The Importance of Zeros: Estimating Treatment Effects when Treatment Causes Zero Earnings

With an Application to the Job Displacement Literature*

Patrick Bennett†

March 2021

Abstract

This paper explains how the widely used log-earnings regression estimates biased treatment effects when the treatment directly affects the probability of having zero earnings. This bias can be large: the log-earnings regression understates long-run earnings losses due to job displacement by 15 percentage points—a factor of 5—compared to an alternative log-like transformation which includes zeros. The bias is so large because displaced workers whose earnings decline the most are excluded from the sample. The findings clarify when earnings regressions in levels will recover the wrong treatment effect assuming a functional form of constant elasticity, and offers practical considerations.

JEL CODES: C1, J63, J65

KEYWORDS: hyperbolic sine, log-transformation, job displacement, treatment effects

1 Introduction

What is the effect of a particular treatment on earnings? Understanding the estimated treatment effect of different policy interventions, labor market shocks, job training programs, and other active labor market policies on earnings are some of the core questions in Economics. Seminal papers such as Ashenfelter (1978), LaLonde (1986), and Heckman et al. (1997) establish causal impacts of job training programs on earnings while Jacobson et al. (1993) estimate the persistence of earnings

*This work was partially supported by the Research Council of Norway through its Centres of Excellence Scheme, FAIR project No. 262675.

†Centre for Applied Research (SNF), Norwegian School of Economics (NHH) and FAIR. Email: Patrick.Bennett@nhh.no

losses after job loss.[1] At the foundation of such an earnings regression is Mincer's influential 1974 earnings regression, a "cornerstone of empirical economics" (Heckman et al., 2003). Such regressions in the program evaluation literature overcome endogeneity concerns and isolate the causal impact of a particular treatment by comparing outcomes across groups with different exposures to treatment.

Taking the log of earnings is a frequently used transformation when estimating earnings regressions. However, as is widely known, the constant elasticity interpretation as given by the log-transformed regression has a serious limitation: the logarithm of zero earnings is undefined.[2] Thus, those with zero earnings are excluded from the estimation sample in the log-transformed regression. This presents a particular problem when the treatment being evaluated directly impacts the probability of having zero earnings. For instance, job training and other active labor market programs may increase labor force attachment, decreasing the probability of having zero earnings. At the same time, negative labor market shocks induce some to exit the labor force entirely, increasing the probability of zero earnings.

This article assesses the importance of excluding those with zero earnings when estimating treatment effects in the log-transformed earnings regression. The paper shows that the estimated treatment effect in the log-earnings regression is biased when the treatment itself directly affects the probability of having zero earnings. By excluding those with zero earnings, this bias can drastically affect the magnitude of the estimated treatment effect compared to alternative log-like transformations which include those with zero earnings. For instance, when treatment increases the probability of having zero earnings, the log-transformation excludes treated individuals whose earnings change the most.

The paper makes important contributions to two separate literatures. First, it establishes the importance of including those with zero earnings when estimating earnings regressions. The paper clarifies how the log-transformation will produce treatment effects of the wrong magnitude when the treatment itself affects the probability of having zero earnings. In addition, the paper explains the conditions when log-like-transformations are preferred to both the log-transformation and the same regression in levels. By providing detailed recommendations as well as guidance from a practical example and a simulation exercise, the paper contributes to a recent emerging literature examining the importance of log-transformation in estimated treatment effects (Silva and Tenreyro, 2006; Ravallion, 2017; Bellemare and Wichman, 2020).

Second, the paper reveals that the extensive literature on job displacement understates the true

[1] See Imbens and Wooldridge (2009) for an overview of the diverse program evaluation literature. Recent contributions to the job training literature include Flores et al. (2012) and Calónico and Smith (2017). Recent contributions estimating the earnings losses due to job displacement include Couch and Placzek (2010) and Lachowska et al. (2020).

[2] In addition to changing the interpretation of the estimated treatment effect relative to a non-transformed measure of earnings, the log-transformation has important considerations such as limiting the influence of outliers, making earnings more closely approximate a normal distribution, and heteroskedasticity considerations.

impacts of job loss on earnings.[3] The paper contrasts how three different measures of earnings affect the estimated magnitude of earnings losses after job loss: (a) log-transformed earnings, undefined at zero, (b) inverse hyperbolic sine (arcsinh)-transformed earnings, a transformation which is defined for zero earnings, and (c) un-transformed earnings measured in levels.[4] The paper's findings reveal that accounting for those who have zero earnings post-displacement matters greatly: such displaced workers experience earnings losses of 100%, yet are excluded from the log-earnings regression. In addition, the paper shows how the standard displacement earnings regression estimated in levels is unlikely to produce the correct treatment effect.

The paper has three main findings. First, it demonstrates how estimating the frequently used log-transformation can produce biased treatment effects by excluding those with zero earnings. Using linked employer-employee for all workers in Norway, the paper estimates earnings losses after job loss by comparing high-tenured displaced workers to similar non-displaced workers. Compared to the estimated treatment effect in the arcsinh-transformed regression, the log-transformation understates long-run earnings losses by 15 percentage points, a decline which is 5 times smaller in magnitude. Such a discrepancy drastically changes how severe the earnings penalty is: while earnings losses recover over time in the log-transformed regression, the same cannot be of the arcsinh-transformed regression. The discrepancy between the two regressions arises *solely* because displaced workers are significantly more likely to have zero earnings post-displacement. Indeed, displacement causes labor force exit (Huttunen et al., 2011) and displaced workers may be more susceptible to future negative earnings shocks (Stevens, 1997). This discrepancy vanishes when the same samples are used across the two regressions. Earnings losses in the aftermath of job loss are identical between the log-transformed and arcsinh-transformed regressions when dropping those with zero earnings post-displacement. In addition, the arcsinh-transformed regression produces similar results to commonly used ad hoc solutions such as the log+1-transformation.

Second, the paper clarifies the conditions when the arcsinh-transformation represents the preferred specification relative to not only the log-transformation but also the un-transformed regression in levels. Indeed, it is common to present results from earnings regressions in levels as well as logs. The paper performs a simulation exercise to show that when the true data generating process is a treatment effect of constant elasticity, the earnings regression in levels is biased in the case of heterogeneous treatment effects. Intuitively, the resulting bias arises because the treatment effect calculated as a fraction of sample average earnings and the average treatment effect are not the same in the presence

[3]Many papers estimate impact of job loss during mass-layoff event on earnings (Jacobson et al., 1993; Couch and Placzek, 2010; Huttunen et al., 2011; Davis and Von Wachter, 2011; Farber, 2015; Raposo et al., 2019; Lachowska et al., 2020).

[4]The arcsinh-transformation of earnings is $ln(earnings + \sqrt{earnings^2 + 1})$.

of heterogeneous treatment effects. The long-run earnings losses due to job loss are also drastically different between the levels and arcsinh-transformed regressions, despite the fact that both regressions include those with zero earnings. While the correct functional form should be guided by economic theory, the paper shows how the impacts of job loss in levels depend on pre-displacement earnings, suggesting a functional form of constant elasticity.

Third, the paper examines how the bias due to excluding those with zero earnings in the log-transformed regression impacts the estimated treatment effects across (i) different samples and (ii) earnings measures with different probabilities of zero earnings post-displacement. Older displaced workers, those aged 45–50, are considerably more likely to have zero earnings relative to younger displaced workers aged 25–34. At the same time, the bias in the long-run earnings losses in the log-transformed regression is even larger among older displaced workers—understating the true decline by 24 percentage points—compared to younger displaced workers—understating the true decline by 13 percentage points. Similar patterns are seen among non-college educated workers, who are considerably more likely to have zero earnings post-displacement relative to college educated workers. Such differences in the probability of having zero earnings across different groups can alter the relative interpretation of how earnings recover in the aftermath of job loss.

In addition, by including benefits in the measure of earnings, the paper shows how excluding those with zero earnings also matters in the log-income regression. Indeed, the probability of having zero income post-displacement is considerably smaller than the probability of having zero earnings, as benefits replace some earnings losses in the aftermath of job loss. Despite this, it remains quantitatively important to account for those with zero income: long-run income losses remain 4 times smaller in the log-earnings regression compared to the arcsinh-transformation.

The paper proceeds as follows. Section 2 describes the data, defines displaced workers and the estimation sample, and compares the log-transformed earnings regression with the arcsinh-transformed earnings regression. Section 3 compares earnings regressions in levels with the arcsinh-transformation, and performs a simulation exercise to further understand the discrepancy between the two regression results. Section 4 examines how much differences in the probability of having zero earnings as a result of being laid off across samples and different earnings measures. Finally, Section 5 concludes and offers practical advice.

2 The Estimated Impact of Job Displacement on Earnings: Log vs. Arcsinh Transformation

The section reveals the importance of zeros when estimating earnings regressions, comparing the log-transformed regression to alternative measurements of earnings. Prior to this, the section describes the data, defines job displacement, and discusses the sample restrictions imposed throughout the paper.

2.1 Norwegian Register Data

The paper makes use of linked employer-employee Norwegian Register Data to assess how the impacts of job loss during a mass-layoff or plant closing event impact earnings. The paper links data across four different sources. Data is linked across the different sources by an anonymous personal identification number. First, the central population register provides data on demographic characteristics such as birth year, gender, and municipality of residence. Data is available at the annual level, and the population register covers all individuals who are resident in Norway.

Second, data on education is extracted from the education register. All educational qualifications which students are enrolled in and complete are reported by schools to Statistics Norway, and schools have a legal mandate to report such information on any student. Throughout the paper, education is classified by whether or not an individual has graduated college, a bachelors degree or above.

Third, labor earnings are measured in the tax register. Earnings are measured as annual wage earnings in Norwegian Kroner (NOK), and detailed earnings data is available from 1993. As an alternative, Section 4.3 replicates the paper's main results measuring income, labor earnings plus any unemployment insurance, social assistance, or disability insurance benefits.[5] Unemployment insurance replaces approximately 70% of earnings for all eligible workers and lasts for 36 months (OECD, 2006).[6] In contrast, the replacement rate for disability insurance depends on the level of workers' earnings, where workers with lower earnings receive a higher fraction of earnings on disability insurance (for a further discussion of the disability insurance system, see Kostol and Mogstad, 2014).

Finally, the paper makes use of linked employer-employee data from the employer and employee register, available from 1986. Crucially, this provides a linkage between workers and their employing plant and firm, which is measured at the end of November. Throughout the paper, the focus is at the plant-level. In addition, the data provides information on hours employed per week, and full-

[5]Those on disability insurance in Norway have capacity to work and return to work following a shift in the incentives to do so (Kostol and Mogstad, 2014). For a discussion of the importance of benefits in post-displacement, see East and Simon (2020).

[6]Given the sample restrictions discussed in Section 2.3, all workers in the sample are eligible for unemployment insurance benefits.

time employment is defined as employment of more than 30 hours per week. Importantly, when combining data across these sources, individuals are observed in each year irrespective of whether they are employed or not.

2.2 Involuntary Job Loss: Defining Job Displacement

To estimate the impacts of job loss on earnings, the job displacement literature isolates involuntary job loss during a mass-layoff or plant closing event among high-tenured employees. Following standard definitions from the job displacement literature, displacement is defined as (1) an employee who transitions into unemployment or to a new job who is (2) employed in a plant which experiences either: (a) a mass-layoff event or (b) a plant closing event. Mass-layoff events are defined as a downsizing in plant employment of 30% or more, as is standard in the literature (Jacobson et al., 1993). Importantly, plant closings exclude false closings, classified as when 80% of the same coworkers all move to the same new plant.

In the regressions that follow, outcomes of displaced workers are compared to those of non-displaced workers. The restrictions described below which isolate a sample of workers in stable employment are all implemented pre-displacement. As such, both displaced and non-displaced workers are followed unconditionally over time. Previous work in the job displacement literature emphasizes the importance of comparing displaced workers to non-displaced workers who are followed irrespective of their employment status (Krolikowski, 2018). Indeed, non-displaced workers may be subject to a displacement event in a future year, though this variation is not exploited in the regressions below.

2.3 Isolating the Impact of Displacement Among High-Tenured Workers

Both future displaced and non-displaced workers are high-tenured workers for whom job loss is involuntary. The sample restrictions to isolate a sample of high-tenured workers are imposed relative to a base year b. Workers are all employed in b by construction, but may transition into non-employment in all future years, $b + 1$ and on. The estimation sample compares base years 1997–2002, a time when the Norwegian labor market had recovered from a financial crisis affecting all the Nordic countries at the start of the 1990s (see Aaberge et al., 2000, for further details). Thus, the first cohort of workers is displaced between 1997 and 1998, while the final cohort of workers is displaced between 2002 and 2003. As mentioned, non-displaced and displaced workers are subject to the exact same sample restrictions.

High-tenured workers are defined using the following the four sample restrictions. First, workers must be high-tenured, defined as a minimum of 1 year of tenure in their employing plant in year b. Second, workers must be strongly attached to the labor force: employed full-time and receive no employment benefits between $b-3$ and b. In addition, workers must have a minimum level of earnings in all years from $b-3$ to b greater than the social security base rate.[7] Third, workers must be employed in a sufficiently large plant, defined as a plant with at least 10 workers in year b. Finally, workers must be between the ages of 25–50 (inclusive) in year b.

Separate samples for each base year b are created, where the separate samples are then stacked relative to the displacement event. Event time t is calculated relative to b. In the event study regressions that follow, $t = year - b$. Workers are followed both prior to and after displacement, from -3 to $+7$, and the sample is subject to displacement between time 0 and $+1$.

2.4 The Estimated Impact of Job Displacement on Earnings

As a starting point, Figure 1a plots the unconditional probability of having zero earnings relative to the displacement event for displaced and non-displaced workers. Prior to job displacement, future displaced and non-displaced workers are all employed by construction and 100% of both samples have positive earnings. After being laid off between 0 and 1 year after displacement, those displaced are more likely to have zero earnings compared to non-displaced workers.[8] This gap increases in the first few years after displacement and 3 years after displacement, 2.6% of displaced workers have zero earnings compared to 1.4% of non-displaced workers. This gap stabilizes over time, and displaced workers are between 1.2–1.3 percentage points more likely to have zero earnings.

While the probability of having zero earnings increases as a result of being laid off, it is equally important to understand where this difference in zero earnings comes from. Are displaced workers more likely to exit the labor force, and have multiple years with zero earnings, or do they experience one or two years of no earnings while searching for a new job? Figure 1b reveals that the majority of displaced workers experience a few years of zero earnings. The impact of displacement on the number of years with zero earnings, as given by the difference between the two bars, is highest for one and two years with zero earnings. However, there are also some workers for whom displacement causes many years with zero earnings. This suggests that displacement also causes labor force exit for a small, but potentially important, fraction of workers.

Clearly, job displacement impacts the probability of having zero earnings. As a result, the log-transformation of earnings will differentially exclude the two groups of workers. Indeed, displaced workers are significantly more likely to be excluded from the estimation sample. This presents a particular problem: displacement itself induces maximum earnings losses among those who have zero earnings, but they are not contributing to the estimated impact of displacement on (log) earnings.

[7]In 1997, this corresponded to 42,500NOK, while in 2002 this was 54,170NOK.

[8]Such differences are statistically significant from +1 (see Figure 7).

Figure 1: The Impact of Job Displacement on The Probability of Having Zero Earnings

(a) The Probability of Zero Earnings Over Time (b) The Number of Years with Zero Earnings

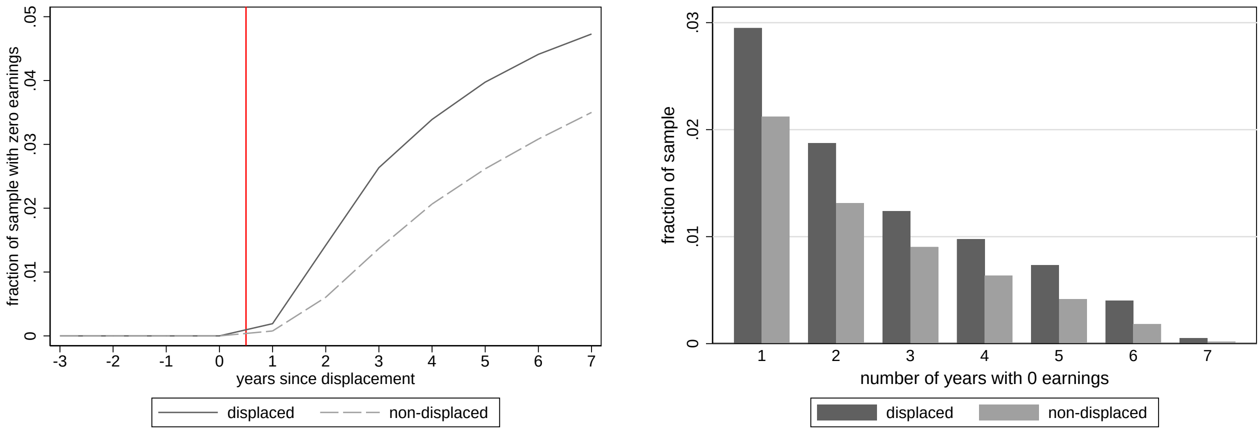


Figure plots the unconditional probability of having zero earnings pre- and post-displacement for displaced workers and non-displaced workers. Panel (a) plots the difference in the probability of zero earnings over time between displaced and non-displaced workers. Panel (b) plots the cumulative number of post-displacement years that workers have zero earnings, separately by displaced and non-displaced workers.

The paper estimates the following event study regression to assess the importance of excluding those with zero earnings from the estimation sample:

$$y_{ibt} = \alpha + \sum_{k=-3}^{+7} \delta_k (D_{ib} \times \tau_t)^k + \tau_t + \gamma D_{ib} + u_{ibt} \tag{1}$$

where $(D_{ib} \times \tau_t)^k = 1$ k years after a worker i in base year b is displaced and 0 otherwise. Displaced workers, for whom $D_{ib} = 1$, are displaced between +0 and +1. δ_k allows the estimated effect of displacement on earnings to vary over time, and corresponds to the average difference in earnings between displaced and non-displaced workers k years after displacement. When $k \leq 0$, δ_k corresponds to the difference in earnings pre-displacement. δ_{-1} is omitted by convention, and δ_k is interpreted relative to this fixed omitted difference between displaced and non-displaced workers. For $k > 0$, δ_k corresponds to the average earnings difference post-displacement in a given time t. τ_t corresponds to event time fixed effects.

The outcome variable y_{ibt} corresponds to one of four measures of earnings: (i) $log(earnings)$, (ii) $arcsinh(earnings)$, (iii) $log(earnings + 1)$, and (iv) $earnings$ measured in monetary units (NOK). While earnings at zero are undefined for (i), all other measures include those with zero earnings in the estimation sample. Throughout the paper, earnings losses are measured as the percent of pre-displacement earnings lost due to displacement. Results estimated in levels are scaled to percentage declines, using the sample average of earnings in -1, while the exact percentage decline is used for the log- and arcsinh-transformations.

Figure 2: The Estimated Impact of Displacement on Earnings Across Different Transformations of Earnings

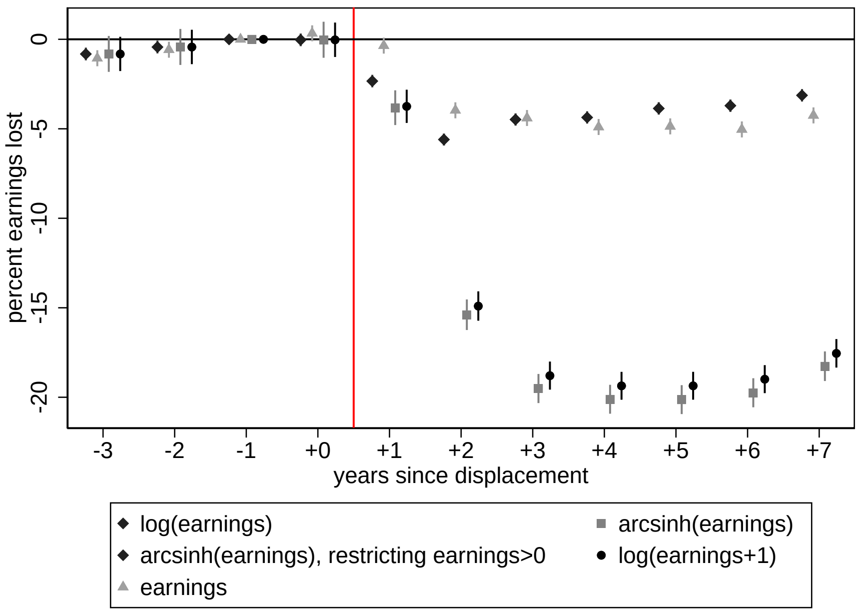


Figure plots δ_k coefficients, the event study estimates in equation 1, for one of five regressions: log(earnings), arcsinh(earnings), arcsinh(earnings) conditioning on positive earnings, log(earnings+1), and earnings measured in levels. The number of observations in the estimation samples of the five regressions are 36,887,063, 37,359,156, 36,887,063, 37,359,156, and 37,359,156 respectively. Estimates are plotted in percent of earnings lost. Percent earnings lost for earnings measured in levels calculated as the fraction of average pre-displacement earnings in -1. Estimated δ_k coefficients interpreted relative to fixed difference in -1, where δ_{-1} is omitted by convention. Displacement event occurs between $+0$ and $+1$. 95% confidence intervals reported, though not always visible due to symbol.

Figure 2 graphs the estimated impact of displacement pre- and post-displacement across the four different earnings measures. Across all regressions, displacement has statistically significant and long-run impacts on earnings, albeit with far different magnitudes of such declines. When log-transforming earnings—and excluding those with zero earnings by construction—displacement decreases earnings by 3–6%.

When transforming earnings by the inverse hyperbolic sine—and including those with zero earnings—the estimated earnings losses are far greater: displacement decreases earnings by 15–20%. Starkly different patterns also emerge over time. The estimated earnings losses between the log- and arcsinh-transformation are similar immediately after job loss in +1, when the difference in the probability of zero earnings between displaced and non-displaced workers is most similar (Figure 1a). Already by +2, there exist large differences in the estimated earnings losses between the two different regressions: 6% with the log-transformation and 15% with the arcsinh-transformation. From 3 years after displacement, the earnings losses estimated in the two regressions continue to diverge. While earnings begin to recover in the log-transformed regression from +2, ultimately reaching a 3% decline in +7, earnings continue to decline in the arcsinh-transformed regression, reaching an 18% decline by +7.

While earnings recover in the log-transformed regression, earnings remain persistently low and exhibit far less recovery in the arcsinh-tranformed regression. Indeed, the log-transformed regression understates earnings losses by 15 percentage points in the long-run. Though the probability of zero earnings increases by just more than 1 percentage point post-displacement, these displaced workers have a large impact on the estimated earnings losses. Including such workers is so important to the estimated earnings losses because earnings decline by 100% in a given year for these workers.

With the same interpretation of the two transformations (see Bellemare and Wichman, 2020), the only difference between the two regressions is the inclusion of those with zero earnings in a given post-displacement year in the estimation sample. Consistent with this, the estimated earnings losses are identical when regressing the arcsinh-transformed earnings on displacement indicators conditioning on a sample of workers who have positive earnings in a given post-displacement year.[9] At the same time, including those with zero earnings in the log-transformed regression, by using the ad hoc solution of taking the log of $earnings + 1$, produces similar results to the arcsinh-transformed regression.

[9]To be clear, this regression conditions on post-event earnings outcomes, and estimates the impact of job loss on earnings excluding those who have zero earnings as a result of displacement.

3 The Estimated Impact of Job Displacement on Earnings: Levels vs. Arcsinh Transformation

Clearly, those with zero earnings post-displacement matter for the estimated earnings losses due to job loss. A remaining question is how using un-transformed earnings, and estimating the same regression in levels, presents a possible solution to the exclusion of zeros in the log-transformed regression. Indeed, previous evidence from the job displacement literature emphasizes the importance of measuring earnings in both levels and log (Lachowska et al., 2020). Such decisions of functional form should be guided by economic theory: if job loss causes earnings losses which are constant in *monetary units*, the levels regression is correctly specified. If, however, job loss causes earnings losses which are constant in *percentage terms*, then the levels regression is misspecified.

Do workers with different levels of pre-displacement earnings suffer the same earnings losses in monetary terms—a constant monetary unit effect—or in percentage terms of pre-displacement earnings—a constant elasticity effect? If wages decline by 20% for all displaced workers, or indeed if time spent in employment declines by 20% while wages remain unchanged, then two workers with different pre-displacement earnings levels experience constant earnings losses in percentage terms, but different earnings losses in monetary units.

Table 1 demonstrates that those with higher pre-displacement earnings have larger long-run earnings losses in monetary units after job loss. Panel A presents the long-run impact of displacement in levels 7 years after job loss, estimated separately by pre-displacement earnings quartiles. The long-run earnings losses in monetary units depend on pre-displacement earnings: those in higher quartiles have higher earnings losses. Those in the bottom quartile of the pre-displacement earnings distribution have the lowest long-run earnings losses in levels, 5,614NOK. The second quartile has higher earnings losses, while the third quartile has even higher earnings losses. Finally, the top quartile has the largest earnings losses in levels, 25,506NOK. These differences are substantial: the long-run earnings losses due to displacement in the highest quartile are 4.5 times as large as earnings losses in the lowest quartile.

While decisions of functional form should be guided by economic theory, Table 1 suggests that displaced workers experience a constant percentage decline in earnings rather than a constant monetary unit decline. This is confirmed in both panel B, which presents results of the arcsinh-transformed regression by the same earnings quartiles, and in panel A, which also reports the percent earnings losses, dividing the long-run earnings losses in levels by the sample average of pre-displacement earnings. Indeed, the four groups have relatively similar earnings losses when measured as a percentage of

Table 1: The Long-Run Impact of Displacement on Earnings by Pre-Displacement Earnings Quartiles

	(1) Bottom 25%	(2) p25–p50	(3) p50–p75	(4) Top 25%
Panel A: Earnings				
Long-run earnings losses (+7)	-5614***	-10340***	-11860***	-25506***
	(549)	(535)	(652)	(2053)
Observations	9298025	9326065	9323970	9411261
Percent Earnings Losses	3.0	4.1	3.9	5.3
Panel B: Arcsinh-transformed Earnings				
Long-run earnings losses (+7)	-0.182***	-0.184***	-0.163***	-0.190***
	(0.012)	(0.010)	(0.009)	(0.009)
Observations	9298025	9326065	9323970	9411261

Table presents the long-run impact of job displacement on earnings, 7 years after displacement. While the Table reports only δ_7 from equation 1, the estimated regression includes all pre- and post-displacement indicators. Each column corresponds to separate regression on sample of different quartiles of the pre-displacement earnings distribution from the bottom 25% (column 1) to the top 25% (column 4). Panel A corresponds to regressing earnings on displacement indicators, panel B corresponds to regressing arcsinh-transformed earnings on displacement indicators. $^{***}\rho < 0.01$, $^{**}\rho < 0.05$, $^{*}\rho < 0.10$.

earnings. This is true despite the fact that within each earnings quartile, there exist workers of many different ages, education levels, gender, and industries of employment. Such differences likely also affect the magnitude of earnings losses post-displacement.

However, as seen in Figure 2, even when scaled in percentage terms, the estimated earnings losses from the levels and arcsinh-transformed regressions differ greatly. This is despite the fact that both regressions include those with zero earnings post-displacement, and the estimation samples are exactly the same. If the correct functional form is one of constant elasticity rather than a constant monetary unit effect, why does the estimated treatment effect of the two earnings regressions differ so greatly?

3.1 Reconciling the earnings regressions in levels and constant elasticity effects: a simulation exercise

To understand this point further, Figures 3 and 4 present results from a simple simulation exercise. The simulation compares the two different estimates of earnings losses: true earnings losses, which are by construction a constant percentage of pre-displacement earnings, and earnings losses assuming a constant monetary unit decline and scaling the decline by the pre-displacement sample average.

In the exercise, there exist two types of workers: high-type and low-type workers ($type = h, l$), where there are an equal amount of both worker types. High-type workers earn an average of 200,000 monetary units with a standard deviation of 20,000 while low-type workers earn an average of 100,000 prior to displacement with a standard deviation of 10,000. Displacement is constructed as a shock in which 10% of all workers are subject to a random chance of being laid off between time +0 and +1, where ρ^{type} is the probability of displacement for each worker type. Consistent with the literature on job displacement, the probability of displacement may vary by worker type: for instance, Farber (2015) reveals that low-educated workers are much more likely to be laid off relative to high-educated workers.

Earnings losses, and subsequent earnings recovery, are determined by the following data generating process. Throughout all different scenarios considered, displacement reduces earnings by 20% immediately after displacement for all type of workers. Following the initial negative shock, earnings recover as time goes on by a constant fraction of pre-displacement earnings, ψ^{type}, in each period. Again, earnings recovery may vary by worker type: when $\psi^h = \psi^l$, there is homogeneous recovery from displacement while when $\psi^h \neq \psi^l$, there exists heterogeneous recovery from displacement. The existing job displacement literature finds evidence of heterogeneity in the recovery of earnings, where low-educated workers suffer significantly longer and more persistent earnings losses compared to high-educated workers (Farber, 2015).

The simulation exercise compares the actual percentage decline in earnings, as given by the data generating process, and the percentage decline in earnings as given by the levels earnings regression. Specifically, the latter decline is calculated as a percentage decline by scaling the monetary unit decline by the sample average pre-displacement level of earnings.[10] Four different scenarios are considered, where alternative choices of ψ^{type} and ρ^{type} exhibit similar patterns:

1a. Homogeneous recovery from displacement of 2% per year, $\psi^h = \psi^l = 0.02$, with an equal probability of displacement for both worker types, $\rho^h = \rho^l$ (Figure 3a)

1b. Homogeneous recovery from displacement of 2% per year, $\psi^h = \psi^l = 0.02$, where low-type are more likely to be displaced, $\rho^h < \rho^l$ (Figure 3b)

2a. Heterogeneous recovery from displacement where high-type workers recover stronger than low-type workers, $\psi^h = 0.03$, $\psi^l = 0.005$, with an equal probability of displacement (Figure 4b)

2b. Heterogeneous recovery from displacement, $\psi^h = 0.03$, $\psi^l = 0.005$, where low-type workers are more likely to be displaced, $\rho^h < \rho^l$ (Figure 4c)

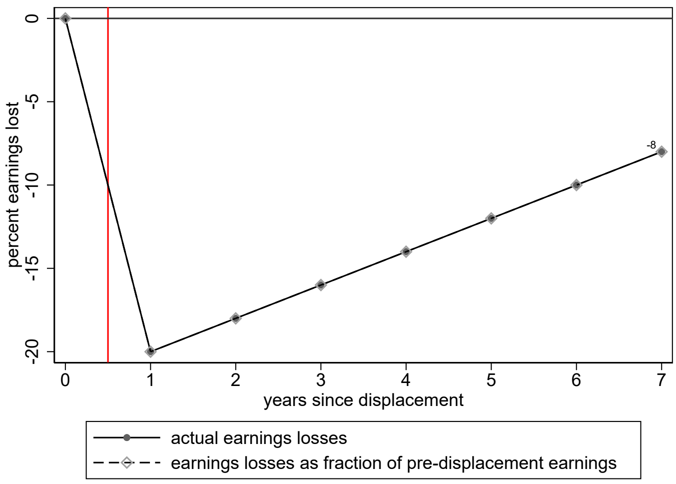
Figures 3a and 3b plot the earnings losses under the first two simulations, 1a and 1b. With homogeneous recovery from negative employment shocks, the two different earnings measures recover

[10] Because the sample contains equal amounts of high- and low-type workers, the sample average pre-displacement earnings is 150,000.

identical treatment effects. Indeed, true earnings losses—as given by the constant percent decline determined by the data generating process—and the earnings losses as a fraction of the sample average pre-displacement earnings are identical when $\psi^h = \psi^l$. This is true when workers of different types are equally likely to be displaced (Figure 3a) and when low-type workers are more likely to be displaced (Figure 3b).

Figure 3: The Simulated Earnings Losses With Homogeneous Recovery from Displacement Shock

(a) Equal pr(displacement), $\rho^h = \rho^l = 0.1$



(b) Pr(displacement) varies by type, $\rho^h < \rho^l$

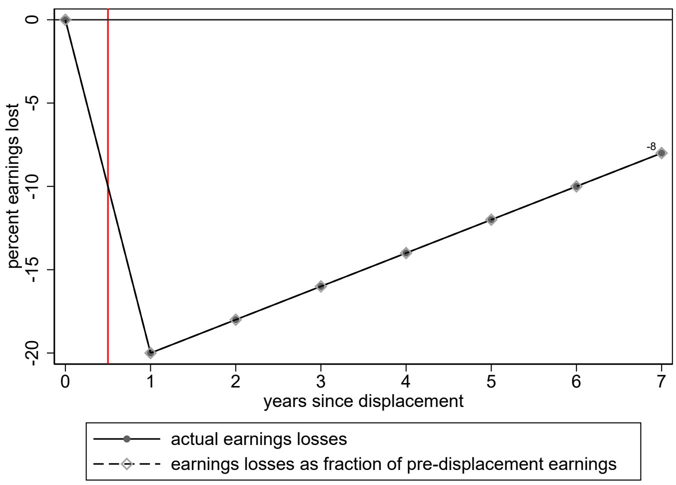


Figure plots the average earnings losses, percent of earnings lost, for simulations 1a and 1b. The solid line plots actual earnings losses, as determined by the data generating process described in Section 3. The dashed line plots earnings losses measured in levels and scaled by the pre-displacement sample average. Figure (a) plots homogeneous recovery from displacement by worker type, with equal probabilities of displacement. Figure (b) plots homogeneous recovery from displacement by worker type, where low-type workers are more likely to be displaced.

In contrast, with heterogeneity in the recovery after job loss, earnings losses in levels recover the wrong treatment effect and understate true earnings losses. Earnings of high-type workers recover much stronger and long-run earnings losses are just 2% compared to low-types whose long-run earnings losses are 17% (see Figure 4a). With heterogeneous recovery, scaling the earnings losses in monetary units by the sample average pre-displacement earnings differs considerably from the true earnings losses. Even with an equal probability of displacement as in Figure 4b, true earnings losses are 9.8% compared to earnings losses in levels which are only 7%. When low-type workers are more likely to be displaced, a similar difference exists: while true earnings losses are 11.6%, earnings losses in levels are only 8.8%.

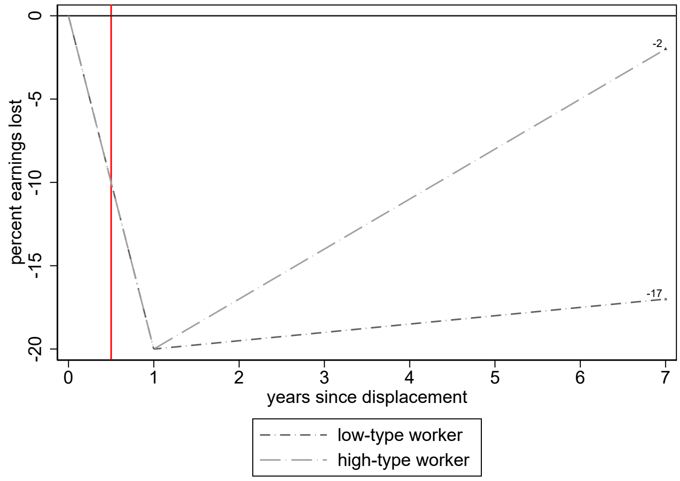
Heterogeneity in the recovery of earnings from job loss is sufficient to produce treatment effects of the wrong magnitude when the correct functional form is one of constant elasticity. Why does the regression in levels not recover the correct treatment effect? The discrepancy between the levels regression and the true data generating process occurs because the average treatment effect, calculated separately for high- and low-type workers, and the treatment effect at the average are not the same. Intuitively, those with the highest pre-displacement earnings suffer the lowest earnings losses (in percentage terms), and dividing by the sample average therefore understates the true effect. If, instead, high-type workers exhibited larger earnings losses than low-type workers, then the regression in levels would overstate the true treatment effect. However, the earnings regression in levels will recover the correct treatment effect if earnings losses are calculated as a fraction of sample average earnings *separately* by worker type. When doing so, the regression in levels scales the heterogeneous recovery in earnings by the correct sample average of pre-displacement earnings.

When the correct functional form is one of constant elasticity, the regression in levels will recover treatment effects of the wrong magnitude when workers of different types with different levels of pre-displacement earnings exhibit different recoveries from displacement. This is true even across two broad types of workers, as the sample average earnings also differs across worker type. Thus, with treatment effect heterogeneity and a constant elasticity functional form, the inverse hyperbolic sine transformation is the preferred measure of earnings as it recovers the correct treatment effect.

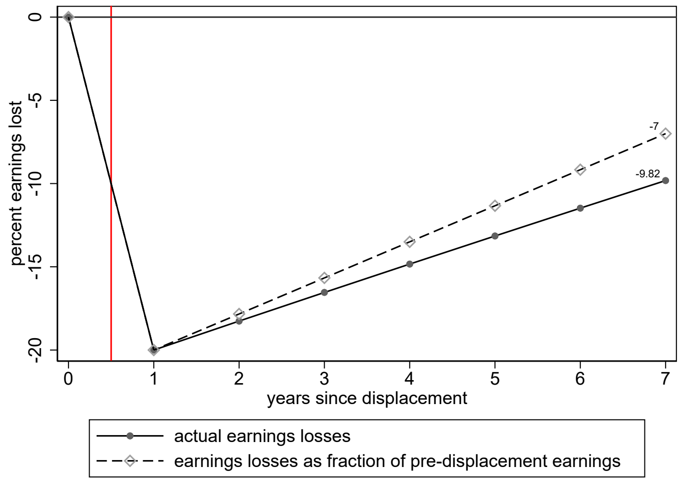
While the paper emphasizes the importance of correctly specified functional form, it also draws attention to moving beyond reporting a single treatment effect using linear estimators. Indeed, even the misspecified model estimated in levels recovers the correct treatment effect when calculated within worker type. Earnings losses in monetary units likely depend on many factors prior to job loss, which linear estimators fail to capture. Løken et al. (2012) demonstrate the importance of moving beyond linear estimators, and allowing for nonlinear effects which depend on different levels of, for instance, income. However, average earnings losses are also an important policy-relevant parameter, and arcsinh-transformation best captures the average effect of displacement compared to regressions in levels and log-transformed earnings.

Figure 4: The Simulated Earnings Losses With Heterogeneous Recovery from Displacement Shock

(a) Heterogeneous Recovery From Displacement Shock by Worker Type, $\psi^h = 0.03$, $\psi^l = 0.005$



(b) Equal pr(displacement), $\rho^h = \rho^l = 0.1$



(c) Pr(displacement) varies by type, $\rho^h < \rho^l$

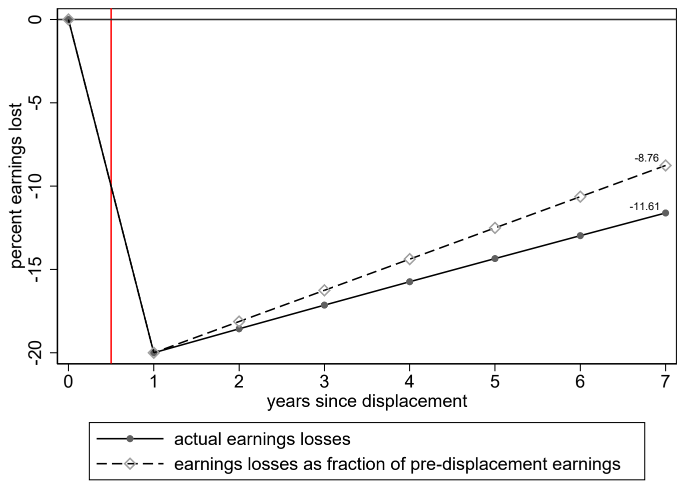


Figure plots the average earnings losses, percent of earnings lost, for simulations 2a and 2b. The solid line plots actual earnings losses, as determined by the data generating process described in Section 3. The dashed line plots earnings losses measured in levels and scaled by the pre-displacement sample average. Figure (a) plots actual earnings losses, as determined by the data generating process, separately for low-type and high-type workers. Figure (b) plots heterogeneous recovery from displacement by worker type, with equal probabilities of displacement. Figure (c) plots heterogeneous recovery from displacement by worker type, where low-type workers are more likely to be displaced.

4 How Much do Differences in the Probability of Zero Earnings Matter?

Figure 2 details the stark differences in the estimated earnings losses due to job displacement when correctly accounting for those with zero earnings. How sensitive are the estimated earnings losses to excluding those with zero earnings in the log-transformed regression? The results below examine how the estimated earnings losses differ between regressions with log- and arcsinh-transformed earnings when the probability of having zero earnings differs across subgroups and how earnings are measured. Section 4.1 examines how the difference in the estimated earnings losses varies across workers of different ages. Similarly, Section 4.2 examines the importance of zero earnings across workers of different education levels. Finally, Section 4.3 examines how the difference in the estimated earnings losses varies depending on whether or not benefits such as unemployment insurance, disability insurance, and social assistance are included in the measure of earnings.

4.1 Heterogeneity in the Probability of Zero Earnings Post-Displacement by Age

Figure 5 examines the importance of zero earnings across displaced workers of different ages: those 25–34 at displacement, those aged 35–44, and those aged 45–50. Older workers who are closer to retirement are significantly more likely to have zero earnings after being laid off (Figure 5a).[11] This is particularly true among the 45–50 group, where displaced workers of these ages are nearly 2 percentage points more likely to have zero earnings in the long-run. Such workers are closer to retirement, and while they cannot retire by +7 as the earliest a worker can retire is age 62 (Johnsen et al., 2020), they remain less likely to return to work.

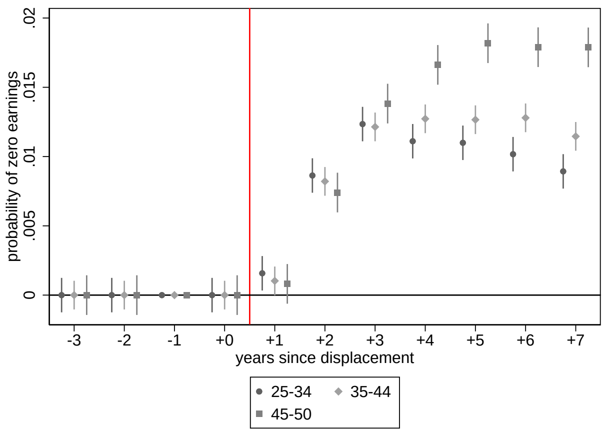
Figure 5b plots the estimated earnings losses across the three age groups measuring the log of earnings while Figure 5c plots the results from the arcsinh-transformed regression. While the discrepancy between the log-transformed regression and the arcsinh-transformed regressions is large for all age groups, it is even larger among older displaced workers. For the 25–34 group, the difference between earnings losses in the log- and arcsinh-transformed regressions is 12.6 percentage points 7 years after displacement. For those aged 35–44, the log-transformed regression understates earnings losses by 15.8 percentage points and for the oldest group, 45–50, this is an even larger 24.4 percentage point difference.

The extent to which the log-earnings regression is biased depends on the probability of zero earnings as a result of being displaced. The log-transformation drastically understates the true earnings losses

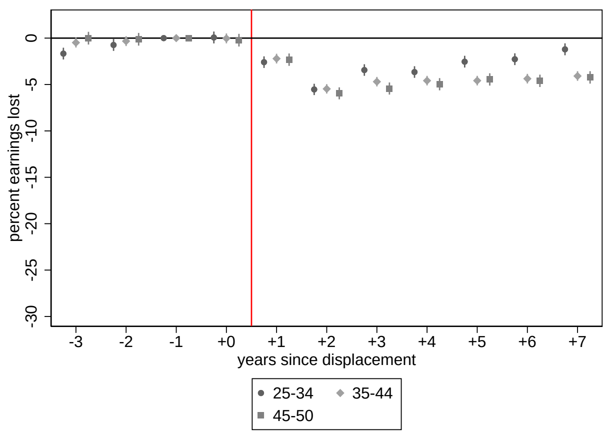
[11] Older men are slightly more likely to have zero earnings relative to older women. As such, including those with zero earnings post-displacement is slightly more important among older men.

Figure 5: The Impact of Job Displacement on the Probability of Having Zero Earnings and Earnings of Different Transformations, Separately by Different Ages

(a) The Probability of Having Zero Earnings



(b) Earnings, Log-transformed



(c) Earnings, Arcsinh-transformed

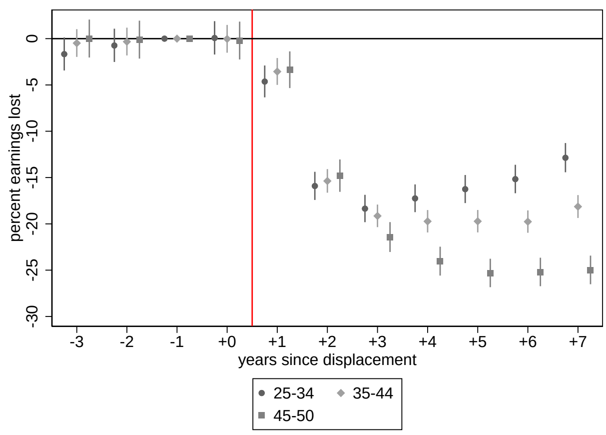


Figure plots δ_k coefficients, the event study estimates in equation 1, for one of three regressions: an indicator for zero earnings (figure a), log-transformed earnings (figure b), and arcsinh-transformed earnings (figure c). Estimates are plotted in percent of earnings lost for figures (b) and (c). Separate regressions estimated for each worker age group, comparing displaced and non-displaced workers within the same age group. Estimated δ_k coefficients interpreted relative to fixed difference in -1, where δ_{-1} is omitted by convention. Displacement event occurs between +0 and +1. 95% confidence intervals reported, though not always visible due to symbol.

for workers who are more likely to have zero earnings, as is the case with older workers. In addition, such differences across samples can substantially alter the interpretation of the results. While the long-run earnings losses of displaced workers aged 35–44 and 45–50 are virtually identical in the log-transformed regression, the arcsinh-transformation exhibits larger earnings losses among older displaced workers which never recover.

4.2 Heterogeneity in the Probability of Zero Earnings Post-Displacement by Education

Figure 6 examines the importance of zero earnings across displaced workers of different education levels. Workers are separated by whether or not they have a bachelors degree or higher. Non-college educated workers are substantially more likely to have zero earnings as a result of being laid off relative to college educated workers: 1.4 percentage points compared to 0.8 percentage points. The stark difference by education level suggests that displaced workers with a college education find new jobs after displacement at higher rates.

Given the differences in zero earnings among the two sets of displaced workers, the discrepancy between the log- and arcsinh-transformed regressions is larger among non-college educated workers. For these lower educated workers, the log-earnings regression understates earnings losses by 18.9 percentage points compared to just 11.3 percentage points for college educated workers. While earnings losses are always smaller among college educated workers across both earnings regressions, excluding those with zero earnings is more important to the estimated earnings losses among lower educated workers.

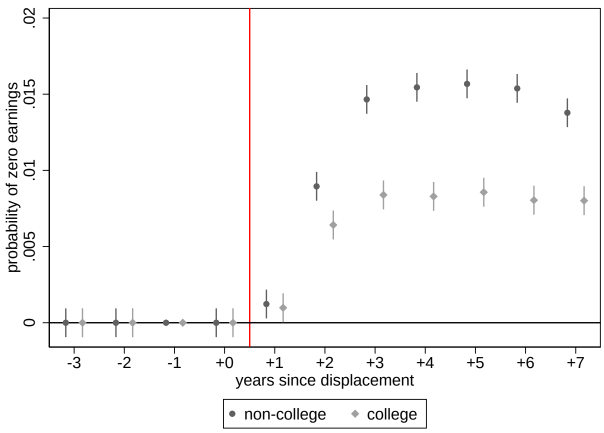
4.3 Measurement of Earnings: Including or Excluding Benefits

In the results of the previous sections, earnings are measured as labor earnings. While labor earnings losses after job loss are undoubtedly of interest, it is also of interest to understand how benefits provide a safety net to earnings losses in the aftermath of displacement. Indeed, both earnings and income including such benefits are frequently of interest in the displacement literature (East and Simon, 2020). While income losses due to job loss will be lower when including benefits such as unemployment insurance, disability insurance, and social assistance, the probability of having zero income in a given year is also lower, as seen in Figure 7.

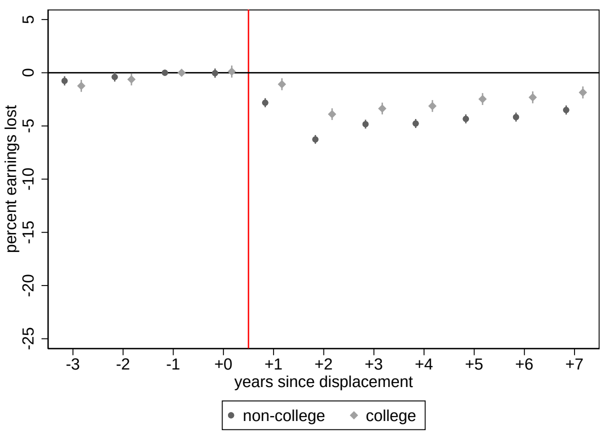
Figure 8 replicates the regressions reported in Figure 2 using income, labor earnings plus benefits, rather than labor earnings. As expected, the estimated income losses are considerably lower than earnings losses across all different measures of earnings, as benefits replace a portion of earnings

Figure 6: The Impact of Job Displacement on the Probability of Having Zero Earnings and Earnings of Different Transformations, Separately by Education

(a) The Probability of Having Zero Earnings



(b) Earnings, Log-transformed



(c) Earnings, Arcsinh-transformed

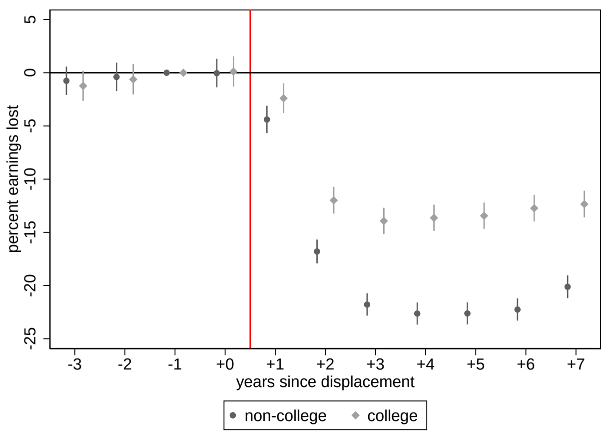


Figure plots δ_k coefficients, the event study estimates in equation 1, for one of three regressions: an indicator for zero earnings (figure a), log-transformed earnings (figure b), and arcsinh-transformed earnings (figure c). Estimates are plotted in percent of earnings lost for figures (b) and (c). Separate regressions estimated for workers with/without a bachelors degree, comparing displaced and non-displaced workers within the same level of education. Estimated δ_k coefficients interpreted relative to fixed difference in -1, where δ_{-1} is omitted by convention. Displacement event occurs between +0 and +1. 95% confidence intervals reported, though not always visible due to symbol.

Figure 7: The Impact of Job Displacement on The Probability of Having Zero Earnings

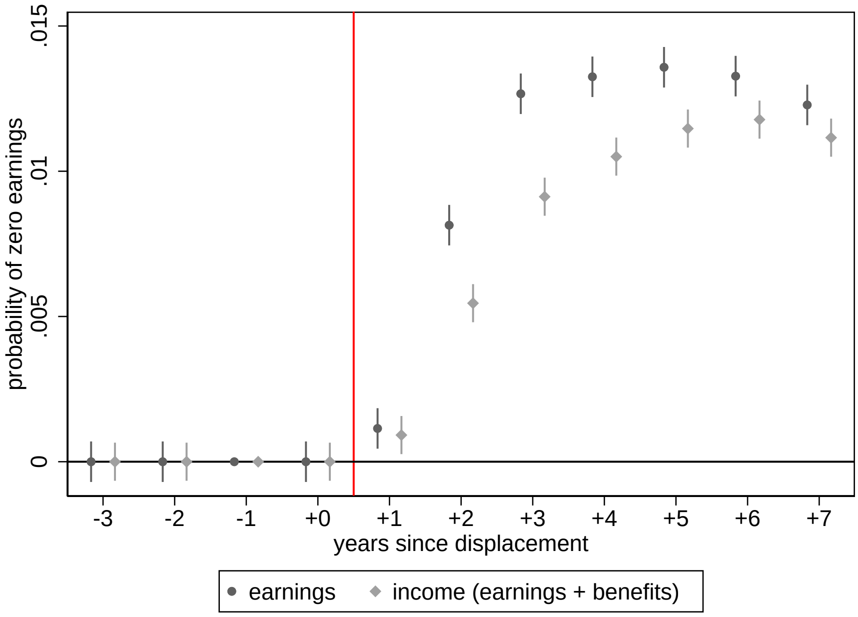


Figure plots the estimated impact of job displacement on the probability of having zero earnings and zero income, the event study estimates from equation 1. Earnings measured as annual labor earnings while income measured as earnings plus unemployment benefits, disability benefits, and social assistance payments. Estimated δ_k coefficients interpreted relative to fixed difference in -1, where δ_{-1} is omitted by convention. Displacement event occurs between +0 and +1. 95% confidence intervals reported.

losses post-displacement. Even though the probability of having zero income is lower than the probability of having zero earnings, there remains a considerable difference between the log-transformed, un-transformed, and arcsinh-transformed regressions. The log-transformed regressions understates income losses by 10.3 percentage points in the long-run when using earnings including benefits. When measuring earnings excluding benefits as in Figure 2, where the probability of having zero earnings is even greater, the difference is an even larger 15 percentage points. While the discrepancy between the log- and arcsinh-transformation matters less when the probability of zero earnings is smaller, the difference in the magnitude of the estimated treatment effect remains considerable.

5 Conclusion

The paper explains how the measurement of earnings has important implications for the estimated treatment effects in standard earnings regressions. First, the paper shows that by excluding those with zero earnings, the log-transformed earnings regression may produce biased treatment effects. This is true when treatment directly impacts the probability of having zero earnings, as is the case when workers are displaced during mass-layoff or plant closing events. This bias can be substantial: earnings losses in the long-run are 5 times smaller when excluding those with zero earnings in the

Figure 8: The Estimated Impact of Displacement on Earnings Across Different Transformations of Earnings

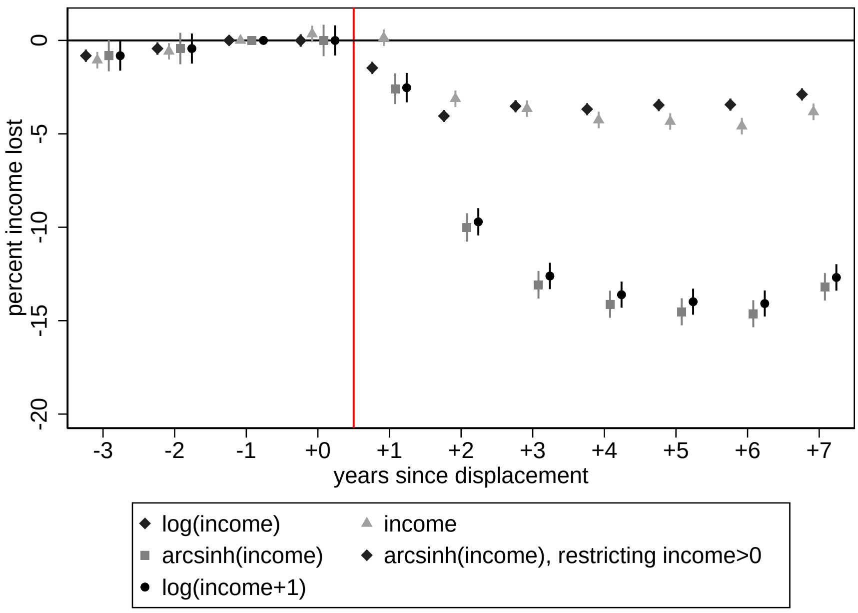


Figure plots δ_k coefficients, the event study estimates in equation 1, for one of five regressions: log(income), income in levels, arcsinh(income), arcsinh(income) conditioning on positive income, and log(income+1). Estimates are plotted in percent of earnings lost. Percent earnings lost for earnings measured in levels calculated as the fraction of average pre-displacement earnings in -1. Estimated δ_k coefficients interpreted relative to fixed difference in -1, where δ_{-1} is omitted by convention. Displacement event occurs between +0 and +1. 95% confidence intervals reported, though not always visible due to symbol. Income measured as earnings plus unemployment benefits, disability benefits, and social assistance payments.

log-earnings regression.

Second, the paper explains the conditions in which log-like transformations, the arcsinh-transformation, are preferred to estimating the same regression in levels. The paper shows how when the correct functional form is one of constant elasticity, earnings regressions in levels will not recover the correct average treatment effect with heterogeneous earnings losses by worker type. Finally, the paper demonstrates how differences in the probability of having zeros post-displacement matters across different samples—where zero earnings are more important among older and non-college educated workers—and across different measures of earnings—including or excluding benefits such as unemployment insurance, social assistance, and disability insurance. Such differences can alter the interpretation of the recovery of earnings after job loss across different groups.

Two important practical considerations emerge. One, the commonly used log-transformation should be avoided when treatment impacts the probability of having zero earnings. The correct earnings transformation matters, as the bias resulting from excluding those with zero earnings can be substantial. This is likely to be more problematic in settings when the probability of having zero earnings is particularly high: for instance, when using quarterly data compared to annual data or in settings when levels of non-employment are particularly high, such as in a severe recession. Two, researchers should think carefully about functional form, and whether they believe a constant monetary or constant elasticity interpretation is correct in the earnings regression. Indeed, the difference is quantitatively important in the presence of treatment effect heterogeneity across different types of worker. When the correct functional form is one of constant elasticity, alternative log-like transformations which are defined for those with zero earnings such as the arcsinh-earnings regression are the preferred specification.

References

Aaberge, R., T. Wennemo, A. Bjorklund, M. Jantti, P. J. Pedersen, and N. Smith (2000). Unemployment shocks and income distribution: How did the nordic countries fare during their crises? *The Scandinavian Journal of Economics 102*(1), 77–99.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics 60*(1), 47–57.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61.

Calónico, S. and J. Smith (2017). The women of the national supported work demonstration. *Journal of Labor Economics 35*(S1), S65–S97.

Couch, K. A. and D. W. Placzek (2010). Earnings losses of displaced workers revisited. *American Economic Review 100*(1), 572–89.

Davis, S. J. and T. M. Von Wachter (2011). Recessions and the cost of job loss. Technical report, National Bureau of Economic Research.

East, C. N. and D. Simon (2020, December). How well insured are job losers? efficacy of the public safety net. Working Paper 28218, National Bureau of Economic Research.

Farber, H. S. (2015, May). Job loss in the great recession and its aftermath: U.s. evidence from the displaced workers survey. Working Paper 21216, National Bureau of Economic Research.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann (2012, 02). Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps. *The Review of Economics and Statistics 94*(1), 153–171.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies 64*(4), 605–654.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2003, May). Fifty Years of Mincer Earnings Regressions. NBER Working Papers 9732, National Bureau of Economic Research, Inc.

Huttunen, K., J. Møen, and K. G. Salvanes (2011). How destructive is creative destruction? effects of job loss on job mobility, withdrawal and income. *Journal of the European Economic Association*.

Imbens, G. W. and J. M. Wooldridge (2009, March). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993, September). Earnings losses of displaced workers. *The American Economic Review 83*(4), 685–709.

Johnsen, J. V., A. Willén, and K. Vaage (2020, September). Interactions in Public Policies: Spousal Responses and Program Spillovers of Welfare Reforms. Discussion Paper Series in Economics 20/2020, Norwegian School of Economics, Department of Economics.

Kostol, A. R. and M. Mogstad (2014, February). How financial incentives induce disability insurance recipients to return to work. *American Economic Review 104*(2), 624–55.

Krolikowski, P. (2018). Choosing a control group for displaced workers. *ILR Review 71*(5), 1232–1254.

Lachowska, M., A. Mas, and S. A. Woodbury (2020, October). Sources of displaced workers' long-term earnings losses. *American Economic Review 110*(10), 3231–3266.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review 76*(4), 604–620.

Løken, K. V., M. Mogstad, and M. Wiswall (2012, April). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics 4*(2), 1–35.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings.* NBER.

OECD (2006). *OECD Employment Outlook 2006 Boosting Jobs and Incomes: Boosting Jobs and Incomes.* OECD Employment Outlook. OECD Publishing.

Raposo, P., P. Portugal, and A. Carneiro (2019). The sources of the wage losses of displaced workers: the role of the reallocation of workers into firms, matches, and job titles. *Journal of Human Resources*.

Ravallion, M. (2017). A concave log-like transformation allowing non-positive values. *Economics Letters 161*, 130–132.

Silva, J. M. C. S. and S. Tenreyro (2006, November). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Stevens, A. H. (1997, January). Persistent effects of job displacement: The importance of multiple job losses. *Journal of Labor Economics 15*(1), 165–88.