EXCERSISES IN APPLIED PANEL DATA ANALYSIS #2

CHRISTOPHER F. PARMETER

1. Introduction

This exercise will help you to think about the underlying economic issues for your panel data models. Given that we will be paying attention to the correlation (or lack therof) between the covariates in the model and the unobserved heterogeneity, as well as the plausibility of strict exogeneity conditional on the unobserved effects we will not be using `R`. Once we have arrived at the appropriate understanding of our modeling assumptions, then we will use `R` to implement the appropriate estimator.

2. Fixed versus Random Effects?

Lets consider several prominent applied papers that investigate the validity of the random effects framework. Keep in mind that under the random effects framework for the unobserved effects model **all** of the covariates are uncorrelated with the individual specific effect.

2.1. Modelling Strike Activity.

Owusu-Gyapong (1986) studies a model of strike activity in Canada across 60 manufacturing industries over the period 1967-1979. His main concern is with differentiating appropriate techniques of estimating pooled regressions. One of the main issues in this approach is to decide amongst the fixed and random effects frameworks. As Owusu-Gyapong (1986, pg. 526) notes "Most studies in the past employing pooled data have relied on a priori reasons to assume a specification of the individual-specific effects ... Such an approach of deciding whether an effect is fixed or random is arbitrary and at best inadequate."

The variables in Owusu-Gyapong's (1986) database include the rate of change in nominal wages, union coverage, the total proportion of female workers, a concentration ratio for the firm, average age of the industry's labor force, export-to-sales ratio, inventory-to-sales ratio, contract expirations, the rate of change in the CPI, aggregate rate of unemployment and a dummy variable capturing wage/price controls in 1976-1978.

With these variables the key issue is if any of them are potentially correlated with the unobserved firm specific effect that influences strike duration. Owusu-Gyapong (1986, Tables 1 and 2) provides estimates from econometric estimators for both effects frameworks as well as their differences.

University of Miami
Date: September 9, 2013.
Christopher F. Parmeter, Department of Economics, University of Miami; e-mail: cparmeter@bus.miami.edu.
*These exercises have been prepared for the "Applied Panel Data Econometrics" course in Dakar, Senegal sponsored by IFPRI and AGRODEP..

However, even if the differences are small this does not provide guidance into the validity of the random effects framework. Based on individual tests of equality of these estimates, along with a formal test across the competing frameworks, Owusu-Gyapong (1986) concludes that the random effects framework is appropriate.

There are several interesting pedagogical issues from this paper. Owusu-Gyapong (1986) never provides an economic rationale for which framework might be appropriate, instead letting the data inform him. This can undermine solid scientific study as economic theory should at least guide a baseline hypothesis, rather than simply estimating the two different frameworks and testing amongst them. Further, my setting a baseline framework, statistical rejection of this framework may lead to greater insight into the nature of the behavior of the effects.

One reason the random effects framework may be found suitable in this setting is that Owusu-Gyapong's (1986) database is highly aggreagted, not being at the firm level, but the industry level. This aggregation may mitigate specific influences that firm specific effects have on the covariates. That is, at the firm level we may expect firm specific effects to be correlated with the determinants of strike activity, such as wages, gender composition, etc. but once we aggregate this data up to an industry level these connections disappear, becoming effectively random 'shocks' to average number of strikes.

2.2. **The Impact of Currency Unions on Trade.** Glick & Rose (2002) use a panel documenting international trade to determine if leaving a currency union reduces trade flows. Deploying an extensive tradebase, that covers bilateral trade pairs for 217 countries starting 1948, they have over 219,000 observations. The covariates that Glick & Rose (2002) use in their gravity model are the product of real GDP, the product of the population of the countries, the distance between the countries, an indicator if the countries share the same language, a count indicating the number of countries in the pair that are landlocked ($\{0, 1, 2\}$), an indicator if the countries share a border, if the countries belong to a free trade agreement, a count indicating the number of countries in the pair that are island nations ($\{0, 1, 2\}$), the product of the area of the countries, if the countries had the same colonizing country, an idicator if both countries are currently colonies, an indicator if the trading pair is colonizer-colony, a binary variable which is unity if both countries remained part of the same nation during the sample (e.g., France and Guadeloupe, or the UK and Bermuda) and the key variable of the study, an indicator if the trading pair are members of a currency union.

The bilateral nature of the data actually gives rise to a three way panel in Glick & Rose's (2002) anaylsis. That is they can control for country specific effects, as well as country pair specific effects. The key issue remains though, are these effects consistent with a random effects specification? Glick & Rose (2002, Table 4) provide estimates under both frameworks and argue that the main parameter of interest, that on the indicator for joint membership in a currency union is robust across the frameworks. While this is suggestive it overlooks the fact that which framework one uses impacts not just the estimator but the variance of that estimator as well. Glick & Rose (2002) provide heteroskedasticity robust standard errors as well to circumvent these issues. Again, as with the Owusu-Gyapong (1986) study, no discussion of the appropriate framework is given *a priori*, instead statistical tests are performed *ex poste* to determine the appropriate framework.

Glick & Rose (2002) also run a series of sensitivity checks to see if this basic result holds. Across a variety of data specifications (see their Table 5) their is little difference in the effect of a currency union on bilateral trade across the fixed and random effects frameworks, though statistical test results are not provided. Further, no discussion is provided why these alternative data specifications (such as dropping island nations, using five year averages for the variables, and eliminating country pairs with low bilateral trade) might have an effect on why one might prefer the random effects framework over the fixed effects framework. It is always good empirical practice to determine one or two 'variables' that are likely to be captured in the unobserved effects as a basis for determining the plausibility of the random effects framework.

2.3. **The Public vs. Private Capital Puzzle.** Munnell (1990) looks at a panel on gross state product across the 48 contiguous United States over the 1970-1986 period. The main variables in her database are the public capital stock of the state, the private capital stock as well as the state unemployment rate and nonagricultural labor input. The approach of Munnell (1990) was extended by Baltagi & Pinnoi (1995). Unlike the previous studies described earlier, Baltagi & Pinnoi (1995, pg. 353) make an attempt to discuss the types of variables which are likely to be contained in the unobserved individual (state) effects. Given the left hand side variable is gross state domestic product, Baltagi & Pinnoi (1995) list endowment of natural resources, the quality of public infrastructure and the physical characteristics of a state as potential variables in these effects.

Both Munnell (1990) and Baltagi & Pinnoi (1995) estimate a range of models. A key insight from Baltagi & Pinnoi (1995) regarding the appropriate framework for the unobserved effects is that first, when public capital enters the model, the appropriate framework switches from random effects to fixed effects, i.e. the aggregate public capital stock is correlated with the unobserved state effects. Second, if public capital enters into the model based on individual components of the public capital stock (roads and highways, water facilities and sewers, etc.) then the fixed effects framework is more roundly suggested. This stems from the fact that there are likely measurement errors in these individual components, that when aggregated are mitigated. Thus, a story similar to Owusu-Gyapong (1986) arises.

The work of Baltagi & Pinnoi (1995) is crucial for applied work because it gives insights into the nature of the impact of measurement error and its impact in panel data analysis. The distinction between the fixed and random effects framework can hinge critically on a given set of covariates being mismeasured. Thus, it is always good practice to think of the plausibility of measurement error in a given dataset as well as how this potential measurement error will impact the statstical framework one works with.

3. Conditionally Strict Exogeneity?

Consider the classic program evaluation panel data model:

$$y_{it} = x'_{it}\beta + \delta prog_{it} + c_i + \varepsilon_i t, \tag{1}$$

where *prog* denotes participation in a given program, such as a job training program. To help fix ideas consider that we only have data on two time periods. At time $t = 1$ no individual is in the program, whereas at time $t = 2$ some subpopulation of individuals is enrolled in the program. Thus, $prog_{i1} = 0\ \forall i$ while $prog_{i2} = 1$ for some individuals. Those individuals for which $prog_{i2} = 0$ are referred to as the control group and those for which $prog_{i2} = 1$ are referred to as the treatment group.

Program evaluation researchers beleive that c_i is present in the model because of unobserved factors of the individuals who are in the data. The concern facing the econometrician is if the decision to participate in the program is correlated with c_i, something known as the self-selection problem. In this case we can be reasonably confident that we should consider this model in a fixed effects framework. However, what about strict exogeneity?

We may feel confident that the program is such that $prog_{it}$ is uncorrelated with ε_{it}, but what about potential correlation between ε_{it} and $prog_{it+1}$? Individuals may chose to participate in the future based on current shocks to their outcome variable. Suppose y represents wages and an individual takes a pay cut in one year (conditional on x). If this pay shock then leads to the individual enrolling in the program the following year, then we no longer have strict exogeneity. Alternatively, if the program administrators select people to join the program based on ε, even after controlling for c_i, see Ham & LaLonde (1996). Alternatively, if program participation has effects beyond the current time period, then we may have to worry about a dynamic structure, including lags of program participation.

An alternative example is to consider the distributed lag model of Hausman, Hall & Griliches (1984). They study the impact that research and development budgets for firms has on patents award to the firm. A simplified version of their model is

$$patents_{it} = x'_{it}\beta + \delta_0 RD_{it} + \delta_1 RD_{it-1} + \cdots + \delta_5 RD_{it-5} + c_i + \varepsilon_{it}. \tag{2}$$

where RD is spending on research and development and *patents* is the number of patents awarded to the firm. We must decide if research and development spending is correlated with unobserved heterogeneity and if shocks to patents influnce future spending on research and development (in which case strict exogeneity is invalid).

Lets consider the following example

$$y_{it} = \beta_1 y_{it-1} + c_i + \varepsilon_{it}. \tag{3}$$

Strict exogeneity conditional on unobserved heterogeneity cannot hold because $E[x_{it+1}\varepsilon_{it}] \neq 0$. In the unobserved effects model with lagged dependent variables this condition is necessarily violated. Thus, particular care must be paid when you have dynamics in your model.

Now lets try for yourself. The model you are thinking of estimating is a model for capital investment in the manufacturing industry and you have access to data at the province/state level over T time periods:

$$invest_{it} = x_{it}'\beta + \delta_1 tax_{it} + \delta_2 disaster_{it} + c_i + \varepsilon_{it}. \tag{4}$$

tax_{it} measures the marginal tax rate on capital facing the industry, $disaster_{it}$ is a dummy variable that identifies if a major flood hit the province/state in year t while $invest_{it}$ is the amount fo investment in capital for the industry. x_{it} captures other variables relevant to capital investments. Discuss whether strict exogeneity conditional on unobserved heterogeneity is reasonable for tax and $disaster$. Assume that neither of these variables have lagged effects on capital investment.

References

Baltagi, B. H. & Pinnoi, N. (1995), 'Public capital stock and state productivity growth: Further evidence from an error components model', *Empirical Economics* **20**, 351–359.

Glick, R. & Rose, A. K. (2002), 'Does a currency union affect trade? The time-series evidence', *European Economic Review* **46**, 1125–1151.

Ham, J. C. & LaLonde, R. J. (1996), 'The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training', *Econometrica* **64**, 175–205.

Hausman, J. A., Hall, B. & Griliches, Z. (1984), 'Econometric models for count data with an application to the patents-r&d relationship', *Econometrica* **52**, 909–938.

Munnell, A. H. (1990), 'How does public infrastructure affect regional economic performance?', *New England Economic Review* **September**, 11–32.

Owusu-Gyapong, A. (1986), 'Alternative estimating techniques for panel data on strike activity', *The Review of Economics and Statistics* **68**, 526–531.