Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 2233
November 2026

Union earnings gaps and membership dynamics

Henry S. Farber
Stephen Machin

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UKRI Economic and Social Research Council

Abstract
This paper revisits estimation of the union-nonunion earnings gap based on the experience of job losers in the 1994-2024 Displaced Workers Survey (DWS). A central focus is placed on earnings changes associated with union status changes for full-time private sector job losers who find a new full-time private sector job. An important feature of the analysis is that we directly account for classification error in union status in order to correct the attenuation of estimates of the earnings gap. This is particularly serious when measuring earnings changes associated with union status changes in longitudinal data. Two results are highlighted. First, in the low union environment in the U.S., losers of union jobs are surprisingly likely to be union members in their new job (on the order of 60 percent). Second, estimated difference-in-difference union earnings gaps that account for classification error are substantial but asymmetric among union joiners and leavers. Union leavers suffer a substantial earnings loss relative to union stayers of about 28 percent, while union joiners enjoy a 19 percent earnings gain relative to those who remain nonunion.

This paper was produced as part of the Labour Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We are grateful to participants at the CReAM annual workshop, the IFS Human Capital and Labor Markets seminar, the Helsinki Wages, Employment, and Inequality III workshop, and the conference Exploring Frontiers in Labor Economics in honor of Christian Dustmann at Humboldt University for comments and suggestions. Farber thanks the Princeton University Industrial Relations Section for funding. Machin thanks the UK Economic and Social Research Council for funding.

Henry S. Farber, Princeton University and NBER. Stephen Machin, Department of Economics and Centre for Economic Performance at the London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© H. S. Farber and S. Machin, submitted 2026.

DOI: 10.21953/researchonline.lse.ac.uk.00142142

1 Introduction

A cornerstone question extensively studied by labor economists has been how much more union workers earn relative to what they would earn in a nonunion job. The research methodology to estimate the union differential has developed very significantly over time, and a large literature has consistently produced evidence that, in a wide range of settings, union workers earn more than otherwise equivalent nonunion workers.[1] This has been true in very early work from over sixty years ago using aggregate data, in subsequent evidence based on individual-level cross-section data, and in more up to date studies using longitudinal data following the same people over time. Lewis's (1963) classic study of union relative wage effects, which relied on aggregated data (e.g., at the industry level) set the stage for work on this problem. Lewis (1986) updated this earlier survey to incorporate work in the intervening years that exploited the newly available individual cross-section data on union status and wages from household and individual surveys. Interestingly, the findings using both aggregate and individual data were consistent in finding a 15-25 percent union wage advantage.

Research has clearly recognized that these estimates do not reflect the wage differential that unions obtain for their members owing to concerns around selection on both observable and unobservable factors that determine who is and who is not in a union job. The research on queues for union jobs (Abowd and Farber, 1982; Farber, 1983) reflects this, as does work by Card (1996) accounting for selection on observables where workers at different points of the skill distribution – measured in his study by a predicted non-union wage – are differentially likely to be working on a union job. Selection on unobservables has also been addressed by looking at wage changes for individuals who change union status in longitudinal data (Card, 1996; Chowdhury and Nickell, 1985; Freeman, 1984; Jakubson, 1991; Kulkarni and Hirsch, 2021; Mellow, 1981). More recent analyses of earnings changes using longitudinal administrative data generally cannot address the union-nonunion earnings gap due to a lack

[1] Note that the presence of unions may not only result in higher wages for unionized workers but also induce nonunion employers to offer higher wages and improved working conditions in order to discourage union organization (the threat effect). For analyses of the threat effect, see, for example, Rosen (1969), Farber (2005), and Fortin, Lemieux, and Lloyd (2023). We do not address the threat effect here.

of information on union status at the individual level.[2]

We extend this earlier work to focus on the earnings changes that accompany changes in union status. The analysis uses longitudinal data on re-employed job losers in the U.S. Displaced Workers Survey (DWS) from 1994-2024. Identification comes from a research design based on involuntary job changers for which we observe earnings and union status in the old and new job.[3]

The focus on involuntary job losers allows us to study in detail union status transitions and the nature of union membership dynamics so as to avoid some of the selection issues that arise when using union status transitions that are the result of a voluntary job change or a change of union status without a change in employer. Workers who change jobs voluntarily do not change jobs and union status randomly. The presence of a union wage advantage means 1) that a union worker will quit to take a non-union job only if the new job has higher wages and 2) that a worker's decision to take a new job, union or non-union, will depend on the wage it pays compared to the initial job.[4] Since job losers do not change jobs voluntarily, we avoid this particular selection problem.[5]

Like some of the other work in the area, where classification errors even of relatively small frequency have been shown to affect union wage gap estimates quite markedly (Card, 1996; Freeman, 1984; Mellow and Sider, 1983), we pay particular attention to correcting for classification error in reported union status. This plays a particularly important role in deriving consistent estimates of wage changes associated with true union status changes in longitudinal data. This is because the low underlying unionization rate in the United States together with even a modest amount of classification error implies that a substantial fraction

---

[2] An important exception is a recent contribution by Beauregard et al (2025), who use Canadian administrative data with information on union status and earnings at the individual level over time to estimate union earnings effects and connections to inter-firm earnings gaps.

[3] This is in contrast to much of the earlier work using longitudinal data which has been unable to condition on job change, voluntary or involuntary, in matched Current Population Survey (CPS) and other similarly structured data. There are a few exceptions, including Raphael (2000) and, more recently, Kulkarni and Hirsch (2021). Both of these studies, like ours, use the DWS data to focus on involuntary job changers. Another example, using Canadian data to focus on displaced workers, is Kuhn and Sweetman (1998).

[4] Of course, we recognize that there are characteristics of jobs other than wages that are relevant to employment and mobility decisions. We focus on earnings here to make the key points clearly.

[5] Selection issues related to the locus of job loss and its relationship to union status remains.

of observed union status changes are spurious. Given an assumed rate of misclassification, we address classification error directly in a univariate analysis by deriving the "true" cross-section unionization rate and the "true" longitudinal union status transition rates and use these to calculate the implied "true" union-nonunion earnings gap.[6] We extend the analysis to a multivariate setting, allowing control for relevant worker characteristics, by specifying a maximum likelihood model based on the assumed misclassification rate. This approach yields estimates that adjust for the attenuation bias induced by classification error and results in substantially larger estimates of the earnings gap than found in much of the earlier literature. The findings of this study also connect to the large literature on earnings changes for displaced workers, a literature that very rarely considers union status in the study of earnings losses from displacement, an omission that our findings show to be important.[7]

The bottom line finding is that union representation is highly valuable in terms of earnings, in spite of unionization levels in the U.S. being quite low by international and historical standards. There are two important facets to this. One emerges from a new finding that union status tends to be highly persistent, as displaced union workers are substantially more likely than displaced nonunion workers to be in a union job when they are re-employed. Those union job losers who are successful in finding a new union job do not suffer substantial earnings losses, but those union job losers who are reemployed in a nonunion job experience large earnings declines. We also find that only a very small fraction of nonunion job losers are reemployed in a union job. Although these "lucky" union joiners do experience an earnings gain, it is much smaller in magnitude than the earnings loss suffered by union leavers. These dynamic aspects of transitions between union and non-union status and the earnings changes they generate are among the key findings reported in the paper, and they refine and add to what we know from the previous longitudinal estimates of union differentials in the United States.

---

[6] This assumed rate of misclassification is, like Card (1996), based on the 1977 validation survey of the CPS analyzed by Mellow and Sider (1983). We describe what we did in more detail below.

[7] For example, there is no mention of unions and any connection to the costs of job loss at all in Couch and Placzek's (2010) review of some of the best known studies of earnings loss following displacement. Closer to home, Farber (2017) does not consider union status in his analysis of the consequences of job loss based on the DWS. The very recent piece by Athey et al (2026) that focuses on heterogeneity in the earnings impact of job loss across workers, establishments, and markets has no role for union status.

The remainder of the paper is structured as follows. Section 2 describes the DWS data. Section 3 contains the analysis of union status and earnings in the cross-section accounting for classification error in union status. Section 4 extends this analysis to longitudinal data from the DWS. The results suggest that a substantial fraction of the union status changes measured in the DWS are spurious due to classification error, resulting in a large downward bias in the estimated earnings gap based on observed union joiners and union leavers. Section 5 contains multivariate analyses of the earnings gap in both the cross-section and longitudinal settings that account for classification error controlling for observable differences across workers. In order to accomplish this, we develop a maximum-likelihood estimator for the joint determination of union status and earnings. Estimation of this model allows us to recover union status and transition probabilities and the union earnings gap accounting for classification error. Our analysis here confirms that classification error in union status substantially attenuates both cross-section and longitudinal estimates of the union-nonunion earnings gap. Section 6 introduces the longitudinal (matched year-to-year) CPS MORG data of the type used by earlier longitudinal analyses and compares estimates using these data to our estimates using the DWS, uncovering important differences. Section 7 provides further analysis to help interpret the estimates of the earnings gap. These include: 1) analysis of the persistence of union status among job losers and 2) evidence that the cross-section earnings gap (on both the lost job and the new job) is strongly related to union status in both periods. Section 8 concludes.

2 Data

Our primary source of data is the Displaced Workers Survey (DWS, a CPS supplement), which focuses on the experience of involuntary job changers. This focus addresses at least part of the selection problem due to the choices of workers who stay with the same employer or are voluntary job changers. The DWS is fielded bi-annually early in even years, and we use the data from the 1994-2024 surveys.[8] The DWS includes individuals who report having

[8] The DWS started in 1984, but data on union status on the lost job were not collected until the 1994 survey.

lost a job in the three calendar years prior to the (January or February) survey date and includes the usual CPS information along with information about the lost job.[9]

Information is collected on weekly earnings for both the lost job and the new job held at the survey date. No information is collected on hours of work, beyond asking whether that job was full-time or part-time. While weekly hours worked at the survey date are collected in the core CPS, there is no source of hours on the lost job. For this reason, we use the available weekly earnings measure and focus on the union earnings gap rather than the union hourly wage gap. Given the lack of hours data on the lost job, we limit our sample to workers who report losing a full-time job and are reemployed on a full-time job (defined as weekly hours ≥ 35).

We note there is a recent analysis (Blanchflower and Bryson, 2025) suggesting that union workers have longer working hours than do nonunion workers. To the extent that this is the case in our data, the union weekly earnings gap will be larger than the hourly union wage gap. However, a simple analysis (not shown here) of data from the MORG CPS that we describe and use later, shows that among full-time workers there is essentially no "hours gap" but that union workers are more likely than nonunion workers to be full-time. In other words, the finding of higher hours for union workers is driven by the fact that union workers are less likely to be part-time and not by higher union hours among full-time workers. We conclude that for our analysis of full-time workers, the weekly union earnings gap is a good proxy for the union hourly wage gap.

While union status on the lost job is observed for all job losers, union status on the new job is observed only for those job losers in the outgoing rotation groups (months 4 and 8) in the DWS survey month. For those in rotation groups 1-3 and 5-7 in the DWS survey month, we match forward month-by-month to find these job losers in an outgoing rotation group with information on union status.

The DWS sample for analyzing union status transitions was developed with the following

[9] Job losers are asked the reason for losing their job, and 6 responses are recorded: 1) plant closing, 2) slack work, 3) position abolished, 4) seasonal job ended, 5) self-employment ended, and 6) other. Detailed information (including information on union status on the lost job) is collected only for those who report losing the job due to reasons 1-3, and we restrict our analysis to these job losers.

criteria:

- Reemployed job losers aged 20-64 who make a full-time to full-time transition.
- Employed in the private sector on the old job and the new job.
- Non-missing information on union status on the lost job.[10]
- Non-missing and non-allocated information on union status on the current job.[11]
- Non-missing information on a standard set of observable demographic characteristics (age, education, sex, marital status, race) and industry of employment of both the lost job and the current job.

The resulting sample for the analysis of union status transitions contains 15368 job losers in the DWS between 1994 and 2024. We also create two cross-section subsamples from the DWS data. One has the full-time private sector observations on the lost job (sample size, n=18957). The other has the full-time private sector observations on the new job (n=16975).

When analyzing earnings, we follow the selection rules outlined above and include observations with non-missing information on weekly earnings on the lost job and weekly earnings on the new job.[12] We follow Autor, Katz, and Kearney (2008) in adjusting the CPS earnings data for top-coding and trimming observations with very low weekly earnings. Top-coded weekly earnings data are multiplied by 1.5 (0.4055 added to log weekly earnings). This affects about 3.5 percent of earnings change observations and about 2.1 percent of cross-section earnings observations. Observations with weekly earnings less than $67 in 1984 dollars (assuming 40 hours of work corresponding to a real hourly wage less than $1.675) are flagged as low observations to be dropped. A "low flag" on either wage results in the observation being

[10] Note that while we would prefer to use "membership or coverage by a collective bargaining agreement" as our measure of union status, we use union membership alone as our measure of union status because coverage information for the lost job is not available in the DWS.

[11] There is no allocation flag for union status on the lost job.

[12] The DWS documentation notes that weekly earnings on the current job could be allocated, but there is no allocation flag provided. There is no such note for earnings on the lost job. It is well known that allocated earnings do not adequately account for union status, tending to bias the estimate of the union wage gap downward.

dropped from the longitudinal analysis of wage changes (173 observations). For cross-section samples, we drop the observations based on the relevant "low flag" (136 observations on the old wage, 124 observations on the new wage). The resulting sizes of estimation samples are 15195 for wage changes, 18821 for the old wage, and 16851 for the new wage.

3 Union Density and the Earnings Gap in the Cross-Section: Accounting for Classification Error

In the low-unionization environment of the current U.S. labor market, classification error in recorded union status in surveys can lead to substantial over-estimation of the cross-section unionization rate. Classification error also implies substantial underestimates of the cross-section union earnings gap. In this section, we introduce and discuss a model of classification error in the cross-section, based on the work of Mellow and Sider (1983) and Card (1996), that has these clear implications. We explore the model's implications for estimation of union density and the union wage gap in the cross-section, and we propose a maximum likelihood approach to estimation of models that account for such classification error.

3.1 Union Density and Classification Error in the Cross-Section

Let U_i be an indicator of the true union status of a worker at a point in time, and denote the true probability that a worker is unionized by γ. We observe union status (U_i^*), where the probability of misclassification is λ and does not depend on true union status.[13] We define the observed unionization rate as γ^*. In this case,

$$P(U_i^* = 1) = \gamma^* = \gamma(1-\lambda) + (1-\gamma)\lambda, \tag{1}$$

where the first term comes from the correct reports of union status by union workers and the second term comes from the incorrect reports of union status by nonunion workers. Rearranging the terms in equation 1, yields

$$\gamma^* = \gamma + (1-2\gamma)\lambda. \tag{2}$$

[13] Card (1996) presents evidence that characterization of the classification error rate as independent of true union status is consistent with the CPS data used by Mellow and Sider (1983).

If $\gamma < 0.5$, as is the case in all data we are considering for the relatively low unionization setting of the contemporary U.S., there are more nonunion workers who make the mistake of reporting they are union than there are union workers who make the mistake of reporting they are nonunion. The result is that γ^* is an upward biased estimate of the unionization rate (γ). Based on equation 2, the bias is

$$\gamma^* - \gamma = (1 - 2\gamma)\lambda., \tag{3}$$

and the true unionization rate as function of the observed unionization rate is

$$\gamma = \frac{\gamma^* - \lambda}{1 - 2\lambda}. \tag{4}$$

A commonly used estimate of the union misclassification rate is based on Card's (1996) analysis of the 1977 CPS data first analyzed by Mellow and Sider (1983).[14] These data contain, for a subset of observations, both worker and employer reports of union status and other measures including hours and wages. The extent of disagreement in reports of union status provides evidence on the rate of misclassification. Card's analysis suggests that the data are consistent with characterizing each report of union status as subject to misclassification at the same rate and that the probability of misclassification is independent of true union status. A simplified approach to this analysis is to recognize that the probability that there is disagreement in reported union status can occur in two directions with equal probability independent of the true union density. A worker report of U and an employer report of N occurs with probability $\lambda(1 - \lambda)$. And a worker report of N and an employer report of U occur with the same probability $\lambda(1-\lambda)$. Either of these observed probabilities in the Mellow-Sider data can be solved for the misclassification rate so that the misclassification rate is over-identified in these data.

Card (1996) uses these data to derive a minimum distance estimate of the misclassification rate (λ) of 0.027, which has become a common consensus number. However, it is important to note that it masks a substantial difference between the public and private sectors. The top panel of table 1 uses the Mellow-Sider data to show the two estimates of the disagreement

[14] Freeman (1984) also analyzes these data to assess misclassification error in union status.

Table 1: Distribution of Worker and Employer Reports of Union Status, Cell Fractions

All Workers, age 20-65, n=3520

	Employer Report		
Worker Report	NU	U	All
NU	0.713	0.035	0.748
U	0.028	0.224	0.252
All	0.742	0.259	1.00

Private Sector Workers, age 20-65, n=2717

	Employer Report		
Worker Report	NU	U	All
NU	0.735	0.017	0.752
U	0.017	0.231	0.248
All	0.752	0.248	1.00

Public Sector Workers, age 20-65, n=803

	Employer Report		
Worker Report	NU	U	All
NU	0.641	0.095	0.736
U	0.065	0.199	0.264
All	0.706	0.294	1.00

Source: Mellow-Sider (1983) from January 1977 CPS data

rate (in the off-diagonals) in the range of 2.8-3.5 percent. The middle panel of the table then shows the breakdown for the private sector workers, with much lower disagreement rates, both virtually identical at 1.7 percent. Solving for λ produces a private sector estimate of the misclassification rate of 0.0174 The disagreement rate is substantially higher in the public sector (bottom panel of table 1). One disagreement rate is 6.5 percent and the other is 9.5 percent, implying high misclassification rates.[15] In what follows, we restrict our analysis to

[15] While we cannot be sure what causes such a high misclassification rate in the public sector, one strong possibility is ambiguity between what is a labor union and what is a professional association. This could be of particular importance in the public sector. School teachers, who make up a sizeable fraction of public sector unionized workers, could belong to the National Education Association (NEA). Historically, many teachers belonged to this organization, but the organization only sometimes (and in some places) engaged in "normal" collective bargaining. The result could be confusion on either or both sides as to whether there was a union in the "normal" sense. Additionally, there is variation across states in what public sector unions can negotiate, and it is well known that unions of Federal Government employees generally cannot negotiate compensation.

Table 2: Union Membership Rates, Cross Sections

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$	Sample Size
DWS, old job	0.081	0.065	18957
DWS, new job	0.070	0.054	16975
CPS-MORG, monis 4	0.081	0.066	1325297

Note: The DWS and CPS data are for the period 1994-2024 for private sector full-time workers, as described above. The CPS-MORG data include rotation group 4 (monis 4) only in order to avoid multiple observations for the same individual. The corrected membership rates are calculated using equation 4.

the private sector, and use an estimate of the misclassification rate of $\lambda = 0.0174$, consistent with our analysis of the private sector Mellow-Sider data.

Equation 4 implies a simple calculation that yields the true unionization rate as a function of the observed unionization rate and the assumed misclassification rate. The adjustment can be substantial, particularly where the observed unionization rate is low. We use our assumed misclassification rate of λ=0.0174 and the DWS 1994-2024 old job and new job cross-section samples to calculate observed union membership rates as in equation 4 to calculate classification error adjusted membership rates. The results are in table 2, and show that the cross-section union membership rate is substantially overestimated when classification error is not accounted for. The calculations in table 2 show overestimates of 25 percent (old job) and 30 percent (new job) in both DWS cross-section samples. This table also presents these statistics for a comparable CPS-MORG sample, and the unionization rates, observed and corrected, are very close to those for the DWS. This suggests that the DWS sample of job losers is not unusual with regard to union status on the lost job compared to private sector full-time workers generally.

A straightforward maximum likelihood approach allows estimation of multivariate models of union status as a function of observed characteristics accounting for the misclassification. This approach is derived directly from equation 1, assuming a functional form for the true probability of unionization (γ). We use a probit model so that $\gamma = \Phi(X\beta)$ where $\Phi(\cdot)$ is the

standard normal CDF. The contribution of the i^{th} observation of a given sample to the log likelihood function is the log probability of observing the reported union status (U_i^*). This is

$$\ell n(L_i) = \ell n[U_i^* P(U_i^* = 1) + (1 - U_i^*)P(U_i^* = 0)] \tag{5}$$

The constituent probabilities are derived directly from equation 1 (noting that $P(U_i^* = 0) = 1 - P(U_i^* = 1)$) and the definition of $\gamma_i = \Phi(X_i\beta)$ in the probit model. These probabilities are

$$P(U_i^* = 1) = (1 - \lambda)\gamma_i + \lambda(1 - \gamma_i) = (1 - \lambda)\Phi(X_i\beta) + \lambda(1 - \Phi(X_i\beta)), \tag{6}$$

and

$$P(U_i^* = 0) = (1 - \lambda)(1 - \gamma_i) + \lambda\gamma_i = (1 - \lambda)(1 - \Phi(X_i\beta)) + \lambda\Phi(X_i\beta), \tag{7}$$

where the first term in each expression is the contribution from correct reporting of union status and the second term is the contribution of misclassified observations.

We estimate both a probit model for union membership not accounting for classification error ($\lambda = 0$) and the ML model accounting for classification error ($\lambda = 0.0174$). We control for education (4 categories), age (5 categories), sex, marital status, the interaction of sex and marital status, and indicators for survey year. We also control for 4 industry categories: 1) construction, 2) manufacturing, 3) transportation, communications, public utilities, and 4) "other". The results for each of the three data sources are shown in Appendix tables A1 to A3, and they are in accordance with expectations and qualitatively identical to the univariate results referred to earlier in this section. Union density is lower for more educated workers, for younger workers, and for females. Union density is highest in construction and lowest in the catch-all other category. Private sector union density declines sharply over time from the 1990s to the 2020s.

3.2 The Earnings Gap and Classification Error in the Cross-Section

Classification error in the cross-section implies that some fraction of observed union workers are, in fact, misclassified nonunion workers. Analogously, some fraction of observed nonunion workers are, in fact, misclassified union workers. Assuming that true union wages are higher than true nonunion wages, the average observed union wage is then an underestimate of

the average true union wage, and the average observed nonunion wage is an overestimate of the average true nonunion wage. The result is that the observed union wage gap is a downward-biased estimate of the true union wage gap.

In order to quantify the cross-sectional bias, consider a true proportional earnings gap of

$$D = \frac{W_u - W_n}{W_n}, \tag{8}$$

where W_u and W_n are the union and nonunion earnings respectively. The usual approximation is that

$$D \approx \ell n W_u - \ell n W_n = w_u - w_n, \tag{9}$$

where w_u is log union earnings and w_n is log nonunion earnings.

The expectation of observed log union earnings in the presence of classification error is

$$w_u^* = \frac{\gamma(1-\lambda)w_u + (1-\gamma)\lambda w_n}{\gamma^*} = \frac{\gamma(1-\lambda)w_u + (1-\gamma)\lambda w_n}{\gamma + (1-2\gamma)\lambda}, \tag{10}$$

using the definition of γ^* in equation 2. The first term in the numerator comes from union workers who correctly report their union status and the second term in the numerator comes from nonunion workers who incorrectly report their union status. Analogously, the expectation of observed log nonunion earnings is

$$w_n^* = \frac{\gamma\lambda w_u + (1-\gamma)(1-\lambda)w_n}{1-\gamma^*} = \frac{\gamma\lambda w_u + (1-\gamma)(1-\lambda)w_n}{1-\gamma-(1-2\gamma)\lambda}, \tag{11}$$

where the first term in the numerator comes from union workers who incorrectly report their union status and the second term in the numerator comes from nonunion workers who correctly report their union status.

Bias in the observed union wage gap (D^*) from the classification error arises almost entirely from the fact that the observed average union wage is substantially contaminated by nonunion workers who are misclassified as union workers. Given the relatively low union density and misclassification rate, there is very little bias in the observed average nonunion wage. A simple calculation using an observed union membership rate on the lost job in the DWS of $\gamma^* = 0.081$ (based on the information in table 2) and a misclassification rate of $\lambda = 0.0174$ implies a true union membership rate of $\gamma = 0.065$ and illustrates this point. Using these values in equations 10 and 11 yields

$$w_u^* = \frac{0.064w_u + 0.016w_n}{0.081} = 0.798w_u + 0.202w_n, \tag{12}$$

Table 3: True Union Density and Wage Gap Ratio by Observed Union Densities

Adjustment for Union Status Classification Error ($\lambda = 0.0174$)

γ^*	γ	D/D^*	γ^*	γ	D/D^*
0.040	0.023	1.739	0.160	0.148	1.102
0.060	0.044	1.384	0.180	0.168	1.087
0.080	0.065	1.255	0.200	0.189	1.076
0.100	0.086	1.189	0.220	0.210	1.067
0.120	0.106	1.149	0.240	0.231	1.059
0.140	0.127	1.122	0.300	0.293	1.024

Note: γ^* = observed union density, γ = true union density from equation 4, and D/D^* is the ratio of the true union wage gap (D) to the observed union wage gap (D^*) based on equation 15.

and

$$w_n^* = \frac{0.001w_u + 0.918w_n}{0.919} = 0.001w_u + 0.999w_n. \tag{13}$$

The example illustrates that the observed union wage is a substantial underestimate of the true union wage while the observed nonunion wage is very close to the true nonunion wage.

Using the simplifying approximation that $w_n^* = w_n$, we can rewrite the observed union wage gap using equation 10 (and the definition of γ^* in equation 2) as

$$D^* \approx w_u^* - w_n = \frac{\gamma(1-\lambda)w_u + (1-\gamma)\lambda w_n}{\gamma^*} - w_n = \frac{\gamma}{\gamma^*}(1-\lambda)(w_u - w_n) = \frac{\gamma}{\gamma^*}(1-\lambda)D. \tag{14}$$

The true union wage gap as a function of the observed differential is

$$D = \frac{\gamma^*}{\gamma(1-\lambda)}D^* = \left[1 + \frac{(1-\gamma)\lambda}{\gamma(1-\lambda)}\right] D^* > D^*. \tag{15}$$

The "inflation factor" needed to adjust the observed wage gap for misclassification is

$$\frac{D}{D^*} - 1 = \frac{\gamma^*}{\gamma(1-\lambda)} - 1 = \frac{(1-\gamma)\lambda}{\gamma(1-\lambda)} \tag{16}$$

and varies inversely with union density.

Table 3 contains, for a range of observed union densities (γ^*) assuming a union status misclassification rate of $\lambda = 0.0174$, calculations of the true union density (γ) and the ratio of the true union wage gap to the observed union wage gap. These calculations show substantial overestimation of union density and underestimation of the union wage gap due

to union status misclassification, particularly at lower union densities. For example, an observed union density of about 8 percent (the rough private sector average from 1994-2024 shown in table 2) implies a true union density of 6.5 percent and a true union wage gap 25.5 percent higher than the observed wage gap. These calculations suggest that reasonable misclassification error rates will lead to very substantial understatement of the union wage advantage, rendering standard cross-section estimates unreliable in the low-union private sector in the U.S.[16]

4 Union Status and Earnings Changes in Longitudinal Data: Accounting for Classification Error

We demonstrated in the previous section that, in the low-unionization environment that exists in the U.S. private sector, classification error in reported union status in surveys leads both to substantial over-estimation of the cross-section unionization rate and to substantial under-estimation of the cross-section union wage gap. In this section, we extend our analysis to longitudinal data where we rely on union status transitions to estimate the union wage gap.

4.1 Union Status Transitions and Classification Error

When using longitudinal data on union status, each linked observation has two possible misclassifications. Given a pair of observations on union status and the wage for a worker at two points in time, there are four possible true union status transitions as well as four possible observed transitions: 1) nonunion to nonunion, 2) nonunion to union, 3) union to nonunion, and 4) union to union. We start by defining the probabilities of each of the true transitions and showing how they map into probabilities of observed transitions where there is classification error.

We extend the structure of classification error we used for the cross-section. There is a

[16] This is not the case in the public sector where the observed union membership rate based on the CPS-MORG is quite a lot higher at about 35 percent over the period studied.

one-period classification error probability of λ, and we assume that the errors are independent across periods. Let θ_{ij} represent the true probability of a two period union sequence ij, where i and j each are either 0 (representing nonunion) or 1 (representing union). Analogously, let θ^*_{ij} represent the observed probability of the two period union sequence ij. The probabilities of the four observed union status sequences as a function of the true union status sequence probabilities and the misclassification probability are, in matrix form,

$$\Theta^* = \begin{pmatrix} \theta^*_{00} \\ \theta^*_{01} \\ \theta^*_{10} \\ \theta^*_{11} \end{pmatrix} = \begin{pmatrix} (1-\lambda)^2 & \lambda(1-\lambda) & \lambda(1-\lambda) & \lambda^2 \\ \lambda(1-\lambda) & (1-\lambda)^2 & \lambda^2 & \lambda(1-\lambda) \\ \lambda(1-\lambda) & \lambda^2 & (1-\lambda)^2 & \lambda(1-\lambda) \\ \lambda^2 & \lambda(1-\lambda) & \lambda(1-\lambda) & (1-\lambda)^2 \end{pmatrix} \begin{pmatrix} \theta_{00} \\ \theta_{01} \\ \theta_{10} \\ \theta_{11} \end{pmatrix} = M\Theta, \qquad (17)$$

where Θ^* is the vector of observed union transition probabilities, M is the 4x4 "classification error rate" matrix, and Θ is the vector of true union transition probabilities. Each row of the matrix M contains the set of probabilities related to misclassification that translates the true probabilities into the observed probability in that row.

To gain some intuition, consider the first element, (θ^*_{00}), which is the observed probability of a nonunion-nonunion transition. This is

$$\theta^*_{00} = (1-\lambda)^2\theta_{00} + \lambda(1-\lambda)\theta_{01} + \lambda(1-\lambda)\theta_{10} + \lambda^2\theta_{11}, \qquad (18)$$

which is a probability-weighted average of the true transition probabilities. The probability weights reflect the number of misclassifications (out of two possible) that would be associated with observing a 0-0 transition if there was a particular true transition. For example, the first term on the right of equation 18 is associated with no classification errors (which happens with probability $(1-\lambda)^2$) and a true 0-0 transition (probability θ_{00}). The other three terms are the contributions of other true transitions, which occur when there is either 1 classification error (0-1 or 1-0 misclassified as 0-0) or 2 classification errors (1-1 misclassified as 0-0).

The system of linear equations in equation 17 can be inverted to solve for the vector of true transition probabilities as a function of the observed transition probabilities and the misclassification rate, yielding

$$\Theta = M^{-1}\Theta^*, \qquad (19)$$

Table 4: Union Membership Transition Rates, DWS

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$
θ_{00}	0.889	0.920
θ_{01}	0.029	0.013
θ_{10}	0.041	0.026
θ_{11}	0.041	0.041

Note: θ_{ij} is the union membership transition rate from i to j. The transition rate adjusted for classification error is defined in equation 19. The data are for the period 1994-2024 for private sector full-time workers, as described above. The transition rates are weighted by the final weight at the DWS survey date. Sample size is 15368 reemployed private-sector full-time involuntary job changers.

The adjustment to the observed transition rates can be substantial, particularly where the observed unionization rate is low.

Observed and corrected union status transition rates based on the DWS information on lost job and new job union membership are presented in table 4. The observed transition rate vector (Θ^*) is presented in column 1 of the table. We then use the observed transition rates together with equation 19, assuming a one-period classification error rate of the magnitude we use in the cross-section ($\lambda = 0.0174$), to calculate the vector of classification-error adjusted union status transition probabilities (Θ), presented in column 2 of the table.

Reflecting the low level overall level of unionization, it is not surprising that about 90 percent of observations are nonunion on both the lost job and the new job. Of particular relevance for using union status transitions to estimate the union wage gap, is the frequency of actual changes in union status. The observed transition rates in table 4 imply that 7 percent of observations in the DWS involve a change in union status. However, the transition probabilities adjusted for classification error using equation 19 show that only 3.9 percent of observations in the DWS involve a true change in union status. These transitions are more likely to be union leavers than joiners. The implication is that our DWS sample contains a very small expected number of true union status transitions. Based on the adjusted transition rates and the sample size in Table 4, we expect only about 600 union status changers (400

union leavers and 200 union joiners) in the DWS sample.

It appears that almost half (about 44 percent) of observed union status changers in the DWS are, in fact, misclassified non-changers (mostly N-N).[17] This level of misclassification will have profound effects on estimates of the union wage gap using longitudinal data since it implies that the observed wage changes of union status changers are badly contaminated with observed wage changes of non-changers. It is, therefore, of central importance to account for the classification error in union status when using longitudinal data.

Union sectoral retention rates among job losers in the DWS are quite a lot higher than expected. If workers were reemployed randomly with respect to union status, then, based on the adjusted estimates of the union membership rate in table 2, about 6 percent of job losers would be reemployed in a union job, regardless of union status on the lost job. The adjusted nonunion sectoral retention rate for the DWS sample in table 4 suggests that almost all nonunion job losers are reemployed in a nonunion job ($\theta_{00}/(\theta_{00} + \theta_{01}) = 0.986$). The analogous calculation of the adjusted union sectoral retention rate suggests that 61 percent of union job losers are reemployed in a union job ($\theta_{11}/(\theta_{10} + \theta_{11}) = 0.612$). Given that the adjusted union membership rate is less than 7 percent, it is surprising that more than half of union job losers find another union job. This could reflect some kind of tie between workers and their union that goes beyond the particular job.[18]

As we did with the cross-section, we present a straightforward maximum likelihood approach that allows estimation of multivariate models of union status transitions as a function of observed characteristics accounting for the misclassification. This approach is derived directly from equation 17, which defines the observed transition probabilities (Θ^*) as a function of the true transition probabilities (Θ) and the misclassification rate (λ) by assuming a functional form for the true union status transition probabilities as a function of observables.

There are four elements of the true transition probability vector for individual i (θ_{00i},

[17] To be precise, given the randomness introduced by the classification error, these fractions reflect one minus the ratios of the adjusted to the unadjusted transition rates rather than identifying the observed transitions of misclassified workers.

[18] This is clearly the case in craft industries like construction, where the unions run hiring halls that can refer unemployed members to new jobs. While industrial unions often do not have formal hiring halls, there can be a less formal job referral process for members who lose a job.

θ_{01i}, θ_{10i}, and θ_{11i}). Given that these elements sum to one, there are three free probabilities to be estimated, and we specify $\theta_{00i} = 1 - \theta_{01i} - \theta_{10i} - \theta_{11i}$. U^*_{jki} is an indicator variable which equals one if individual i is observed to make a union status transition $j \to k$ (where j and k are either zero or one).

The contribution of the i^{th} observation to the likelihood function is the probability of observing a transition U^*_{jki} as a function of the true transition rate vector (Θ and the misclassification rate (λ). This contribution is

$$\begin{aligned} L_i \quad = \quad & U^*_{00i}[(1-\lambda)^2\theta_{00i} + \lambda(1-\lambda)\theta_{01i} + \lambda(1-\lambda)\theta_{10i} + \lambda^2\theta_{11i}] \\ + \quad & U^*_{01i}[\lambda(1-\lambda)\theta_{00i} + (1-\lambda)^2\theta_{01i} + \lambda^2\theta_{10i} + \lambda(1-\lambda)\theta_{11i}] \\ + \quad & U^*_{10i}[\lambda(1-\lambda)\theta_{00i} + \lambda^2\theta_{01i} + (1-\lambda)^2\theta_{10i} + \lambda(1-\lambda)\theta_{11i}] \\ + \quad & U^*_{11i}[\lambda^2\theta_{00i} + \lambda(1-\lambda)\theta_{01i} + \lambda(1-\lambda)\theta_{10i} + (1-\lambda)^2\theta_{11i}]. \end{aligned} \tag{20}$$

The implied log likelihood function is maximized with respect to the three free θ_{jk}, yielding estimates of the true transition probabilities adjusted for classification error.

To allow the estimated transition probabilities and the wage gap to vary with observable characteristics, we estimate a multivariate model assuming a functional form for the transition probabilities. Among several options, we selected a bivariate probit model where the two discrete options are union status in the two periods.[19]

4.2 The Earnings Gap and Classification Error in Union Status Transitions

The union wage gap estimated with longitudinal data depends on the wage changes experienced by workers who make a union status transition. There are two possible difference-in-difference estimates of the union earnings gap using earnings changes of individuals who change union status:

1. The average change in earnings of nonunion job losers who transition to a union job

[19] Note that this introduces a constraint since the 4 transition probabilities are determined by two potentially correlated underlying probabilities. Some other possibilities are multinomial logit, independent univariate probits, and independent linear probability models. Each have their own statistical limitations.

(joiners) relative to the average change in earnings of nonunion job losers who remain nonunion.

2. The(negative) average change in earnings of union job losers who transition to a nonunion job (leavers) relative to the change in earnings of union job losers who remain union.

We are using job losers who do not change union status as control groups for those who do change union status.

Classification error implies that some fraction of workers who are observed to make a particular union status transition are, in fact, workers who are misclassified as making a different transition. As we showed earlier, assuming a misclassification rate of $\lambda = 0.0174$ and based on the unadjusted and adjusted transition probabilities from the DWS in table 4, of those workers observed to make a union status transition, only 56 percent actually change union status. Given the low unionization rate in these data, most of those misclassified actually were nonunion at both points in time. It is clear that using earnings changes for those observed to make a N-U or U-N transition will result in a substantial underestimate in magnitude of the union earnings gap.

We present the algebra for computing the classification-error corrected estimates of the average earnings gap in longitudinal data in Appendix I. These corrected estimates, as presented in equation 39, are a function of the observed union earnings gap, the observed union status transition rates, and the assumed classification error rate.

To get a sense of the magnitude of the bias due to classification error, we use summary statistics from the DWS sample described in section 2, from 1994-2024 for 15195 private sector full-time job losers reemployed on a private sector full-time job. Equation 21 contains these estimates:

$$\Delta^*_{DWS} = \begin{pmatrix} \Delta^*_{00} \\ \Delta^*_{01} \\ \Delta^*_{10} \\ \Delta^*_{11} \end{pmatrix} = \begin{pmatrix} -0.048 \\ 0.031 \\ -0.265 \\ -0.032 \end{pmatrix} \qquad \Delta_{DWS} = \begin{pmatrix} \Delta_{00} \\ \Delta_{01} \\ \Delta_{10} \\ \Delta_{11} \end{pmatrix} = \begin{pmatrix} -0.048 \\ 0.132 \\ -0.411 \\ -0.028 \end{pmatrix} \tag{21}$$

The average observed log real earnings changes by transition for DWS job losers (left

part of equation 21) indicate small average declines for those job losers who do not change union status, a small increase for union joiners, and a very large average earnings decline of 26.5 log points (23.3 percent) for union leavers. These effects of union status changes are substantially increased when accounting for classification error (right part of equation 21). Union joiners have an average 13.2 log point (14.1 percent) increase in earnings and union leavers, on average, have a very large decrease in earnings of 41.1 log points (33.7 percent).

Since job loss tends to reduce earnings independent of union status, it is appropriate to calculate the difference-in-difference estimates of the earnings changes associated with union status changes. As described above, these are 1) a comparison of the earnings change of union joiners with the earnings change of nonunion job losers who remain nonunion and 2) a comparison of the earnings change of union leavers with the earnings change of union job losers who remain union. These D-i-D estimates are

$$\Delta\Delta^*_{DWS} = \begin{pmatrix} \Delta^*_{01} - \Delta^*_{00} \\ \Delta^*_{10} - \Delta^*_{11} \end{pmatrix} = \begin{pmatrix} 0.079 \\ -0.233 \end{pmatrix} \quad \Delta\Delta_{DWS} = \begin{pmatrix} \Delta_{01} - \Delta_{00} \\ \Delta_{10} - \Delta_{11} \end{pmatrix} = \begin{pmatrix} 0.180 \\ -0.383 \end{pmatrix} \tag{22}$$

Given that job losers who retain their union status suffer earnings declines on average, these D-i-D estimates magnify the increase in earnings for union joiners and reduce the decrease in earnings for union leavers. The substantial asymmetry between leavers and joiners persists, with (corrected for classification error) average earnings gains for union joiners of 18 log points (19.7 percent) and average earnings losses for union leavers of 38.3 log points (31.8 percent). Our estimates of the union earnings gaps, for both union joiners and union leavers, are considerably larger than earlier longitudinal estimates.[20]

5 Multivariate Analysis of the Earnings Gap

We have established that classification error in union status results in attenuated estimates of the union earnings gap in both cross-section and longitudinal data. In the cross-section, we

[20] The earlier literature, based on the usual matched longitudinal data, generally finds close to symmetric estimates (the average earnings gain of union joiners is approximately equal the average earnings loss of union leavers). Our analysis of the MORG data in section 6 shows results consistent with this observation.

were able to solve for the corrected union status probability and the corrected union earnings gap. In longitudinal data, we were able to solve for the corrected union status transition rates and the corrected union earnings gap. However, it is clear that union status/transition probabilities and earnings vary across workers and over time in systematic ways that may be correlated with union status and union status transitions. In this section we specify maximum likelihood techniques that allow us to estimate multivariate models of the corrected union wage gap jointly with true union status/transition probabilities.

5.1 The Union Earnings Gap in the Cross-Section

We assume a standard earnings generation process for individual i of

$$w_i = X_i\beta + \delta U_i + \epsilon_i, \tag{23}$$

where w_i is the log of weekly earnings, U_i an indicator for true union status, X_i is a vector of observed individual characteristics (e.g., education, age, race, sex, industry, year), and the random component, ϵ_i, is distributed normally with zero mean and variance σ^2. The parameter δ is the log earnings gap.

The probability (γ^*) that individual i is observed to be union ($U_i^* = 1$) is defined in equation 1 as a function of the true union probability (γ) and the classification error rate (λ). The implied observed earnings function with classification error is

$$w_i = X_i\beta + \delta U_i^* + \epsilon_i^*, \tag{24}$$

where $\epsilon_i^* = \delta(U_i - U_i^*) + \epsilon_i$.[21]

We further assume that true union status is the result of a standard probit latent variable model such that $U_i = 1$ when an unobserved latent variable $Y_i = Z_i\alpha + \eta_i$ is positive, where Z_i is a vector of observed individual characteristics and η_i is an error term with a standard normal distribution. For tractability, we assume independence of the random components in the earnings and union status functions (ϵ_i and η_i respectively).

[21] Clearly, ϵ_i^* is negatively correlated with U_i^*, so that the OLS estimate of the union wage gap ($\hat{\delta}$) in equation 18 based on observed union status is a downward biased estimate of the true union wage gap (δ). This is consistent with our earlier discussion.

Table 5: ML Estimates of Cross-Section DWS Union Earnings Differentials (δ)

	(1) Observed $\lambda = 0$ No Controls	(2) Observed $\lambda = 0$ Controls	(3) Corrected $\lambda = 0.0174$ Controls	Sample Size
δ, old job	0.174 (0.016)	0.183 (0.015)	0.252 (0.018)	18821
δ, new job	0.169 (0.019)	0.193 (0.018)	0.284 (0.022)	16851

Note: Robust standard errors are in parentheses. Controls include indicators for 5 age categories, 5 education categories, female, married, female x married, nonwhite, 4 industry categories, and 16 DWS survey years from 1994-2024. All models are weighted by the CPS final sample weights. Full model estimates for columns 2 and 3 with controls are contained in Appendix Tables A4 and A5.

Based on these assumptions, we derive in Appendix II the joint probability density of observing ϵ_i^* and $U_i^* = 0$ for workers observed to be nonunion and 2) the joint probability density of observing ϵ_i^* and $U_i^* = 1$ for workers observed to be union. These probability densities (equations 40 and 41 in Appendix II) are the components of a likelihood function for observed earnings and observed union status that accounts for classification error. The implied likelihood function can be maximized with respect to the model parameters (α, β, δ, and σ). As before, we rely on our analysis of the Mellow and Sider (1983) data and use a value of $\lambda = 0.0174$ for our private sector sample.

Table 5 contains maximum likelihood estimates of cross-section earnings gap (δ) based on the DWS for both the lost job (first row) and the new job (second row). Column 1 contains the estimated union earnings gaps with no correction for misclassification and no controls. These estimates are about 17 log points (18.5 percent) on both the old and new job. Adding the controls in column 2 yields slightly larger estimates of about 19 log points (21 percent). Accounting for misclassification in union status as well as the controls in column 3 yields substantially higher estimates of the earnings gap of about 27 log points (31 percent). This is consistent with our analytical finding that classification error in union status leads to substantial downward bias in the union earnings gap.

5.2 The Union Earnings Gap in Longitudinal Data

We define a pair of latent variables, Y_0 and Y_1, that determine union membership in first and second periods respectively. A positive Y_t implies union membership in period t. These latent variables are specified as $Y_0 = X_0\beta_0 + \nu_0$ and $Y_1 = X_1\beta_1 + \nu_1$, where the X's are vectors of observable variables, the β's are parameters to be estimated, and ν_0 and ν_1 are random components that have a bivariate normal distribution with variances equal to one and correlation ρ. Denote the CDF of this bivariate normal distribution by $\Psi(\cdot,\cdot,\rho)$. The probabilities of the four possible true union membership status transitions are specified as

$$\theta_{00} = P(Y_0 < 0, P(Y_1 < 0) = P(\nu_0 < -X_0\beta_0, \nu_1 < X_1\beta_1) = \Psi(-X_0\beta_0, -X_1\beta_1, \rho), \quad (25)$$

$$\theta_{01} = P(Y_0 < 0, P(Y_1 > 0) = P(\nu_0 < -X_0\beta_0, \nu_1 > -X_1\beta_1) = \Psi(-X_0\beta_0, X_1\beta_1, -\rho), \quad (26)$$

$$\theta_{10} = P(Y_0 > 0, P(Y_1 < 0) = P(\nu_0 > -X_0\beta_0, \nu_1 < -X_1\beta_1) = \Psi(X_0\beta_0, -X_1\beta_1, -\rho), \quad (27)$$

$$\theta_{11} = P(Y_0 > 0, P(Y_1 > 0) = P(\nu_0 > -X_0\beta_0, \nu_1 > -X_1\beta_1) = \Psi(X_0\beta_0, X_1\beta_1, \rho). \quad (28)$$

With this specification for union status transition probabilities in hand, we can estimate the classification-error adjusted earnings changes associated with the four possible union status transitions via maximum likelihood. This is analogous to the maximum likelihood approach we outlined to estimate the classification error adjusted union earnings gap using cross-section data. To be concrete, define the change in log earnings for individual i (Δw_i) as

$$\Delta w_i = X_i\beta + (\delta_{01} - \delta_{00})T_{01i} + (\delta_{10} - \delta_{00})T_{10i} + (\delta_{11} - \delta_{00})T_{11i} + \epsilon_i, \quad (29)$$

where i indexes individuals. The T_{jki} are indicator variables that equal 1 if the true union status transition for individual i is from j to k. The parameter δ_{jk} is the average wage change (controlling for X) associated with a transition from i to j in union status. The difference-in-differences estimate of the effect of union transitions, relative to the appropriate base, are $D_{01} = \delta_{01} - \delta_{00}$ for union joiners and $D_{10} = \delta_{10} - \delta_{11}$ for union leavers. The D-i-D estimate of the earnings changes for union stayers relative to "never union" is $D_{11} = \delta_{11} - \delta_{00}$. In order to obtain these D-i-D estimates, we rewrite equation 29 as

$$\begin{aligned} \Delta w_i &= X_i\beta + (\delta_{01} - \delta_{00})T_{01i} + (\delta_{10} - \delta_{11})T_{10i} + (\delta_{11} - \delta_{00})(T_{11i} + T_{10i}) + \epsilon_i \\ &= X_i\beta + T_{01i}D_{01} + T_{10i}D_{10} + U_{0i}D_{11} + \epsilon_i, \end{aligned} \quad (30)$$

where $U_{0i} = (T_{11i} + T_{10i})$ is union status on the lost job. Our estimates of D_{01} and D_{10} are the D-i-D estimates of the union earnings gap for union joiners and union leavers, respectively. They are the multivariate analogs of the summary measures in equation 22.

Assuming that the error, ϵ_i in equation 30, is distributed normally with zero mean and variance σ^2, the probability density of the error is

$$f(\epsilon_i) = \frac{1}{\sigma}\phi\left(\frac{\Delta w_i - X_i\beta - T_{01i}D_{01} - T_{10i}D_{10} - U_0 D_{11}}{\sigma}\right), \tag{31}$$

where $\phi(\cdot)$ is the standard normal PDF. If we observed the true union status transitions (without classification error), we could specify and maximize a likelihood function with respect to β, D_{01}, D_{10}, D_{11}, and σ directly from equation 31, and this would be equivalent to running the OLS regression implied by equation 30.

In order to account for classification error in union status, we use the assumption that misclassification occurs with probability λ on both jobs and is independent across jobs. There are four possible true earnings changes associated with each observed union status transition, one for each of the true underlying union status transitions. We calculate the probability of observing a particular wage change, Δw_i, in each of the four observed transitions (indicated by T^*_{jki}). Each probability is a different weighted average of the probability-densities of the wage changes associated with the four true transitions.[22] The contribution to the likelihood function for individual i is the joint probability for individual i of the observed wage change associated together with an observed j to k union status transition is L_{ijk}. These probabilities are presented in Appendix III and are the basis of the likelihood function.

Maximization of the log likelihood function yields consistent D-i-D estimates of the wage changes associated with each true union status transition. Of particular interest are the D-i-D estimates of the average wage change associated with a U-N transition relative to a U-U transition (D_{10}) and the average wage change associated with a N-U transition relative to a N-N transition (D_{01}). Note that, as a consequence of the classification error all observed union status transitions contribute to the estimation of each of these wage changes.

Table 6 contains maximum likelihood D-i-D estimates of union earnings differentials based

[22] The weights associated with each observed transition are defined in the appropriate row of matrix M (equation 17).

Table 6: Estimates of Longitudinal DWS Union Earnings Gaps ($\Delta\Delta$)

	(1) Observed $\lambda = 0$ No Controls	(2) Observed $\lambda = 0$ Controls	(3) Corrected $\lambda = 0.0174$ Controls
Difference-in-differences estimates (base keep same U)			
$\Delta\Delta_{01}$ $(\delta_{01} - \delta_{00})$	0.080 (0.026)	0.075 (0.025)	0.175 (0.048)
$\Delta\Delta_{10}$ $(\delta_{10} - \delta_{11})$	-0.234 (0.029)	-0.213 (0.030)	-0.334 (0.042)
$\Delta\Delta_{11}$ $(\delta_{11} - \delta_{00})$	0.017 (0.020)	0.029 (0.021)	0.035 (0.022)
P-values			
Test of $(\delta_{01} - \delta_{00}) = -(\delta_{10} - \delta_{11})$	0.000	0.001	0.013
Test of $\delta_{11} = 0$	0.412	0.182	0.106

Note: Robust standard errors are in parentheses. Controls include indicators for 5 age categories, 5 education categories, female, married, female x married, nonwhite, 4 industry categories, and 16 DWS survey years from 1994-2024. All models are weighted by the CPS final sample weights. Full model estimates are contained in Appendix Table A7.Sample size of 15195 reemployed full-time involuntary job changers.

on the earnings changes of union status changers relative to the earnings changes of those workers who did not change union status. Column 1 contains estimated union earnings gaps with no correction for misclassification and no controls. Consistent with earlier calculations, the estimate of the earnings gap for observed union leavers far exceeds the estimate of the earnings gap for observed union joiners. Union joiners have an average earnings increase of 8 log points (8.3 percent), while observed union leavers have an average earnings decline of 23.4 log points (21 percent). The p-value of the test of equality of magnitude easily rejects the null. Adding controls in column 2 yields slightly smaller estimates of both, and again, the null hypothesis of equality of magnitude of the wage gap for union leavers and union joiners is easily rejected. Correcting for misclassification (column 3) as well as controlling for worker characteristics and year not surprisingly yields substantially higher estimates of the earnings gap of about 17.5 log points (19.1 percent) for union joiners and 33.4 log points (28.4 percent) for union leavers. Again, this confirms our analytical finding that classification error in union status leads to substantial downward bias.

To summarize, there are at least two important findings based on the earnings changes of job losers who change union status, accounting for classification error. First, the earnings

decline of union leavers is substantially larger than the earnings increase of union joiners. Second, the magnitudes of both of these adjusted earning gaps are substantial, and the estimate based on union leavers is considerably larger than the usual cross-section estimate of the earnings gap.

6 A Comparison of the MORG with the DWS

We referenced in the introduction the existing literature using longitudinal data to estimate union wage/earnings gaps based on the earnings changes of workers reporting a change in union status. With few exceptions, this literature does not focus on job changers, voluntary or involuntary, and includes the wage changes experienced by workers with the same employer at both points in time along with the experience of those who change jobs for any reason. As we noted, changes in union status in these data can be the result of involuntary job changes, of voluntary job changes (within or between employers), or classification error.[23] Given the relatively wide use of general longitudinal data on workers' earnings and union status, it is worth comparing our results using displaced workers to prototypical results using longitudinal data from the matched CPS MORG. This is the task to which we turn in this section.

6.1 The MORG Data

We use data from the Merged Outgoing Rotation Group CPS (MORG) from 1994-2024. The MORG provides data on wages and union status for a large sample of workers.[24] To generate a cross-section sample for the analysis of union status in each year, we use only

[23] There is insufficient organizing or decertification activity based on data from the NLRB to account for the observed rate of union status changes among those who do not change jobs in these data. Farber and Western (2001) present evidence on the very low level of union organizing through NLRB elections relative to the size of the nonunion workforce. From 1985-1997, a period with much higher private sector union membership rates than more recent years, new union organization through the NLRB election process was about 0.1 percent of private sector nonunion employment each year.

[24] The rotation group structure of the CPS is 4 months in sample (rotation groups 1-4), 8 months out of sample, and 4 months in sample (rotations groups 5-8). Rotation groups 4 and 8 are termed the outgoing rotation groups, and respondents in these two groups are asked questions about earnings and union status.

observations from rotation group 4 to avoid having two observations on the same individual (one from rotation group 4 and another from rotation group 8 the following year).[25] We select individuals aged 20-64 who are employed full-time in a private sector wage or salary job and for whom union status is not allocated. This results in a total sample size for the analysis of union status of 1325297.

For the cross-section earnings analysis, we further select this sample by including only observations on weekly earnings that can be calculated without imputation of earnings or hours (not allocated) and by trimming the sample (following again Autor, Katz, and Kearney (2008)) to eliminate any observations with real weekly earnings less than $67 in 1984 dollars. As we did with the DWS, following Autor Katz, and Kearney (2008), we adjust for topcoding of earnings by adding 50 percent to top-coded earnings observations. The result is a total sample size for the cross-section earnings analysis of 916155 in rotation group 4.

The rotation group structure of the CPS allows matching of rotation group 4 in month j in year $t-1$ to rotation group 8 in month j in year t, resulting in a longitudinal sample with information at both points in time on union status and earnings.[26] We match the observations in our cross-section samples year-to-year (e.g., rotation group 8 in January 2000 is matched with rotation group 4 in January 1999).[27] For consistency with the DWS, we include only workers aged 20-64 in the second year (rotation group 8) who are employed full-time in private-sector wage and salary jobs at both points in time. To generate the sample for the analysis of union status transitions, we include only matched observations where union status is not allocated in either year. The resulting sample contains 644514 matched pairs. For the longitudinal earnings analysis, we further select this sample by including only observations for which earnings in both years can be calculated without imputation of earnings or hours. We also adjust the top-coded cross-section earnings observations and trim the lower tail of the cross-section earnings observations as described above. The result is a

[25] We have replicated our cross-section MOGRG analyses using only rotation group 8 a year later. The results are virtually identical to those using rotation group 4.

[26] Unfortunately, we do not know if there is a job change in the interim or the cause of any job change.

[27] Given the change in 1994 to computer-aided interviewing, it is difficult to match the CPS forward year-to-year in 1994 through August 1995. As a result, our longitudinal analysis of the MORG uses 1996m9–2024, matching back one year.

Table 7: ML Estimates of Cross-Section Earnings Gaps (δ)

	(1) $\lambda = 0$ No Controls	(2) $\lambda = 0$ Controls	(3) $\lambda = 0.0174$ Controls	Sample Size
MORG, monis 4	0.142 (0.002)	0.141 (0.002)	0.195 (0.003)	916155
DWS, old job	0.174 (0.016)	0.183 (0.015)	0.252 (0.018)	18821
DWS, new job	0.169 (0.019)	0.193 (0.018)	0.284 (0.022)	16851

Note: Robust standard errors are in parentheses. Controls include indicators for 5 age categories, 5 education categories, female, married, female x married, nonwhite, 4 industry categories, and 29 survey years from 1996-2024. All models are weighted by the CPS final sample weights. Full model estimates are contained in Appendix Table A6.

total sample size of 367874 matched pairs for the longitudinal wage analysis. These data are meant to be roughly comparable to the sort of longitudinal data used by others to estimate union wage gaps (e.g. Mellow, 1981; Freeman, 1984; Jakubson, 1991; Card, 1996).

6.2 Analysis of the MORG Data

We already calculated the observed and adjusted cross-section union membership rates in the CPS MORG sample. These numbers are presented in table 2 and show a virtually identical corrected union membership rate of about 6.6 percent in both the CPS-OGRG data and on the lost lost-job in the DWS. The rate is somewhat lower, at 5.4 percent on the new job in the DWS due to the fact that many union job losers are reemployed in a nonunion job.

The cross-section earnings gaps estimated with the MORG (rotation group 4) along with those estimated from the DWS (old job and new job in turn) are presented in table 7.[28] Both the adjusted and unadjusted earnings gaps are smaller in the MORG than in the DWS. The adjusted earnings gap in column 3 of table 7 show an earnings gap in the DWS (averaged across the old and new jobs) that is substantially larger than the earnings gap in the MORG. These earnings gaps are 26.8 log points (30.7 percent) in the DWS compared to 19.5 log points (21.5 percent) in the MORG.

The corrected union status transition rates in the MORG sample are calculated as a

[28] The DWS estimates are included in the table to facilitate comparison. These estimates were presented earlier in table 5.

Table 8: Union Status Transition Rates, MORG and DWS

	(1) MORG Observed $\lambda = 0$	(2) MORG Corrected $\lambda = 0.0174$	(3) DWS Observed $\lambda = 0$	(4) DWS Corrected $\lambda = 0.0174$
θ_{00}	0.893	0.925	0.889	0.920
θ_{01}	0.024	0.007	0.029	0.013
θ_{10}	0.023	0.007	0.041	0.026
θ_{11}	0.060	0.061	0.041	0.041

Note: θ_{ij} is the union membership transition rate from i to j. The transition rate adjusted for classification error is defined in equation 19. The data are for the period 1994-2024 for the DWS and 1996m10-2024 for the MORG and includes observations for private sector full-time workers, as described above. The sample sizes are 15368 for the DWS and 644514 for the matched MORG. The DWS fractions are weighted by the final weight at the DWS survey date. The MORG fractions are weighted by the average of the final weights in the two months.

function of the observed transition rates and the misclassification rate using equation 19. Columns 1 and 2 of table 8 contain the observed and corrected union status transition rates for our MORG sample. To facilitate comparison, columns 3 and 4 of this table contain the analogous transition rates from our DWS sample (shown earlier in table 4). Again, we focus on the rates corrected for misclassification.

The basic patterns are the same, but there are important differences. About 92 percent of each sample is nonunion in both periods, but, relative to workers in the DWS, workers in the MORG are about 2 percentage points (50 percent) more likely to be union in both periods (6.1 percent vs. 4.1 percent). This substantial difference reflects the fact that most workers in the matched MORG data likely did not change jobs, so that union workers in the first period are quite likely to remain union. The most important difference for estimation of the union earnings gap is that the incidence of union status changes is much smaller in the MORG than in the DWS. Union joiners make up 1.3 percent of job losers in the DWS but only 0.7 percent in the MORG. The difference is even sharper for union leavers, where 2.6 percent of job losers in the DWS are union leavers while only 0.7 percent of workers in the matched MORG are union leavers. The very low corrected rates of union status changes in the MORG implies a substantial persistence of union status. Union retention rate in the

MORG is $0.9 = \theta_{11}/(\theta_{11} + \theta_{10})$), compared with about 0.6 for job losers in the DWS.

Estimation of the union wage gap based on longitudinal data fundamentally relies on within-person changes in the wage associated with a change in union status (nonunion-to-union or union-to-nonunion). Given the low level of new union organization (and, relatedly, the low level of decertification of existing union jobs) in the U.S. over the period we are studying implies that most changes in union status will occur when there is a job change. As we noted earlier, a weakness of the usual longitudinal data used to estimate union wage gaps, including our matched MORG data, is that these samples do not focus on job changers or the reason for any job changes. Observed union status changes may be due to within-job changes resulting from new union organization or decertification of an existing union (rare events), job reassignment, the result of involuntary job changes, the result of quits by workers perhaps related to offered wages, or classification error.

For comparison with the earlier literature using matched longitudinal data or panel data to estimate union earnings gaps, we use our matched MORG data from 1996-2024 for private sector full-time workers, described earlier (n=367874), to calculate the same earnings changes we presented for the DWS in equation 21. As before, we show this first uncorrected for classification error and then we show the corrected estimates. These estimates are:

$$\Delta^*_{MORG} = \begin{pmatrix} \Delta^*_{00} \\ \Delta^*_{01} \\ \Delta^*_{10} \\ \Delta^*_{11} \end{pmatrix} = \begin{pmatrix} 0.030 \\ 0.067 \\ -0.019 \\ 0.018 \end{pmatrix} \qquad \Delta_{MORG} = \begin{pmatrix} \Delta_{00} \\ \Delta_{01} \\ \Delta_{10} \\ \Delta_{11} \end{pmatrix} = \begin{pmatrix} 0.030 \\ 0.171 \\ -0.152 \\ 0.018 \end{pmatrix} \tag{32}$$

The estimates based on the MORG and adjusted for classification error implies similar union earnings gaps of 17.1 log points (18.7 percent) for union joiners and 15.2 log points (14.1 percent) for union leavers. The estimated earnings gap for union joiners in the MORG is similar to the gap for union joiners in the DWS (13.2 log points (14.1 percent) in the DWS). But the earnings gap for union leavers in the MORG is substantially smaller than the gap for union leavers in the DWS (41.1 log points (33.7 percent) in the DWS).

Since earnings change over time independently of union status, it is appropriate to calculate the difference-in-differences estimates of union status changes for the MORG as we did for the DWS. We compare the union joiners to those nonunion workers who remain nonunion,

and we compare the union leavers to those union workers who remain union. These D-i-D estimates for the MORG are

$$\Delta\Delta^*_{MORG} = \begin{pmatrix} \Delta^*_{01} - \Delta^*_{00} \\ \Delta^*_{10} - \Delta^*_{11} \end{pmatrix} = \begin{pmatrix} 0.037 \\ -0.037 \end{pmatrix} \quad \Delta\Delta_{MORG} = \begin{pmatrix} \Delta_{01} - \Delta_{00} \\ \Delta_{10} - \Delta_{11} \end{pmatrix} = \begin{pmatrix} 0.141 \\ -0.170 \end{pmatrix} \quad (33)$$

Given that year-to-year, workers who retain their union status realize some growth in earnings, these D-i-D estimates reduce the earnings gap based on union joiners and increase the union earnings gap based on union leavers. The corrected D-i-D estimates for the MORG data show little to no asymmetry between the average earnings gains for union joiners (14.1 log points, 15.1 percent) relative to the average earnings loss suffered by union leavers (17 log points, 15.6 percent). This is contrast to the D-i-D estimates of the earnings gap for union joiners and union leavers in the DWS (equation 22), where there is substantial asymmetry, with union leavers suffering a much larger earnings loss (38.3 log points, 31.8 percent) and union joiners experiencing a smaller earnings gain (18 log points, 20 percent). The smaller earnings penalty for union leavers in the MORG relative to the DWS likely reflects the fact that the MORG estimates include workers who do not change jobs as well as those who change jobs voluntarily. As such, they ignore the underlying selection process driving union status transitions.

Finally, we account for variation in earnings and union status related to observables by implementing the full ML model on union transitions and earnings changes with the MORG data, replicating what we did for the DWS with one notable exception. There are computation problems estimating the model with the assumed misclassification rate of $\lambda = 0.0174$ because of very low union status transition rates in the MORG data (table 8).[29] We "solve" this problem by cutting the assumed misclassification rate in half to $\lambda = 0.0087$.

These ML results for the MORG are contained in table 9 along with estimates of the ML model for the DWS with the new lower misclassification rate. The first two columns of the table contain the estimates for the two samples uncorrected for classification error. The last two columns contain the estimates for the two samples corrected for misclassification

[29] A manifestation of this problem is that the ML optimization algorithm drives the correlation of the two errors in the bivariate probit union status model to 1 when we assume $\lambda = 0.0174$.

Table 9: ML D-i-D Estimates of Longitudinal Union Earnings Gaps ($\Delta\Delta$)

Comparison of MORG and DWS

	(1) Observed $\lambda = 0$ MORG	(2) Observed $\lambda = 0$ DWS	(3) Corrected $\lambda = 0.0087$ MORG	(4) Corrected $\lambda = 0.0087$ DWS
Δ_{01} $(\delta_{01} - \delta_{00})$	0.037 (0.004)	0.075 (0.025)	0.099 (0.006)	0.122 (0.029)
Δ_{10} $(\delta_{10} - \delta_{11})$	-0.043 (0.004)	-0.213 (0.030)	-0.112 (0.006)	-0.243 (0.023)
Δ_{11} $(\delta_{11} - \delta_{00})$	-0.005 (0.002)	0.029 (0.021)	-0.005 (0.002)	0.031 (0.019)
p-value H_0: $\Delta_{01} = -\Delta_{10}$	0.327	0.001	0.459	0.001
p-value H_0: $\Delta_{11} = 0$	0.024	0.182	0.027	0.106
Sample size	367874	15195	367874	15195

Note: Controls in all models are included in the two bivariate probit union status equations and in the earnings change function. These include indicators for 5 age categories, 5 education categories, female, married, female x married, nonwhite, 4 industry categories, and survey year years (16 in DWS from 1994-2024, 29 in MORG from 1994-2024. All models are weighted by the CPS final sample weights. Robust standard errors are in parentheses. MORG full model estimates are contained in Appendix Table A8.

with $\lambda = 0.0087$. Even with the lower misclassification rate, the corrected D-i-D estimates of the effect of union status transitions (columns 3 and 4) are substantially larger than the uncorrected estimates (columns 1 and 2). Due to the smaller assumed misclassification rate, the attenuation in the estimated earnings gaps is smaller than we have seen earlier, but it is still substantial.

The estimates in table 9 continue to show the symmetry in earnings changes between union joiners and union leavers in the MORG together with a substantial asymmetry for job losers in the DWS. The fact that at least some union status transitions in the MORG are the result of voluntary choices of workers related to offered wages could account for the contrast between the symmetry in the MORG and the asymmetry in the DWS. This suggests that it is dangerous to interpret earnings changes associated with union status changes in the usual longitudinal data (e.g. the MORG) as consistent estimates of the union earnings gap. To the extent that the DWS estimates are closer to the underlying effects, a substantive conclusion is that involuntary loss of union status has an immediate substantial negative effect on earnings, while the finding a union job after loss of a nonunion job has a smaller positive effect on earnings.

7 Understanding DWS Estimates of the Earnings Gap

Our estimates of the union earnings gap based on the earnings changes of job losers who change union status addresses some of the selection issues related to the use of longitudinal data. Specifically, it eliminates the selection induced by individuals making voluntary job/position changes (within and between firms) that may involve a change in union status. In this section, we further investigate selection issues regarding union status and heterogeneity in the earnings gap across workers.

7.1 The Persistence of Union Status among Job Losers

We noted earlier while presenting union transition probabilities that there is substantial persistence of union status after a job loss. Despite the fact that only a very small fraction of jobs in the economy are organized, union job losers are quite likely to be reemployed in a union job. At least part of the success union job losers have in finding a new union job is likely due to a strong union referring job losers to other union jobs. A particular example is the set of craft unions common in the construction industry. Construction jobs are typically of relatively short duration (construction projects are completed and the workers are laid off). Craft workers in these unions are strongly attached to their union, and less so to their employer. Unions in the construction industry play a central role in referring unemployed members to unionized firms with new or ongoing projects. Unions in other industries often play some (perhaps less formal) role in matching unemployed members to union jobs, contributing to the persistence of union status among job losers.

We examine this further by calculating union status transition rates and union persistence probabilities by industry. Table 10 contains, separately for 4 groups of industries, the corrected union density (on the lost job) in column 1, union status transition rates for job losers in columns 2-5, and the union status persistence rate, $\theta_{11}/(\theta_{10} + \theta_{11})$, in column 6. The last row of the table presents the numbers for all workers in our DWS sample. These numbers confirm the low aggregate union density among job losers (6.7 percent overall) and the low rates of union status changes and the high union status persistence rate (61.5 percent overall). Most workers are nonunion on both jobs, and only 3.9 percent change union status.

Table 10: Union Density and Transition Probabilities, by Industry on Lost Job Corrected for Classification Error ($\lambda = 0.0174$)

Industry	(1) γ	(2) θ_{00}	(3) θ_{01}	(4) θ_{10}	(5) θ_{11}	(6) $\theta_{11}/(\theta_{10}+\theta_{11})$	(7) N
Construction	0.202	0.777	0.021	0.034	0.168	0.832	1802
Manufacturing	0.110	0.869	0.012	0.063	0.047	0.428	3679
Trans/Comm/Util	0.099	0.890	0.011	0.052	0.048	0.481	1310
Other	0.015	0.977	0.009	0.004	0.011	0.732	8577
All	0.067	0.920	0.013	0.026	0.041	0.615	15368

The value of γ is the corrected fraction of job losers who lost a union job (based on equation 4. The θ_{ij} are corrected estimates of the probability of making a union status transition from i to j (based on equation 19 and shown in table 10). The data are from the 1994-2024 DWS. The Sample includes full-time private sector job losers who are reemployed full-time in private sector jobs at the DWS survey date. Weighted by CPS final weights at the DWS survey date.

We break the sample of job losers into four groups by broad industry on the lost job: 1) Construction, 2) Manufacturing, 3) Transportation, Communications, and Public Utilities (TCPU), and 4) Other. More than half the workers are in Other. Taken as a group of industries, the corrected union density in Other is very small at 1.5 percent, and there is no subset of industries of sufficient size with reasonably higher union density to feasibly examine on its own. The three specific industry groups we examine directly have substantially higher union density. Most prominent is Construction, with a union density of 20 percent. The other two specific industries (Manufacturing and TCPU), have union density of about 10 percent. Not surprisingly, almost all job losers in Other industries make a N-N transition (98 percent), while only 77 percent of construction job losers are N-N and 17 percent are U-U. Manufacturing and TCPU are intermediate between Construction and Other.

The persistence of union membership is calculated as the fraction of union job losers who are reemployed in a union job and is presented by industry in column (6) of table 10. Persistence rates are calculated directly from the corrected transition rates in the table as $\Pr[U_{new} = 1 | U_{old} = 1] = \theta_{11}/(\theta_{10} + \theta_{11})$. Union membership persistence is highest in the Construction industry at 83 percent, reflecting the strong ties of construction workers to their craft union, The persistence rate is smaller for Manufacturing and TCPU job losers, at

43 and 48 percent respectively. All of these persistence rates are much larger than would be expected if union and nonunion job losers were randomly assigned union status in the new job from the pool of private sector jobs in the economy.[30]

7.2 Cross-Section Earnings and Union Membership Dynamics

The persistence of union status among job losers raises issues of selection related to the individual and/or jobs. A key question is whether workers who change their union status after a job loss are different from those who maintain their union status in ways related to earnings at a point in time. In order to address this, we revisit the cross-section earnings functions we estimated in section 5.1. We augment the cross-section earnings function specified in equation 23 with a full set of longitudinal union status transition indicators. This is

$$w_{0i} = X_i\beta + \psi^0_{01}T_{01i} + \psi^0_{10}T_{10i} + \psi^0_{11}T_{11} + \epsilon_i, \tag{34}$$

where w_{0i} is log earnings on the old job for individual i, T_{jk} is an indicator variable for a union status transition from j to k and ψ^0_{jk} is the difference in log earnings on the old job of a worker who makes a j to k union status transition relative to the log earnings of a worker who is nonunion in both jobs. The base case, where only union status on the old job affects old job earnings, implies a pair of constraints on equation 34: 1) $\psi^0_{10} = \psi^0_{11}$ and 2) $\psi^0_{01} = 0$. The analog for earnings on the new job is

$$w_{1i} = X_i\beta + \psi^1_{01}T_{01i} + \psi^1_{10}T_{10i} + \psi^1_{11}T_{11} + \epsilon_i, \tag{35}$$

where w_{1i} is log earnings on the new job for individual i. The analogous constraints are 1) $\psi^1_{01} = \psi^1_{11}$ and 2) $\psi^1_{10} = 0$. As we show below, both of these pairs of constraints are decisively rejected, implying that there is selection of workers on unobservables that affects union status across jobs.

Table 11 contains estimates of the coefficients (ψ_{01}, ψ_{10}, and ψ_{11}) of the union status transition indicators in equations 34 and 35. Column (1) contains OLS estimates based on

[30] The high persistence rate in the Other category is not very reliable given the very small share (1.5 percent) of job losers in Other that are union.

Table 11: Cross-Section Earnings Gaps, With Union Status Transitions

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$
(1) Old job – current U		
Union on old job (ψ^0)	0.185 (0.016)	0.251 (0.019)
(2) Old job – transitions		
ΔUnion joiner (ψ^0_{01})	-0.019 (0.025)	-0.017 (0.024)
ΔUnion leaver (ψ^0_{10})	0.084 (0.020)	0.136 (0.026)
ΔUnion stayer (ψ^0_{11})	0.292 (0.024)	0.347 (0.025)
(3) New job – current U		
Union on new job (ψ^1)	0.206 (0.019)	0.289 (0.021)
(4) New job – transitions		
ΔUnion joiner (ψ^1_{01})	0.055 (0.027)	0.095 (0.033)
ΔUnion leaver (ψ^1_{10})	-0.107 (0.022)	-0.105 (0.021)
ΔUnion stayer (ψ^1_{11})	0.310 (0.024)	0.378 (0.025)

Note: ψ = standard union earnings gap. ψ_{jk} = earnings gap of job losers making a j to k union status transition relative to an N-N transition. All specifications include controls comprised of indicators for 5 age categories, 5 education categories, female, married, female x married, nonwhite, 4 industry categories, and 16 DWS survey years from 1994-2024. All models are weighted by the CPS final sample weights. Robust standard errors are in parentheses. Sample size of 15195 reemployed full-time involuntary job changers.

observed union status on the two jobs. Column (2) contains ML estimates corrected for misclassification. We focus our discussion on the corrected estimates and start with earnings on the old (lost) job. Panel 1 of the table contains standard cross-section estimates of the cross-section earnings gap on the old job, and the estimate of 25.1 log points (28.5 percent) is consistent with the estimate of the same quantity in table 5. Panel 2 of the table contains estimates of the cross-section earnings gap on the old job for the union status transitions relative to the earnings of workers making a N-N transition. These estimates show a sharp contrast between the cross-section old job earnings gap of union leavers (ψ^0_{10}) at 13.6 log points (14.6 percent) and union stayers (ψ^0_{11}) at 34.7 log points (41.5 percent) despite the fact that both union leavers and union joiners are union on the old job. In other words, among workers who lose a union job, those who find a new union job earned substantially

more on the old job than those who find a new nonunion job.[31] There is no significant difference between the cross-section earnings of union joiners and workers who are nonunion on both job (ψ^0_{11}). These findings suggest there is unobserved heterogeneity in worker or job quality on the old job that is positively related to union status over time. The fact that union joiners do not have a wage advantage on the old job suggests that there is more complicated selection process in operation. It is not simply worker quality that is related to union status.

Panels 3 and 4 of table 11 contains analogous results for cross-section estimation of the earnings gap on the new job. Panel 3 of the table contains standard cross-section estimates of the cross-section earnings gap on the new job., and the estimate of 28.9 log points (33.5 percent) is in line with the estimate of the same quantity in panel 1 of the table and in table 5. Panel 4 of the table contains estimates of the cross-section earnings gap on the new job for the union status transitions relative the earnings of workers making a N-N transition. These estimates show a sharp contrast between the cross-section new job earnings gap of union stayers (ψ^1_{11}) at 37.8 log points (45.9 percent) and union joiners (ψ^1_{01}) at 9.5 log points (10 percent) despite the fact that both union stayers and union joiners are union in the new job. In other words, among job losers who find a new union job, workers who were in a union on the lost job earn substantially more on the new job than those job losers who were not unionized on the old job. Interestingly, union job losers who are reemployed on a nonunion job earn significantly less on the new job than nonunion job losers who are reemployed in a nonunion job ($\hat{\psi}^1_{10} = -10.5$ log points (9 percent)). These findings suggest that losers of union jobs who are unable to find a new union job may be of lower unobserved quality than the average worker who is nonunion on both jobs.

8 Conclusions

This study contains new estimates of the union-nonunion earnings gap based on the experience of reemployed involuntary job losers who make union status transitions between the

[31] The hypothesis that the estimated cross-section wage gaps on the old job are equal for those are union leavers and those who are union stayers ($\psi^0_{10} = \psi^0_{11}$) is sharply rejected.

lost job and a new job. The core analysis uses the Displaced Workers Survey (DWS) in the U.S. over the 1994-2024 period. The focus is on losers of full-time private-sector jobs who are reemployed on full-time private-sector jobs. Using involuntary job losers addresses at least part of the selection problem in the determination of union status that is an issue in most existing longitudinal analyses of the earnings gap.

A central feature of our analysis is that we account explicitly for classification error in union status. We are not the first to note that misclassification of union status can lead to seriously attenuated estimates of the earnings gap. This is a particularly serious problem in the low union density environment that characterizes the U.S. private sector. The attenuation effect on the earnings gap is magnified when considering union status transitions, where there are two possible classification errors in point-in-time union status that can result in a misclassified transition. We address this directly by specifying a probability model for the misclassification that yields an algebraic solution for deriving corrected union status and transition probabilities and corrected estimates of the earnings gap as a function of the observed quantities. We further specify a maximum likelihood model to estimate corrected union status and transition probabilities jointly with corrected estimates of the earnings gap that allows controlling for important observable characteristics of workers.

One noteworthy finding is that union status is highly persistent in that reemployed losers of union jobs are more likely than not to be in a new union job. This is true despite that fact that less than 10 percent of private sector workers in the U.S. are union members in the period studied. Related to this, our estimates show that, even in the low union setting of the US private sector, union status is highly valuable for individual workers in terms of higher earnings. Critical to this are the dynamics of union membership and earnings. When displaced union workers change jobs, if they retain union membership status in their new job, they enjoy a sizable cross-sectional earnings advantage on the old job relative to those union job losers who are reemployed in nonunion jobs. There is strong evidence of selection of those who retain union membership after job loss, reflected in both the persistence of union membership status and our finding that the workers who gain most from union status are those who are union members in both their old and new jobs.

Another finding is that there is asymmetry in the earnings changes experienced by union

leavers (a U-N transition) and that experienced by union joiners (a N-U transition). Union leavers on average lose about 28 percent in earnings relative to union stayers, while union joiners on average gain about 19 percent in earnings relative to those who remain nonunion. This asymmetry is not consistent with most earlier work using longitudinal data to estimate the earnings gap. These studies, using data like the CPS MORG, consider all union status changers, including those with the same employer and voluntary job changers as well as involuntary job changers. The finding in these studies is that there is rough parity in the earnings gain realized by union joiners and the earnings loss experienced by union leavers. Our analysis of union transitions and earnings gaps in the matched MORG data shows an earnings gap of about 10 percent for both leavers and joiners accounting for classification error, confirming the earlier findings. The finding of lack of symmetry of earnings changes among displaced workers is an empirically important feature of the cost of job loss that has gone largely unnoticed in the very sizable literature studying wage and earnings losses from job displacement.

A future research challenge is to reconcile the different earnings trajectories associated with union status changes for involuntary job movers as compared to voluntary job movers. As this paper shows, issues of misclassification in longitudinal data sources are an important feature to be addressed in such an analysis. A way forward will be to study earnings differentials and membership dynamics in longitudinal data where wage changes associated with union status changes can be studied separately for within-employer transitions, voluntary job changes, and involuntary job changes.

References

Abowd, John M. and Henry S. Farber."Job Queues and the Union Status of Workers," *Industrial and Labor Relations Review*, Vol 35, No. 3 (April 1982), pp. 354-367.

Athey, Susan, Lisa K. Simon, Oskar Nordstrom Skans, Johan Vikstrom, and Yaroslav Yakymovych "The Heterogeneous Earnings Impact of Job Loss Across Workers, Establishments, and Markets", National Bureau of Economic Research Working Paper #34946 (March 2026).

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, Vol 90, No. 2 (May 2008), pp. 300-323.

Beauregard, Pierre-Loup, Thomas Lemieux, Derek Messner and Raffaele Saggio."Why Do Union Jobs Pay More? New Evidence from Matched Employer-Employee Data", National Bureau of Economic Research Working Paper #33740 (May 2025).

Blanchflower, David G. and Alex Bryson. "Unions, Wages, and Hours," *British Journal of Industrial Relations*, Volume 63, No. 3 (September 2025), pp. 1-24).

Card, David. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* Vol 64, No. 4 (July 1996), pp. 957-979.

Chowdhury, Gopa and Stephen Nickell. "Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness, and Unemployment using PSID Data," *Journal of Labor Economics* Vol 3, No. 1 (January 1985), pp. 38-69.

Couch, Kenneth and Dana Placzek. "Earnings Losses of Displaced Workers Revisited," *American Economic Review* Vol 100, No. 1 (March 2010), pp. 572-589.

Farber, Henry S. "The Determination of the Union Status of Workers," *Econometrica*, Vol 51, No. 5 (September 1983), pp. 1417-1437.

Farber, Henry S. "Nonunion Wage Rates and the Threat of Unionization," *Industrial and Labor Relations Review*, Vol 58, No. 3 (April 2005), pp. 335-352.

Farber, Henry S. "Employment, Hours, and Earnings Consequences of Job Loss: U.S. Evidence from the Displaced Workers Survey," *Journal of Labor Economics*, 35(S1) July 2017, pp. S235-S272.

Farber, Henry S. and Bruce Western. "Accounting for the Decline of Unions in the Private Sector, 1973-1998," *Journal of Labor Research*, Summer 2001. pp. 459-485.

Fortin, Nicole, Thomas Lemieux, and Neil Lloyd, "Labor Market Institutions and the Distribution of Wages: The Role of Spillover Effects," *Journal of Labor Economics* Vol 39, No. S2 (2021), pp. S369-S412.

Freeman, Richard B. "Longitudinal Analyses of the Effects of Trade Unions," *Journal of Labor Economics* Vol 2, No. 1 (January 1984), pp. 1-26.

Jakubson, George. "Estimation and Testing of the Union Wage Effect Using Panel Data," *Review of Economic Studies* Vol 58, No. 5 (October 1991), pp. 971-991.

Kuhn, Peter and Arthur Sweetman. "Vulnerable Seniors: Unions, Tenure, and Wages following Permanent Job Loss," *Journal of Labor Economics* Vol 17, No. 4, pt. 1 (October 1999), pp. 671-693.

Kulkarni, Abhir and Barry T. Hirsch. "Revisiting Union Wage and Job Loss Effects Using the Displaced Worker Surveys," *Industrial and Labor Relations Review* Vol 84, No. 4 (August 2021), pp.948-976.

Lewis, H. Gregg. *Unionism and Relative Wages in the United States: An Empirical Inquiry.* Chicago: University of Chicago Press, 1963.

Lewis, H. Gregg. *Union Relative Wage Effects: A Survey.* Chicago: University of Chicago Press, 1986.

Mellow, Wesley. "Unionism and Wages: A Longitudinal Analysis," *Review of Economics and Statistics* Vol 63, No. 1 (February 1981), pp.43-52.

Mellow, Wesley and Hal Sider. "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics* Vol. 1, No. 4 (October 1983), pp. 331-344.

Raphael, Stephen "Estimating the Union Earnings Effect Using a Sample of Displaced Workers," *Industrial and Labor Relations Review* Vol 55, No. 3 (April 2000), pp.513-521.

Rosen, Sherwin, "Trade Union Power, Threat Effects, and the Extent of Organization," *Review of Economic Studies* Vol 36, No 2 (April 1969), pp. 185-196.

Appendix

I – Correcting the Longitudinal Wage Gap for Classification Error

The algebra to derive the classification-error adjusted wage changes is an extension of what we developed earlier for the true union status transition probabilities. We define two 4x4 diagonal matrices to simplify the notation. First is a matrix with the observed union sequence probabilities on the diagonals. This is

$$\Theta_d^* = \begin{pmatrix} \theta_{00}^* & 0 & 0 & 0 \\ 0 & \theta_{01}^* & 0 & 0 \\ 0 & 0 & \theta_{10}^* & 0 \\ 0 & 0 & 0 & \theta_{11}^* \end{pmatrix}. \tag{36}$$

The second is a matrix with the true union sequence probabilities on the diagonals. This is

$$\Theta_d = \begin{pmatrix} \theta_{00} & 0 & 0 & 0 \\ 0 & \theta_{01} & 0 & 0 \\ 0 & 0 & \theta_{10} & 0 \\ 0 & 0 & 0 & \theta_{11} \end{pmatrix}. \tag{37}$$

The vector of observed log earnings changes associated with each observed union transition is a linear function of the vector of true log earnings changes defined by

$$\Delta^* = (\Theta_d^*)^{-1} M \Theta_d \Delta. \tag{38}$$

The system of four linear equations defined in equation 38 can be inverted to solve for the true average log earnings changes (the Δ vector), yielding

$$\Delta = \Theta_d^{-1} M^{-1} \Theta_d^* \Delta^*, \tag{39}$$

where the matrix M is defined in equation 17 as a function of the assumed classification error rate (λ), the elements of the diagonal matrix Θ_d^* are the observed union status transition rates, the elements of the vector Δ^* are the observed log earnings changes associated with each observed union status transition, and the elements of the diagonal matrix Θ_d are the union status transition probabilities corrected for classification error (defined in equation 19 as a function of λ and Θ^*).

II – Contributions to the likelihood function for observed earnings levels and union status in the cross-section

In the case where a worker is observed nonunion ($U_i^* = 0$), $\epsilon_i^* = w_i - X_i\beta$ if there has been no classification error ($U_i = 0$) and $\epsilon_i^* = w_i - X_i\beta - \delta$ if there has been a classification error ($U_i = 1$). The probability density of observing ϵ_i^* and $U_i^* = 0$ is

$$\begin{aligned} f(\epsilon_i^*, U_i^* = 0) &= P(U_i = 0, U_i^* = 0)\phi\left(\frac{w_i - X_i\beta}{\sigma}\right) + P(U_i = 1, U_i^* = 0)\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right) \\ &= (1-\lambda)(1-\gamma_i)\phi\left(\frac{w_i - X_i\beta}{\sigma}\right) + \lambda\gamma_i\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right) \\ &= (1-\lambda)(1-\Phi(Z_i\alpha))\phi\left(\frac{w_i - X_i\beta}{\sigma}\right) + \lambda\Phi(Z_i\alpha)\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right), \quad (40) \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal CDF and $\phi(\cdot)$ is the standard normal PDF.

In the case where the worker is observed union ($U_i^* = 1$), $\epsilon_i^* = w_i - X_i\beta - \delta$ if there has been no classification error ($U_i = 1$) and $\epsilon_i^* = w_i - X_i\beta$ if there has been a classification error ($U_i = 0$). The probability density of observing ϵ_i^* and $U_i^* = 1$ is

$$\begin{aligned} f(\epsilon_i^*, U_i^* = 1) &= P(U_i = 1, U_i^* = 1)\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right) + P(U_i = 0, U_i^* = 1)\phi\left(\frac{w_i - X_i\beta}{\sigma}\right) \\ &= (1-\lambda)\gamma_i\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right) + \lambda(1-\gamma_i)\phi\left(\frac{w_i - X_i\beta}{\sigma}\right) \\ &= (1-\lambda)\Phi(Z_i\alpha)\phi\left(\frac{w_i - X_i\beta - \delta}{\sigma}\right) + \lambda(1-\Phi(Z_i\alpha))\phi\left(\frac{w_i - X_i\beta}{\sigma}\right). \quad (41) \end{aligned}$$

The probabilities defined in equations 40 and 41 are the basis of a log-likelihood function for the joint probability density of observed earnings and observed union status that can be maximized with respect to the model parameters (δ, α, β, and σ). As before, we rely on our analysis of the Mellow and Sider (1983) data and use a value of $\lambda = 0.0174$ for the private sector.

III – Contributions to the likelihood function for observed earnings changes and union transitions in longitudinal data

1. Observed nonunion to nonunion transition ($T^*_{00i} = 1$):

$$\begin{aligned} L_{00i} &= P(T_{00i}=1, T^*_{00i}=1)\phi\left(\frac{\Delta w_i - X_i\beta}{\sigma}\right) + P(T_{01i}=1, T^*_{00i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{01}}{\sigma}\right) \\ &+ P(T_{10i}=1, T^*_{00i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{10} - D_{11}}{\sigma}\right) + P(T_{11i}=1, T^*_{00i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{11}}{\sigma}\right). \end{aligned} \tag{42}$$

2. Observed nonunion to union transition ($T^*_{01i} = 1$):

$$\begin{aligned} L_{01i} &= P(T_{00i}=1, T^*_{01i}=1)\phi\left(\frac{\Delta w_i - X_i\beta}{\sigma}\right) + P(T_{01i}=1, T^*_{01i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{01}}{\sigma}\right) \\ &+ P(T_{10i}=1, T^*_{01i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{10} - D_{11}}{\sigma}\right) + P(T_{11i}=1, T^*_{01i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{11}}{\sigma}\right). \end{aligned} \tag{43}$$

3. Observed union to nonunion transition ($T^*_{10i} = 1$):

$$\begin{aligned} L_{10i} &= P(T_{00i}=1, T^*_{10i}) = 1\phi\left(\frac{\Delta w_i - X_i\beta}{\sigma}\right) + P(T_{01i}=1, T^*_{10i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{01}}{\sigma}\right) \\ &+ P(T_{10i}=1, T^*_{10i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{10} - D_{11}}{\sigma}\right) + P(T_{11i}=1, T^*_{10i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{11}}{\sigma}\right). \end{aligned} \tag{44}$$

4. Observed union to union transition ($T^*_{11i} = 1$):

$$\begin{aligned} L_{11i} &= P(T_{00i}=1, T^*_{11i}=1)\phi\left(\frac{\Delta w_i - X_i\beta}{\sigma}\right) + P(T_{01i}=1, T^*_{11i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{01}}{\sigma}\right) \\ &+ P(T_{10i}=1, T^*_{11i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{10} - D_{11}}{\sigma}\right) + P(T_{11i}=1, T^*_{11i}=1)\phi\left(\frac{\Delta w_i - X_i\beta - D_{11}}{\sigma}\right). \end{aligned} \tag{45}$$

Each of the 16 joint probabilities in equations 42-45, $P(T_{jki}, T^*_{mni})$, can be expressed using Bayes rule as

$$P(T_{jki}, T^*_{mni}) = P(T^*_{mni}|T_{jki})P(T_{jki}). \tag{46}$$

The conditional probability of observing a particular transition given a true transition, $P(T^*_{mni}|T_{jki})$, is strictly a function of the assumed misclassification rate λ using appropriate elements of the matrix M in equation 17. These conditional probabilities depend on the number of misclassifications implied by the number of differences between m and i and between n and j.[32] The true probabilities of a particular union status transitions ($P(T_{jki})$) are objects for estimation as a function of observed variables. When analyzing union status transitions, we implement multivariate models of these probabilities ($P(T_{jki})$) as bivariate probit models as specified in equations 25-28.

[32] The conditional probability, $P(T^*_{mni}|T_{jki})$, where there are no misclassifications ($m = i$ and $n = j$) is $(1-\lambda)^2$. Where there is one misclassification ($m \neq i$ or $n \neq j$ but not both inequalities), the conditional probability) is $\lambda(1-\lambda)$. Where there are two misclassifications ($m \neq i$ and $n \neq j$), the conditional probability) is λ^2.

Online Appendix

I - Union Status Cross-Section Estimates

Table A1: Estimates of Old Job DWS Union Membership Status

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$
Age 25-34	0.037 (0.010)	0.040 (0.012)
Age 35-44	0.052 (0.010)	0.058 (0.012)
Age 45-54	0.058 (0.010)	0.062 (0.012)
Age 55-64	0.056 (0.012)	0.061 (0.013)
High school	0.037 (0.008)	0.038 (0.008)
Some college	0.032 (0.008)	0.036 (0.008)
College degree	-0.022 (0.010)	-0.030 (0.011)
Advanced degree	-0.042 (0.014)	-0.085 (0.033)
Female	-0.024 (0.007)	-0.023 (0.008)
Married	0.009 (0.006)	0.011 (0.005)
Female x Married	-0.007 (0.010)	-0.019 (0.011)
Nonwhite	0.017 (0.006)	0.009 (0.007)
Construction	0.116 (0.006)	0.118 (0.007)
Manufacturing	0.078 (0.005)	0.087 (0.006)
Trans/Comm/Util	0.077 (0.007)	0.085 (0.008)
Year dummies (15)	Yes	Yes
Sample size	18957	18957

Note: Column 1 contains marginal effects from a union status probit estimated without accounting for classification error. Column 2 contains maximum likelihood estimates of marginal effects from union status probit accounting for classification error. The base group consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

Table A2: Estimates of New Job DWS Union Membership Status

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$
Age 25-34	0.028 (0.010)	0.024 (0.011)
Age 35-44	0.032 (0.010)	0.031 (0.011)
Age 45-54	0.037 (0.010)	0.036 (0.011)
Age 55-64	0.032 (0.011)	0.032 (0.012)
High school	0.018 (0.008)	0.023 (0.008)
Some college	0.014 (0.008)	0.023 (0.009)
College degree	-0.032 (0.010)	-0.035 (0.012)
Advanced degree	-0.043 (0.014)	0.120 (0.143)
Female	-0.014 (0.007)	-0.016 (0.008)
Married	0.007 (0.005)	0.009 (0.005)
Female x Married	-0.026 (0.010)	-0.034 (0.013)
Nonwhite	0.015 (0.006)	0.012 (0.007)
Construction	0.109 (0.006)	0.106 (0.006)
Manufacturing	0.059 (0.005)	0.066 (0.006)
Trans/Comm/Util	0.066 (0.007)	0.070 (0.008)
Year dummies (15)	Yes	Yes
Sample size	16975	16975

Note: Column 1 contains marginal effects from a union status probit estimated without accounting for classification error. Column 2 contains maximum likelihood estimates of marginal effects from union status probit accounting for classification error. The base group consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

Table A3: Estimates of CPS-OGRG Union Membership Status

	(1) Observed $\lambda = 0$	(2) Corrected $\lambda = 0.0174$
Age 25-34	0.025 (0.001)	0.026 (0.001)
Age 35-44	0.041 (0.001)	0.041 (0.001)
Age 45-54	0.057 (0.001)	0.062 (0.001)
Age 55-64	0.062 (0.001)	0.067 (0.001)
High school	0.029 (0.001)	0.031 (0.001)
Some college	0.023 (0.001)	0.026 (0.001)
College degree	-0.016 (0.001)	-0.023 (0.001)
Advanced degree	-0.016 (0.001)	-0.039 (0.002)
Female	-0.014 (0.001)	-0.021 (0.001)
Married	0.006 (0.001)	0.006 (0.001)
Female x Married	-0.012 (0.001)	-0.016 (0.001)
Nonwhite	0.021 (0.001)	0.021 (0.001)
Construction	0.080 (0.001)	0.078 (0.001)
Manufacturing	0.056 (0.001)	0.059 (0.001)
Trans/Comm/Util	0.097 (0.001)	0.095 (0.001)
Year dummies (31)	Yes	Yes
Sample size	1325297	132597

Note: Column 1 contains marginal effects from a union status probit estimated without accounting for classification error. Column 2 contains maximum likelihood estimates of marginal effects from union status probit accounting for classification error. The base group consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

II - Full Cross-Section Earnings Estimates

Table A4: Estimates of Old Job Cross-Section DWS Union Earnings Function

Corresponding to the old job estimates of δ in the first row of Table 5

	(1)	(2)	(3)
	OLS ($\lambda = 0$)	MLE ($\lambda = 0.0174$)	
	Earnings	Earnings	Union
δ, old job	0.183 (0.015)	0.252 (0.018)	
Age 25-34	0.230 (0.015)	0.227 (0.015)	0.042 (0.012)
Age 35-44	0.418 (0.016)	0.413 (0.016)	0.059 (0.012)
Age 45-54	0.466 (0.017)	0.462 (0.017)	0.063 (0.012)
Age 55-64	0.480 (0.019)	0.476 (0.019)	0.063 (0.013)
High school	0.265 (0.015)	0.261 (0.015)	0.035 (0.008)
Some college	0.437 (0.015)	0.434 (0.015)	0.033 (0.008)
College degree	0.803 (0.017)	0.804 (0.017)	-0.037 (0.011)
Advanced degree	1.059 (0.023)	1.062 (0.023)	-0.156 (0.110)
Female	-0.128 (0.013)	-0.128 (0.013)	-0.024 (0.008)
Married	0.175 (0.012)	0.174 (0.012)	0.009 (0.005)
Female x Married	-0.143 (0.017)	-0.140 (0.017)	-0.019 (0.011)
Nonwhite	-0.100 (0.012)	-0.099 (0.012)	0.007 (0.007)
Construction	0.125 (0.014)	0.114 (0.014)	0.117 (0.007)
Manufacturing	0.043 (0.010)	0.037 (0.010)	0.087 (0.006)
Trans/Comm/Util	0.124 (0.016)	0.120 (0.016)	0.084 (0.008)
Year dummies (15)	Yes	Yes	Yes
Sample size	18821	18821	

Note: Column 1 contains OLS estimates of earnings equation estimated via OLS without accounting for classification error. Columns 2 and 3 contain maximum likelihood estimates of earnings function (column 2) and marginal effects from union status probit (column 3) estimated jointly accounting for classification error. The base group for union status consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. The base group for earnings additionally consists of nonunion workers. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

Table A5: Estimates of New Job Cross-Section DWS Union Earnings Function

Corresponding to the new job estimates of δ in the second row of Table 5

	(1)	(2)	(3)
	OLS ($\lambda = 0$)	MLE ($\lambda = 0.0174$)	
	Earnings	Earnings	Union
δ, new job	0.193 (0.018)	0.284 (0.022)	
Age 25-34	0.157 (0.016)	0.156 (0.015)	0.022 (0.010)
Age 35-44	0.293 (0.016)	0.291 (0.016)	0.029 (0.010)
Age 45-54	0.309 (0.017)	0.306 (0.017)	0.034 (0.010)
Age 55-64	0.298 (0.020)	0.296 (0.020)	0.029 (0.012)
High school	0.258 (0.015)	0.254 (0.015)	0.022 (0.008)
Some college	0.420 (0.015)	0.415 (0.015)	0.022 (0.008)
College degree	0.815 (0.018)	0.815 (0.018)	-0.039 (0.012)
Advanced degree	1.077 (0.024)	1.079 (0.024)	-0.383 (0.020)
Female	-0.110 (0.014)	-0.109 (0.014)	-0.017 (0.008)
Married	0.170 (0.013)	0.168 (0.013)	0.009 (0.005)
Female x Married	-0.134 (0.019)	-0.132 (0.017)	-0.032 (0.013)
Nonwhite	-0.102 (0.013)	-0.102 (0.013)	0.012 (0.007)
Construction	0.172 (0.016)	0.156 (0.016)	0.106 (0.006)
Manufacturing	0.070 (0.011)	0.063 (0.012)	0.067 (0.006)
Trans/Comm/Util	0.094 (0.019)	0.087 (0.019)	0.071 (0.008)
Year dummies (15)	Yes	Yes	Yes
Sample size	16851	16851	

Note: Column 1 contains OLS estimates of earnings equation estimated via OLS without accounting for classification error. Columns 2 and 3 contain maximum likelihood estimates of earnings function (column 2) and marginal effects from union status probit (column 3) estimated jointly accounting for classification error. The base group for union status consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. The base group for earnings additionally consists of nonunion workers. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

Table A6: Estimates of Cross-Section MORG Union Earnings Function

Corresponding to the estimates of δ in the first row of Table 7

	(1)	(2)	(3)
	OLS ($\lambda = 0$)	MLE ($\lambda = 0.0174$)	
	Earnings	Earnings	Union
δ	0.141 (0.002)	0.195 (0.003)	
Age 25-34	0.200 (0.002)	0.199 (0.002)	0.025 (0.002)
Age 35-44	0.349 (0.002)	0.347 (0.002)	0.044 (0.002)
Age 45-54	0.389 (0.002)	0.386 (0.002)	0.060 (0.002)
Age 55-64	0.372 (0.003)	0.369 (0.003)	0.064 (0.002)
High school	0.279 (0.002)	0.277 (0.002)	0.028 (0.001)
Some college	0.468 (0.002)	0.466 (0.002)	0.022 (0.001)
College degree	0.851 (0.002)	0.852 (0.002)	-0.039 (0.002)
Advanced degree	1.157 (0.003)	1.161 (0.003)	-0.069 (0.003)
Female	-0.151 (0.002)	-0.150 (0.002)	-0.022 (0.001)
Married	0.173 (0.002)	0.172 (0.002)	0.005 (0.001)
Female x Married	-0.127 (0.002)	-0.125 (0.002)	-0.017 (0.002)
Nonwhite	0.106 (0.002)	-0.107 (0.002)	0.022 (0.001)
Construction	0.141 (0.002)	0.135 (0.002)	0.079 (0.001)
Manufacturing	0.102 (0.002)	0.098 (0.002)	0.061 (0.001)
Trans/Comm/Util	0.152 (0.002)	0.144 (0.002)	0.096 (0.001)
Year dummies (29)	Yes	Yes	Yes
Sample size	916155	916155	

Note: Column 1 contains OLS estimates of earnings equation estimated via OLS without accounting for classification error. Columns 2 and 3 contain maximum likelihood estimates of earnings function (column 2) and marginal effects from union status probit (column 3) estimated jointly accounting for classification error. The base group for union status consists of single white males aged 20-24 with less than a high-school education employed in "other" industry in 1994. The base group for earnings additionally consists of nonunion workers. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

III - Full Longitudinal Estimates

Table A7: Estimates of Longitudinal DWS Union Earnings Gap

Corresponding to estimates of $\Delta\Delta$ in Table 6

	(1)	(2)	(3)	(4)
	OLS ($\lambda = 0$)	MLE ($\lambda = 0.0174$)		
	Changes in earnings	Changes in earnings	Union, old job	Union, new job
Difference-in-differences estimates (base keep same U)				
$\Delta\Delta_{01}$ $(\delta_{01} - \delta_{00})$	0.075 (0.025)	0.175 (0.048)		
$\Delta\Delta_{10}$ $(\delta_{10} - \delta_{11}))$	-0.213 (0.030)	-0.334 (0.042)		
$\Delta\Delta_{11}$ $(\delta_{11} - \delta_{00})$	0.029 (0.021)	0.035 (0.022)		
Age 25-34	-0.062 (0.016)	-0.060 (0.016)	0.042 (0.014)	0.021 (0.011)
Age 35-44	-0.126 (0.017)	-0.122 (0.017)	0.066 (0.014)	0.027 (0.011)
Age 45-54	-0.170 (0.017)	-0.166 (0.017)	0.068 (0.014)	0.031 (0.011)
Age 55-64	-0.203 (0.019)	-0.199 (0.019)	0.064 (0.016)	0.026 (0.012)
High school	-0.021 (0.015)	-0.020 (0.015)	0.031 (0.009)	0.018 (0.008)
Some college	-0.031 (0.016)	-0.031 (0.015)	0.030 (0.009)	0.019 (0.008)
College degree	-0.007 (0.017)	-0.007 (0.017)	-0.042 (0.012)	-0.040 (0.011)
Advanced degree	0.014 (0.021)	0.014 (0.021)	-0.061 (0.024)	-0.073 (0.022)
Female	0.015 (0.013)	0.016 (0.013)	-0.027 (0.009)	-0.023 (0.008)
Married	-0.005 (0.011)	-0.005 (0.011)	0.009 (0.006)	0.006 (0.005)
Female x Married	0.013 (0.016)	0.012 (0.016)	-0.015 (0.012)	-0.021 (0.012)
Nonwhite	-0.009 (0.011)	0.011 (0.011)	0.000 (0.008)	0.011 (0.007)
Construction, old	-0.098 (0.017)	-0.089 (0.016)	0.106 (0.007)	
Manufacturing, old	-0.072 (0.012)	-0.064 (0.012)	0.084 (0.007)	
Trans/Comm/Util, old	-0.059 (0.017)	-0.051 (0.017)	0.084 (0.008)	
Construction, new	0.121 (0.018)	0.111 (0.018)		0.093 (0.006)
Manufacturing, new	0.087 (0.011)	0.081 (0.011)		0.056 (0.006)
Trans/Comm/Util, new	0.058 (0.018)	0.052 (0.018)		0.063 (0.008)
Union status correlation ρ			0.917 (0.015)	
Year dummies (15)	Yes	Yes	Yes	Yes
Sample size	15195		15195	
P-values				
Test of $(\delta_{01} - \delta_{00}) = -(\delta_{10} - \delta_{11})$	0.001	0.013		
Test of $\delta_{11} = 0$	0.182	0.106		

Note: Column 1 contains OLS estimates of changes in earnings equation estimated via OLS without accounting for classification error. Columns 2, 3 and 4 contain maximum likelihood estimates of changes in earnings function (column 2) and marginal effects from union status bivariate probit (column 3, old job; column 4, new job; with correlation coefficient ρ) estimated jointly accounting for classification error. The base group for union status consists of single white males aged 20-24 with less than a high-school education in 1994 employed in "other" industry on both the old and new jobs. The base group for earnings changes additionally consists of workers who are nonunion on both jobs. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.

Table A8: Estimates of Longitudinal MORG Union Earnings Gap

Corresponding to estimates of $\Delta\Delta$ in Table 9

	(1)	(2)	(3)	(4)
	OLS ($\lambda = 0$)	MLE ($\lambda = 0.0087$)		
	Changes in earnings	Changes in earnings	Union, monis 4	Union, monis 8
Difference-in-differences estimates (base keep same U)				
$\Delta\Delta_{01}$ $(\delta_{01} - \delta_{00})$	0.037 (0.004)	0.099 (0.006)		
$\Delta\Delta_{10}$ $(\delta_{10} - \delta_{11}))$	-0.043 (0.004)	-0.112 (0.006)		
$\Delta\Delta_{11}$ $(\delta_{11} - \delta_{00})$	-0.005 (0.002)	0.005 (0.002)		
Age 25-34	-0.027 (0.003)	-0.027 (0.003)	0.031 (0.003)	0.022 (0.003)
Age 35-44	-0.047 (0.003)	-0.046 (0.003)	0.038 (0.003)	0.037 (0.003)
Age 45-54	-0.055 (0.003)	-0.055 (0.003)	0.064 (0.003)	0.052 (0.003)
Age 55-64	-0.063 (0.003)	-0.062 (0.003)	0.069 (0.003)	0.058 (0.003)
High school	0.002 (0.002)	0.002 (0.002)	0.028 (0.002)	0.026 (0.002)
Some college	0.006 (0.002)	0.006 (0.002)	0.019 (0.002)	0.018 (0.002)
College degree	0.014 (0.002)	0.014 (0.003)	-0.042 (0.002)	-0.044 (0.002)
Advanced degree	0.016 (0.003)	0.016 (0.003)	-0.055 (0.002)	-0.057 (0.003)
Female	0.003 (0.002)	0.003 (0.002)	-0.027 (0.002)	-0.027 (0.002)
Married	-0.003 (0.002)	-0.003 (0.002)	-0.000 (0.002)	0.001 (0.001)
Female x Married	0.000 (0.003)	0.000 (0.003)	-0.011 (0.002)	-0.012 (0.002)
Nonwhite	0.000 (0.002)	-0.000 (0.002)	0.022 (0.002)	0.023 (0.002)
Construction, old	-0.027 (0.004)	-0.024 (0.004)	0.067 (0.002)	
Manufacturing, old	-0.022 (0.003)	-0.021 (0.003)	0.048 (0.001)	
Trans/Comm/Util, old	-0.015 (0.004)	-0.012 (0.004)	0.082 (0.001)	
Construction, new	0.028 (0.004)	0.025 (0.004)		0.068 (0.002)
Manufacturing, new	0.017 (0.003)	0.016 (0.003)		0.049 (0.001)
Trans/Comm/Util, new	0.013 (0.004)	0.010 (0.004)		0.082 (0.001)
Union status correlation ρ			0.983 (0.001)	
Year dummies (29)	Yes	Yes	Yes	Yes
Sample size	367874		367874	
P-values				
Test of $(\delta_{01} - \delta_{00}) = -(\delta_{10} - \delta_{11})$	0.327	0.459		
Test of $\delta_{11} = 0$	0.024	0.027		

Note: Column 1 contains OLS estimates of changes in earnings equation estimated via OLS without accounting for classification error. Columns 2, 3 and 4 contain maximum likelihood estimates of changes in earnings function (column 2) and marginal effects from union status bivariate probit (column 3, monis 4; column 4, monis 8; with correlation coefficient ρ) estimated jointly accounting for classification error. The base group for union status consists of single white males aged 20-24 with less than a high-school education in 1994 employed in "other" industry on both the old and new jobs. The base group for earnings changes additionally consists of workers who are nonunion on both jobs. Robust standard errors are in parentheses. All models are weighted by the CPS final sample weights.