UNIVERSITY OF ALBERTA
ALBERTA SCHOOL OF BUSINESS

BUEC 311: Business Economics, Organization and Management
Estimating Demand and Supply

Slide Pack

Andrew Leach

September 20, 2021

Outline

1. Elasticity Refresher
    - Own Price
    - Cross-Price
    - Income
    - Elasticities along the demand curve
    - Arc elasticity
2. Using regression analysis to estimate market behaviour
    - Choosing explanatory variables
    - Sample selection
    - Bias, identification, causality
    - Using regressions to forecast
3. The identification problem in practice

Elasticity of Demand

- Recall that the price elasticity of demand is the percentage change in quantity demanded for a given percentage change in price;

$$\epsilon = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in price}} = \frac{\Delta Q/Q}{\Delta p/p} = \frac{\Delta Q}{\Delta p}\frac{p}{Q}$$

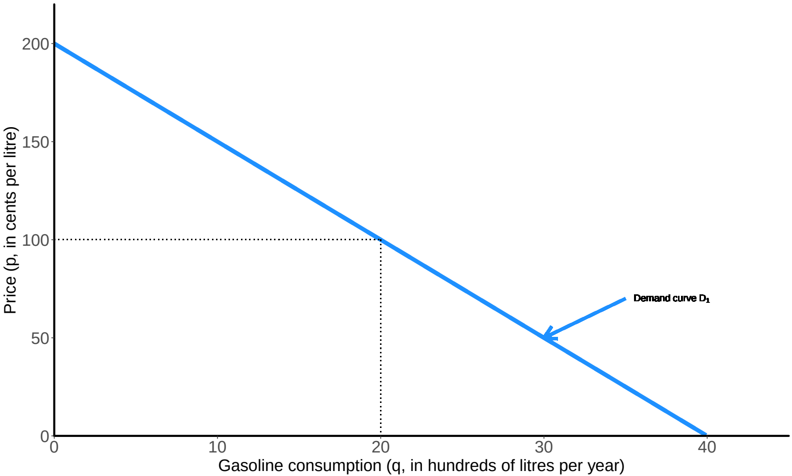
- The elasticity of supply is calculated similarly as the percentage change in quantity supplied for a given percentage change in price;
- Hint: more *elastic* is more responsive:
  - *inelastic* demand or supply means an elasticity less than one (i.e. the relative magnitude of the quantity response is smaller than the price change)
  - *perfectly inelastic* demand has an elasticity of zero
  - *perfectly elastic* demand has undefined elasticity

Elasticity along the demand curve

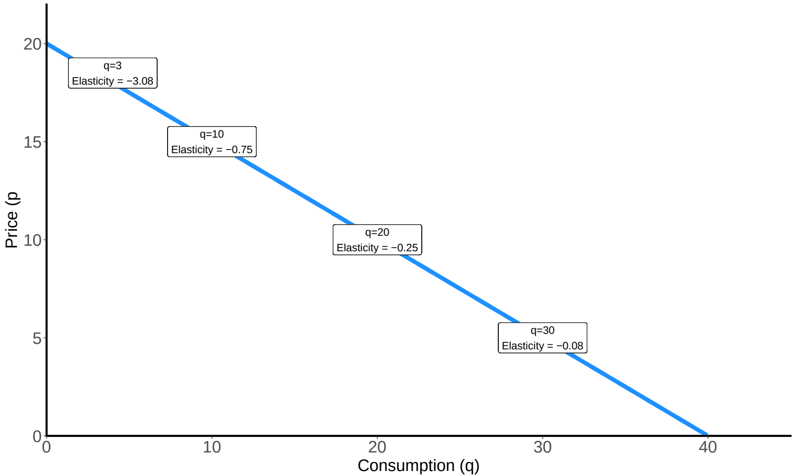
- Elasticity varies as we move along the demand curve
- Why? Because (in our case), the demand curve is linear while elasticity measures percentage changes

- Recall that the euqation of the line is gien by $Q = a + bp$;
- We can use this in the elasticity formula, since we know that $\frac{\Delta Q}{\Delta p} = b$
- We therefore know that $\epsilon = \frac{\Delta Q}{\Delta p}\frac{p}{Q} = b\frac{p}{Q}$
- We will, later in the class, look at constant elasticity relationships
- for a given change in price, the change in quantity you observe gives you the arc elasticity, but may not be representative of elasticity at other (price,quantity) pairs
  - what does this mean for empirical estimates?

Graph comparison



Graph comparison



Estimation of elasticities

- We can use real world data to estimate elasticities
- Examples:
  - Do emissions respond to carbon pricing? (BC carbon tax)
  - Do wages affect labour supply? (Alberta's minimum wage changes)
  - Do the prices of complements affect labour supply? (Quebec child care)
  - Do changes in production technology or regulation affect supply (US light tight oil)
- Beware:
  - Correlation vs causation
  - Omitted variables bias
  - Sample selection bias

Estimation of demand

What is a general manager paying for when they sign a hockey player?

- goals?
- assists?
- points?

What other attributes might matter the an orginization's willingness to pay for a player?

You might have a sense of where this is going...

`https://www.nytimes.com/video/movies/100000001341145/clip-moneyball.html`

Economic Moneyball

Much of the estimation of demand is trying to determine, with statistical tools, the marginal willingness to pay for just one more:

- run
- year of education
- year of experience
- goal
- square footage of floor space
- tenth of a point of GPA
- horsepower
- kWh
- hour of work

Economic Moneyball

Much of the estimation of demand is trying to determine, with statistical tolls, the marginal willingness to pay for just one more:

- run (SABR-metrics)
- year of education (human capital)
- year of experience (human capital)
- goal (hockey analytics)
- square footage of floor space (real estate)
- tenth of a point of GPA (labour markets)
- horsepower (vehicle pricing)
- kWh (vehicle pricing)
- hour of work (labour markets)

Let's try a couple out

- Start with NHL hockey salaries and performance statistics from the 2018-2019 season `https://www.capfriendly.com/`
- Choose only players making more than $1 million per year
- Build a model for the willingness to pay based on salaries as a function of performance

A basic, univariate regression

NHL Salary Regression Models

	Dependent variable:
	salary
Constant	1.840^{***} (0.176)
points	0.065^{***} (0.004)
Observations	440
Adjusted R^2	0.354
Residual Std. Error	2.090 (df = 438)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

- The regression tells us that the average salary is $1.845 million with *one more point* being worth $65,372.400, the precise value of the reported coefficient on points, 0.065, converted to thousands of dollars.

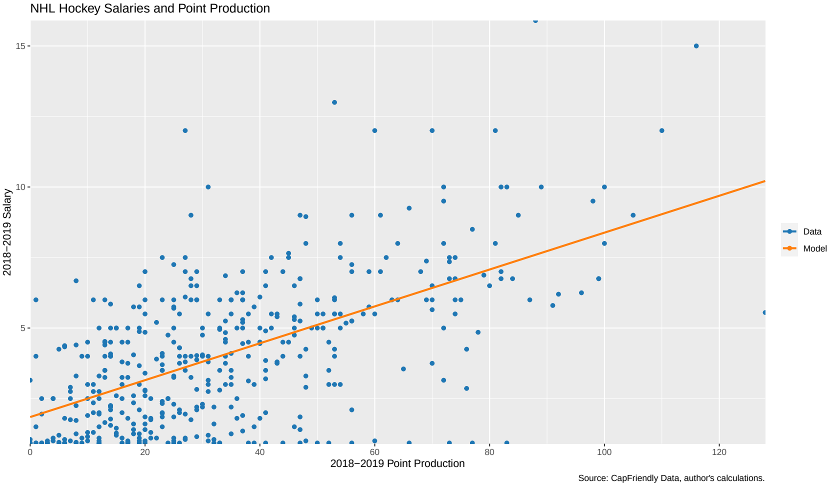
A basic, univariate regression

NHL Salary Regression Models

	Dependent variable:
	salary
Constant	1.840*** (0.176)
points	0.065*** (0.004)
Observations	440
Adjusted R^2	0.354
Residual Std. Error	2.090 (df = 438)
Note:	*p<0.1; **p<0.05; ***p<0.01

- The standard errors tell us how precisely the coefficients are estimated. A consistent relationship or a larger sample size will give you smaller standard errors.
- The regression tells us that *one more point* is worth \$65,372.403 $\pm$ \$8,242.740, 19 times out of 20.

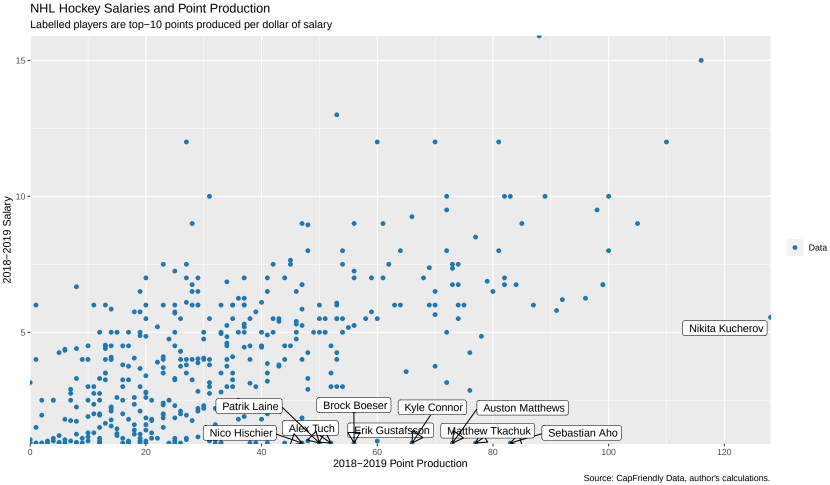
Graph comparison



Goodness of fit

- R-squared (R^2) statistic measures the goodness of fit, or the share of the variation in the dependent variable (salary) that is explained by the regression.
- The statistic must lie between 0 and 1.
  - 1 indicates that 100% of the variation is explained by the regression.
- Adjusted R-squared is a modified version of R-squared
  - Adjusted for the number of predictors in the model.
  - The adjusted R-squared increases only if the new term improves the model more than would be expected by chance.
- $R^2 = 1 - \frac{SS_{Res}}{SS_{Total}}$, where SS_{Res} and SS_{Total} are the sum of squared residuals and total dependent variable sum of squares.
- $Adj\ R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$, where k is the number of parameters.

Residuals and what they can tell us



Now who are the most underpaid players?

label
#1: Sebastian Aho
#2: Matthew Tkachuk
#3: Auston Matthews
#4: Kyle Connor
#5: Erik Gustafsson
#6: Nikita Kucherov
#7: Brock Boeser
#8: Alex Tuch
#9: Patrik Laine
#10: Nico Hischier

What are we missing

- Multivariate regression
- Sample selection bias
- Omitted variables bias
- Causality vs correlation

Multivariate regression

- Multivariate Regression uses two or more explanatory variables.
  - e.g. $s = a + bX_1 + cX_2 + \epsilon$.
- We can estimate (in our case) NHL salaries as a function of point production but also other explanatory variables
  - a, b, and c are **coefficients** to be estimated
  - ϵ is a random error or residual.
- When we use least squares regression to estimate a, b, and c, we get a prediction of the salary for combination of measurable outputs.
- Least squares regression minimizes the sum of squared residuals.
  - We can also learn from residuals
- With multivariate regression, we can isolate the individual **marginal effects** of each explanatory variable **holding the other explanatory variables constant**.

A slightly better regression

NHL Salary Regression Models

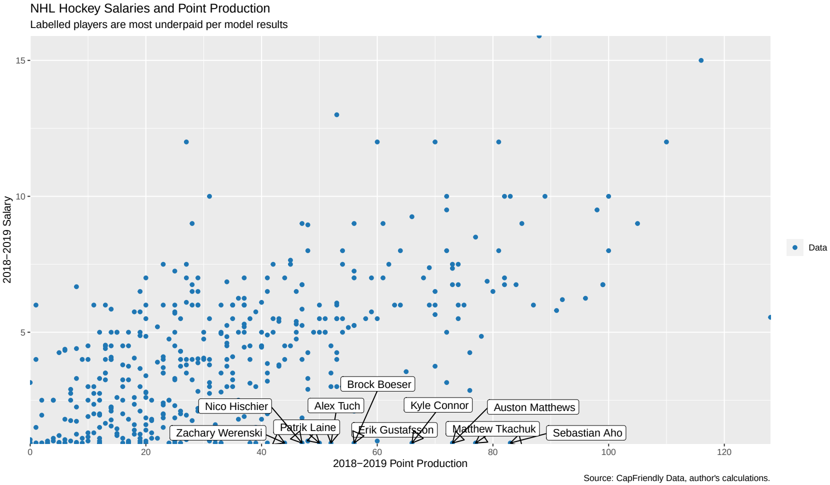
	Dependent variable:	
	salary	
	(1)	(2)
Constant	1.840*** (0.176)	−1.220*** (0.460)
points	0.065*** (0.004)	0.050*** (0.005)
toi		0.206*** (0.029)
Observations	440	440
Adjusted R^2	0.354	0.420
Residual Std. Error	2.090 (df = 438)	1.980 (df = 437)
Note:		*p<0.1; **p<0.05; ***p<0.01

- The second regression tells us that the unadjusted salary is $-1.219 million with *one more point* being worth $50,411.400, and *one more minute of ice time* being worth $205,631.820.
- How can a negative unadjusted salary value possibly make sense? Have I done something wrong?

Now who are the most underpaid players?

label
#1: Sebastian Aho
#2: Erik Gustafsson
#3: Matthew Tkachuk
#4: Auston Matthews
#5: Kyle Connor
#6: Zachary Werenski
#7: Brock Boeser
#8: Nico Hischier
#9: Patrik Laine
#10: Alex Tuch

Now who are the most underpaid players?



Adjust for position

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	1.840*** (0.176)	−1.220*** (0.460)	−2.290*** (0.810)
points	0.065*** (0.004)	0.050*** (0.005)	0.038*** (0.008)
toi		0.206*** (0.029)	0.309*** (0.066)
positionD			−0.575 (1.630)
toi:positionD			−0.011 (0.106)
points:positionD			−0.003 (0.017)
Observations	440	440	440
Adjusted R^2	0.354	0.420	0.423
Residual Std. Error	2.090 (df = 438)	1.980 (df = 437)	1.970 (df = 434)
Note:			*p<0.1; **p<0.05; ***p<0.01

- The third regression tells us that the unadjusted salary is $-2.289 million with *one more point* being worth $38,245.700, and *one more minute of ice time* being worth $38,245.700.
- We also adjust for differential values for each for defence vs forwards.

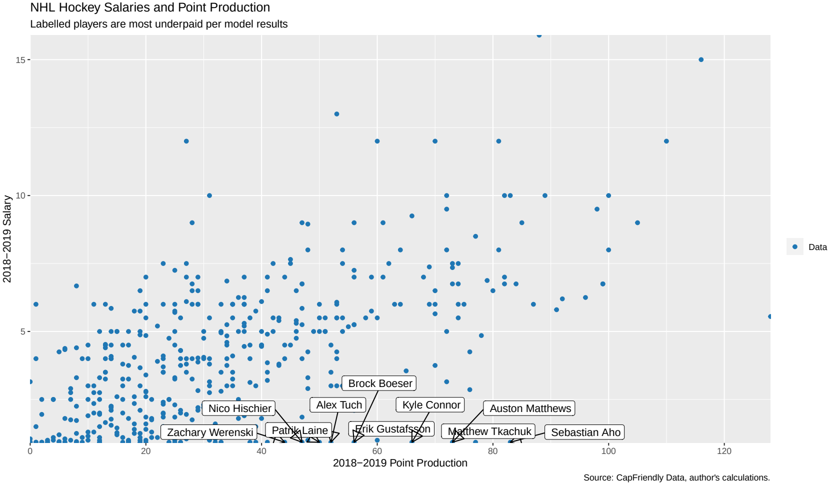
Adjust for position

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	1.840*** (0.176)	−1.220*** (0.460)	−2.290*** (0.810)
points	0.065*** (0.004)	0.050*** (0.005)	0.038*** (0.008)
toi		0.206*** (0.029)	0.309*** (0.066)
positionD			−0.575 (1.630)
toi:positionD			−0.011 (0.106)
points:positionD			−0.003 (0.017)
Observations	440	440	440
Adjusted R^2	0.354	0.420	0.423
Residual Std. Error	2.090 (df = 438)	1.980 (df = 437)	1.970 (df = 434)
Note:			*p<0.1; **p<0.05; ***p<0.01

- Note here that, as we add variables to the regression which help explain the salary levels, we're seeing the other coefficient values change. Why?

Now who are the most underpaid players?



Now who are the most underpaid players?

label
#1: Sebastian Aho
#2: Auston Matthews
#3: Kyle Connor
#4: Matthew Tkachuk
#5: Erik Gustafsson
#6: Brock Boeser
#7: Zachary Werenski
#8: Nico Hischier
#9: Patrik Laine
#10: Alex Tuch

Omitted variables bias

NHL Salary Regression Models

	Dependent variable: salary (1)	(2)	(3)
Constant	1.840*** (0.176)	−1.220*** (0.460)	−2.290*** (0.810)
points	0.065*** (0.004)	0.050*** (0.005)	0.038*** (0.008)
toi		0.206*** (0.029)	0.309*** (0.066)
positionD			−0.575 (1.630)
toi:positionD			−0.011 (0.106)
points:positionD			−0.003 (0.017)
Observations	440	440	440
Adjusted R^2	0.354	0.420	0.423
Residual Std. Error	2.090 (df = 438)	1.980 (df = 437)	1.970 (df = 434)
Note:			*p<0.1; **p<0.05; ***p<0.01

- In regression (1), I tried to explain salaries using only points. What variables did I omit?
- TOI is obviously important, and so by omitting it, I'd been forcing points to explain other things that are correlated with it

Non-linear effects

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	1.845*** (0.176)	−1.219*** (0.460)	0.360 (1.501)
points	0.065*** (0.004)	0.050*** (0.005)	0.037*** (0.013)
points2			0.0002 (0.0001)
toi		0.206*** (0.029)	0.035 (0.179)
toi2			0.005 (0.005)
Observations	440	440	440
Adjusted R^2	0.354	0.420	0.421
Residual Std. Error	2.087 (df = 438)	1.977 (df = 437)	1.975 (df = 435)
Note:			*p<0.1; **p<0.05; ***p<0.01

- Note here that, as we add variables to the regression which help explain the salary levels, we're allowing for non-linear effects of points or toi. Why?

Omitted variables bias

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	1.840*** (0.176)	−1.220*** (0.460)	−2.290*** (0.810)
points	0.065*** (0.004)	0.050*** (0.005)	0.038*** (0.008)
toi		0.206*** (0.029)	0.309*** (0.066)
positionD			−0.575 (1.630)
toi:positionD			−0.011 (0.106)
points:positionD			−0.003 (0.017)
Observations	440	440	440
Adjusted R^2	0.354	0.420	0.423
Residual Std. Error	2.090 (df = 438)	1.980 (df = 437)	1.970 (df = 434)
Note:			*p<0.1; **p<0.05; ***p<0.01

- TOI is obviously important, and so by omitting it, I'd been forcing points to explain other things that are correlated with it like TOI
- TOI, all else equal, correlates with points, so I was forcing points to explain both the direct (points) effect and the TOI (omitted) effect

Sample selection bias

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	-1.220^{***} (0.460)	0.830 (0.551)	3.190^{***} (0.695)
points	0.050^{***} (0.005)	0.049^{***} (0.004)	0.041^{***} (0.005)
toi	0.206^{***} (0.029)	0.130^{***} (0.032)	0.059 (0.037)
Observations	440	324	223
Adjusted R^2	0.420	0.379	0.287
Residual Std. Error	1.980 (df = 437)	1.770 (df = 321)	1.680 (df = 220)
Note:			*p<0.1; **p<0.05; ***p<0.01

- Regression (1) is players earning more than \$1 million, Regression (1) is players earning more than \$2 million, Regression (3) is players earning more than \$4 million.
- The estimated values of one more point and one more minute of TOI are different in the different samples

Causation and correlation

NHL Salary Regression Models

	Dependent variable:		
	salary		
	(1)	(2)	(3)
Constant	-1.220^{***} (0.460)	0.830 (0.551)	3.190^{***} (0.695)
points	0.050^{***} (0.005)	0.049^{***} (0.004)	0.041^{***} (0.005)
toi	0.206^{***} (0.029)	0.130^{***} (0.032)	0.059 (0.037)
Observations	440	324	223
Adjusted R^2	0.420	0.379	0.287
Residual Std. Error	1.980 (df = 437)	1.770 (df = 321)	1.680 (df = 220)
Note:			*p$<$0.1; **p$<$0.05; ***p$<$0.01

- None of these results can possibly explain base salaries in 2018-2019. Any idea why?

Causation and correlation

NHL Salary Regression Models

	Dependent variable: salary (1)	(2)	(3)
Constant	-1.220^{***} (0.460)	0.830 (0.551)	3.190^{***} (0.695)
points	0.050^{***} (0.005)	0.049^{***} (0.004)	0.041^{***} (0.005)
toi	0.206^{***} (0.029)	0.130^{***} (0.032)	0.059 (0.037)
Observations	440	324	223
Adjusted R^2	0.420	0.379	0.287
Residual Std. Error	1.980 (df = 437)	1.770 (df = 321)	1.680 (df = 220)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

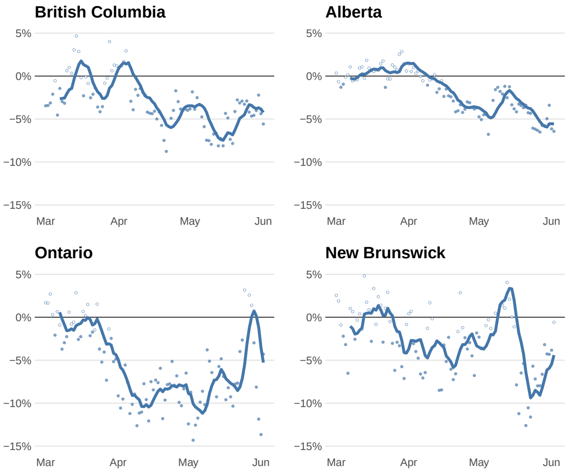
- It's temporally impossible for 2018-2019 points to explain base salaries paid in that season since the salaries are negotiated before the season starts
- How could I rectify this problem with my data?

Another example - the demand for electricity in COVID Times

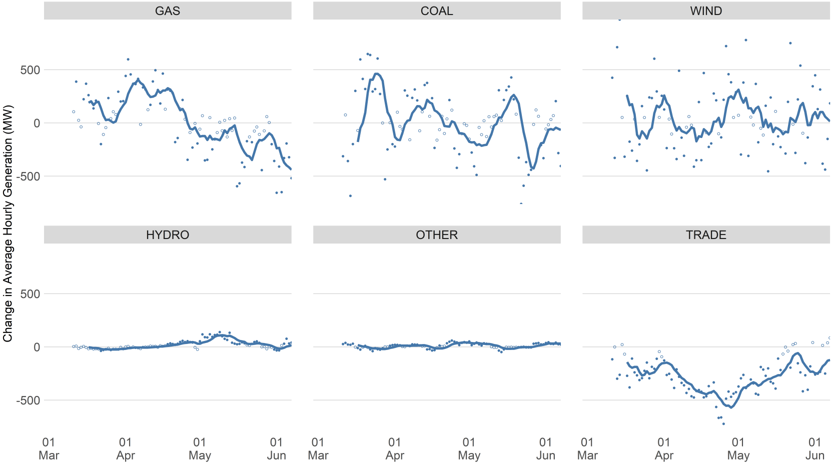
Andrew Leach, Nic Rivers, and Blake Shaffer. 2020. Canadian Electricity Markets during the COVID-19 Pandemic: An Initial Assessment. *Canadian Public Policy* 46:S2, S145-S159

- apologies for the self-citation, but it's work that I can talk about pretty easily
- we were interested in how the electricity market in 4 provinces had changed after the COVID-19 pandemic started
- we were mostly interested in this as a real-time proxy for economic activity for which we often get data with a lag
- our technique used (basically) the same approach as the NHL hockey salaries, asking how much lower is demand today than it would be expected to be adjusting for day, weekday, month, temperatures, and other factors

Demand for electricity in COVID Times



Alberta's demand for electricity in COVID Times



Ontario Industrial demand for electricity in COVID Times