The Training Benefits Program – Methods for Panel Data and the P-20 Database

To: Pacific Northwest Regional Economic Conference
By: Jonathan Adam Lind
Date: 05/12/16

Employment Security Department
WASHINGTON STATE

Agenda

- 1) A brief overview of the Training Benefits (TB) program.
- 2) A quick glance at the raw data assessing the effects of TB program participation on earnings.
- 3) The nature of these data – a quick introduction to panel data.
- 4) A method for working with panel data that allows us to better leverage the P-20 data.

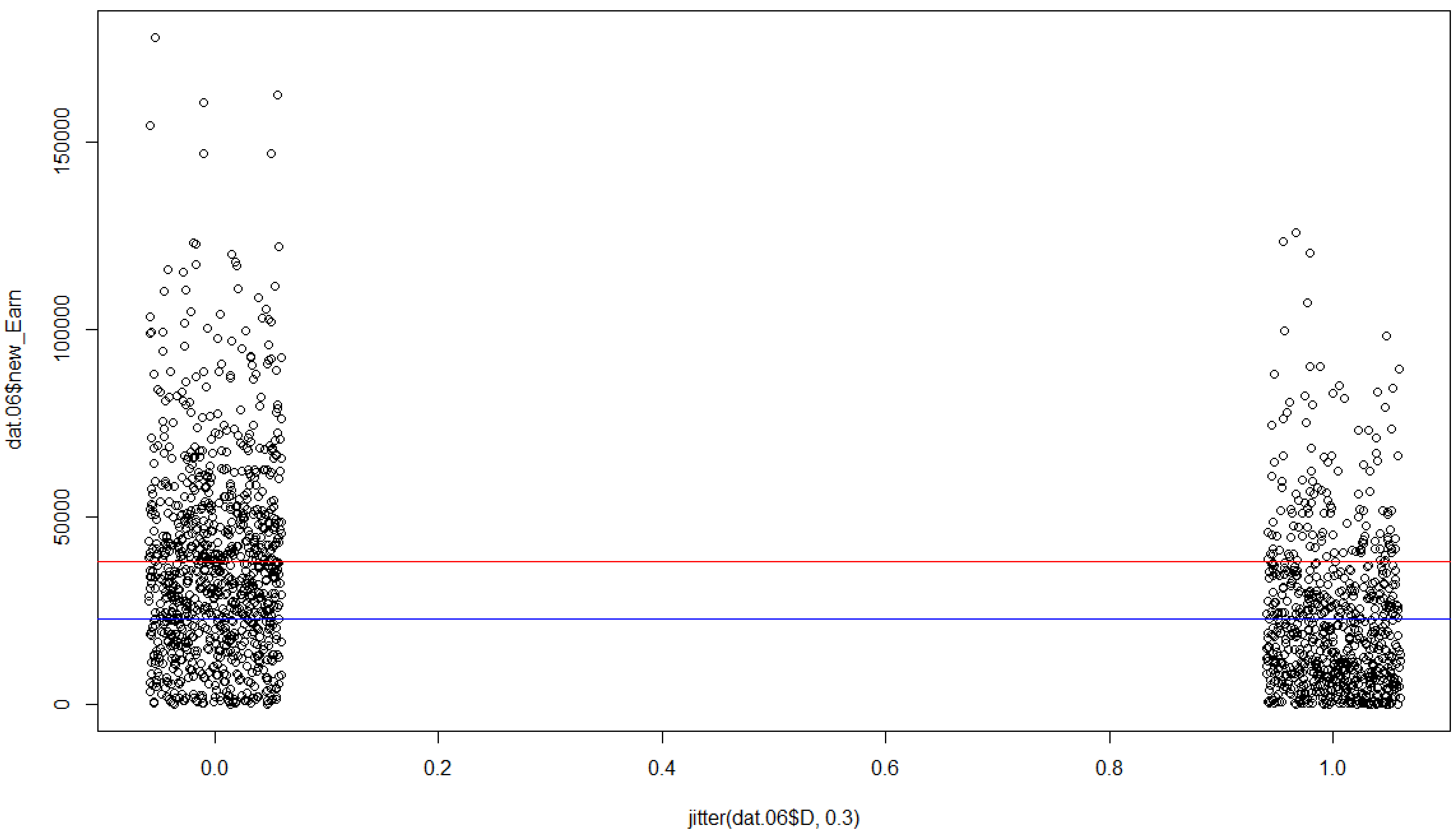
Employment Security Department
WASHINGTON STATE

Training Benefits (TB) Program Overview

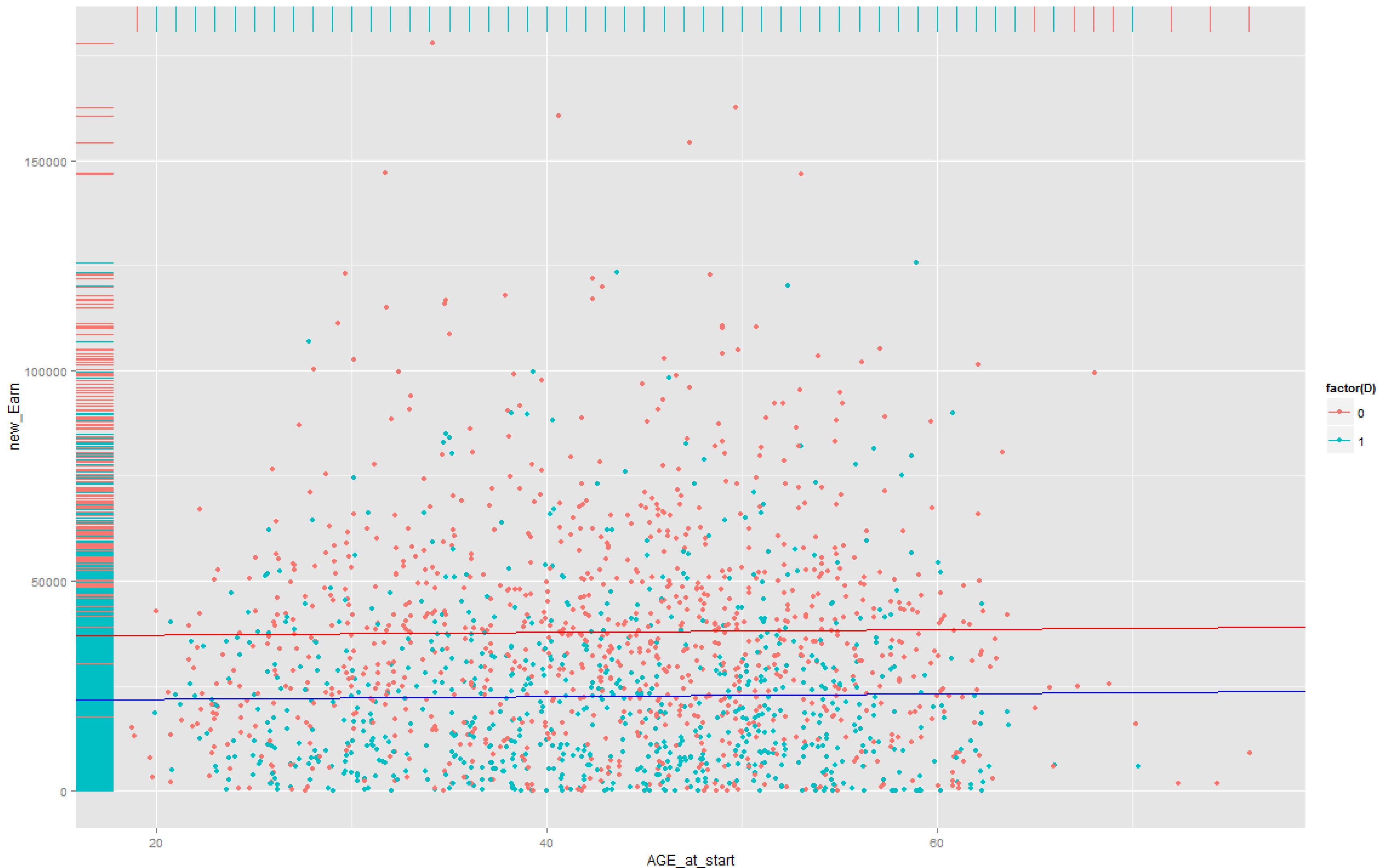
- The Training Benefits (TB) Program was established by the Washington State Legislature in 2000.
- The goal of the program is to retrain unemployed individuals who qualify for unemployment benefits but whose skills are no longer in demand.
- Approximately 1,650 individuals were approved for the TB program in fiscal year 2015.
- TB participants are drawn from a broad spectrum of the overall UI population, but tend to be more female, more urban, and slightly more educated than the general UI population.

The Raw Data – TB Participation and Earnings

A Bivariate OLS Regression of Earnings on TB Participation



A Multivariate OLS Regression of Earnings on TB Participation and Age



Omitted Variable Bias – The Achilles Heel of Methods Such as OLS

- The classical linear regression model (OLS) assumes that the error term (u) is 1) unrelated to the dependent variable and 2) uncorrelated with any of the explanatory variables.

$$Y = \beta_0 + \beta_1 X + u$$

- This is, of course, a rather restrictive assumption in that it requires us to include *everything* in our model that is correlated with *any* dependent or explanatory variable.
- This is fine if we have measures of all the factors that may be correlated with both earnings and all other explanatory variables and there are indeed many things we can (and do!) measure that fit this criteria.
  - For example, we have measures of age, gender, ethnicity, occupation, STEM credits, Health Credits… many others.

- But… we can't measure everything. Indeed, some things that may be related to earnings that we simply *cannot* measure.
  - For example, it would be difficult to measure things like motivation, or "people skills" – yet we still might expect these things to be associated with earnings.
  - These sorts of unmeasurable concepts are often referred to as "unobserved heterogeneity".

Employment Security Department
WASHINGTON STATE

One way to mitigate the effects of omitted variable bias - Panel Data

- One distinguishing feature of our data presented is that it “follows” individuals as they progress through time.

- Such a structure, of course, allows us to describe change over time – but panel data allows us to do much more than simple description.

- One of the key benefits of panel data is that it allows us to use methods that can help to mitigate the effects of omitted variable bias.

- Panel data allows us to use methods that can better account for unobserved heterogeneity – as long as the unobserved factors correlated with X do not change, at the individual level, over time.

The Fixed Effects Model (AKA the "within" estimator)

- The within transformation can control for unobserved, time-invariant factors via a transformation that time-demeans the data.

- Part of this transformation involves a composite error term.

$$e = a_i + u_{it}$$

- Where a_i represents all *time-invariant* unobserved factors, and u_{it} represents stochastic error.

The Fixed Effects Model (AKA the "within" estimator)

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it} \qquad (1)$$

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \ldots + \beta_k \bar{x}_{ik} + a_i + \bar{u}_i \qquad (2)$$

$$(y_{it} - \bar{y}_i) = \beta_1 (x_{it1} - \bar{x}_i) + \ldots + \beta_k (x_{itk} - \bar{x}_{ik}) + (u_{it} - \bar{u}_i) \qquad (3)$$

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it1} + \ldots + \beta_k \ddot{x}_{itk} + \ddot{u}_{it} \qquad (4)$$

Again, the key idea is that any change in (y) cannot be caused by (a) because (a) does not change between time periods (t), (even though it may be correlated with the x's).

How the within transformation can help us shed new light on differential effects of the TB Program.

- We interacted program participation on a number of time-invariant measures such as gender, low-income status, race/ethnicity, and U.S. veteran status to assess whether the effects of TB participation vary by sub-group.

- We found evidence suggesting certain sub-groups (e.g. low-income persons, persons pursuing a health-intensive course of study) disproportionately benefited from TB participation…
- … and other sub-groups (e.g. persons entering the program during the Great Recession) did not benefit as much from TB participation.

Employment Security Department WASHINGTON STATE

In summary

- Simpler modelling strategies (such as OLS) can help us better understand some research questions, but they suffer from the problem of omitted variable bias.
  - How can you control for things you cannot even observe!
- Panel data, or data that tracks people over time, can be exploited to get around the limitation.
  - If something about an individual does not change over-time, then it cannot be responsible for over-time changes in individual-level outcomes… even if that something cannot be observed by the researcher!
- The “within” transformation allows us to control for unobserved heterogeneity as well as assess the effect of treatment on different sub-sets of the sample.
- The P-20 data can be used in conjunction with the “within” method to generate estimates of outcomes for different groups of individuals who pursue different programs of study.