ECON 35101 International Macroeconomics and Trade: Assignment 2

Xianglong (Sean) Kong*

October 31, 2021

Table 1

Table 1: Log-linear estimation

Panel A: driving distance

	(1) reg	(2) xtreg	(3) areg	(4) reghdfe
log(driving distance)	-0.407***	-0.407***	-0.407***	-0.407***
	(0.00120)	(0.00120)	(0.00120)	(0.00120)
Observations	358160	358160	358160	358159
R^2	0.502	0.482	0.502	0.502
T	62.61	34.13	28.34	1.724

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel B: driving time

	(1) reg	(2) xtreg	(3) areg	(4) reghdfe
log(driving time)	-0.861***	-0.861***	-0.861***	-0.861***
	(0.00218)	(0.00218)	(0.00218)	(0.00218)
Observations	358160	358160	358160	358159
R^2	0.542	0.524	0.542	0.542
T	64.38	34.64	25.32	1.489

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1. *Are the point estimates and standard errors numerically identical across the different estimators? Should they be?*

   Yes, and they should be.

2. *Are the number of observations and R-squared statistics identical? Should they be?*

   reghdfe drops singletons, which is why it has one less observation than the other three. *xtreg* reports different R^2, see Stata FAQ page for explanations.

*Kenneth C. Griffin Department of Economics, University of Chicago. Email: xlkong@uchicago.edu.

3. *Compare the relative computation times of these estimators.*

   From the slowest to the fastest: *reg* > *xtreg* > *areg* > *reghdfe*.

4. *Are the coefficients on the distance and time covariates the same? Should they be?*

   No, and they shouldn't be. Intuitively, driving time is affected by driving distance as well as other factors (e.g. traffic).

Table 2

Table 2: Zeros

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(driving distance)	-0.407***	-0.309***	-0.328***	-1.057***	-4.616***	-0.372***	-0.372***	-0.280***
	(0.00120)	(0.000891)	(0.000586)	(0.00243)	(0.0127)	(0.00306)	(0.00306)	(0.00190)
Observations	358159	358159	1352569	1352569	1243585	1350243	1350243	358159
R^2	0.502	0.516	0.526	0.429	0.401			
T	1.502	2.939	3.046	3.114	3.334	164.1	23.71	6.702

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1. *Are your results sensitive to the omission of zeros?*

   Yes. For log-linear regression, the results from (1)-(2) (where we drop zeros) are quite different from those from columns (3)-(5) (where we keep zeros in various ways). Moreover, Poisson models give different estimates compared with log-linear models, whether we keep zeros or not.

2. *How well does transforming the dependent variable to be* $\log(x+1)$, $\log(x+0.01)$, *or* $10^{-12}X_{jj}$ *if zero work? Is the result sensitive to the choice of transformation?*

   None of them works well, and the result is quite sensitive to the choice of transformation.

3. *Examine the residuals from your log-linear regression. Are they heteroskedastic? Report a Breusch–Pagan test statistic and a scatterplot of the residuals that addresses this question.*

   Yes, the errors are heteroskedastic. The Breusch–Pagan test statistic is 90992.82 (for the regression underlying column (1) in Table 2). See Figure 1 for a scatterplot (heatplot) of the residuals.

4. *How do the computation times compare?*

   Computing times are faster with less observations. Columns (1)-(5) use *reghdfe*, which is overall faster, with or without zeros. Poisson models are slower, partly because they involve maximizing likelihoods.

Table 3

1. *Verify that reghdfe, FixedEffectModels, and fixest return identical point estimates. Are the standard errors identical?*

   Yes, point estimates and standard errors are identical.

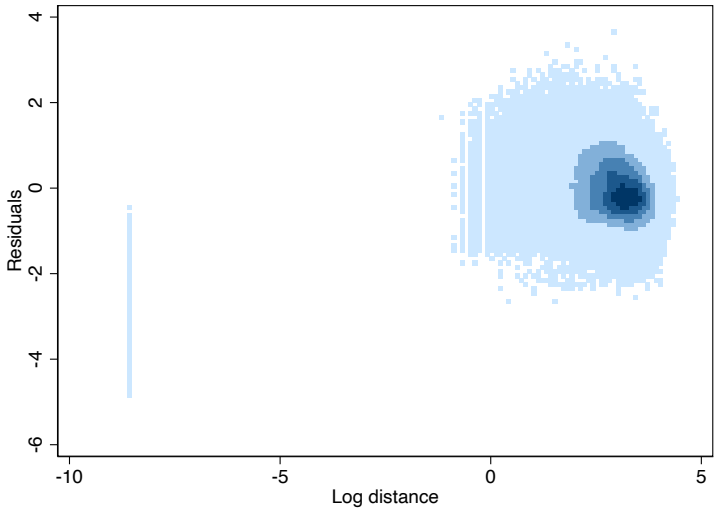

Figure 1: Residuals

Table 3: Comparing Stata's *reghdfe*, Julia's *FixedEffectModels*, R's *fixest*

	(1) Stata	(2) Julia	(3) R
log(driving distance)	-0.407***	-0.407***	-0.407***
	(0.003)	(0.003)	(0.003)
Observations	358159	358159	358160
R^2	0.502	0.502	0.502
T	1.054	0.729	0.447

Heteroskedasticity-robust standard-errors in parentheses
***: 0.01, **: 0.05, *: 0.1

2. *Which estimator is faster? By what magnitude?*

   R's *fixest* is faster than Julia's *FixedEffectModels*, which is faster than Stata's *reghdfe*, although differences in run times are within an order of magnitude.