

Exam 3

December 16, 2010: 50 points

Consider an always “hot-button” issue of health insurance. Suppose that we model the demand for health insurance in the United States over the previous 20 years – 1991-2010. Assume that insurance demand is measured by the number of people who purchase coverage.

1. What would be your strategy in modeling insurance demand? That is, how would you choose regressors and what might be some of those regressors? (8 points)
2. Why would you use a multiple regressor model? Why not simply use k -number of simple regression models, where k is the number of independent variables used to explain variation in insurance demand? (5 points)
3. What happens if you omit a relevant variable from the model? Why? (5 points)
4. Suppose that your OLS regression is estimated, and the following coefficients represent marginal effects of the included variables:

Coefficient	Represented variable
β_0	Constant
$\beta_1 - \beta_4$	Four variables from part (1)
$\beta_{age\ male}$	Average age of males
$\beta_{age\ female}$	Average age of females
$\beta_{child\ ill}$	Number of children that suffer from serious chronic health problems

It has been shown that older females are more likely to bear children who are more likely to suffer from serious chronic health problems. All variables intuitively belong in the model, but your estimation results produce the following output:

Parameter	Estimate	Standard Error	t-statistic	p-value
β_0	25.224	5.44	4.64	4.0E-06
β_1	1.245	1.9	0.66	0.512
β_2	0.975	0.274	3.56	3.9E-04
β_3	-3.674	1.32	-2.78	5.5E-03
β_4	0.117	0.069	1.70	0.090
$\beta_{age\ male}$	0.023	0.0074	3.11	0.002
$\beta_{age\ female}$	0.034	0.0011	31.00	8.8E-148
$\beta_{child\ ill}$	-0.134	0.234	-0.57	0.567

$R^2 = 0.7854$	$n = 1,000$	F -statistic = 45.334 (p-value = 0.0001)
----------------	-------------	---

What is surprising about these results? What do you infer from this? If there is anything unintuitive about the empirical results, what other methods might you implement to check if there are problems? How might you go about dealing with the problem, if it exists? (8 points)

5. What may be another issue associated with your time-series data? What would this issue do to the form of the error variance? That is, how is the form of $Var(\varepsilon|\mathbf{X})$ under the Gauss-Markov assumptions different from the form when the Gauss-Markov assumption is violated? (5 points)
6. How would you set up a statistical test for determining if the issue in part (5) exists? Suppose that your statistical test shows the presence of the issue. Plot a hypothetical scatter plot of the OLS regression residuals against the lagged OLS regression residuals. (4 points)
7. Suppose you want to account for the introduction of the 2010 Patient Protection and Affordable Care Act (healthcare reform bill), which seeks for every American have health insurance. How would you incorporate the introduction of this policy into your model? How would you interpret the associated coefficient? (5 points)
8. How would you add a term that answers the question of whether parents with children suffering serious chronic health problems increased their demand for health insurance after the health bill? (3 points)
9. Suppose that you would like to test whether older females have a greater effect on insurance demand than older males. Set up the null and alternative hypotheses as well as the associated R , β , and q matrices. Explain how you would test the null hypothesis. (7 points)