Applied Quantitative Analysis II
Lab 6

Eric G. Zhou

Spring 2020

NYU Wagner School of Public Service
NYU Grossman School of Medicine

General reminder

Next week is the mid-term week – half way through the semester

You should start analyzing the data for your project – the most important part

But greatness comes in entirety – other parts are as important

Contents for today

Quiz

Briefly on the introduction part of your paper

Brief recap & introduction to dimensionality

Interaction effects

Non-parametric methods

Stata Programming

AQA II: Lab 6

Quiz questions

1. When would you use ordered or multinomial logit/probit regression? Think about other factors beside whether your outcome is ranked.

2. Many times we already have a great regression results, why bother making a graph of the estimates over some other covariates?

3. What is a simple test for the sensitivity of your regression analysis, especially when your sample size is not great?

Quiz questions

1. When would you use ordered or multinomial logit/probit regression? Think about other factors beside whether your outcome is ranked.

Quiz questions

2. Many times we already have a great regression results, why bother making a graph of the estimates over some other covariates?

Quiz questions

3. What is a simple test for the sensitivity of your regression analysis, especially when your sample size is not great?

Quiz answers

1. When would you use ordered or multinomial logit/probit regression? Think about other factors beside whether your outcome is ranked.

Sample size matters marginally as the mlogoit needs more degree of freedom. The assumption of mlogit (IIA) is hard to justify, whereas the assumption of ologit (PLA) is testable and easy to adjust using the generalized ordered logit models.

Sometimes, we can retreat to run a truncated/discretized continuous variable with a ologit, e.g., income brackets, education brackets, etc. Although you would need to work hard to justify your choice.

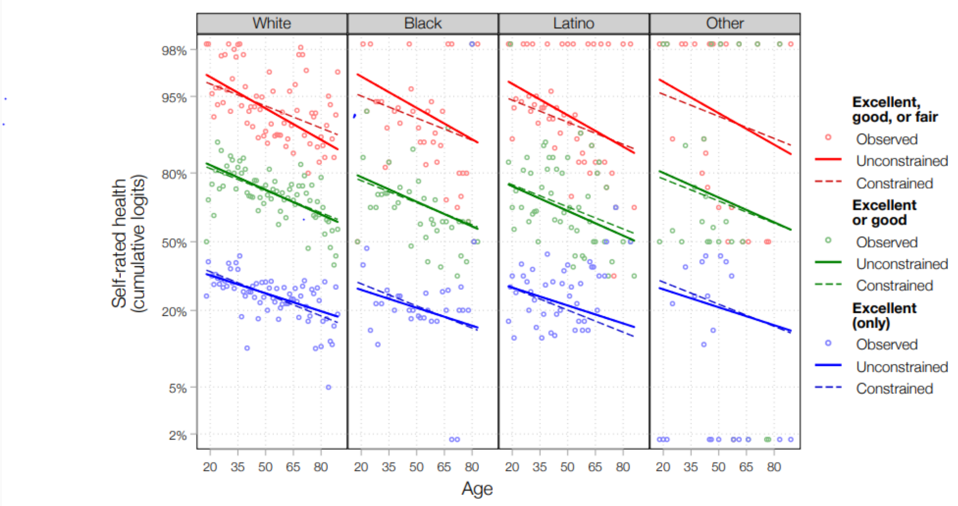
Quiz answers

2. Many times we already have a great regression results, why bother making a graph of the estimates over some other covariates?

It is very often the better to tell a story with graph, and sometimes plotting a graph is the only way to tell a story as in most non-parametric methods.

Besides, a graph would help you to understand/verify if any patterns in your estimates, untold by your regression estimates, have regional over/under fitting and local curvature, etc.

AQA II: Lab 6



White
Black
Latino
Other
98%
95%
80%
50%
20%
5%
2%
Self-rated health (cumulative logits)
20 35 50 65 80
Age
Excellent, good, or fair
Observed
Unconstrained
Constrained
Excellent or good
Observed
Unconstrained
Constrained
Excellent (only)
Observed
Unconstrained
Constrained

Quiz answers

3. What is a simple test for the sensitivity of your regression analysis, especially when your sample size is not great?

You could simply subset your data randomly and run the same specification to see if the estimates are sensitive.

Another simple approach, add interactions!

If you are into prediction modeling, check out Cross Validation, which I would like to cover if we have time later on.

Briefly on the introduction part of your paper

Motivate with a puzzle or a problem (1 - 2 paragraphs)

Clearly state your research question (1 paragraph)

Empirical approach (1 paragraph)

Detailed results (3 - 4 paragraphs)

Value-added relative to related literature (1 - 3 paragraphs)

Optional paragraphs: robustness checks, policy/theoretical relevance, limitations

Roadmap (1 paragraph)

Briefly recap & intro to dimensionality

See the white broad.

Interaction effects of regression models

Let's consider an example drawn from Riphahn, Wambach, & Million (2003, JAE). Their model to estimate income is

$$\begin{aligned}\text{Income} = \exp(&\beta_1 + \beta_2\text{Age} + \beta_3\text{Age}^2 + \beta_4\text{Educ} + \beta_5\text{Female} \\ &+ \beta_6\text{Female x Education} + \beta_7\text{Age x Education}) + \epsilon\end{aligned}$$

This is apparently a non-linear model in its current form as it is fitted exponentially. Though the income distribution is often believed to be log-normal, the interpretation can be very tricky. Why?

Interaction effects of regression models

	Nonlinear Least Squares			Linear Least Squares		
Variable	*Estimate*	*Std. Error*	*t*	*Estimate*	*Std. Error*	*t*
Constant	−2.58070	0.17455	14.78	−0.13050	0.06261	−2.08
Age	0.06020	0.00615	9.79	0.01791	0.00214	8.37
Age^2	−0.00084	0.00006082	−13.83	−0.00027	0.00001985	−13.51
Education	−0.00616	0.01095	−0.56	−0.00281	0.00418	−0.67
Female	0.17497	0.05986	2.92	0.07955	0.02339	3.40
Female × Educ	−0.01476	0.00493	−2.99	−0.00685	0.00202	−3.39
Age × Educ	0.00134	0.00024	5.59	0.00055	0.00009394	5.88
$\mathbf{e}'\mathbf{e}$		106.09825			106.24323	
s		0.15387			0.15410	
R^2		0.12005			0.11880	

Interaction effects of regression models

This is apparently a non-linear model in its current form as it is fitted exponentially. Though the income distribution is often believed to be log-normal, the interpretation can be very tricky. Why?

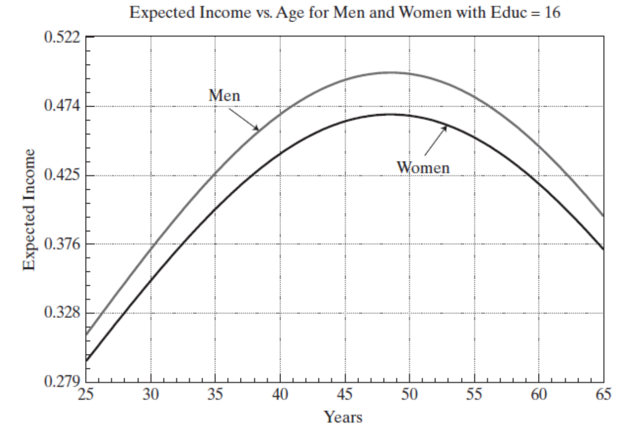
The coefficient can be thought of as the marginal effect of certain regressors and it should be:

$$\partial E[y|x]/\partial x_k = \exp(\alpha + x'\beta) \times \partial(\alpha + x'\beta)/\partial x_k \tag{1}$$

As a result, the second part is not equal to the coefficient. So what we should do if we really like this way of fitting the models?

Interaction effects of regression models

It's good to plot them and we have shown how to in Stata. Relationship between age and predicted income stratified by gender **when educ = 16**.



Interaction effects of regression models

If you really like to talk about numbers or drawing too many graphs isn't helping. We could either fit the model linearly, or do log transformations.

In most social sciences areas, you do not need to discuss the error structure when you do such things.(except in Econ)

And you should be very familiar with how to interpret the interaction effects. So, let's practice!

Interaction effects of regression models

	Nonlinear Least Squares			*Linear Least Squares*		
Variable	*Estimate*	*Std. Error*	t	*Estimate*	*Std. Error*	t
Constant	−2.58070	0.17455	14.78	−0.13050	0.06261	−2.08
Age	0.06020	0.00615	9.79	0.01791	0.00214	8.37
Age^2	−0.00084	0.00006082	−13.83	−0.00027	0.00001985	−13.51
Education	−0.00616	0.01095	−0.56	−0.00281	0.00418	−0.67
Female	0.17497	0.05986	2.92	0.07955	0.02339	3.40
Female × Educ	−0.01476	0.00493	−2.99	−0.00685	0.00202	−3.39
Age × Educ	0.00134	0.00024	5.59	0.00055	0.00009394	5.88
$\mathbf{e}'\mathbf{e}$		106.09825			106.24323	
s		0.15387			0.15410	
R^2		0.12005			0.11880	

Non-parametric methods

Let's throw in some definitions first. What's a parametric vs non-parametric method?

Technically speaking, a parametric model is a family of probability distributions that has a finite number of parameters.

$$\mathbb{P} = \{P_\theta | \theta \in \Theta\}$$

We say the mode is a parametric model if $\Theta \subseteq \mathbb{R}^k$ for some positive integer k. Whereas, for non-parametric methods, the number of parameters goes to infinity when your sample size does.

What about semi-parametric?

Something in between. They combine a parametric component, greatly reducing the dimensionality, with a nonparametric component.

Non-parametric methods

An example of a (in)famous parametric model:

the Weibull translation model with a 3-D parameter $\theta = (\lambda, \beta, \mu)$:

$$\mathcal{P} = \{f_\theta(x) = \frac{\beta}{\lambda}(\frac{x-\mu}{\lambda})^{\beta-1}\exp(-(\frac{x-\mu}{\lambda})^{\beta})\mathbf{1}_{x>\mu} | \lambda > 0, \beta > 0, \mu \in \mathbb{R}\}$$

This is actually part of the property proofs we need for the non-linear least square estimates earlier.

Non-parametric methods

What's the biggest implication of having these methods separated?

It's all about assumptions. We make minimum assumptions about non-parametric method, while we assume functional forms and DGP in parametric methods. So, in general, the non-parametric method is more flexible.

The asymptotic theories are different for these two methods too. Can you guess which converge more slowly?

Examples of Non-parametric methods

Essentially, we are interested in variables, for instance Y or X of any given dimensions.(observables)

One basic yet important group of methods are called **Kernel Density Estimations**, and they include:

- Histogram
- Kernel density functions – many variations
- Local near regression
- Lowess
- Spline
- Series estimator

Examples of Non-parametric methods
Histograms

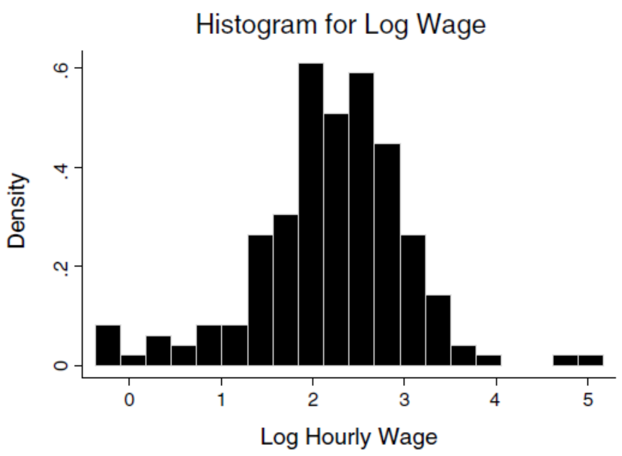
For a sample $x_i, i = 1, ..., N$, the histogram estimator is :

$$\hat{f}_{HIST}(x_0) = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{1}(x_0 - h < x_i < x_0 + h))}{2h}$$

In other words, you choose a h and cut data into equal bins and plot them up with the frequency in each bin.

Examples of Non-parametric methods

Histograms: an example



Examples of Non-parametric methods

Kernel density estimations The Kernel density estimations, introduced by Rosenblatt (1956) can be in a general form as

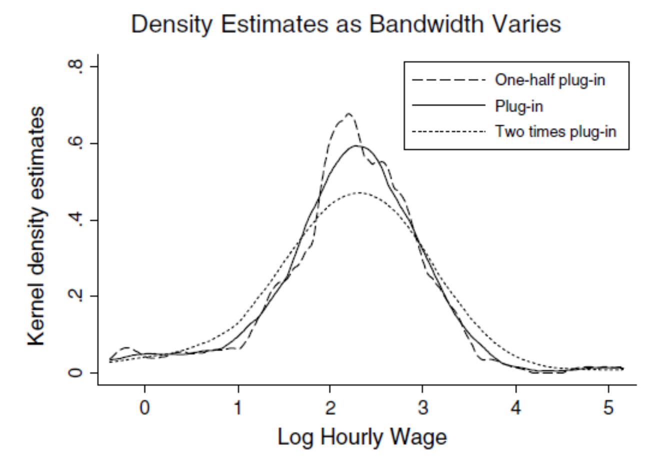
$$\hat{f}(x_0) = \frac{1}{Nh}\sum_{i=1}^{N} K(\frac{x_i - x_0}{h})$$

where the $K(\cdot)$ is called a Kernel function that satisfies some restrictions, and the parameter h is called the bandwidth. The change of either one could result in different estimates.

Let's see some examples!

Examples of Non-parametric methods

Kernel density estimations: an example with varying bandwidths



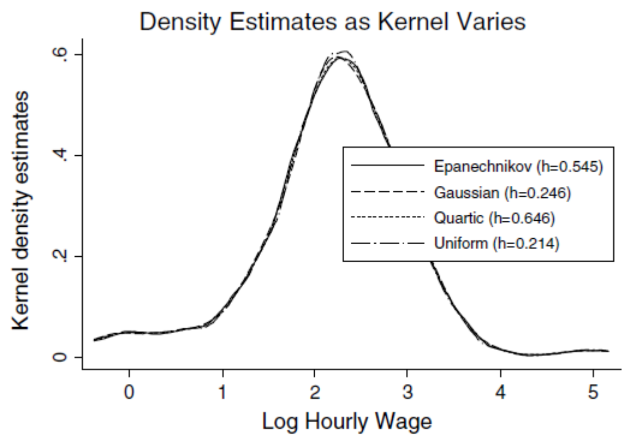
Examples of Non-parametric methods

Kernel density estimations: an example with varying Kernel functions

Kernel	Kernel Function $K(z)$	δ
Uniform (or box or rectangular)	$\frac{1}{2} \times \mathbf{1}(\lvert z \rvert < 1)$	1.3510
Triangular (or triangle)	$(1 - \lvert z \rvert) \times \mathbf{1}(\lvert z \rvert < 1)$	–
Epanechnikov (or quadratic)	$\frac{3}{4}(1 - z^2) \times \mathbf{1}(\lvert z \rvert < 1)$	1.7188
Quartic (or biweight)	$\frac{15}{16}(1 - z^2)^2 \times \mathbf{1}(\lvert z \rvert < 1)$	2.0362
Triweight	$\frac{35}{32}(1 - z^2)^3 \times \mathbf{1}(\lvert z \rvert < 1)$	2.3122
Tricubic	$\frac{70}{81}(1 - \lvert z \rvert^3)^3 \times \mathbf{1}(\lvert z \rvert < 1)$	–
Gaussian (or normal)	$(2\pi)^{-1/2} \exp(-z^2/2)$	0.7764
Fourth-order Gaussian	$\frac{1}{2}(3 - z)^2(2\pi)^{-1/2} \exp(-z^2/2)$	–
Fourth-order quartic	$\frac{15}{32}(3 - 10z^2 + 7z^4) \times \mathbf{1}(\lvert z \rvert < 1)$	–

Examples of Non-parametric methods

Kernel density estimations: an example with varying Kernel functions



Examples of Non-parametric methods

Nonparametric local regressions