Home price index prediction for greater Toronto

Queen's University MMAI 823 – Section 2
Professor: Matthew Thompson

Team Alfred:
Kelsey Pericak, Shilpa Vishwanath, Mohammad Raza, Shayaan Mehdi, Sky Tang, Nezil Gopinadh, Theebak Sothilingam & Yvan Large

Single family home price index

Home price index (HPI)

- CREA
- Sophisticated formula
- Greater Toronto
- 1 unit = 1% from base price (100)
- January 2005 base

Two main objectives

1. Predict HPI 1 month ahead for greater Toronto area using historical data
2. Learn which features have relationship with/ impact on real estate prices in greater Toronto

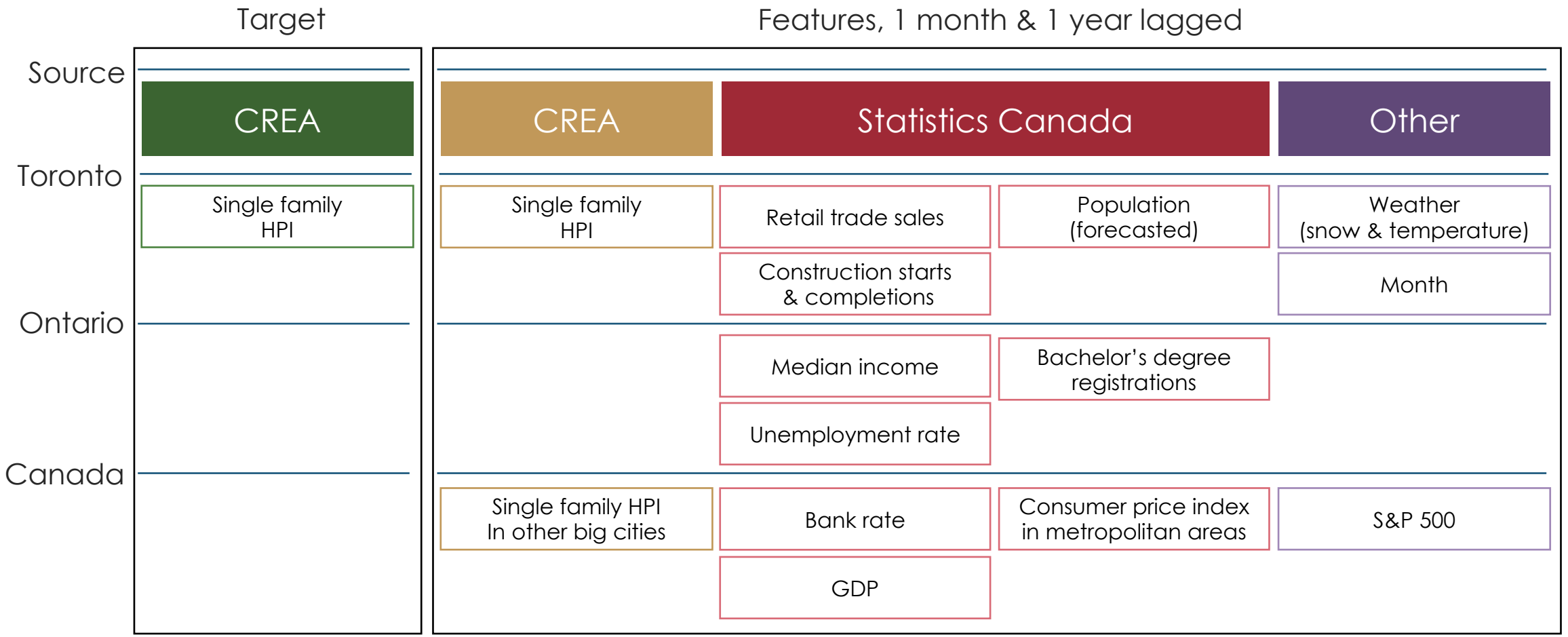
Why does it matter?

- When to buy
- When to sell
- Use trends to plan for future investments/ policy making
- Evaluate potential return on investment
- Real estate agent in-the-know

Our modeling process

1. Dataset creation
2. Exploratory analysis
3. Data preparation
4. Time sensitive split
   - Train (85%) from Jan. 2006 to Sept. 2018
   - Test (15%) split from Oct. 2018
5. Base model creation
6. Feature selection
7. Hyper parameter tuning
8. Prediction

35 features in dataset from 2006 to 2020



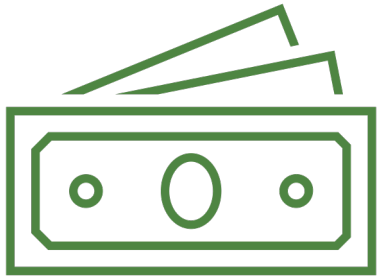
Examples of features built from hypotheses

Consumer price index	---------	↑ inflation = price ↑
Bank rate	---------	↓ rate = ↑ demand for mortgage
Weather	---------	↓ temperatures = ↓ interest in moving
Unemployment	---------	↑ rate = ↑ wealth gap OR ↓ buying
Construction	---------	↑ completions = ↑ new homes @ high prices

Building the dataset required some feature engineering

Calculated average snowfall and temperature with daily data

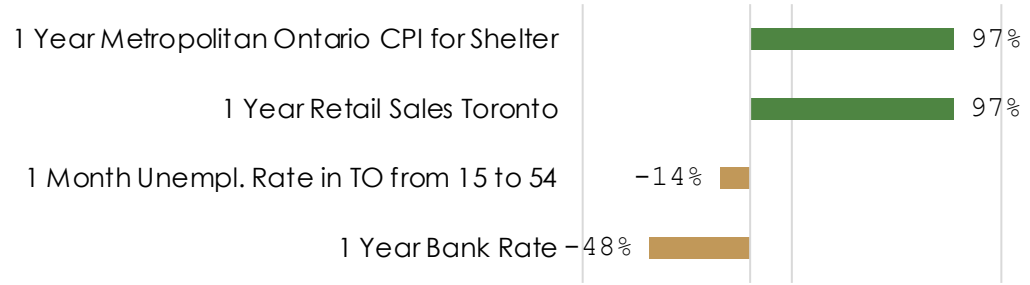
Forecasted population posted annually



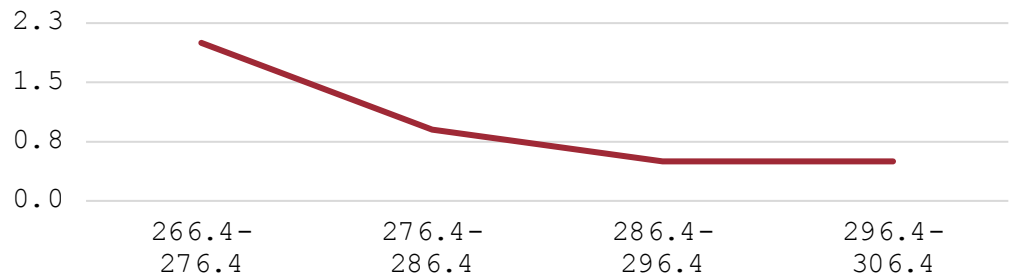
Imputed mean values for median family income for couples and lone parents

Insights about single family home price index

Interesting Correlations with Target



Average 1 Month Lag Bank Rate (%) vs HPI Ranges in 2020



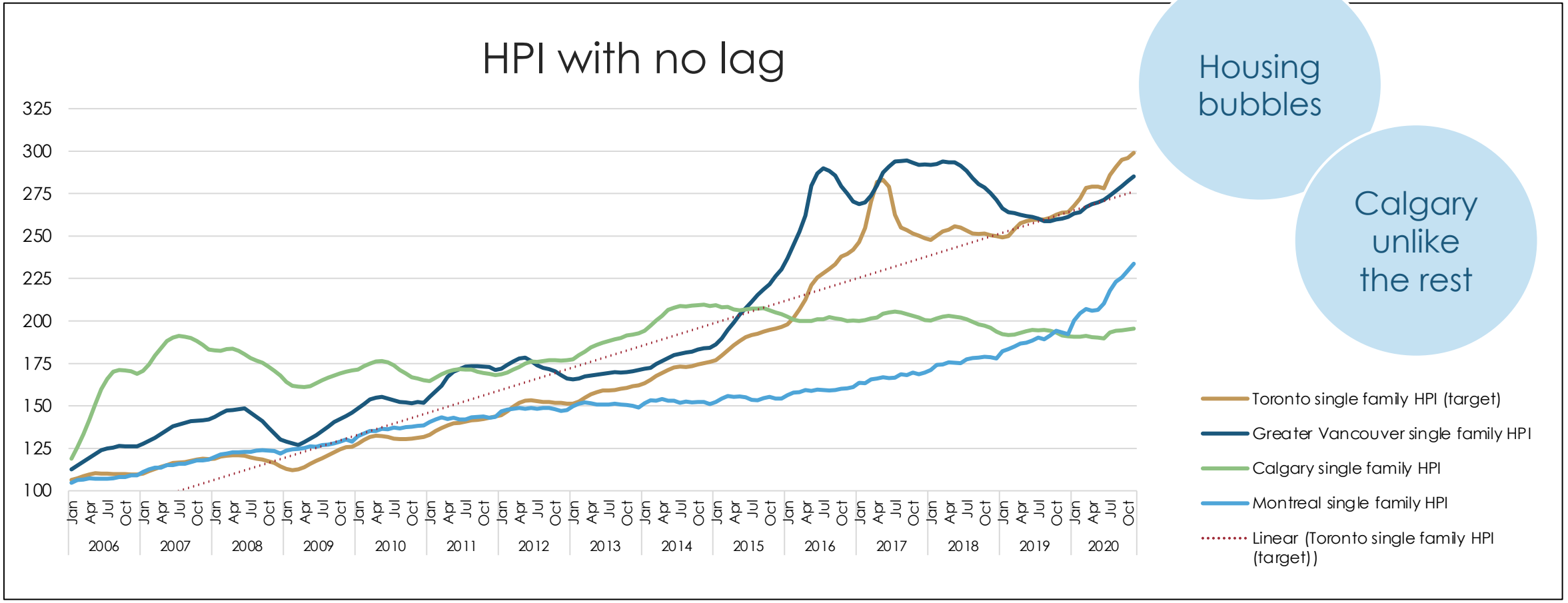
1.5X HPI
Target Increase From 2015 to 2020

>17%
Growth in avg annual HPI in 2016 & 2017

Seasonal
Troughs or stability in HPI during colder months

19
Base features with >.9 correlation with target

Comparing HPI indices for single family

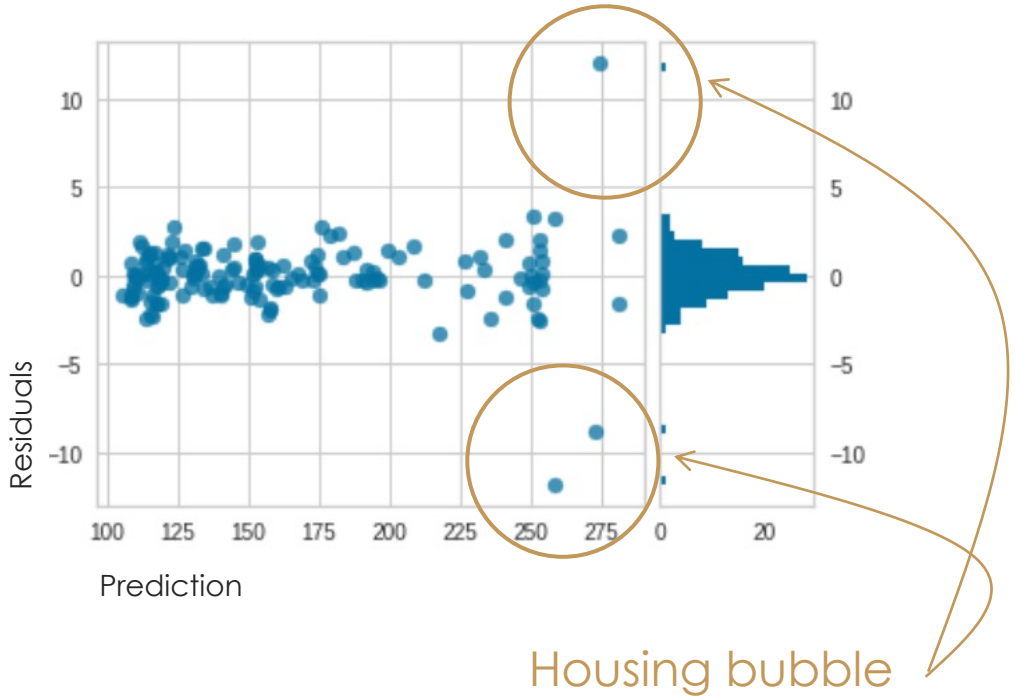


Three main data cleaning steps beyond dataset creation and before feature selection

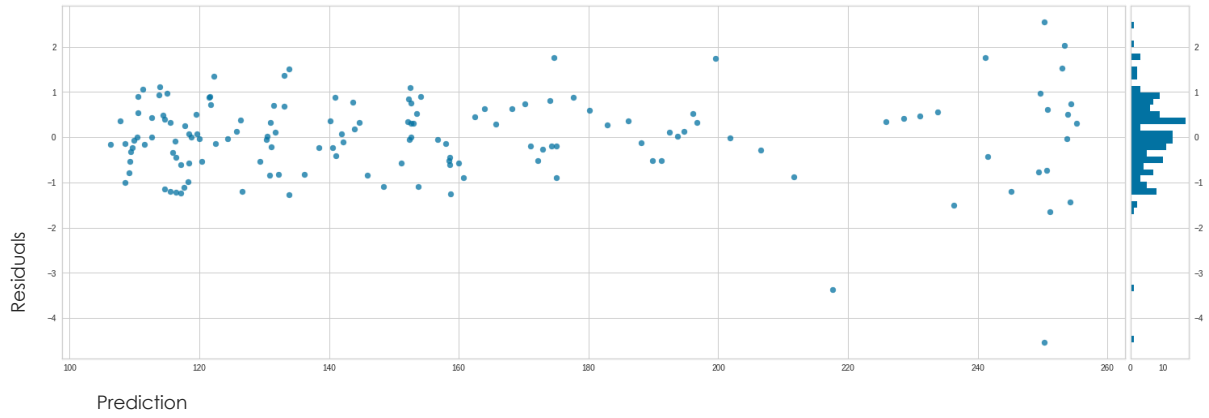
1. Dropped Calgary
2. Eliminated bubble time range
3. Capped construction outliers

Normality testing and outlier detection

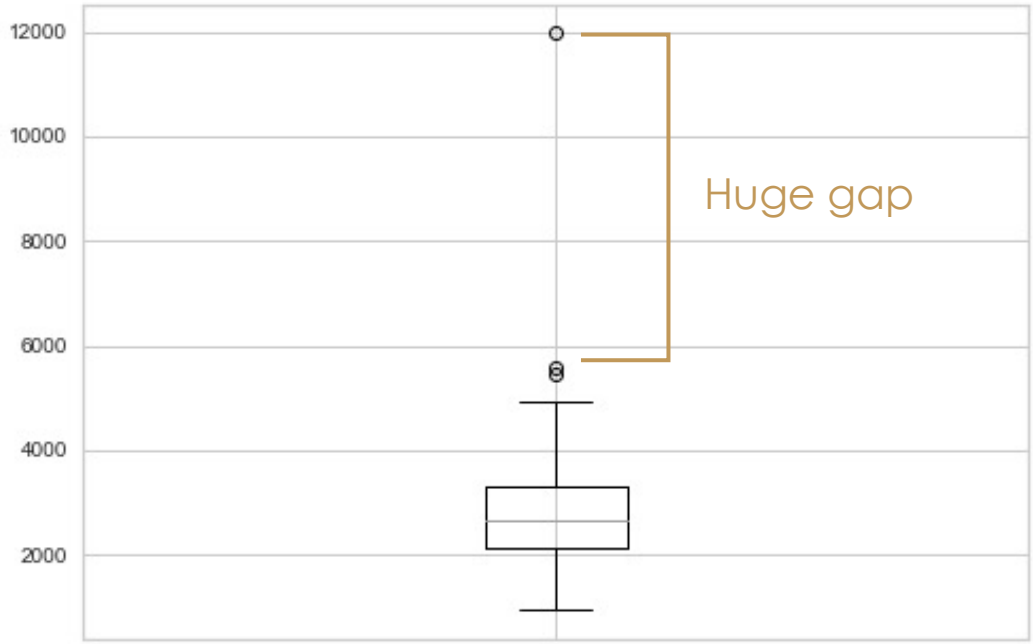
All data



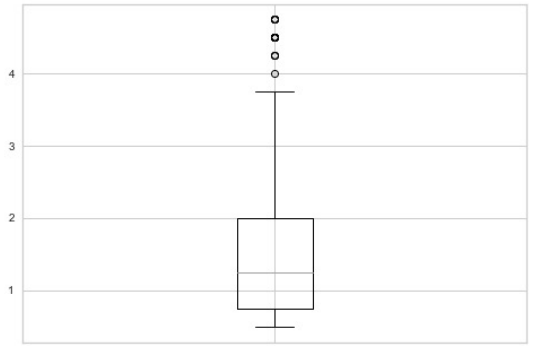
Cleaned data



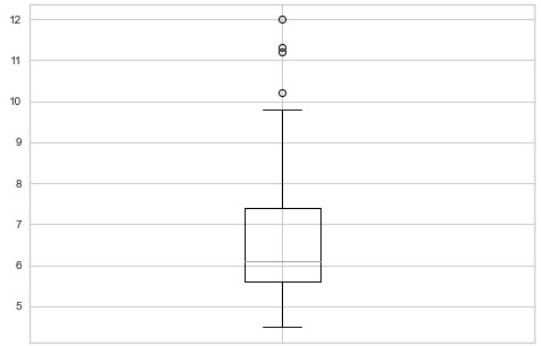
Features with outliers



1 month lagged Toronto construction completion units



1 month lagged bank rate



1 month lagged unemployment rate in Toronto from 15 to 54

Bayesian model selected

No feature selection or tuning

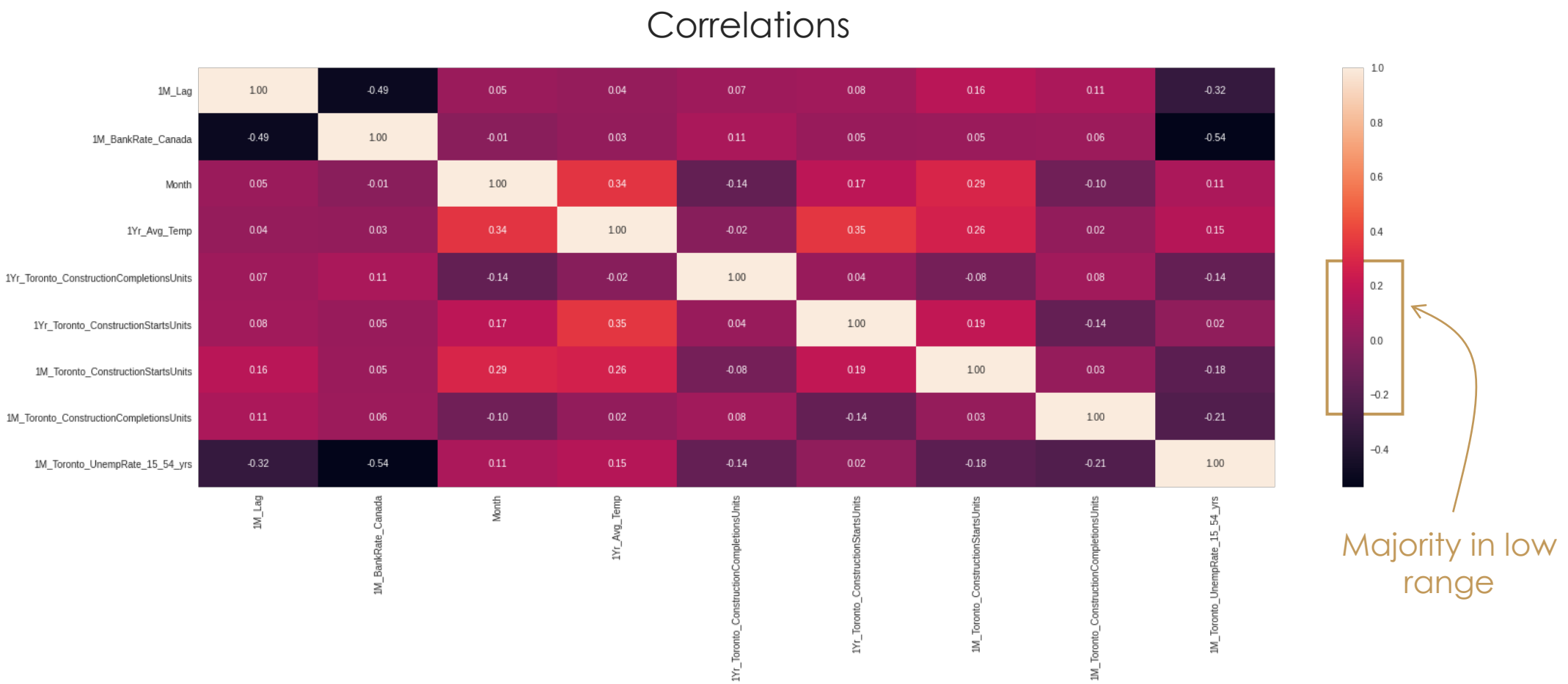
	ARD	Linear Regression	Bayesian
R^2 on training data	0.99	0.99	0.99
R^2 on test data	0.97	0.76	0.74
RMSE on test data	2.76	8.0	8.43

Feature selection and tuning

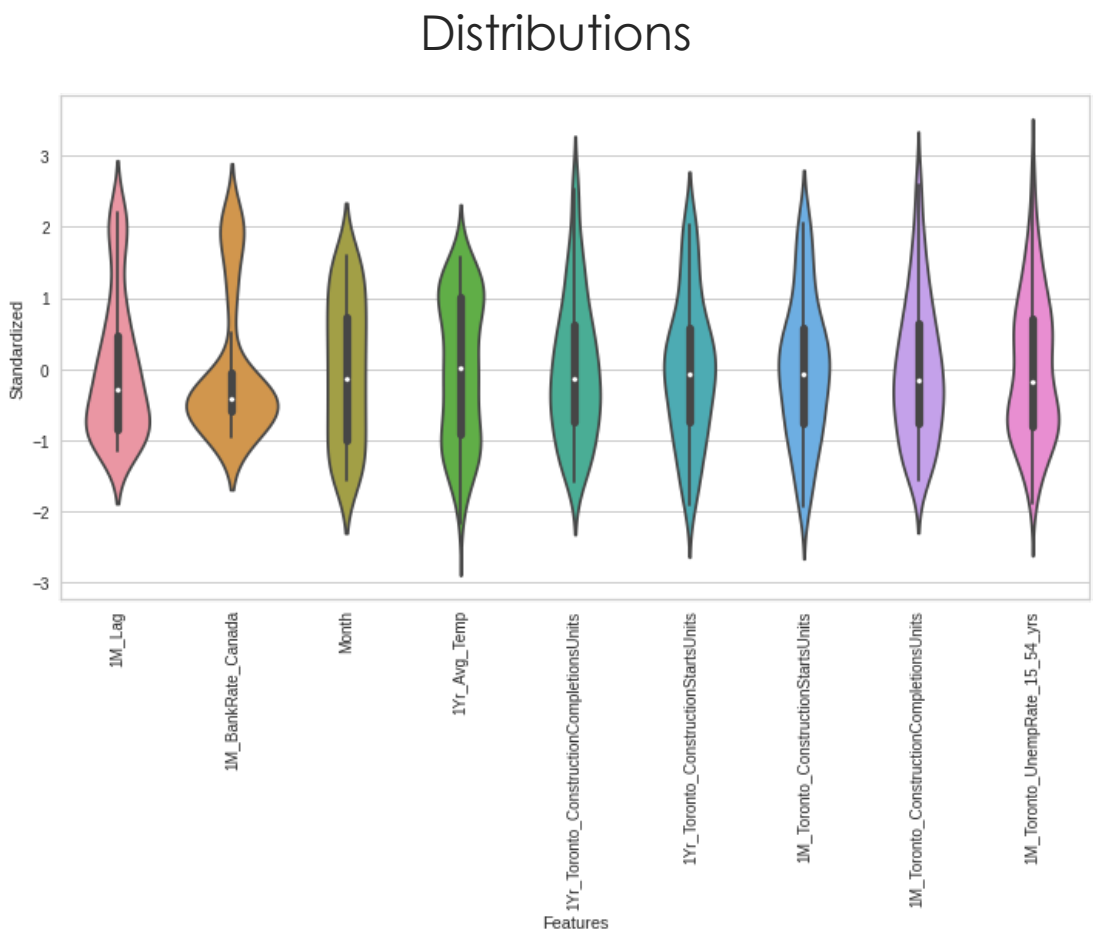
Eliminated	Eliminated	Selected

• Only 2 features
  • Last month HPI
  • Weather
• 0.980 R^2
• Skeptic during economic shifts
	• Gave coefficients for all features selected
	• High performance
	• 0.984 R^2 on test
	• Few hyperparameters to tune over time
		• Gave coefficients for all features selected
		• High performance
		• 0.984 R^2 on test
		• Many hyperparameters

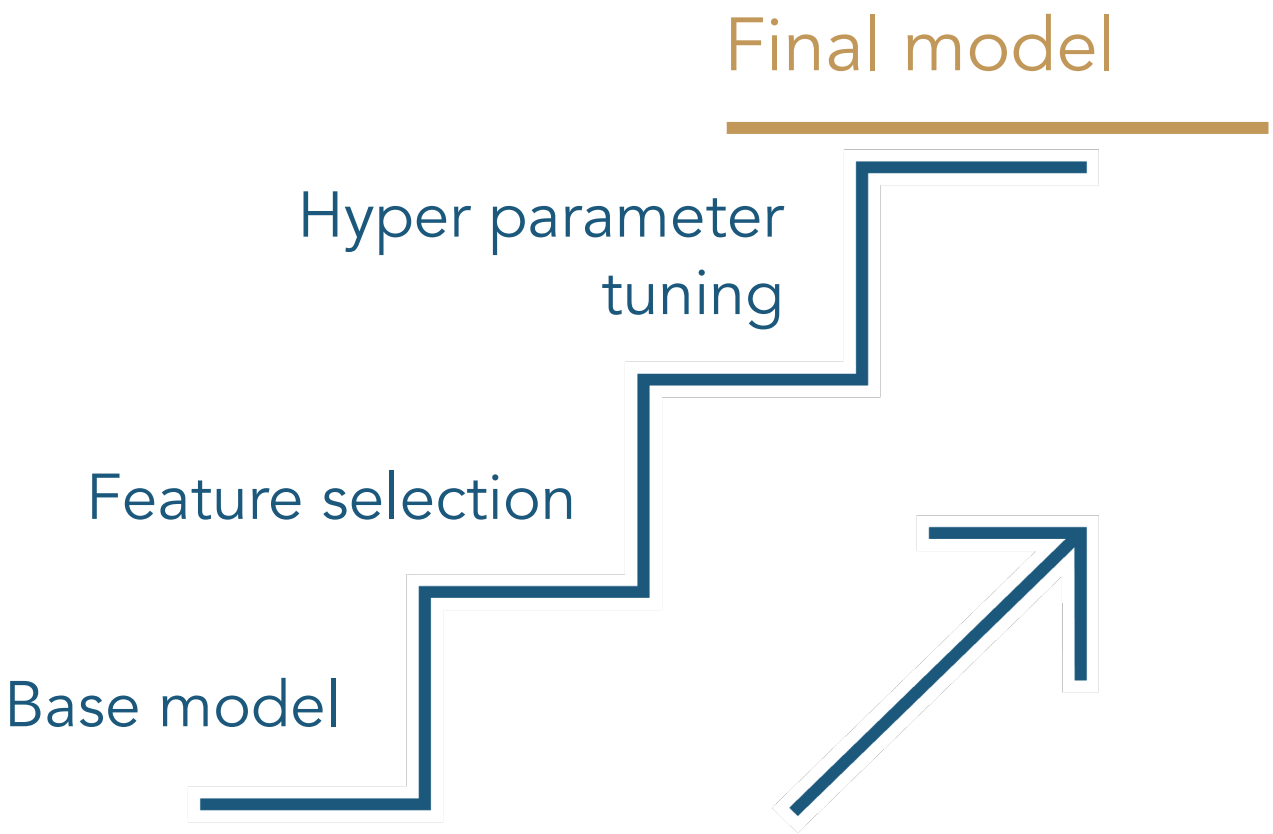
Multicollinearity removed



Selected features are nicely distributed



Bayesian regression evolution



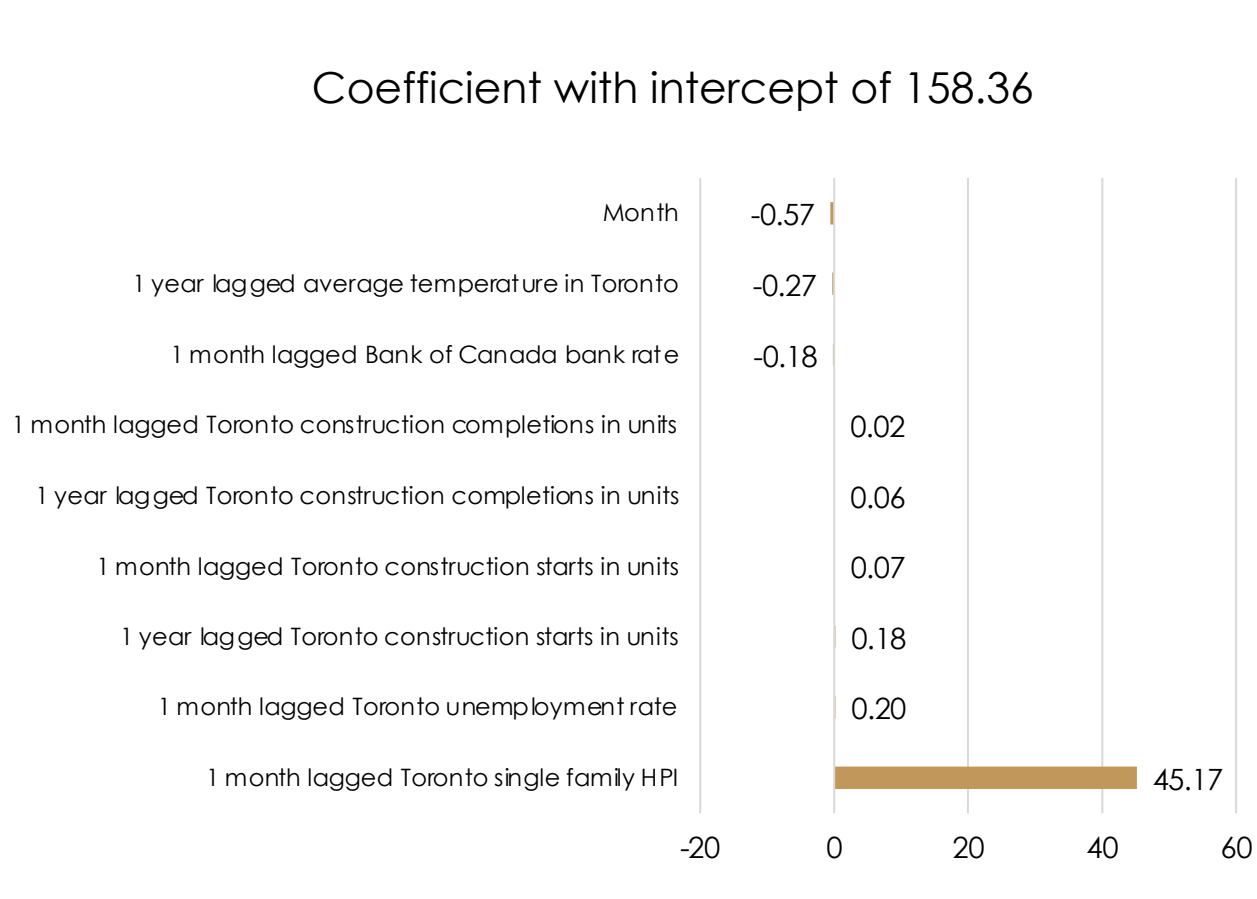
R^2 on test: 0.984
RMSE on test: 2.08
Features: 9
Parameters:

- Fit intercept = True
- Alpha 2 = 0.001
- Alpha 1 = 10
- N iterations = 1,000
- Normalize = True

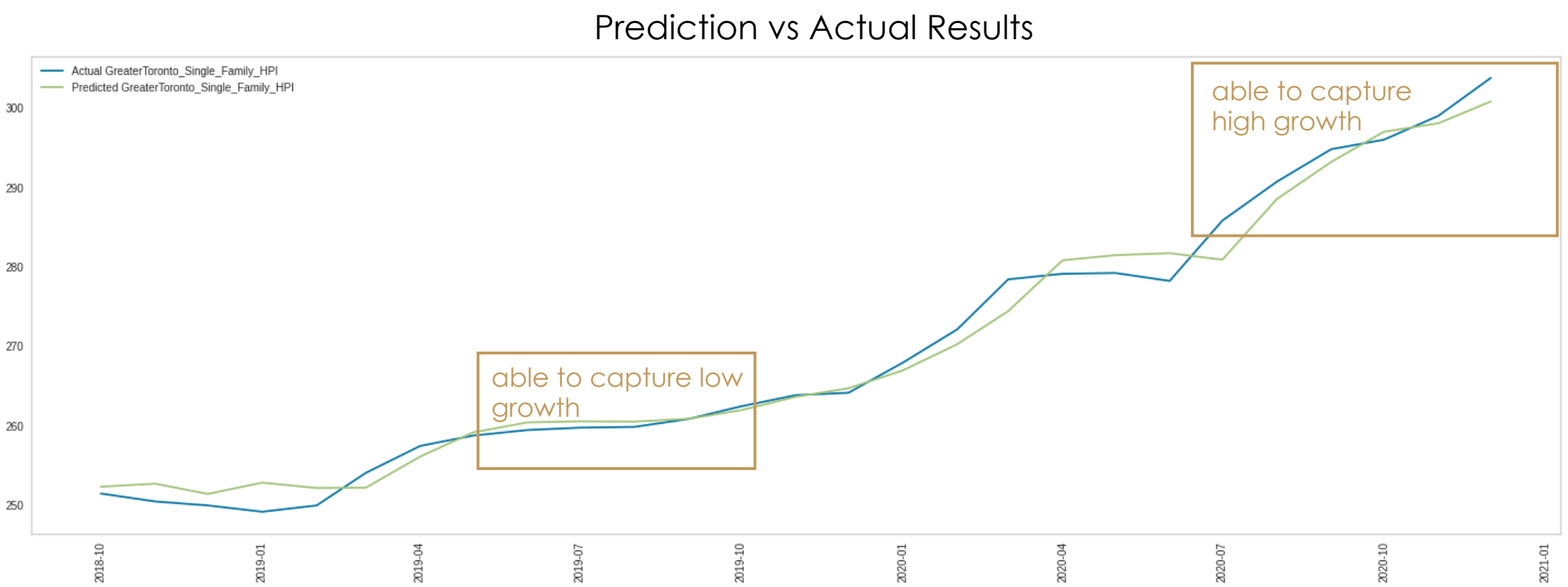
Parsimonious equation (rounded & standardized)

Monthly HPI Prediction =

No change in R^2 or RMSE whether standardized or not (normalized in both cases)



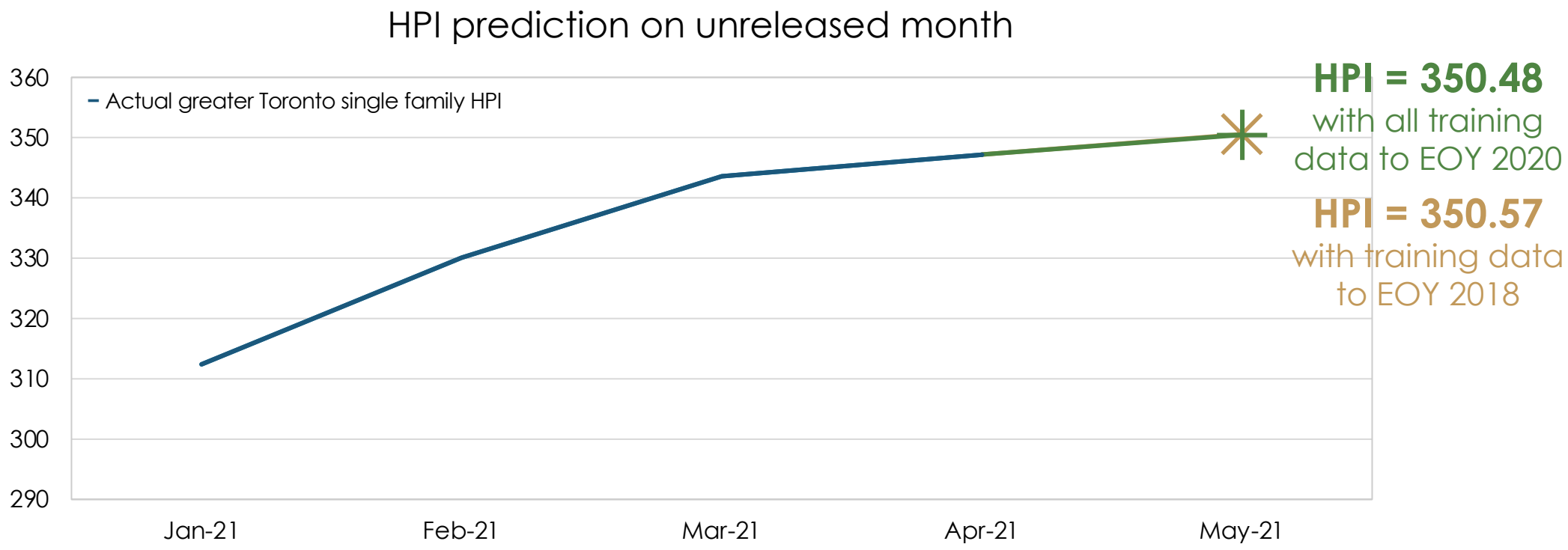
Model output on test data



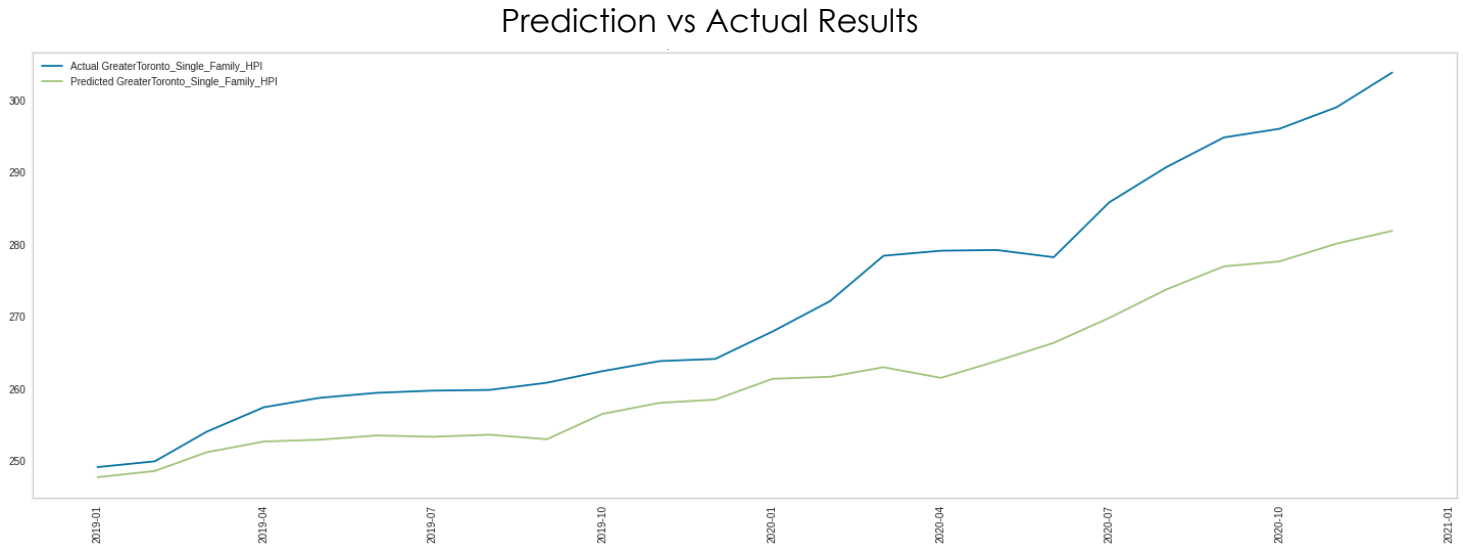
Prediction for May (unreleased)using model on normalized but not standardized data

Feature	Coefficient	Value	Product
Intercept	-1.12	1	-1.12
1 month lagged Toronto single family HPI	1.01	347.20	350.32
Month	-0.16	5	-0.82
1 month lagged bank of Canada bank rate	-0.13	0.50	-0.07
1 month lagged average temperature in Toronto	-0.03	7.89	-0.22
1 year lagged construction completion units in Toronto	6.71e-05	2,942	0.20
1 month lagged construction starts units in Toronto	7.31e-05	2,802	0.21
1 year lagged construction starts units in Toronto	0.00019	2,558	0.49
1 month lagged construction completion units in Toronto	2.55e-05	2,822	0.08
1 month lagged unemployment in Toronto	0.18	8.50	1.51
			350.57

2021 single family HPI with 2 predictions showing little to no data drift



Model with no self lag and standardized data



R^2 on test: 0.432
RMSE on test: 12.02
Features: all
Dropped:

- 2016 & 2017

Recommendations

Consumers
Strategize on best-time to refinance, purchase or sell a home

Banks
Anticipate higher demand for mortgages when lending rate lowers

Home Repair/Renovation
Adjust prices proportional to HPI and plan for material demand

Policy Makers
Education of consumer impacts

Mortgage/Real Estate Agents
Optimize marketing spend and personnel for high seasons

Developers
Adjust prices proportional to HPI index

Property Tax
Evaluate if re-assessment is required in the future

Policy Makers
Proactive action to various changes in factors (such as employment rate)

Model implications

- Retraining required if base year for index changes for either or both HPI and CPI
- Monitor data drift during abnormal times and retrain if model predicting poorly
- Can only predict one month in advance since using prior month's data
- Relies on CREA publishing prior month's data in a timely manner

Model improvements

- Additional features: for example, acquire immigration data from CIC and learn how new immigrants moving to Toronto drives real estate demand
- Using web scraping for sentiment analysis to determine what public feels about housing trends or to capture policy changes which could influence housing bubbles
- Use similar model to predict HPIs for other cities and home types

Thank you!

Appendix A: All columns

Index(['Date', 'Target_GreaterToronto_Single_Family_HPI', '1Yr_MetroOntario_CPI_All',
'1M_MetroOntario_CPI_all', '1Yr_MetroOntario_CPI_Shelter', '1M_MetroOntario_CPI_Shelter', 'Month',
'1Yr_Lag', '1M_Lag', '1Yr_GreaterVancounver_Single_Family_HPI',
'1M_GreaterVancounver_Single_Family_HPI', '1Yr_Calgary_Single_Family_HPI',
'1M_Calgary_Single_Family_HPI', '1Yr_Montreal_Single_Family_HPI', '1M_Montreal_Single_Family_HPI',
'1M_RealEstate_GDP_Trading_AdjustedConstant_BasePrice',
'1Yr_RealEstate_GDP_Trading_AdjustedConstant_BasePrice', '1Yr_RetailSales_Toronto',
'1M_RetailSales_Toronto', '1M_Avg_Temperature', '1Yr_Avg_Temp', '1M_Snow_onGround',
'1Yr_Snow_onGround', '1Yr_Toronto_Median_Annual_Income_Couple_Families',
'1Yr_Toronto_Median_Annual_Income_Lone_Parent_Families',
'1M_Toronto_ConstructionCompletionsUnits', '1Yr_Toronto_ConstructionCompletionsUnits',
'1M_Toronto_ConstructionStartsUnits', '1Yr_Toronto_ConstructionStartsUnits',
'1M_Toronto_UnempRate_15_54_yrs', '1Yr_BankRate_Canada', '1M_BankRate_Canada', '1M_Population',
'1Y_S&P500_Close', '1M_S&P500_Close', '1Y_Bachelors_Degree_Ontario_NewRegEducation',
'1Y_Bachelors_Degree_Ontario_NewRegAll'], dtype='object')

Appendix B: Coefficients trained to 2020

0 intercept
{'fit_intercept': False, 'alpha_2': 1e-05, 'alpha_1': 0.0001}

[('1M_Lag', 1.008367052199725), ('1M_BankRate_Canada', -0.1365842044000074), ('Month', -0.1792894291317665 7), ('1Yr_Toronto_ConstructionCompletionsUnits', 6.61131129536024e-05), ('1M_Toronto_ConstructionCompletionsUnits', -0.0001469348391244013 8), ('1Yr_Avg_Temp', -0.01886744494095134 5), ('1Yr_Toronto_ConstructionStartsUnits', 5.02705070930094 5e-05), ('1M_Toronto_ConstructionStartsUnits', 2.97415824051453 2e-05), ('1M_Toronto_UnempRate_15_54_yrs', 0.185251197568790 2)]

Appendix C: Coefficients no self lag on all standardized data

Intercept 150.36969696969837

[('1Yr_MetroOntario_CPI_All', 4.3359339530660534), ('1M_MetroOntario_CPI_all', 1.058312664546516), ('1Yr_MetroOntario_CPI_Shelter', 0.7337013540851673), ('1M_MetroOntario_CPI_Shelter', 2.525820407753989), ('Month', 5.042981919409141), ('1Yr_GreaterVancounver_Single_Family_HPI', 2.0421408473418263), ('1M_GreaterVancounver_Single_Family_HPI', 15.499695971193294), ('1Yr_Montreal_Single_Family_HPI', 3.030558814575963), ('1M_Montreal_Single_Family_HPI', -0.3703416158921996), ('1M_RealEstate_GDP_Trading_AdjustedConstant_BasePrice', 13.097832233141274), ('1Yr_RealEstate_GDP_Trading_AdjustedConstant_BasePrice', 0.19156564556744485), ('1Yr_RetailSales_Toronto', 2.245223467007025), ('1M_RetailSales_Toronto', -1.3489615333798972), ('1M_Avg_Temperature', 0.5475206285935481), ('1Yr_Avg_Temp', -1.8959158054863605), ('1M_Snow_onGround', 0.015421834192089808), ('1Yr_Snow_onGround', 0.22586881853668353), ('1M_Toronto_ConstructionCompletionsUnits', 0.15368612655643293), ('1Yr_Toronto_ConstructionCompletionsUnits', 0.061603607509710 3), ('1M_Toronto_ConstructionStartsUnits', -0.008778186200790662), ('1Yr_Toronto_ConstructionStartsUnits', 0.09451417591770342), ('1M_Toronto_UnempRate_15_54_yrs', 0.6475051730984084), ('1Yr_BankRate_Canada', -2.290543693106973), ('1M_BankRate_Canada', 1.325407759711066), ('1M_Population', -74.56530972420299), ('1Y_S&P500_Close', -2.021108263069027), ('1M_S&P500_Close', 3.151273147348706), ('1Y_Bachelors_Degree_Ontario_NewRegAll', 1.339805027762643), ('Year', 68.19038720103602)]

Appendix D: Normality no self lag