Family Detached property type with Interest Rate Incorporated Kalman Smoother

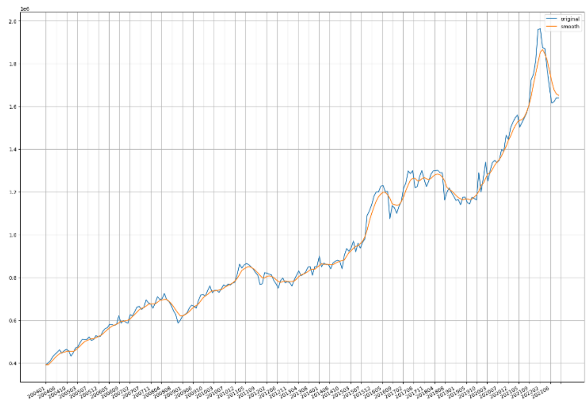


Fig. 10(a): Vancouver CMA Single Family Detached property type with Original Kalman Smoother

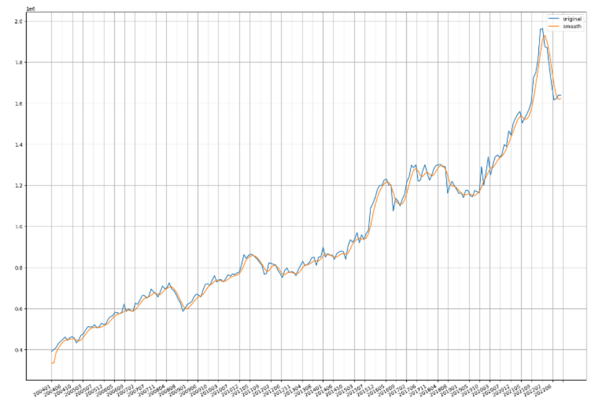


Fig. 10(b): Vancouver CMA Single Family Detached property type with Interest Rate Incorporated Kalman Smoother

7. Results

The extensive quality analysis performed was to ensure the introduction of the IR feature was a more accurate reflection of house price trends across all stratifications by blind comparisons of HPI estimations to actual market transactions. That testing showed at periods of high market fluctuation the RPS HPI model with enhanced sensitivity remains accurate.

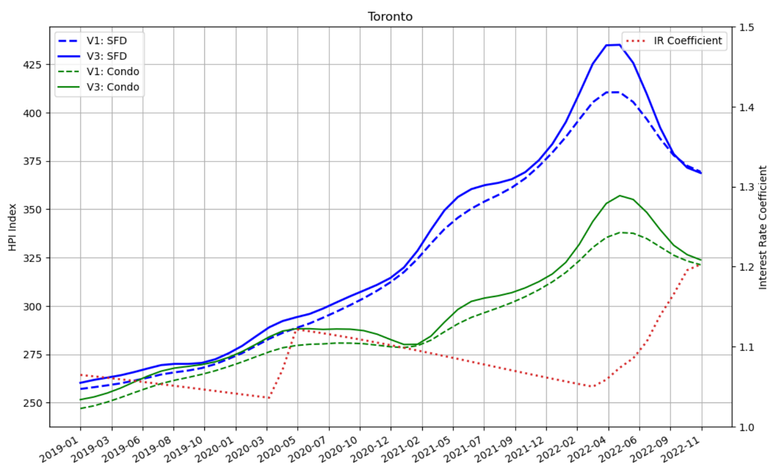
It should be noted that the introduction of IR feature does not affect the smoothness of the HPI output. The plotted results of the enhanced HPI model appear to show that smoothness is reduced (volatility increases), however the perceived volatility is the result of the enhanced adaptability of the Kalman Gain weighting of observed / measured data corresponding to the IR feature and represents a more accurate reflection of house price trends. At periods where interest rate changes occur, the HPI model will now be proportionally more inclined to use observed data by triggering the IR feature to offset any less trustworthy HPI estimations calculated by the Kalman filter the IR feature.

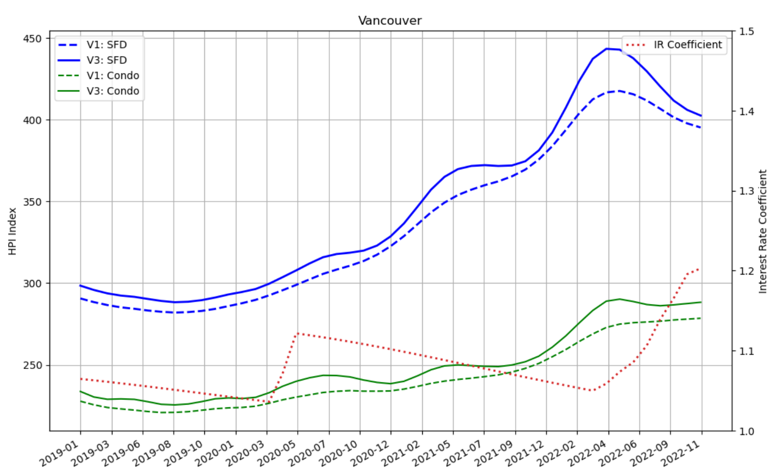
The IR feature (interest rate coefficient) implementation in the RPS HPI model is denoted as “V3” in the following charts and compared to the current commercially available HPI version denoted as “V1”.

The charts provide a comparison of the HPI index estimatsions for various CMAs across Canada for both Single Family Detached (SFD) and Condominium Apartment (Condo) property types.

The IR feature is plotted synchronous to the HPI model estimations for each period. As a Gaussian Probability Density Function, and referring back to Figure 7, we can see that the IR feature has continuous influence on the Kalman Gain.

While interest rates can change in one period and remain the





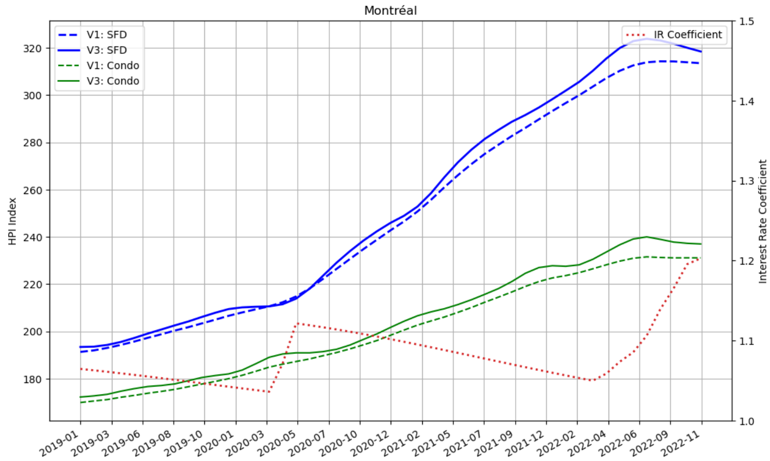
same over time, the IR feature provides continual input to ensure an optimized Kalman Gain is applied.

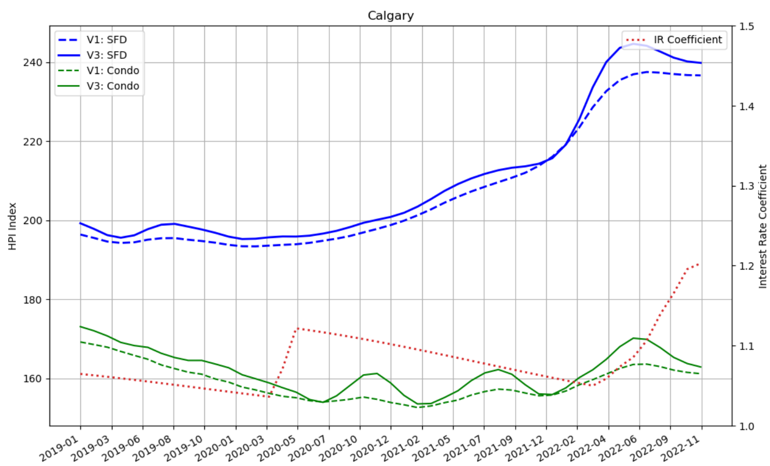
As mentioned above, these charts show what appears to be a reduction in smoothness / increase in volatility with the V3 HPI model, but it is in fact improved accuracy in house price trends as the Kalman Gain adapts to interest rate changes.

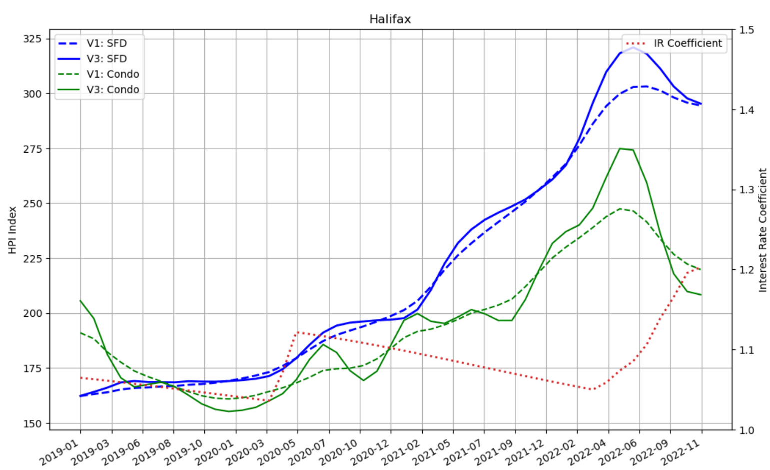
The overall trends in house prices remain directionally similar, but the V3 estimatations signal more responsive outputs.

Also apparent is the diversity of sensitivity to interest rate change between markets and property types. As macro-economic impacts affect local markets differently, so does the impact of interest rate change.

These nuances can be seen in the distance at each time period between the V1 and V3 model estimations – Montreal's sensitivity does not mimic that of Toronto, evidence that the IR feature reamins accurate regardless of localization.







For each of th CMAs and property types plotted above, the V3 model not only provided better overall sensitivity but also provides improved accuracy. To demonstrate the improved accuracy, the Mean Absolute Percentage Error

(MAPE) was used to compare the outputs of the V1 and V3 models at the CMA stratification. MAPE is a measure of prediction accuracy of a forecasting method as a ratio defined by:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \qquad (6)$$

where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value A_t. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points n. MAPE is commonly used as a loss function for regression problems and in model evaluation, because of its very intuitive interpretation in terms of relative error. The use of the MAPE as a loss function for regression analysis is feasible both practically and theoretically since the existence of an optimal model and the consistency of the empirical risk minimization can be proved.

The lower the MAPE value the more accurate the model outputs are. In general tems, MAPE results <5% indicate very high accuracy, values <10% indicate high accuracy, values in the 10% to 20% range indicate good accuracy[13].

CMA	**MAPE (2019-01 - 2022-11) Single Family Detached**	**MAPE (2019-01 - 2022-11) Condo Apartment**	**CMA**	**MAPE (2019-01 - 2022-11) Single Family Detached**	**MAPE (2019-01 - 2022-11) Condo Apartment**
Toronto	V1: 8.84068 V3: 8.03961	V1: 3.36822 V3: 3.01239	Edmonton	V1: 4.59387 V3: 3.57725	V1: 11.52741 V3: 12.32125
Vancouver	V1: 10.37285 V3: 9.27689	V1: 5.88158 V3: 4.79128	Winnipeg	V1: 4.75344 V3: 4.39323	V1: 11.72290 V3: 11.56494
Montreal	V1: 7.17554 V3: 5.61797	V1: 8.89078 V3: 7.41882	Hamilton	V1: 4.51758 V3: 4.78920	V1: 6.63493 V3: 7.45540
Calgary	V1: 5.93981 V3: 4.94076	V1: 5.70609 V3: 5.30780	Victoria	V1: 6.68698 V3: 5.56756	V1: 7.51327 V3: 7.86651
Halifax	V1: 5.59582 V3: 5.43434	V1: 10.35722 V3: 9.31607	Regina	V1: 5.27109 V3: 5.50169	
Ottawa-Gatineau	V1: 4.03467 V3: 3.72718	V1: 6.38067 V3: 6.35865	St. John’s	V1: 5.59582 V3: 5.43434	
Quebec	V1: 2.71558 V3: 2.65108	V1: 5.35652 V3: 5.34849			

As seen in the sample above of the largest CMAs across Canada, MAPE results improved by 8.46% for single family detached property type indices, and 3.69% for condominium apartment property type indices.

The RPS Enhanced Sensitivity HPI model V3 (IR feature) will be commercially available in January 2023.

References

[1] Kai Xiao, MS, PhD – Data Scientist
RPS Real Property Solutions Inc.

[2] Robert Maughan – SVP Technology & Data Solutions
RPS Real Property Solutions Inc.

[3] Bank of Canada https://www.bankofcanada.ca/rates/banking-and-financial-statistics/posted-interest-rates-offered-by-chartered-banks/ CANSIM V80691311 - The weekly posted prime interest rates offered by the six major chartered banks in Canada. The methodology for calculating the typical rate is based on the statistical mode of the rates posted by the six largest banks, with the following rules: (a) in the case of no mode, select the rate closest to the simple 6-bank average, (b) in the case of multiple modes, select the mode closest to the simple 6-bank average, and (c) in the case of two modes at equal distance from the simple 6-bank average, select the mode whose banks have the largest value of assets booked in CAD. (Data source: latest M4 return data released on Office of the Superintendent of Financial Institutions' website.). The prime rate, or prime lending rate, is the interest rate a financial institution uses as a base to determine interest rates for loan products. Each financial institution sets its own prime rate, as a function of its cost of funding, which, in turn, is influenced by the target for the overnight rate set by the Bank of Canada.

[4] Elliot Anenberg, Edward Kung, *Interest Rates and Housing Market Dynamics in a Housing Search Model*, (October 30, 2017). Available at SSRN: https://ssrn.com/abstract=3062113

[5] Sean J. Taylor, Benjamin Letham, *Forecasting to Scale*, The American Statistician, 2018, 72(1), pp. 37 - 45.

[6] Bryan Lima, Sercan O. Arlk, Nicolas Loeff, Tomas Pfister, *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting*, 2021, 37(4), pp. 1748 - 1764.

[7] Paul Zarchan, Howard Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*, American Institute of Aeronautics and Astronautics (2000), ISBN 978-1-56347-455-2.

[8] Kalman filter and smoother, Jeff Miller, *Lecture Note of Advanced Stochastic Modeling*, Harvard University, https://jwmi.github.io/ASM/6-KalmanFilter.pdf

[9] Peter Swerling, *A proposed stagewise differential correction procedure for satellite tracking and prediction*, Rand Corporation, 1958.

[10] Rudolf Emil Kálmán, *A new approach to linear filtering and prediction problems*, Journal Basic Engineering, 82(1), pp. 35-45, 1960.

[11] Aurthur Gelb, *Applied optimal estimation*, MIT Press, 1974.

[12] Statsmodels – a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration - https://www.statsmodels.org

[13] Lewis, C.D., *International and Business Forecasting Methods*, Butterworths, 1982