Problem Set 2

IMPORTANT Solve the following exercises in a Stata do-file. The first line in the do-file should set your current directory (CD)—make sure I can run your do-file without error if I simply replace the CD with my own (save the pdmv_sl.dta in a sub-folder called data, see exercise I). Some functions necessary to solve these exercises may not be explained on the slides, use other resources (e.g., the STATA manual, the STATA help, and Google) to find a solution. Send your do-file to alexander.ahammer@jku.at. The email's subject line **must read**

PDMV PS2 [Last name]

where you replace [Last name] with the last name written exactly as it is in KUSSS. I cannot accept your do-file if you use a different subject line. Deadline is **Sunday, November 29, 2020,** midnight sharp.

Exercise I Go through the simple event study we discussed step-by-step in the course. Replicate the code on slide 44. Check the data browser or list a portion of the data after every step, and try to understand exactly what each command does. In the end, perform the same exercise, but calculate firing probabilities for the first 4 *quarters* after the end of the sick leave (instead of the first 5 months as on the slides).¹

Exercise II Use the data in pdmv_sl and solve the following problems:

- (a) Analyse the missing observations in p_educ. Do you observe a pattern?
- (b) You want to control *flexibly* for annual wage in a regression. Find an appropriate categorical representation of e_wage that allows you to do that. Regress sl_dur on the categorical variable and compare the resulting partial wage effect with a regression of sl_dur on wage in continuous form.
- (c) Write a summarize command that provides the average sick leave duration for every possible combination of patient gender and GP gender.² What do you find?
- (d) Generate a variable sl_avgdev as the difference between sl_dur and the average sick leave duration *per firm and year* the worker is in at the time of observation. How can you interpret this variable? Why would we be interested in it?
- (e) In a panel regression you want to control for the number of *previous* sick leaves a worker already had in the given year. Construct such a variable.

¹Try to code the quarterly event study from scratch. If that is too difficult or you get stuck, use the original code on the slides and replace the relevant commands.

²Note that such an exercise is more easily done with cross-tabulation commands such as tabstat. Here you should train how to work with by-groups.