

Measurement of Nominal Wages and Payroll Schedules in Administrative Earnings Data*

Seth Murray

July 13, 2020

Abstract

The economic questions that can be addressed using many of the large administrative employer-employee linked data sets with workers' earnings have been limited by the absence of information on workers' base wages, variable compensation, hours or weeks worked, and other factors determining workers' earnings. This paper presents a set of machine learning methods that identify each worker's unobserved persistent base wages, paydays weeks, and annual bonuses from the worker's quarterly earnings. I then implement and evaluate the quality of these methods using quarterly earnings data in the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) dataset, an employer-employee linked dataset for the United States. Using the estimated nominal wages of workers in 30 U.S. states, I document four patterns of nominal wage adjustment: i) estimated persistent wage changes exhibit downward nominal wage rigidity, ii) optimal real wage cuts are suppressed by downward nominal wage rigidity, iii) workers' nominal raises follow a Taylor-like pattern, with the probability of a wage raise spiking every four quarters, and iv) the timing of workers' annual raises are synchronized within the firm.

*I am deeply indebted to my dissertation committee: Borağan Aruoba, John Haltiwanger, Judy Hellerstein, and Henry Hyatt and my graduate director, John Shea, for their support and guidance. I also want to thank Katharine Abraham, Joonkyu Choi, Leland Crane, Cynthia Doniger, Joanne Gaskell, Jessica Goldberg, Ethan Kaplan, Erika McEntarfer, Michael Murray, Felipe Saffie, Kristin Sandusky, Matthew Staiger, Lesley Turner, Larry Warren, the seminar participants at the University of Maryland, the Bureau of Labor Statistics, Bates College, and the U.S. Census Bureau, and the LEHD group at the Center for Economic Studies in the U.S. Census Bureau for engaging with me on my research. I am grateful for the financial support of the Center for Retirement Research and the Betancourt Fellowship. The opinions expressed herein are those of the author alone and do not necessarily reflect the view of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential data are disclosed. See U.S. Census Bureau Disclosure Review Board bypass numbers: DRB-B0073-CED-20190910, DRB-B0069-CED-20190725, DRB-B0037-CED-20190327, CBDRB-2018-CDAR-061. Email: murrase@umd.edu

1 Introduction

Over the last three decades, many countries have created administrative data sets that track workers' earnings at many or nearly all firms. The rich data contained in these employer-employee linked data sets have enabled researchers to examine the impact of policy changes, worker and firm decision-making, labor market dynamics, and many other topic areas. However, this research has been limited by the fact that most employer-employee linked data sets fail to decompose workers' reported earnings into the base wages, variable compensation, hours or weeks worked, and other factors that determine workers' earnings.

To overcome this limitation, I develop a set of machine learning methods that identify each worker's unobserved payday schedule and nominal wage series from the worker's observed quarterly earnings at an employer. First, I note that variation in the number of paydays due to a worker's payday schedule can generate fluctuations of $\pm 15\%$ in the worker's quarter-over-quarter earnings. Although earnings fluctuations generated by changes in the number of paydays appear as noise at an individual level, these payday-related earnings changes can be identified because they are large and common to many workers at the firm. I begin by identifying the individual-level payday schedule that minimizes the implied residual variance of the worker's observed quarterly earnings. I then identify the payday schedule(s) in operation at the firm using a clustering algorithm that determines which payday schedules are common to many workers at the firm. Having identified the payday schedule(s) used by the firm, I then determine which payday schedule (from this more limited set) minimizes the implied residual variance of each worker's quarterly earnings. Knowing the payday schedule allows me to estimate each worker's number of payday weeks in any given period, and thus control for transitory fluctuations in quarterly earnings generated by the worker's payday schedule.

I reframe the problem of identifying persistent changes in each worker's base wage as one of identifying structural breaks in the worker's observed earnings. Each persistent wage change is equivalent to a structural break in the time series of a worker's log earnings over the worker's job spell at a firm. By reframing the problem, I can use methods developed by the extensive literature on identifying structural breaks in time-series data.¹

¹See Casini and Perron (forthcoming) for a review of recent advances in the literature on structural break identification. In the downward nominal wage rigidity literature, Gottschalk (2005) addressed measurement error in survey respondents' reported base wages by applying structural break identification procedures proposed by Bai and Perron (1998). Barattieri, Basu and Gottschalk (2014) further extended Gottschalk's method to account for Type I and Type II error in the identification of nominal wage changes.

The post-Lasso estimation procedure is one such method for identifying structural breaks in time-series data.² When used to identify structural breaks in workers' earnings series at a firm, the post-Lasso estimation procedure allows for persistent wage changes to occur in any period of a worker's employment history. The procedure minimizes an objective function that has two components. The first component optimizes the model fit by choosing the wage change estimates that minimize the Euclidean distance between the predicted persistent log wage history and the observed log earnings history (this is the same as standard OLS minimization of the sum of squared residuals). The second component addresses model over-fitting by including a penalty parameter for the sum of the absolute value of the estimated log wage changes. This penalty parameter causes the post-Lasso procedure to set the estimated persistent wage change to zero for many periods, which is consistent with workers not receiving base wage changes every quarter.

I implement these methods using a 10% random sample of firms from 30 states in the U.S. Census Bureau's LEHD data set, an employer-employee linked administrative data set covering approximately 96% of employment in each state. Because the LEHD dataset, like most administrative employer-employee linked datasets, contains earnings series for millions of workers, the scalability of the post-Lasso procedure makes it a particularly attractive method for identifying structural breaks in the worker's log base wage (i.e. persistent wage changes).

I use the estimated persistent wage changes to identify four patterns of nominal wage adjustment: i) estimated persistent wage changes exhibit downward nominal wage rigidity, ii) real wage cuts that would be optimal in a frictionless environment are suppressed by downward nominal wage rigidity, iii) workers' nominal raises follow a Taylor-like pattern, with the probability of a wage raise spiking every four quarters, and iv) the timing of workers' annual raises is synchronized within the firm.

2 Data

This paper uses the U.S. Census Bureau's LEHD data set - an employer-employee linked data set with quarterly earnings for approximately 96% of all employment in a state. The quarterly earnings data in the LEHD is derived from firms' mandatory unemployment insurance filings. This earnings data is complemented with both worker characteristics (age, sex, race, and education) and firm characteristics (industry, firm age, and firm size) from other data sources. Individuals are uniquely identified by a Protected Identification Key (PIK) that allows each individual to be

²Notable studies that use Lasso estimation for structural break identification include Harchaoui and Lévy-Leduc (2010) and Ciuperca (2014).

tracked across different employers and locations. The LEHD identifies employers at the level of a state employer identification number (SEIN). Firm age and firm size are derived by aggregating one or more SEINs (potentially across states) to the level of the federal employer identification number (EIN). For simplicity, I refer to each SEIN as a firm.

This paper extracts two samples from the LEHD data set. The primary sample consists of a 10% random sample of SEINs from thirty states covering the period from 1998:Q1 to 2017:Q1.³ I chose these thirty states because there are no gaps in reported quarterly earnings for any of these states over the sample period. I also employ a secondary sample that is a 10% random sample of SEINs from the four states that also reported quarterly hours paid during the period from 2011:Q1 to 2018:Q1 (MN, OR, RI, and WA). I use this secondary sample because the hours-paid data helps quantify the degree of measurement error in my baseline post-Lasso estimates of workers' nominal wages.

3 Estimation of Wages and Payday Weeks

A key drawback of the LEHD data set is that the LEHD generally only reports workers' quarterly earnings, which can vary due to fluctuations in overtime pay, bonuses, payday weeks, average weekly hours paid, or the base wage.⁴ Many of these components of quarterly earnings are transitory, and thus are unlikely to affect employment decisions in long-term employment relationships (see Appendix B). To overcome this limitation of the LEHD, I develop a set of novel machine learning tools that identify persistent changes in workers' unobserved nominal base wages from their observed nominal quarterly earnings. This section describes these machine learning methods and evaluates the quality of the resulting estimated nominal wage changes.

To more formally distinguish between persistent versus transitory components of quarterly earnings, I express each worker's quarterly earnings as a function of the worker's base wage, the number of payday weeks in the quarter, the number of regular and overtime hours worked, and any variable

³The states included in the primary sample are: CA, CO, CT, FL, GA, HI, ID, IL, IN, KS, LA, MD, ME, MT, NC, ND, NJ, NM, NV, OR, PA, RI, SC, SD, TN, TX, VA, WA, WI, and WV.

⁴The LEHD data set does have total quarterly hours paid for four states from 2011 forward. Although these hours-paid data are helpful for identifying patterns in workers' wage changes, only Washington state has hours-paid data prior to 2009. For more details on patterns of adjustment in the non-wage components of quarterly earnings, see Appendix C.

compensation paid to the worker (e.g. annual bonuses, tips, and commissions).⁵ Specifically,

$$y_{ikt} = w_{ikt} (n_{ikt} \bar{h}_{ikt} + \theta_{ikt}^o n_{ikt} \bar{h}_{ikt}^o) v_{ikt} \epsilon_{ikt} \quad (1)$$

where y_{ikt} is worker i 's total nominal earnings at firm k in quarter t , w_{ikt} is the worker's base hourly wage, n_{ikt} is the number of payday weeks in the quarter, \bar{h}_{ikt} is the worker's average weekly hours paid, θ_{ikt}^o is the overtime premium for overtime pay (typically 1/2), \bar{h}_{ikt}^o is the worker's average weekly overtime hours paid, v_{ikt} is the worker's variable compensation as a percent of the base wage, and ϵ_{ikt} is measurement error (such as dropping or adding a decimal place in the recorded quarterly earnings).⁶

3.1 Estimation of Payday Weeks

Payroll schedules generate significant fluctuations in quarterly earnings because of variation in the number of pay periods from quarter to quarter. For instance, workers who are paid bi-weekly typically experience $\pm 15\%$ fluctuations in quarterly earnings from one quarter to the next as the number of quarterly paydays switches between six and seven. Although earnings fluctuations generated by changes in the number of payday weeks appear as noise at an individual level, these payday-related changes can be identified because they are common to many workers at the firm. Knowing the changes that are induced by payroll schedules is useful because such changes are directly related to the number of weeks worked, and thus allow estimation of a worker's average weekly earnings.

The method for identifying a firm's payroll schedule(s) exploits three empirical regularities. First, there are a limited number of potential payday schedules: seven weekly, fourteen bi-weekly, one monthly, and one semimonthly payroll schedule. Second, each of these payday schedules has a distinct time series of payday weeks from quarter to quarter. Importantly, the time series of quarterly payday weeks for each payday schedule can be determined from the annual calendar. And third, firms tend to use a small number of payroll schedules for their employees (typically only one or two, see Burgess (2014)), so the fluctuations in quarterly earnings caused by payday

⁵When measuring workers' wages, I only consider "full-quarter" earnings - where the worker has positive earnings from the same SEIN in the quarter immediately before and after the current quarter. This has the benefit of reducing fluctuations in quarterly earnings that result from new hires and job separators working only part of the quarter in which they are hired or separate.

⁶For salaried employees who are exempt from overtime pay requirements, the average reported weekly hours paid always equals 40 (no matter how many hours are actually worked) and the average weekly overtime hours paid is always zero.

schedules are common to many workers at the firm.

Estimating a worker's payday weeks first requires identifying the set of payroll schedules used by the worker's employer. This firm-specific set of payroll schedules is determined by iteratively selecting the payroll schedule that best fits the observed quarter-over-quarter earnings changes for the largest number of workers. For each worker at the firm and potential payroll schedule p , I construct the worker's payroll-schedule-adjusted log earnings change, Δy_{ikt}^p , defined as:

$$\Delta y_{ikt}^p = (\ln(y_{ikt}) - \ln(y_{ikt-1})) - (\ln(n_t^p) - \ln(n_{t-1}^p)) \quad (2)$$

where n_t^p is the number of payday weeks for the payroll schedule p in quarter t based on the annual calendar.⁷

The best fitting payroll schedule for a given worker, p_{ik}^* , is the payroll schedule that implies the lowest variance of Δy_{ikt}^p . The intuition behind this rule is that for the true payroll schedule, the subtracted change in the payroll schedule's payday weeks, $(\ln(n_t^p) - \ln(n_{t-1}^p))$, is perfectly negatively correlated with the true payday weeks change component of the quarterly earnings change. This perfect negative correlation, combined with an assumption that changes in wages, weekly hours, and variable compensation are independent of the number of payroll weeks, implies that the true payday schedule minimizes the variance in Equation 2 as the duration of the job spell approaches infinity. Once each worker's best-fitting payroll schedule has been identified, I select the payroll schedule that is best for the largest number of workers, where each worker's best schedule is given a weight equal to the duration of the worker's job spell (this accounts for differences in the precision of the individual-level variance estimates caused by differences in the job spell duration).

After each iteration, I remove any workers for whom the selected payroll schedule was the best-fitting payday schedule. I then run another iteration for the remaining workers. This process continues until either four payroll schedules have been selected or no remaining payroll schedule is the best-fitting payroll schedule for five or more workers. Having identified the set of potential payday schedules at the firm, I select from this constrained set the worker-specific payroll schedule that minimizes the variance of the worker's payroll-schedule-adjusted change in log earnings (Δy_{ikt}^p).

⁷One complication from using the annual calendar is that January 1st, New Years Day, is a holiday that occurs at the transition between Q4 and Q1. From the data, it is apparent that some firms shift weeks paid from Q1 to Q4 when the weekday of payment falls on New Years Day.

3.2 Post-Lasso Estimation of Persistent Nominal Wage Changes

I next use a post-Lasso procedure to extract persistent changes in each worker's unobserved base wage (w_{ikt}) from their payday-adjusted quarterly earnings (y_{ikt}). This procedure involves four steps. First, I express each worker's base wage in any given period t as a recursive formulation of the worker's starting wage (w_{ik1}) and all base wage percentage changes up to the current period (Δ_{iks}^w):

$$w_{ikt} = w_{ik1} \prod_{s=2}^t (1 + \Delta_{iks}^w) \quad (3)$$

Inserting this recursive formulation into the quarterly earnings decomposition in Equation 1 and then taking the natural log allows observed quarterly earnings to be rewritten as:

$$\ln(y_{ikt}) = \ln(w_{ik1}) + \sum_{s=2}^t \ln(1 + \Delta_{iks}^w) + \ln(n_{ikt}) + \ln(\bar{h}_{ikt} + \theta_{ikt}^o \bar{h}_{ikt}^o) + \ln(v_{ikt}) + \ln(\epsilon_{ikt}) \quad (4)$$

Second, although I do not observe any of the right-hand side variables, I can express Equation 4 as a standard linear regression model. Specifically, if we observe T periods of employment for the worker at the firm, then the current period quarterly earnings can be expressed as:

$$\ln(y_{ikt}) = \beta_{ik}^1 d_{ikt}^1 + \sum_{s=2}^T \beta_{ik}^s d_{ikt}^s + \alpha_{ik} \ln(n_{ikt}) + \underbrace{\ln(\bar{h}_{ikt} + \theta_{ikt}^o \bar{h}_{ikt}^o)}_{\tau_{ikt}} + \ln(v_{ikt}) + \ln(\epsilon_{ikt}) \quad (5)$$

where d_{ikt}^s is a set of T indicator variables that, in any given period t , take on a value of 1 only if $t \geq s$. Thus, d_{ikt}^1 corresponds to the intercept term and its coefficient, β_{ik}^1 , represents the log starting weekly wage of the worker: $\ln(w_{ik1})$. The coefficient on each subsequent indicator variable, β_{ik}^s , represents the persistent nominal wage change experienced by the worker in period s : $\ln(1 + \Delta_{iks}^w)$.

Although I do not observe the number of payday weeks for each worker (n_{ikt}), controlling for the number of payday weeks is important because it generates substantial noise in each worker's quarterly earnings series. Since firms tend to use only one or two payday schedules for all of their workers (see Burgess (2014)), a large share of workers at a firm may exhibit similar persistent earnings changes simply because of their number of payday weeks. To address this concern, I develop a clustering method that estimates the number of payday weeks for each worker. This clustering method exploits the fact that each firm uses only a small number of payday schedules, which are themselves selected from a total universe of 23 payday schedules. Critically, each of these potential payday schedules has a distinct time series of payday weeks from quarter to quarter -

where this time series can be determined from the annual calendar. Thus, the clustering algorithm identifies payday schedules at the firm (and then for a worker) based on patterns of quarter-over-quarter earnings changes that are both common to many workers at the firm and align with one of the potential payday schedules. Section 3.1 contains a complete description of the clustering method for estimating workers' payday weeks (\hat{n}_{ikt}).

The number of payday weeks is common to many workers at the firm. The remaining unobserved components, representing average weekly hours paid, overtime, and variable compensation, are included in the τ_{ikt} error term. While I would ideally observe these components as well, their absence is less troubling since persistent changes in these unobserved components of earnings are relatively rare.

In the third step, I identify the quarters in which a worker received a wage change. It is impossible to estimate the regression model in Equation 5 using standard methods because there are $T + 1$ explanatory variables and only T observations. Instead, I exploit the fact that there are presumably many quarters in which a worker has no change in their base wage. This implies that $\beta_{ik}^s = 0$ in those quarters, making the Lasso variable selection procedure, first proposed by Tibshirani (1996), an ideal method for identifying quarters in which a worker has a persistent wage change (i.e. has a non-zero β_{ik}^s coefficient).

The Lasso estimation procedure selects variables to include in a regression model by trading off the improvement in the explanatory power of the model when the variable is allowed a non-zero coefficient (the standard OLS minimization of the sum of squared residuals) against a penalty for the absolute distance of the coefficient from zero. Thus, for every worker-firm job spell, I use the Lasso estimation procedure to select the set of non-zero β_{ik}^s that solve the following minimization problem:⁸

$$\min_{\beta_{ik}^1, \dots, \beta_{ik}^T, \alpha_{ik}} \left(\sum_{t=1}^T \ln(y_{ikt}) - \sum_{s=1}^T \beta_{ik}^s d_{ikt}^s - \alpha_{ik} \ln(\hat{n}_{ikt}) \right)^2 + \lambda_{ik} \left(\|\alpha_{ik}\| + \sum_{s=1}^T \|\beta_{ik}^s\| \right) \quad (6)$$

where \hat{n}_{ikt} is the estimated number of payday weeks from the procedure described in Section 3.1. The λ_{ik} penalty parameter is set using 10-fold cross validation for each worker-firm job spell.⁹ If

⁸For the 4-state sample with quarterly hours-paid data, I replace $\ln(\hat{n}_{ikt})$ with the log of the reported hours paid ($\ln(n_{ikt}\bar{h}_{ikt})$), which changes the meaning of the β_{ik}^1 coefficient to be each worker's initial starting hourly wage rather than their starting weekly wage.

⁹ X -fold cross validation randomly partitions the worker's wage history into X distinct subsets. Each subset then serves as a holdout group that evaluates the prediction quality of the estimated wage changes generated using the other $X - 1$ subsets. The optimal λ_{ik} penalty parameter is chosen to maximize the prediction quality on the X

the job spell has fewer than ten full quarters of employment, then I instead use leave-one-out cross validation.

The error term of this Lasso minimization problem includes any error from the estimation of payday weeks plus deviations of log weekly hours, overtime hours, variable compensation, and measurement error from the variables' averages over the job spell. Minimizing the Lasso objective function identifies the quarters for which assigning a non-zero wage change coefficient significantly improves the model fit in both the current period and all subsequent periods. Thus, the Lasso procedure identifies persistent changes in the worker's weekly earnings.¹⁰ Although these persistent changes in workers' weekly earnings will often come from changes in their base wages, the Lasso procedure will also pick up persistent changes in workers' average weekly hours worked (e.g. going from part to full-time) or persistent changes in their variable compensation (e.g. a permanent change in the sales commission rate), which will mistakenly be attributed to changes in a worker's persistent base wage.

By penalizing non-zero coefficients based on their absolute distance from zero, the Lasso estimation procedure generates attenuation bias in the coefficient estimates. Thus, the fourth and final step of the wage estimation process addresses this bias by estimating a standard post-Lasso OLS regression model for every worker-firm job spell that only includes the variables with non-zero coefficients selected by the Lasso procedure in Step 3. The resulting coefficient estimates from the post-Lasso OLS regression serve as my estimates of each worker's persistent nominal wage changes.

3.2.1 Quality Evaluation of Post-Lasso Estimated Nominal Wage Changes

Because I do not observe workers' true base wages, it is difficult to validate this post-Lasso estimation procedure. That said, I can evaluate the quality of the post-Lasso estimation procedure in two ways.

First, in the four states with hours-paid data from 2011:Q1 to 2018:Q1, I evaluate how often the post-Lasso procedure identifies wage changes for workers with hourly earnings at or near the old minimum wage in quarters in which the state changes its minimum wage. I begin by identifying quarters in which any of the four states changed their minimum wage. After constructing each worker's average hourly earnings in a given quarter by dividing their reported nominal earnings by

hold-out subsets.

¹⁰This Lasso procedure is very similar in spirit to the structural break identification procedure that Gottschalk (2005) adapted from Bai and Perron (1998) in order to correct for measurement error in self-reported wages from survey data. Barattieri, Basu and Gottschalk (2014) further improve upon Gottschalk's method by explicitly accounting for Type I and Type II errors in the error correction process.

their hours paid, I include in the sample all workers in a state whose hourly earnings in $t - 2$ (where t is the quarter of the state’s minimum wage change) were between the old and the new minimum wage.

The Lasso estimation procedure identifies that 55.0% of these minimum-wage workers received a wage raise in the quarter of the state’s minimum wage change. An additional 33.3% of minimum wage workers are identified as receiving a wage change in the quarter immediately before the minimum wage change, but approximately half of these are workers who are also identified as receiving a wage change in the same quarter as the state’s minimum wage change.¹¹ Aggregating across the quarters immediately before and during the minimum wage change, I find the post-Lasso procedure identifies wage changes for 70.0% of minimum-wage workers in the quarter of or immediately before a state’s minimum wage change. Thus, for minimum wage workers in these four states, 30.0% is a reasonable estimate of the Lasso procedure’s Type II error rate (the failure to detect a wage change when there is a change). In the quarter immediately after the state’s minimum wage change takes effect, the Lasso estimation procedure identifies a nominal wage change for 14.7% of workers. This provides an upper-bound estimate of the Type I error rate (detecting a wage change when there is none), since some minimum wage workers may have received both a raise in the state’s minimum wage change quarter and an additional wage change in the quarter immediately following the minimum wage change.

I also evaluate the accuracy of the post-Lasso estimation procedure by comparing the frequency of quarterly wage raises identified by the post-Lasso procedure against the frequency of quarterly nominal wage changes reported for U.S. workers by Barattieri, Basu and Gottschalk (2014), which uses the SIPP, and Grigsby, Hurst and Yildirmaz (2019) (GHY), which uses ADP administrative payroll data.¹² Table 1 reports the frequency of nominal wage raises, freezes, and cuts at both a quarterly and annual frequency. GHY serves as the best benchmark for my post-Lasso persistent wage change estimates because they use administrative data from a large sample of U.S. firms’ payroll records for which they observe the workers’ true base wages. The only downside to using their results as a benchmark is that their sample includes only firms with 50+ workers. Since GHY

¹¹The post-Lasso procedure may identify a worker as receiving wage changes in two, back-to-back quarters if the true wage change occurred in the middle of the first quarter. In this case, many of these workers with back-to-back estimated wage changes would have received their true wage change in the month or two immediately before the mandated minimum wage change.

¹²Most studies of nominal wage rigidity examine annual changes in workers’ wages. This paper, on the other hand, focuses on quarter-over-quarter wage changes because the proposed quasi-experiment requires estimates of nominal wage changes at a sub-annual frequency to identify calendar quarters in which firms tend to raise wages.

find that smaller firms are less likely to raise workers’ wages at an annual frequency, the quarterly estimates for firms with 50+ workers overestimates the frequency of nominal wage raises in the broader population (and vice-versa for nominal wage freezes).

GHY find that 18.5 percent of workers at 50+ employee firms receive a nominal wage raise in any given quarter. The post-Lasso procedure identifies 26.5% fewer nominal wage raises, estimating that only 13.6 percent of workers receive a nominal raise each quarter. This difference is similar to the 30% Type II error rate upper bound from the earlier analysis of minimum-wage workers. Although it is not clear how much of the difference is due to GHY’s exclusion of small firms (small firms employed 28.2% of workers in 2014 according to the Census Bureau’s Quarterly Workforce Indicators), I believe that the post-Lasso estimation procedure fails to identify a non-trivial share of nominal wage raises.

3.3 Comparison of Estimated Persistent Nominal Wage Changes to Literature

The resulting post-Lasso estimated persistent base wages exhibit patterns very similar to the base wage change patterns identified by Grigsby, Hurst and Yildirmaz (2019) using the ADP administrative payroll records. Table 1 shows that the post-Lasso procedure estimates that 84.9% of workers experience no quarter-over-quarter persistent wage change, 13.6% receive a nominal raise, and 1.6% receive a nominal cut. Examining workers at firms with 50+ employees (which tend to raise workers’ wages more often than smaller firms), Grigsby, Hurst and Yildirmaz (2019) determine that 80.6% of workers experience quarter-over-quarter nominal wage freezes, 18.5% receive a nominal raise, and 0.9% receive a nominal cut.

Table 1: Comparison of Measures of Nominal Compensation Changes

Quarter-over-Quarter Change in Log Nominal Compensation Measure					
Compensation Measure	Data Source	Period	Raise	Freeze	Cut
Hourly Earnings					
Raw Data	LEHD 4 States	2011-2018	55.5%	5.0%	39.5%
No Hours Rounding	LEHD 4 States	2011-2018	46.4%	22.2%	31.4%
No Bonus + Rounding	LEHD 4 States	2011-2018	36.9%	45.3%	17.8%
Persistent Base Wage	Post-Lasso LEHD 4 States	2011-2018	13.6%	84.9%	1.6%
Base Wage	ADP Payroll (GHY 2019)	2008-2016	18.5%	80.6%	0.9%
Base Wage	SIPP Survey (BBG 2014)	1996-2000		78.4-84.8%	

Note: The Hourly Earnings measures are the quarter-over-quarter change in log hourly earnings. The *Raw Data* is calculated from the LEHD quarterly earnings and hours-paid data with no adjustments. The *No Hours Rounding* sets to zero any change that could be accounted for by hours rounding. The *No Bonus + Rounding* also smooths single-period bonuses in addition to correcting for hours rounding errors. See Appendix C for details on how each hourly earnings measure is constructed. *Persistent Base Wage* refers to the post-Lasso estimation results when controlling for the log estimated payday weeks using the secondary sample. Freezes are periods where the Lasso estimation procedure sets the wage change estimate to zero. The *Base Wage* estimates from ADP Payroll Records are from Grigsby, Hurst and Yildirmaz (2019). GHY examine payroll records from 2008-2017 for firms with 50+ workers who use ADP Payroll Services, reweighting the observations to match the firm characteristics of 50+ worker firms in the Census Bureau’s Longitudinal Business Database. The *Base Wage* estimates from the SIPP Survey are from Barattieri, Basu and Gottschalk (2014), which use linked SIPP records from 1996-2000. To address measurement error in self-reported wages, the BBG study employs the structural break identification procedure proposed by Gottschalk (2005) and then further corrects for the frequency of Type I and Type II errors. U.S. Census Bureau Disclosure Review Board bypass numbers DRB-B0037-CED-20190327 and CBDRB-2018-CDAR-061.

4 Patterns of Nominal Wage Adjustment

This section provides evidence regarding four patterns of nominal wage adjustment: i) the estimated persistent wage changes exhibit downward nominal wage rigidity, ii) real wage cuts that would be optimal in a frictionless environment are suppressed by downward nominal wage rigidity, iii) workers' nominal raises follow a Taylor-like pattern, with the probability of a wage raise spiking every four quarters, and iv) the timing of workers' annual raises are synchronized within the firm.

4.1 Wages Exhibit Downward Nominal Rigidity

Consistent with previous studies,¹³ the post-Lasso estimated persistent base wages exhibit significant downward nominal wage rigidity. Figure 1 shows the histogram of annual post-Lasso estimated nominal wage changes within ± 25 log points of zero. The histogram of estimated persistent base wage changes indicates that a large mass of workers have no change in their nominal wages year-over-year, and that the distribution of nominal wage changes is missing mass to the left of zero nominal change.

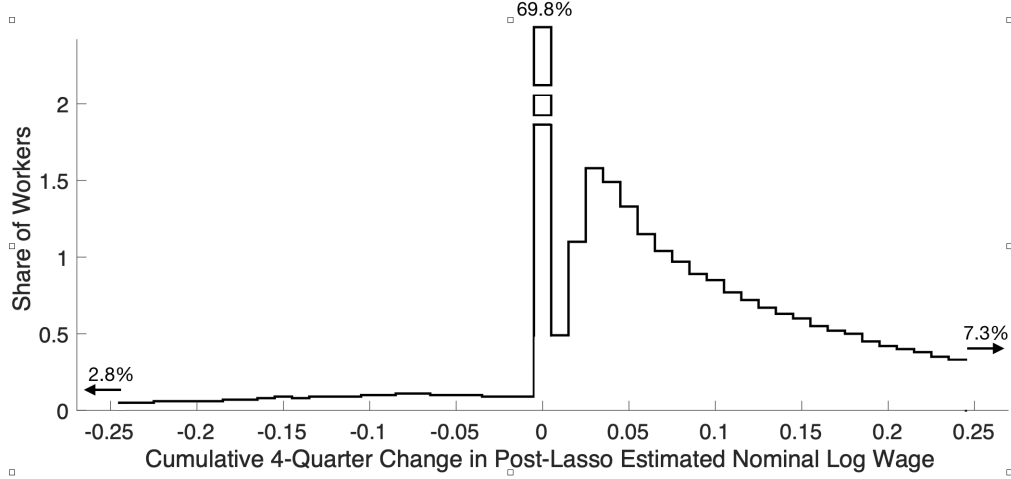
DNWR would imply that there is a discontinuous drop in the frequency of nominal wage adjustment immediately below zero nominal change. Such a discontinuity is apparent in Figure 2, which shows the histogram of nominal changes in 0.1 log point bins within 3 log points of zero nominal change. In Appendix A, I more formally test for the presence of DNWR using a series of regression discontinuity models that check for a discontinuity in the distribution of nominal wage changes at zero nominal change. For all model specifications (changing both bandwidths and polynomials in the running variable), I find evidence of a discontinuity in the distribution at zero nominal change - implying the existence of downward nominal wage rigidity.

4.2 Downward Nominal Wage Rigidity Suppresses Real Wage Changes

Given the finding that workers' persistent wage changes exhibit downward nominal wage rigidity, a natural question is how many real wage changes are suppressed when the change requires a nominal

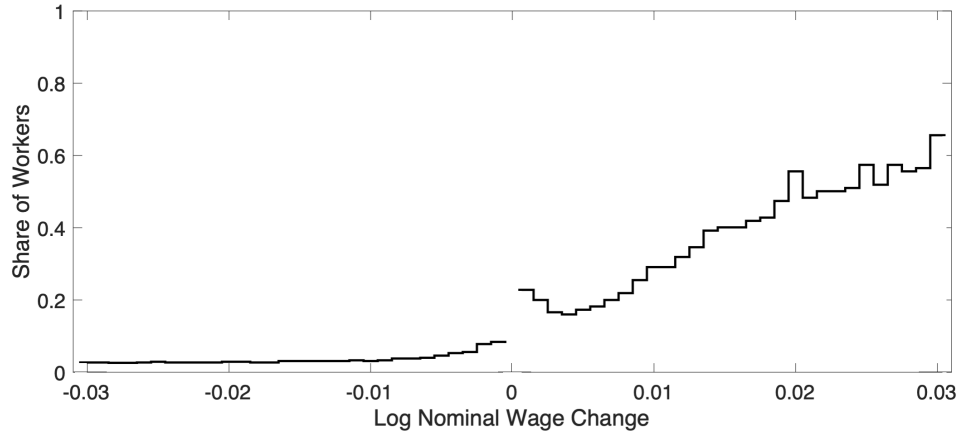
¹³Studies documenting downward nominal wage rigidity using the Panel Study of Income Dynamics (PSID) include: McLaughlin (1994); Lebow, Stockton and Wascher (1995); Akerlof, Dickens, Perry, Gordon and Mankiw (1996); Card and Hyslop (1997); Kahn (1997); Altonji and Devereux (2000); and Dickens, Goette, Groshen, Holden, Messina, Schweitzer, Turunen and Ward (2007). Studies using the Current Population Survey (CPS) include: Card and Hyslop (1997); Daly and Hobijn (2014); Elsby, Shin and Solon (2016); and Jo (2019). Studies using the Survey on Income and Program Participation (SIPP) include: Gottschalk (2005); and Barattieri, Basu and Gottschalk (2014). Studies using the Employer Cost Index (ECI) survey include: Lebow, Saks and Wilson (2003); and Fallick, Lettau and Wascher (2016). Studies using unemployment insurance administrative data include: Kurmann and McEntarfer (2019) and Jardim, Solon and Vigdor (2019). Grigsby, Hurst and Yildirmaz (2019) use ADP payroll data. Hazell and Taska (2018) using job posting data from Burning Glass.

Figure 1: Histogram of Cumulative 4-Quarter Change in Persistent Log Nominal Wage



Notes: Histogram of workers' four-quarter cumulative changes in their post-Lasso estimated persistent log nominal wage. Estimated from the Primary Sample after restricting to workers with at least five full quarters of non-zero earnings. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0037-CED-20190327.

Figure 2: Histogram of Persistent Nominal Wage Changes at Annual Frequency Near Zero



Notes: Frequency of four-quarter cumulative change in workers' post-Lasso estimated persistent log nominal wage grouped into 0.1 log point change bins. Sample is restricted to workers with at least five full-quarters of non-zero earnings. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0037-CED-20190327.

wage cut. This is a useful empirical moment for calibrating models with downward nominal wage rigidity. To estimate the suppression of real wage changes caused by downward nominal wage rigidity, I employ a variant of the method proposed by Kahn (1997) for measuring the effect of downward nominal rigidity on the wage change distribution.¹⁴

¹⁴There are two main differences between Kahn's proposed method and what I do. First, she proposes to use the median of the full wage change distribution, whereas I use the mode of the non-zero changes. I do this because

Kahn’s method compares the frequency of similar magnitude real wage changes across different periods, distinguishing between when the same real change corresponds to a positive versus negative nominal change. I begin by generating histograms of the nominal wage change distribution from different periods. The critical assumption of Kahn’s method is that the modal (or in her case median) nominal wage change in each period corresponds to the same optimal modal “real” wage change. Under this assumption, absent any rigidities or frictions, the proportion of wage changes in the histogram bins that are the same distance r from the period-specific modal nominal wage change bin should be the same across all periods.

To measure the degree to which DNWR suppresses real wage changes, I calculate p_{rt} - the proportion of all wage changes observed in period t (including nominal wage freezes) that fall into the r -distance bin from the modal nominal wage change bin for period t (where the distances are in 0.1 log points). If DNWR suppresses real wage changes, then, when a given r -distance bin requires a nominal wage cut, we should expect the proportion of wage changes in that bin to fall by some percent. Thus, I estimate the following regression model:

$$\ln(p_{rt}) = \sum_{x=-5}^5 \alpha_x d_{rt}^x + \beta_1 d^{\Delta w-} + \beta_2 d^{\Delta w, small+} + \epsilon_{rt} \quad (7)$$

d_{rt}^x is an indicator variable equal to one if $x = r$, which captures the assumption that, absent rigidities and frictions, the proportion of nominal wage changes in the r distance bin should be constant over time. $d^{\Delta w-}$ is an indicator variable equal to one if the r -distance bin in period t corresponds to a nominal wage cut, which captures the effect of downward nominal wage rigidity. And $d^{\Delta w, small+}$ is an indicator variable equal to one if the r -distance bin in period t corresponds to a small positive nominal raise (wage change bins between +0.1 and +0.9 log points), which identifies if small changes are suppressed. For this regression, I use the secondary LEHD sample (which enables me to use hours-paid data in the post-Lasso wage change estimation procedure). I calculate the proportion of nominal wage changes that fall into 0.1 log point bins of the quarterly nominal wage change distributions for each quarter from 2011:Q3 through 2017:Q3.

the infrequency of wage changes would mean that the median wage change is always zero (at both quarterly and annual frequencies). Using the mode of the non-zero changes provides a more consistent “real” wage change measure since the median of the non-zero changes will change significantly depending on the share of wage changes that are frozen. Second, she includes in the regression model described in Equation 7 the zero nominal change bin in each period, along with a build up from the suppressed bins with nominal wage cuts. Including the zero nominal change bin with this build up imposes a restriction that the suppressed nominal wage changes are necessarily wage freezes, whereas excluding the zero nominal change bin relaxes this restriction. By relaxing this assumption, I can estimate the relationship using OLS with a log specification, as opposed to requiring a non-linear estimation procedure.

As shown in Table 2, the proportion of wage changes in a given real change bin falls by 55% when the change requires a nominal wage cut versus a nominal wage raise. For comparison, this estimate is slightly above Kahn’s estimate using the PSID survey data that, when a given real change requires a nominal wage cut, DNWR suppresses 47.3% of hourly workers’ wage changes. The estimate is even further above Kahn’s estimate of 38.1% suppression of salaried workers’ wage changes, but within the 95% confidence interval of her estimate.

The finding that the likelihood of observing a real wage change falls significantly when the wage change requires a nominal cut has two implications for economic models that rely on assumptions about the wage adjustment process. First, that such a large share of wage changes are suppressed by DNWR lends empirical support to the various models that examine the role of DNWR in explaining the asymmetric response of employment and output to contractionary versus expansionary shocks (Kim and Ruge-Murcia (2009); Schmitt-Grohé and Uribe (2016); Evans (2018); Mineyama (2018), Dupraz, Nakamura and Steinsson (2019), and Chodorow-Reich and Wieland (forthcoming)).

Second, Calvo (1983) proposed modeling nominal wage rigidity as the random arrival of opportunities to adjust wages. One testable implication of the Calvo wage adjustment process (and the staggered wage adjustment process proposed by Taylor (1980)) is that the frequency of wage adjustments should not respond to the aggregate state. This implication no longer holds if the wage adjustment process is modified such that the opportunity to cut a worker’s nominal wage arrives less frequently than the opportunity to raise the worker’s wage. With this minor modification, the frequency of wage adjustments falls in response to negative aggregate shocks because a greater share of jobs will have an optimal wage that requires a nominal wage cut. Since nominal wage cuts are even less likely to occur relative to nominal wage raises, the frequency of wage adjustment will decrease disproportionately for a negative shock versus a similar magnitude positive shock - thus making “time-dependent” Calvo and Taylor wage setting processes also state-dependent.

Table 2 also indicates that the likelihood of observing a given real change rises when it requires a small positive nominal change (versus a large positive nominal change). This has two implications for models of the wage adjustment process. First, this finding is consistent with Elsby (2009), which argues that DNWR would generate downward compression in the distribution of positive wage changes since DNWR makes it harder to reverse wage raises. Second, this finding runs counter to the adjustment cost model of nominal wage changes proposed by Rotemberg (1982). If nominal wage changes incur adjustment costs (which could explain Taylor-style staggering of nominal wage

changes), then small nominal wage changes that are close to zero should be suppressed. Kahn’s estimation procedure allows for testing of the adjustment cost theory of wage changes by checking whether the proportion of wage changes that fall into a particular real change bin falls when that real change bin corresponds to a small positive versus a large positive nominal change. The results of the post-Lasso estimation procedure are the exact opposite of what we would expect with an adjustment cost model – specifically there are more wage changes in a given real bin when the wage change requires a small nominal raise versus a larger nominal raise. This finding is unlikely to be an artifact of the post-Lasso estimation procedure because the post-Lasso estimation procedure penalizes small wage changes with little explanatory power for a worker’s subsequent nominal wage – which implies that the post-Lasso is more likely to under-report small nominal wage changes.

Table 2: Suppression of Wage Changes Due to Downward Nominal Wage Rigidity

Log Proportion of Nominal Wage Changes in 0.1 Percentile Bin		
	(1)	(2)
Nominal Cut	-0.56*** (0.06)	-0.79*** (0.07)
Small Nominal Raise [0.1,1.0]	0.12** (0.04)	0.12** (0.04)
Small Nominal Cut [-1.0,-0.1]		0.24*** (0.04)
R-Squared	0.958	0.958
Observations	12,000	12,000

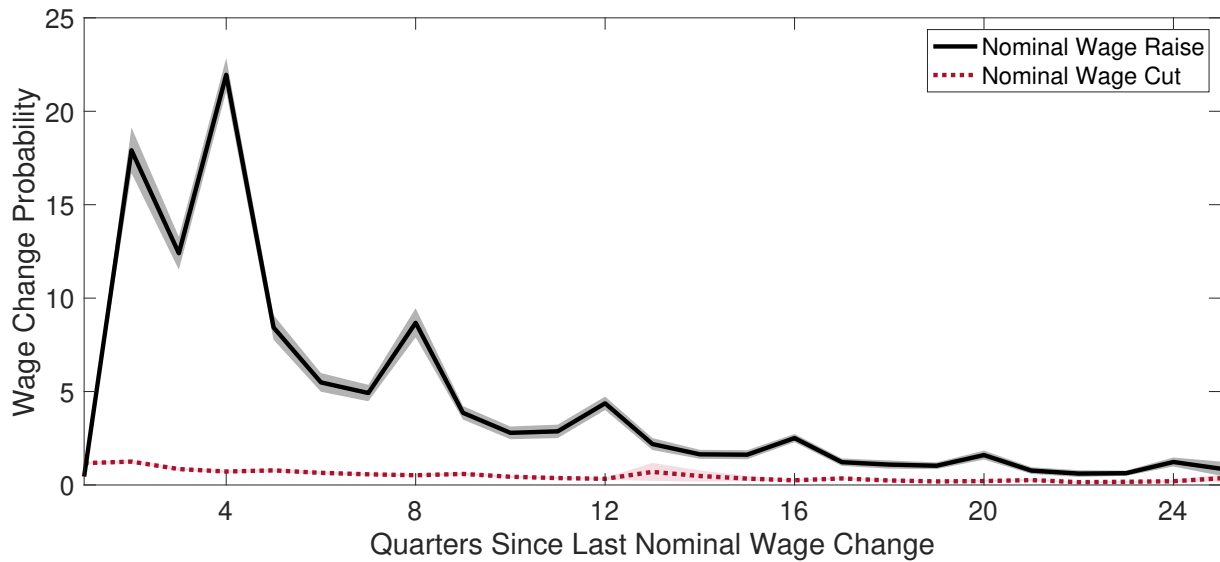
Note: Outcome variable is the log proportion of wage changes in a given nominal wage change bin that is r -distance from the modal nominal wage change bin (excluding wage freezes). Distinct nominal wage change distributions are generated for every quarter between 2011:Q3 and 2017:Q3 and all 0.1 percentage point bins between -5.0% and 5.0% (excluding 0.0) are included in the regression sample. Coefficient estimates correspond to the log change in the proportion of wage changes that fall within a given r -distance bin when the change requires a nominal cut, small nominal raise, or small nominal cut. ***, **, * indicate statistical significance at the 0.1%, 1.0%, and 5.0% levels, respectively. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

4.3 Annual Schedules of Nominal Raises

I find workers’ nominal wages in the LEHD data set broadly follow an annual schedule of nominal raises. This finding is consistent with Taylor (1980) which theorizes that workers receive wage

adjustments at regularly scheduled intervals. Figure 3 shows that a worker’s nominal wage raise hazard rate spikes four quarters after the worker’s last nominal wage change and every subsequent four-quarter anniversary. On the other hand, nominal wage cuts do not exhibit a Taylor-style annual schedule. The figure plots the coefficient estimates (and their 95% confidence intervals) from regressing an indicator variable equal to one if a worker receives a nominal wage raise on a set of dummy variables for the number of quarters since the worker’s last wage change (and similarly for nominal wage cuts). The result that nominal wage raises follow an annual adjustment schedule is consistent with the finding of Barattieri, Basu and Gottschalk (2014) that the wage change hazard rate in the SIPP spikes twelve months after the last wage change (although they did not find a similar pattern at subsequent annual anniversaries of the last wage change, and they did not separately examine wage cuts).

Figure 3: Probability of Wage Raise or Cut by Quarters Since Last Wage Change



Notes: Coefficient estimates from a linear regression model of the probability of an increase (raise) or decrease (cut) in the post-Lasso estimated nominal persistent base wage given the number of quarters since the worker’s last wage change. Shaded areas correspond to 95% confidence intervals using robust standard errors clustered at the SEIN level. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

That nominal raises exhibit a Taylor-style annual raise schedule while nominal cuts do not may have implications for asymmetries in the effectiveness of monetary policy. Dixon and Le Bihan (2012) show that Calvo and Taylor-style wage adjustment assumptions generate different output and employment responses to monetary policy shocks, with greater persistence in the response under the Calvo-style wage adjustment. Thus, if only nominal raises follow a Taylor-style adjustment

schedule, then the dynamics and persistence of responses to contractionary versus expansionary monetary policy shocks may also differ.

4.4 Annual Raise Schedules are Synchronized Within the Firm

Workers’ annual nominal raise schedules are synchronized within firms. A worker is twice as likely to receive a nominal wage raise in the firm’s “typical raise quarter” - the calendar quarter in which coworkers have historically tended to receive nominal wage raises at the firm.¹⁵ I identify typical raise quarters for the firms of 79.6% of workers. This is likely to be an underestimate of the prevalence of within-firm synchronization of annual raises because the procedure for identifying typical raise quarters is underpowered for firms with relatively few observed nominal wage changes. These typical raise quarters exhibit some seasonality, with the plurality of workers having Q3 as their typical raise quarter (26%), the fewest having Q2 (10%), approximately the same share in Q1 and Q4 (16% and 15% respectively), and 12.6% of workers having two typical raise quarters.

That nominal wage raises are more likely to occur in the second half of the year aligns with the hypothesis of Olivei and Tenreyro (2007), which used this fact to show that the effectiveness of monetary policy differs over the calendar year. Olivei and Tenreyro found when monetary policy shocks occur in the first half of the year, wages are slower to adjust and output responds more strongly to the shock.¹⁶

To determine whether workers’ annual schedules of wage raises are coordinated within firms, I test whether the probability that a worker receives a nominal raise can be predicted using the historical typical raise quarter of coworkers. I focus only on a typical raise quarter measure based on data from previous years so as to avoid the concern that firm-wide shocks generate contemporaneous correlation in coworkers’ raise frequencies. Thus, I estimate the following relationship using OLS:

$$\mathbb{1} [\Delta_{ikt}^w > 0] = \alpha d_{ikt}^{RQ} + \mathbf{X}_{ikt}\beta + \epsilon_{ikt} \quad (8)$$

where $\mathbb{1} [\Delta_{ikt}^w > 0]$ is an indicator variable equal to one if the worker receives a nominal wage raise;

¹⁵I classify a firm as having a particular calendar quarter as its “typical raise quarter” if two criteria are met. First, at least 33% of raises in previous years occurred in the given calendar quarter. Second, given the observed number of raises in this calendar quarter and all calendar quarters, I reject the null hypothesis that raises are randomly distributed with equal probability across the four calendar quarters (I use a one-sided hypothesis test at the 5% significance level for a binomial distribution with $p = 0.25$).

¹⁶A more recent study by Björklund, Carlsson and Nordström Skans (2019) examines the implications of seasonal nominal wage rigidity in Sweden and finds that monetary policy was more effective in periods when the staggered timing of union contracts meant that a larger share of workers had rigid nominal wages.

d_{ikt}^{RQ} is an indicator variable equal to one if the calendar quarter in t corresponds to the firm’s typical raise quarter for coworkers in previous years; and \mathbf{X}_{ikt} is a set of control variables that includes the worker’s age, tenure, quarters since their last wage change, and earnings quintile dummy variables.

The regression results shown in Table 3 indicate that the probability that a worker receives a nominal raise increases 9.3 percentage points in the firm’s typical raise quarter, more than doubling the baseline 7.3% probability that a worker receives a nominal raise in any given quarter. This implies that workers’ annual nominal raise schedules are strongly coordinated within the firm.

Table 3: Probability of Nominal Wage Change & Typical Raise Quarter

Dependent Variable:	Raise Probability			Cut Probability	
	(1)	(2)	(3)	(4)	(5)
Typical Raise Quarter	8.2*** (0.9)		9.3*** (0.9)		-0.3*** (0.03)
Baseline (Calendar Quarter I)		9.3*** (0.6)	7.3*** (0.8)	1.2*** (0.07)	1.1*** (0.10)
Calendar Quarter II		-0.24 (0.37)	-0.09 (0.32)	-0.18*** (0.04)	-0.18*** (0.04)
Calendar Quarter III		1.94 (1.14)	1.15 (0.81)	-0.28*** (0.03)	-0.20** (0.05)
Calendar Quarter IV		0.55 (0.32)	0.59 (0.28)	0.14 (0.07)	-0.09 (0.10)
Observations	17.4M	10.6M	10.6M	10.6M	10.6M

Note: Outcome variables are indicator variables equal to one if a worker has a nominal wage raise (1-3) or cut (4-5) in the quarter. Models 1, 3, and 5 include an indicator if the quarter qualifies as the firm’s typical raise quarter. Models 2-3 and 4-5 include a set of calendar quarter dummy variables (with the intercept representing the calendar quarter I). All models include as control variables: worker age, tenure, quarters since last wage change, and earnings quintile dummy variables. Robust standard errors clustered at the firm level. ***, **, * indicate statistical significance at the 0.1%, 1.0%, and 5.0% levels, respectively. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

5 Summary and Concluding Remarks

This paper presents a set of machine learning methods that expand the set of questions that can be answered using large employer-employee linked data sets. By identifying each worker’s unobserved persistent base wages, paydays weeks, and annual bonuses from the worker’s observed

quarterly earnings, the methods presented in this paper allows for the examination of questions related to wage rigidity, rent sharing, and the structure of worker compensation. I then implement and evaluate the quality of these machine learning methods using quarterly earnings data in the U.S. Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) dataset, an employer-employee linked dataset for the United States. Using the estimated nominal wages of workers in 30 U.S. states, I document four patterns of nominal wage adjustment: i) estimated persistent wage changes exhibit downward nominal wage rigidity, ii) when a nominal wage cut is required, approximately 55% of optimal real wage changes do not occur, iii) workers’ nominal raises follow a Taylor-like pattern, with the probability of a wage raise spiking every four quarters, and iv) the timing of workers’ annual raises are synchronized within firms.

References

- Akerlof, George A and Janet L Yellen**, “The fair wage-effort hypothesis and unemployment,” *The Quarterly Journal of Economics*, 1990, 105 (2), 255–283.
- , **William T Dickens, George L Perry, Robert J Gordon, and N Gregory Mankiw**, “The macroeconomics of low inflation,” *Brookings Papers on Economic Activity*, 1996, 1996 (1), 1–76.
- Altonji, Joseph G and Paul J Devereux**, “The extent and consequences of downward nominal wage rigidity,” in “Research in labor economics,” Emerald Group Publishing Limited, 2000, pp. 383–431.
- Bai, Jushan and Pierre Perron**, “Estimating and testing linear models with multiple structural changes,” *Econometrica*, 1998, pp. 47–78.
- Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk**, “Some evidence on the importance of sticky wages,” *American Economic Journal: Macroeconomics*, 2014, 6 (1), 70–101.
- Björklund, Maria, Mikael Carlsson, and Oskar Nordström Skans**, “Fixed-wage contracts and monetary non-neutrality,” *American Economic Journal: Macroeconomics*, 2019, 11 (2), 171–92.
- Burgess, Matt**, “How frequently do private businesses pay workers?,” Technical Report 11 2014.
- Calvo, Guillermo A**, “Staggered prices in a utility-maximizing framework,” *Journal of Monetary Economics*, 1983, 12 (3), 383–398.
- Card, David and Dean Hyslop**, “Does inflation” grease the wheels of the labor market”?,” in “Reducing inflation: Motivation and strategy,” University of Chicago Press, 1997, pp. 71–122.
- Casini, Alessandro and Pierre Perron**, “Structural breaks in time series,” *Oxford Research Encyclopedia of Economics and Finance*, forthcoming.
- Chodorow-Reich, Gabriel and Johannes Wieland**, “Secular labor reallocation and business cycles,” *Journal of Political Economy*, forthcoming.
- Ciuperca, Gabriela**, “Model selection by LASSO methods in a change-point model,” *Statistical Papers*, 2014, 55 (2), 349–374.

- Daly, Mary C and Bart Hobijn**, “Downward nominal wage rigidities bend the Phillips curve,” *Journal of Money, Credit and Banking*, 2014, 46 (S2), 51–93.
- Dickens, William T, Lorenz Goette, Erica L Groshen, Steinar Holden, Julian Messina, Mark E Schweitzer, Jarkko Turunen, and Melanie E Ward**, “How wages change: micro evidence from the International Wage Flexibility Project,” *Journal of Economic Perspectives*, 2007, 21 (2), 195–214.
- Dixon, Huw and Hervé Le Bihan**, “Generalised Taylor and generalised Calvo price and wage setting: micro-evidence with macro implications,” *The Economic Journal*, 2012, 122 (560), 532–554.
- Dupraz, Stephane, Emi Nakamura, and Jon Steinsson**, “A plucking model of business cycles,” 2019. Working Paper.
- Elsby, Michael W and Gary Solon**, “How prevalent is downward rigidity in nominal wages? International evidence from payroll records and pay slips,” *Journal of Economic Perspectives*, 2019, 33 (3), 185–201.
- Elsby, Michael WL**, “Evaluating the economic significance of downward nominal wage rigidity,” *Journal of Monetary Economics*, 2009, 56 (2), 154–169.
- , **Donggyun Shin, and Gary Solon**, “Wage adjustment in the Great Recession and other downturns: Evidence from the United States and Great Britain,” *Journal of Labor Economics*, 2016, 34 (S1), S249–S291.
- Evans, Christopher**, “Optimal monetary policy in the New Keynesian model with downward nominal wage rigidity,” 2018. Working paper.
- Fallick, Bruce, Michael Lettau, and William Wascher**, “Downward nominal wage rigidity in the United States during and after the Great Recession,” 2016. Working paper.
- Gottschalk, Peter**, “Downward nominal-wage flexibility: real or measurement error?,” *Review of Economics and Statistics*, 2005, 87 (3), 556–568.
- Grigsby, John, Erik Hurst, and Ahu Yildirmaz**, “Aggregate nominal wage adjustments: New evidence from administrative payroll data,” Technical Report, NBER Working Paper March 2019.

- Harchaoui, Zaid and Céline Lévy-Leduc**, “Multiple change-point estimation with a total variation penalty,” *Journal of the American Statistical Association*, 2010, *105* (492), 1480–1493.
- Hazell, Jonathon and Bledi Taska**, “Posted wage rigidity,” 2018. Working Paper.
- Jardim, Ekaterina S., Gary Solon, and Jacob L. Vigdor**, “How prevalent is downward rigidity in nominal wages? Evidence from payroll records in washington state,” Technical Report, National Bureau of Economic Research 2019.
- Jo, Yoon J**, “Downward nominal wage rigidity in the United States,” Technical Report 2019.
- Kahn, Shulamit**, “Evidence of nominal wage stickiness from microdata,” *The American Economic Review*, 1997, *87* (5), 993–1008.
- Kim, Jinill and Francisco J Ruge-Murcia**, “How much inflation is necessary to grease the wheels?,” *Journal of Monetary Economics*, 2009, *56* (3), 365–377.
- Kurmann, André and Erika McEntarfer**, “Downward wage rigidity in the United States: New evidence from administrative data,” Technical Report, CES Working paper 2019.
- Lebow, David, David Stockton, and William Wascher**, “Inflation, nominal wage rigidity, and the efficiency of labor markets,” Technical Report, Board of Governors of the Federal Reserve System-Finance and Economics Discussion Series no. 95-45 1995.
- Lebow, David E, Raven E Saks, and Beth Anne Wilson**, “Downward nominal wage rigidity: Evidence from the employment cost index,” *Advances in Macroeconomics*, 2003, *3* (1).
- McLaughlin, Kenneth J**, “Rigid wages?,” *Journal of Monetary Economics*, 1994, *34* (3), 383–414.
- Mineyama, Tomohide**, “Downward nominal wage rigidity and inflation dynamics during and after the Great Recession,” 2018. Working Paper.
- Olivei, Giovanni and Silvana Tenreyro**, “The timing of monetary policy shocks,” *American Economic Review*, 2007, *97* (3), 636–663.
- Rotemberg, Julio J**, “Sticky prices in the United States,” *Journal of Political Economy*, 1982, *90* (6), 1187–1211.

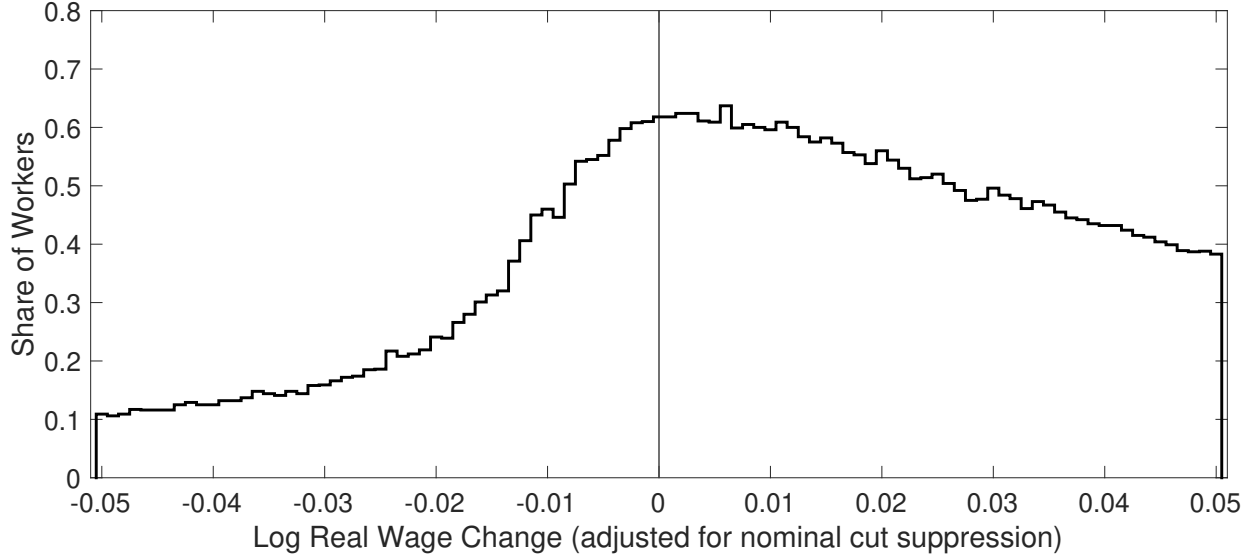
- Schmitt-Grohé, Stephanie and Martin Uribe**, “Downward nominal wage rigidity, currency pegs, and involuntary unemployment,” *Journal of Political Economy*, 2016, 124 (5), 1466–1514.
- Taylor, John B**, “Aggregate dynamics and staggered contracts,” *Journal of Political Economy*, 1980, 88 (1), 1–23.
- Tibshirani, Robert**, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58 (1), 267–288.

A RD Tests for Downward Real and Nominal Rigidity

To more formally test for the existence of a discontinuity at zero nominal change, I use a standard regression discontinuity specification to test for a sharp break in the proportion of workers receiving nominal wage changes immediately below, versus above, zero nominal change. (I exclude nominal wage freezes from this analysis because DNWR implies a discontinuity in realized nominal wage changes at zero nominal change.) As shown in Table 4, using a variety of polynomials in the running variable (nominal wage change) and bandwidths around zero nominal change, I always find a large and statistically significant break in the histogram of nominal wage changes at zero. This finding of a sharp break in the distribution of nominal wage changes at zero nominal change is consistent with the extensive literature on downward nominal wage rigidity.

I also evaluate whether there is any downward real rigidity after taking into account the downward nominal rigidity. This is of particular interest because the fair-wage theory of efficiency wages proposed by Akerlof and Yellen (1990) implies that workers are reluctant to accept wages that are below some reference wage. Any formulations of the “fair-wage” efficiency wage theory with the reference wage denoted in real terms would imply that there should be a discontinuity in the real wage change distribution at zero real wage change. To determine whether there is downward real wage rigidity after taking into account the nominal wage rigidity, I identify all instances where a worker’s nominal wage changes and then use the change in the Employment Cost Index between the current period and the period that the worker last received a wage change to calculate the real wage change. The resulting histogram of real wage changes in 0.1 log point bins is shown in Figure 4. Unlike the nominal wage change distribution, the histogram exhibits no apparent discontinuity at zero real wage change. As shown in Table 5, estimating a similar set of regression discontinuity models with various bandwidths and polynomials in the running variable delivers ambiguous results, confirming that there is no strong evidence of a discontinuity in the real wage change distribution at zero, and thus it is unlikely that there is any downward real wage rigidity after accounting for downward nominal wage rigidity.

Figure 4: Histogram of Persistent Real Wage Changes at Quarterly Frequency Near Zero



Notes: Frequency of post-Lasso estimated real persistent wage change grouped into 0.1 log point bins between -5.0% and 5.0%. The real magnitude of the change is calculated as the nominal change since the workers' last wage change. The nominal change is deflated using the BLS Employment Cost Index (ECI). U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

Table 4: Regression Discontinuity Test at Zero Nominal Wage Change

Polynomial Order	Nominal Wage Change Bandwidth Window		
	[-1.5%, 1.5%]	[-2.0%, 2.0%]	[-3.0%, 3.0%]
First	2.3* (0.90)	1.9* (0.77)	3.0*** (0.60)
Second	4.9*** (0.79)	3.9*** (0.88)	1.9* (0.93)
Third		5.5*** (1.10)	3.8** (1.13)

Note: Outcome variable is the share of workers with a nominal wage change within each 0.1 percentile bin of the nominal wage change distribution. The percentile bins are constructed from the post-Lasso estimation using hours-paid data from the secondary sample for each quarter from 2011:III to 2017:IV. The RD specification allows for distinct polynomials in the nominal wage above and below zero nominal change. ***, **, * indicate statistical significance at the 0.1%, 1.0%, and 5.0% levels, respectively. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

Table 5: Regression Discontinuity Test at Zero Real Wage Change

Polynomial Order	Real Wage Change Bandwidth Window		
	[-1.5%, 1.5%]	[-2.0%, 2.0%]	[-3.0%, 3.0%]
First	-3.4* (1.27)	-2.7* (1.01)	2.4 (1.36)
Second	-0.2 (1.56)	-2.9 (1.57)	-5.9*** (1.4)
Third		1.0 (1.88)	-2.4 (1.57)

Note: Outcome variable is the share of workers with a real wage change within each 0.1 percentile bin of the real wage change distribution (conditional on a nominal wage change). The real wage change distribution includes all nominal wage changes that are then converted into the real wage change using the change in the Employment Cost Index since the worker's last wage change. The percentile bins are constructed from the post-Lasso estimation using hours-paid data from the secondary sample for each quarter from 2011:III to 2017:IV. The RD specification allows for distinct polynomials in the real wage above and below zero real change. ***, **, * indicate statistical significance at the 0.1%, 1.0%, and 5.0% levels, respectively. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0069-CED-20190725.

B Permanent Versus Transitory Earnings Changes

If the quasi-experiment were to measure workers' compensation changes using changes in log earnings, then the transitory components of quarterly earnings would present two problems. First, the frequency of earnings changes makes it difficult to identify a firm's typical raise quarter. As shown in Table 1, even when controlling for quarterly hours paid, workers experience a change in log earnings almost every quarter. Only 5.0% of workers have no quarter-over-quarter change in log hourly earnings, whereas 55.5% receive a nominal raise and 39.5% receive a nominal cut.¹⁷ In contrast, Grigsby, Hurst and Yildirmaz (2019) use ADP payroll data to examine workers' base wages and they find that 80.6% of workers have no quarter-over-quarter change in their base nominal wage, 18.5% receive a nominal base wage raise, and only 0.9% receive a nominal cut. They further show both that the bonus component of hourly earnings drives much of this difference between hourly earnings and base wage changes and that bonuses exhibit little persistence. That workers' quarterly earnings change so often makes it difficult to use quarterly earnings changes to identify a dominant calendar quarter in which a firm tends to raise its workers' compensation.

Second, most fluctuations in quarterly earnings are unlikely to persist into future periods. As a result, historical patterns of workers' quarterly earnings changes are less predictive of firms' start-of-quarter worker compensation costs. To demonstrate the lack of persistence in workers' quarterly earnings changes, I evaluate the relative importance of permanent versus transitory earnings changes using the autocorrelation of workers' four-quarter change in log hourly earnings ($\Delta_{ik,t,t-4}^{yH}$).¹⁸ The four-quarter log hourly earnings change can be decomposed into the sum of the persistent quarterly changes in hourly earnings in $t-3$, $t-2$, $t-1$, and t ($\Delta_{ik,t-a}^P$), plus the transitory change in hourly earnings in t (Δ_{ikt}^T).

$$\Delta_{ik,t,t-4}^{yH} = \ln \left(\frac{y_{ikt}}{h_{ikt}} \right) - \ln \left(\frac{y_{ikt-4}}{h_{ikt-4}} \right) = \Delta_{ikt}^T + \Delta_{ikt}^P + \Delta_{ikt-1}^P + \Delta_{ikt-2}^P + \Delta_{ikt-3}^P \quad (9)$$

Notice that the four-quarter change and its one-period lag ($\Delta_{ik,t-1,t-5}^{yH}$) share three of these per-

¹⁷The relatively small number of quarter-over-quarter freezes in log hourly earnings is consistent with the findings of Kurmann and McEntarfer (2019) and Jardim, Solon and Vigdor (2019) regarding the frequency of four-quarter changes in log hourly earnings. Both studies use quarterly earnings and hours-paid data from Washington state to show hourly earnings exhibit far less year-over-year rigidity relative to the frequency of nominal wage freezes in the survey literature. Grigsby, Hurst and Yildirmaz (2019) uses ADP payroll data to confirm this large difference in the relative frequency of nominal changes in base wages versus hourly earnings. I find that even when controlling for measurement error due to rounding in hours paid and annual bonuses, the share of workers with no quarter-over-quarter change in log hourly earnings remains low at 22.2% (adjusting for rounding in hours paid) and 45.3% (adjusting for annual bonuses and rounding in hours paid).

¹⁸This uses data for workers in the four states with hours-paid data.

sistent components - namely Δ_{ikt-1}^P , Δ_{ikt-2}^P , and Δ_{ikt-3}^P . Assuming that the persistent change components are distributed iid, then the autocorrelation of $\Delta_{ik,t,t-4}^{yH}$ is 0.75 if there are no transitory changes. However, I find that the autocorrelation of the four-quarter change in log hourly earnings is only 0.284. Under a set of strong assumptions,¹⁹ I can estimate the relative magnitude of the permanent and transitory changes in hourly earnings as follows:

$$Corr\left(\Delta_{ik,t,t-4}^{yH}, \Delta_{ik,t-1,t-5}^{yH}\right) = \frac{\mathbb{E}\left[\Delta_{ik,t,t-4}^{yH} \Delta_{ik,t-1,t-5}^{yH}\right]}{Var\left(\Delta_{ik,t,t-4}^{yH}\right) Var\left(\Delta_{ik,t-1,t-5}^{yH}\right)} \quad (10)$$

Assuming that i) the persistent change components for each quarter follow the same iid distribution, ii) similarly, the transitory change components for each quarter follow an iid distribution, and iii) the persistent and transitory components are independent then

$$Corr\left(\Delta_{ik,t,t-4}^{yH}, \Delta_{ik,t-1,t-5}^{yH}\right) = \frac{3Var\left(\Delta_{ik}^P\right)}{4Var\left(\Delta_{ik}^P\right) + Var\left(\Delta_{ik}^T\right)} \quad (11)$$

This implies that the relative variation in quarterly earnings from the transitory versus the persistent changes is:

$$\frac{Var\left(\Delta_{ik}^T\right)}{Var\left(\Delta_{ik}^P\right)} = \frac{3 - 4Corr\left(\Delta_{ik,t,t-4}^{yH}, \Delta_{ik,t-1,t-5}^{yH}\right)}{Corr\left(\Delta_{ik,t,t-4}^{yH}, \Delta_{ik,t-1,t-5}^{yH}\right)} \quad (12)$$

Thus, the autocorrelation estimate implies that the variance of the transitory component of the quarterly measure of hourly earnings is 6.6 times greater than the variance of the permanent component. This indicates that transitory changes account for 86% of the quarter-over-quarter fluctuations in hourly earnings.

It is reassuring to note that the post-Lasso procedure identifies persistent wage changes as occurring in periods in which this autocorrelation measure indicates that log earnings changes are more persistent. Using the same metric of the autocorrelation of the four-quarter change in log hourly earnings, I find that the autocorrelation is 0.512 if the post-Lasso procedure detects a wage change in $t - 1$, whereas it is only 0.123 if no wage change is detected by the post-Lasso procedure in $t - 1$. The estimated wage changes from the post-Lasso procedure are much more likely to persist, and thus affect the firms' start-of-quarter wage bills in future periods.

¹⁹Namely the three assumptions are i) that persistent change components are distributed iid, ii) that temporary change components are distributed iid, and iii) that the temporary and permanent change components are independent within any 5-quarter window.

C Measures of Wage Compensation

For every worker i , firm k , and quarter t combination, the LEHD data set provides either: i) quarterly earnings (primary sample), or ii) both quarterly earnings (y_{ikt}^Q) and quarterly hours paid (h_{ikt}^Q) (secondary sample). For simplicity, I begin by describing various measures of wage compensation that can be constructed when quarterly hours paid is observed.

The literature on nominal wage rigidity has tended to focus on rigidity in workers' base wage (w_{ikt}). However, with the LEHD data, I observe workers' quarterly earnings (y_{ikt}^Q) or their quarterly-averaged hourly earnings (\bar{y}_{ikt}^H , hereafter *hourly earnings*). Two recent working papers, Kurmann and McEntarfer (2019) and Jardim, Solon and Vigdor (2019), use administrative UI records data to explore the degree of nominal rigidity in hourly earnings. Both studies find hourly earnings are much less rigid than base wages.²⁰ Thus, it will be useful to deconstruct the relationship between these two measures of wage compensation. First, note that the quarterly earnings measure can be decomposed as:

$$y_{ikt}^Q = w_{ikt} \left(n_{ikt} \bar{h}_{ikt}^W + \frac{1}{2} n_{ikt} \bar{h}_{ikt}^o \right) v_{ikt} \epsilon_{ikt} \quad (13)$$

where n_{ikt} is the number of payroll weeks in quarter t , \bar{h}_{ikt}^W is the average number of hours worker per week, \bar{h}_{ikt}^o is the average number of overtime hours worked per week, v_{ikt} is any non-overtime variable compensation paid in period t , and ϵ_{ikt} captures measurement error (e.g. the rounding of hours worked to integer digits or order-of-magnitude errors in hours worked). As I show in Appendix 3.1, including the number of payroll weeks in the quarter proves to be quite useful since, depending on the payroll schedule in effect at the firm, the number of payday weeks can fluctuate among 12, 13, or 14 weeks from one quarter to the next. These fluctuations in payday weeks from the payday schedules result in substantial variation in quarterly earnings ($\pm 7 - 15\%$) simply.

Since overtime pay only applies to hours worked in excess of 40 hours per week, I will approximate total quarterly overtime hours as:

$$n_{ikt} \bar{h}_{ikt}^o = \max [0, n_{ikt} (\bar{h}_{ikt}^W - 40)] \quad (14)$$

This approximation and the decomposition of quarterly earnings in Equation 13 imply that a

²⁰This finding of greater nominal wage flexibility in worker earnings (relative to findings from surveys and administrative payroll records) is echoed in the survey of Elsby and Solon (2019). They consolidate findings of many recent international studies of nominal wage rigidity and argue that about 15-25% of workers receive year-over-year nominal wage cuts.

worker's base wage is related to her hourly earnings as follows:

$$y_{ikt}^H = y_{ikt}^Q / h_{ikt}^Q \approx w_{ikt} \left(1 + \frac{1}{2} \frac{\max \left[0, h_{ikt}^Q - 40n_{ikt} \right]}{h_{ikt}^Q} \right) \frac{v_{ikt}\epsilon_{ikt}}{h_{ikt}^Q} \quad (15)$$

It is evident from this decomposition that difference in the degree of rigidity between measures of workers' hourly earnings and their base wages could come from three potential sources: overtime compensation ($\max \left[0, h_{ikt}^Q - 40n_{ikt} \right]$), non-overtime variable compensation (v_{ikt}), or measurement error (ϵ_{ikt}). As to the relevance of these three sources of variation for measuring true rigidity in wage compensation, I discount the importance of fluctuations in hourly earnings caused by measurement error and overtime compensation. Fluctuations in hourly earnings caused by measurement error are simply spurious. Fluctuations in hourly earnings due to changes in overtime compensation are unrelated to the persistence of the worker's base wage, but instead reflect a temporary change in the worker's utilization. Thus, it will be useful to explore the relative importance of each of these three sources in the degree of measured rigidity in hourly earnings.

Table 6 shows the proportion of quarter-over-quarter ($\ln(\bar{y}_{ikt}^H) - \ln(\bar{y}_{ikt-1}^H)$) and four-quarter ($\ln(\bar{y}_{ikt}^H) - \ln(\bar{y}_{ikt-4}^H)$) raises / freezes/ cuts in log hourly earnings for individuals who are employed for the full-quarter in both periods (i.e. they were employed at both the start and the end of the quarter at the same firm). The log hourly earnings exhibit significant variability, with only 5.0% of workers having the same hourly earnings from one quarter to the next.

C.1 Measurement Error: Rounding in Hours Paid

Although I cannot identify all instances of measurement error in hourly earnings changes, I can control for changes in hourly earnings that may be due to state unemployment insurance agencies' instructions that firms report hours paid after rounding them to whole numbers. This instruction has significant implications for the frequency of wage freezes. To measure the impact of this rounding rule, I set to zero change all observed changes in log hourly earnings that could be due to reported hours paid being misreported by ± 1 in the current and/or the lagged quarter. This provides an upper-bound on the degree to which flexibility in hourly earnings could be due to measurement error from rounding hours paid to whole numbers. I find that resetting to zero any hourly earnings change that could be due to hours-rounding causes the frequency of quarter-over-quarter wage freezes to increase more than four-fold, from 5.0% to 22.2%.

C.2 Overtime Compensation

Although the hours paid reported in the LEHD data set makes no distinction between overtime hours and regular pay hours, I use the following method to identify earnings changes that could be due to overtime pay. First, if a worker is reported as having worked 480 or fewer hours in the quarter ($h_{ikt}^Q \leq 480$), then I do not consider whether they worked any overtime because even with the fewest number of payday weeks (12), they could still have attained this number of hours without working any overtime.²¹ Focusing, instead, on quarters in which a worker had 481 or more hours paid ($h_{ikt}^Q > 480$), I consider three alternative numbers of payday weeks ($n_{ikt} \in (12, 13, 14)$). Since overtime hours can be approximated as $\max[0, h_{ikt}^Q - 40n_{ikt}]$, each of these three scenarios corresponds to a different number of overtime hours worked. I conclude that overtime hours have been worked if a particular overtime adjustment results in the hourly pay in period t being within 3 cents of the hourly earnings in at least three of the surrounding four quarters (allowing for adjustments to the hourly earnings in the surrounding quarters for overtime pay).

C.3 Variable Compensation: Annual Bonuses

A worker's non-overtime variable compensation, v_{ikt} , can include tips, commissions, and bonuses. I describe here a method for estimating one form of variable compensation: annual bonuses. An annual bonus that occurs in a particular quarter will appear as a single-quarter spike in hourly earnings. Thus, I identify quarters in which a worker received an annual bonus as any quarter in which $y_{ikt}^H > \max[y_{ikt-1}^H, y_{ikt+1}^H]$. For any such period, I construct an estimate of the bonus as:

$$\hat{b}_{ikt} = \frac{y_{ikt}^H}{\max[y_{ikt-1}^H, y_{ikt+1}^H]} h_{ikt}^Q = \frac{w_{ikt} \left(1 + \frac{1}{2} \frac{\bar{h}_{ikt}^o}{40}\right) \frac{\epsilon_{ikt}}{h_{ikt}^Q}}{\max_{s=t-1, t+1} w_{iks} \left(1 + \frac{1}{2} \frac{\bar{h}_{iks}^o}{40}\right) \frac{\epsilon_{iks}}{h_{iks}^Q}} b_{ikt} \quad (16)$$

The estimated \hat{b}_{ikt} is an accurate measure of the true annual bonus, b_{ikt} , in cases where there is no measurement error or overtime hours worked in either the annual bonus period or the comparison period and the base wage is constant across periods. When I exclude the estimated annual bonuses from the measure of hourly earnings and set changes to zero when they are potential hours rounding errors, the frequency of quarter-over-quarter freezes in log hourly earnings increases to 45.3% from 5% in the raw hourly earnings measure (and from 22.2% in the hourly earnings that exclude

²¹This assumption misses some overtime hours since overtime hours are calculated on a weekly basis. Thus, it is possible for a worker to work more than 40 hours in one week (thus receiving overtime pay) and fewer in another week of the same quarter, and yet still have fewer than 481 hours of pay in the quarter.

potential rounding errors).

Table 6: Hourly Earnings Change Comparison

Quarter-over-Quarter Nominal Log Hourly Earnings Change						
	Source	Period	Raise	Freeze	Cut	
Raw	LEHD 4 States	2011-2018	55.5%	5.0%	39.5%	
Rounding Adj	LEHD 4 States	2011-2018	46.4%	22.2%	31.4%	
Bonus & Rounding Adj	LEHD 4 States	2011-2018	36.9%	45.3%	17.8%	
4-Quarter Log Nominal Hourly Earnings Change						
	Source	Period	Raise	Freeze	Cut	
Rounding Adj	LEHD 4 States	2011-2018	68.1%	11.8%	20.1%	
Kurmann and McEntarfer (2019)	LEHD WA State	1990-2014		8-16%	20-25%	
Jardim, Solon and Vigdor (2019)	UI WA State	2005-2015		2.5-7.7%	20.4-33.1%	
Grigsby, Hurst and Yildirmaz (2019)	ADP 50+ Workers	2008-2016	75.3%	9.0%	15.7%	

Note: ‘Raw’ indicates log change calculated from the original quarterly earnings and hours paid. ‘Rounding Adj’ sets to zero any nominal change that could be explained by adjusting the hours paid by ± 1 hour to account for potential rounding errors. ‘Bonus & Rounding Adj’ first smooths bonuses in the hourly earnings and then sets to zero any nominal change that could be accounted for by a ± 1 change in hours paid. U.S. Census Bureau Disclosure Review Board bypass number DRB-B0037-CED-20190327.