I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13509

Reconciling Occupational Mobility in the Current Population Survey

Christian vom Lehn
Cache Ellsworth
Zachary Kroff

JULY 2020

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13509

Reconciling Occupational Mobility in the Current Population Survey

Christian vom Lehn
Brigham Young University and IZA

Cache Ellsworth
Columbia Business School

Zachary Kroff
U.S. Census Bureau

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Reconciling Occupational Mobility in the Current Population Survey*

Measuring occupational mobility from the Current Population Survey using recall (retrospective) or linked panel responses (longitudinal) generates substantially different outcomes, both in levels and trends. Using a generalized method of moments technique, we estimate the actual level of occupational mobility and the measurement error in both of these measures for 1981-2018. Measurement error in longitudinal measures is large and has been worsening over time. However, actual occupational mobility is approximately 70% higher than retrospectively measures. Our estimated corrections imply workers in tradable occupations are less likely to switch occupations than previously believed, implying potentially lower welfare gains from trade.

JEL Classification: J62, C83, F16

Keywords: occupational switching, worker mobility, current population survey, measurement error, trade adjustment

Corresponding author:
Christian vom Lehn
Brigham Young University
Department of Economics
435 CTB
Provo, UT 84602
USA

E-mail: cvomlehn@byu.edu

* We are grateful for comments and suggestions from Enghin Atalay, Erhan Artuc, Jeff Denning, Eliza Forsythe, Brigham Frandsen, Giovanni Gallipoli, Aspen Gorry, Aaron Hedlund, Lars Lefgren, Christos Makridis, Erin Wolcott, Abigail Wozniak, Ming Xu, and participants in the 2018 Winter Utah Macro Workshop and the 2019 QSPS Workshop. We are also grateful to Gueorgui Kambourov for sharing his corrected occupational mobility data from the PSID. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

1 Introduction

Workers' occupations are frequently used as a lens for understanding a range of economic phenomena — for example, technological change, trade, gender differences, and worker displacement.[1] Part of the appeal in using occupations is that switching occupations is costly, in part because a substantial amount of human capital is occupation-specific (see, for example, Kambourov and Manovskii (2009b)). Thus, worker adjustment across occupations in response to changes in the economic environment is often sluggish. As a result, having accurate measures of the rate at which workers switch occupations and how this has changed over time is relevant for the answers to many economic questions.

This paper aims to improve and better understand estimates of occupational mobility — the fraction of employed workers whose present occupation is different from their occupation a year ago — obtained from the Current Population Survey (CPS). There are two ways of measuring annual occupational mobility within the CPS — retrospectively, using supplement data where workers are asked about their current occupation and their past occupation, or longitudinally, using the limited panel structure of the CPS to link individual responses a year apart. Our starting observation is that these two methods generate very different levels of occupational mobility, with differences of almost an order of magnitude. For example, in the year 2018, retrospective measures of occupational mobility imply rates of 2-5%, whereas longitudinal measures imply rates ranging from 23-45%. We document that not only do the levels of occupational mobility differ across these two measurement methods, but the trends in these rates differ as well. Retrospective measures of occupational mobility have been trending downward since the 1990s, whereas longitudinal measures have been trending upward. Given that both measures of mobility may be subject to significant measurement error – recall bias in retrospective responses and coding errors in longitudinal responses – this generates substantial uncertainty about the actual level of occupational mobility.[2]

This paper reconciles this discrepancy in occupational mobility measures and estimates an accurate measure of annual occupational mobility over time. We use a generalized method of moments approach to estimate actual occupational mobility, using the relationship between occupational switching and observed labor market outcomes and the fact that we can observe multiple

[1] Examples of research on these topics using occupations include: technological change (Acemoglu and Autor (2011), Autor and Dorn (2013)), trade exposure (Ebenstein et al. (2014), Traiberman et al. (2017)), labor market differences by gender (Cortes et al. (2018), Erosa et al. (2017)), and worker displacement (Poletaev and Robinson (2008), Huckfeldt (2016)).

[2] As we discuss further in Section 2, the coding errors in longitudinal responses arise from a worker needing to provide a job description at two points in time, which can then be subjectively interpreted differently by coders in determining the corresponding occupational code. Retrospective measures in the CPS are not subject to this concern since they practice dependent coding, using one's current occupation as a reference point in determining whether an occupation has changed. See Section 2 for more details.

measures of occupational switching for the same person in the data. With a key assumption for identification is that the measurement error in each of these measures is conditionally independent with each other and the labor market outcome, we are able to identify actual occupational mobility and the measurement error in each measure, both on average and over time. A virtue of our estimation strategy is that our assumptions of conditionally independent measurement error lead to our estimation being overidentified, and thus an overidentification test allows us to test this assumption.

We find that annual occupational mobility has been falling over time, consistent with the trend observed in retrospective measures. However, while the trend obtained from the retrospective measures is correct, the level of occupational mobility in retrospective measures is biased downward. On average, from 1980-2018, we find that the actual rate of occupational mobility is 2-3 percentage points or 70% higher than the rate of occupational mobility from retrospective measures. These findings imply that longitudinal measures of annual occupational mobility are biased upward and that this bias is increasing over time, given that longitudinal measures show an increasing rate of occupational mobility. These findings are robust to a range of empirical specifications, including the level of occupational aggregation and the choice of labor market outcomes and individual characteristics used in estimation. We fail to reject the null hypothesis that our estimation is overidentified, providing additional credence for our identification strategy.

We are also able to identify which individual characteristics are most correlated with measurement error in reported occupational mobility. Measurement error in longitudinal occupational switches is more likely among male, nonwhite, Hispanic, and younger workers as well as workers in certain occupations. However, neither changes in the composition of the workforce over time nor the estimated changes in the relationships between observable worker characteristics and measurement error can account for the upward trend in false positives in the longitudinally linked data.

Given our estimates of occupational mobility and measurement error, we illustrate the importance of these findings in two applications, both in the context of studying the worker level impacts of trade shocks. We first consider the model of worker mobility and trade adjustment of Artuç et al. (2010), the parameters of which are estimated using worker flows data from the CPS. While Artuç et al. (2010) make adjustments to the worker flows data to address the possibility of measurement error, we find that they inflate worker flows from retrospective measures by too much. Using implied worker flow rates from our measurement error corrections, we estimate higher moving costs for workers in their model. When we conduct the trade liberalization experiment in their paper with these different parameters, the welfare gains to manufacturing workers are 50% smaller on impact and adjustment to steady state after a trade shock takes almost twice as long.

We also revisit some of the findings of Ebenstein et al. (2014). Using longitudinally linked data, Ebenstein et al. (2014) construct empirical measures of exposure to trade and offshoring, and use an IV strategy to show that workers who were in tradable occupations and subsequently switched occupations experienced wage losses of 12-17%. We jointly estimate their specifications with our model of measurement error, and, correcting for measurement error in occupational switching, we find the opposite result – workers in tradable occupations who switched occupations saw substantial wage *gains*. This result flips because of measurement error in the first stage of their specification. They estimate that workers in tradable occupations are *more* likely to switch occupations. However, once we correct for measurement error, these workers are *less* likely to switch. While we are skeptical that the estimated wage changes for occupational switchers represent the causal impact of displacement from trade shocks, our finding that workers in tradable occupations are less likely to switch suggests the possibility that some workers may become "stuck" in their jobs and be less flexible in adjusting to trade shocks.

Our paper focuses on the CPS because it is a large, representative survey administered over many years. However, Kambourov and Manovskii (2008, 2013) argue that the Panel Study of Income Dynamics (PSID) may be preferable for measuring occupational switching, especially given a concerted effort to reduce occupational coding errors prior to 1980. We compare our corrected time series of annual occupational mobility to the corrected measures generated in Kambourov and Manovskii (2008) over the period 1981-1997. We find lower levels of occupational mobility over this time period than in their study, but our estimated levels are closer to their results than any other measure obtained from the CPS. As a result, since the PSID data is only available every two years since 1997, the corrected CPS time series presents an appealing alternative for understanding recent patterns in occupational mobility. We make available online our code and data in the hopes that they will be useful to researchers interested in occupational mobility.

This paper naturally relates to the large literature studying occupational mobility or using it to understand other economic phenomena.[3] In particular, the paper is closely related to the literature on measurement error in occupational mobility, including Moscarini and Thomsson (2007), Kambourov and Manovskii (2008), and Kambourov and Manovskii (2013). Moscarini and Thomsson (2007) address measurement error in the monthly CPS by trying to identify "suspicious" switches, where "suspicious" switches are those where either workers did simultaneously switch some other labor market outcome (industry, employer, etc.) or if their history of occupational switches appears unusual. Using our annual estimates of occupational mobility, we construct corrections to monthly occupational mobility, and obtain a time series with very similar dynamics to that from their work. We prefer our monthly time series, as it requires less strict assumptions about occupational switch-

[3]Examples include Kambourov and Manovskii (2009a), Lalé (2012), Groes et al. (2014), Papageorgiou (2014), Molloy et al. (2014), Gorry et al. (2014), Artuç and McLaren (2015), Wiczer (2015), Guvenen et al. (2015), Cortes (2016), Gervais et al. (2016), Cortes and Gallipoli (2017), Cubas and Silos (2017), Robinson (2018), Forsythe (2018), Xu (2019), and Carrillo-Tudela et al. (2019).

ing and fewer judgment calls about what constitutes spurious and legitimate switching. However, we show that absent any corrections to monthly occupational mobility, the raw data indicates a spurious upward trend since the year 2007. We also show both monthly time series imply that simple time aggregation methods mapping monthly switches to longer time horizons will likely substantially overstate occupational mobility.

Kambourov and Manovskii (2008) use the retrospective files on occupational mobility in the PSID to obtain more accurate occupational mobility measures prior to 1980, and then estimate corrections for an affine shift in the level of measurement error thereafter. Instead of correcting for measurement error using an external dataset, we estimate the degree of measurement error using relationships between observables and occupational switching and assumptions regarding the nature of that error. Further, given the divergent trends in occupational switching measures in the CPS, we cannot make simple corrections to occupational mobility measures using a set of affine shift parameters. Kambourov and Manovskii (2013) also document level differences across the two measures of occupational mobility within the CPS, and argue that the retrospective measures of mobility from the March supplement to the CPS may not accurately measure annual mobility. However, they do not document the divergent trends in these two occupational mobility rates, nor establish a strong conclusion about which time series is preferable. Absent corrections for measurement error, we find retrospective measures of occupational mobility to be preferable, with closer levels and trends to the actual levels of mobility and a smaller bias when used in regression contexts.

Finally, this paper also relates to the literature on measurement error in survey data, for which Bound et al. (2001) provides an early summary. The type of measurement error we consider, misclassification error, is generally nonclassical, as errors in discrete variables generally don't satisfy the classical assumptions (as pointed out in Aigner (1973)). A common approach used to address this error is considered in Hausman et al. (1998), who use nonlinear functional form assumptions to estimate misclassification error in job changes observed in CPS and PSID data. Relative to Hausman et al. (1998), our situation is somewhat unique in that for each individual, we have two measures of our data moment of interest, occupational switching, and thus instead of relying on nonlinear functional forms to estimate the misclassification error in each measure, our assumption of conditional independence of the measurement error provides identification. This methodology we use to estimate misclassification error is based on the approach used in Kane et al. (1999). They consider the identification of misclassification error when there are two noisy signals of an individual's education, with the ultimate intention of estimating the ultimately aiming the relationship between other observables.[4] Relative to their framework, we do not require that

[4]Importantly, Kane et al. (1999) also point out that with misclassification error, using one noisy signal as an instrument for the other in a regression framework will not generate consistent estimates of parameters of interest.

the measurement error be independent of worker characteristics, but extend their methodology to allow for a estimable correlation between individual characteristics and measurement error.

2 Measuring Occupational Mobility in the Current Population Survey

2.1 Measurement Methods in the CPS

Occupational mobility — defined as the fraction of workers who were employed this year and last year and whose present occupation is different from their occupation a year ago — can be measured from the Current Population Survey (CPS) in two conceptually different ways.[5] The first way of measuring occupational mobility is what we term "retrospective" occupational mobility – measured using supplemental surveys, where in the same survey, an individual is asked both about her current occupation as well as her occupation in the past year. The second way of measuring occupational mobility is what we term "longitudinal" occupational mobility — measured by comparing the occupation of a worker in two different survey responses at two different points in time, linked together using the limited panel component of the CPS.[6]

Measuring occupational mobility in each way has distinct advantages and disadvantages.[7] The most common way to measure occupational mobility retrospectively is to use the annual March socioeconomic supplement to the CPS, where individuals are asked both about their present occupation and the occupation of their longest job during the prior year. This method is convenient and commonly applied because it does not require any linking of samples and allows the researcher to observe the occupational switching patterns of individuals who have recently moved across geographic locations. One concern about this approach is that using the longest job of the prior year may not actually represent the annual time horizon intended to be measured. Further, there are potential concerns about biased responses from imperfect recollection of the past (see Mathiowetz and Ouncan (1988), for example).

An important feature of retrospective measures of occupational mobility is that in the March

[5]By way of background, the Current Population Survey (CPS) is a monthly survey of 60,000 households which provides the key input for labor market outcomes on unemployment, labor force participation, and other worker level outcomes. The CPS also administers supplemental surveys less frequently, covering topics such as displaced workers, fertility, and socioeconomic status. Because occupation is often not recorded in administrative and firm-level data, the CPS is also a key source of data regarding occupations and occupational outcomes. While the National Longitudinal Survey of Youth (NLSY) and Survey of Income Programs and Participation also provide data on occupations and occupational mobility, the large, representative, consistent sample of the CPS makes it a preferred source of occupation data for many applications.

[6]Households interviewed by the CPS are interviewed for four consecutive months, out of the sample for eight months, and then interviewed again for four more consecutive months.

[7]A further discussion of these advantages and disadvantages can be found in Kambourov and Manovskii (2013).

supplement, since 1970, responses about the job and occupation of the prior year are obtained via a dependent coding procedure. In this procedure, individuals are first asked if their employer and job duties in the past year are the same as in the present. If so, then the individual's past occupation is coded as being identical to their present occupation. The advantage of this procedure is that it avoids recoding occupations, which can be a major source of measurement error, as how the coder maps the (potentially imprecise) provided description of job tasks and duties to an occupational code can be subjective. On the other hand, this also puts the burden of identifying an occupational switch on the individual responding, which may also be subjective.[8]

In contrast, when measuring annual occupational mobility longitudinally, there is no dependent coding procedure applied. Since 1994, the monthly CPS has used dependent coding for responses across the four consecutive months a household is interviewed, but when the household is interviewed a year later, occupation is independently coded again. This creates significant potential for measurement error, as the description of job tasks given by the same worker at two points in time for the same job may be different and may be interpreted differently by the interviewer who must convert that description into a numeric occupational code. An example of the magnitudes of this error is presented in Mathiowetz (1992), which documents large differences (25-50%) in how different coders assign numeric occupation values to the same job description. It is not difficult, for example, to envision how an academic economist's job description – writing research papers, teaching students, providing administrative service – could be classified as being an economist/market researcher, college subject instructor, or even technical writer or a manager/administrator in education and related fields, all of which correspond to different occupation codes.

Measuring occupational mobility longitudinally also has the disadvantage of being unable to identify occupational switching associated with moving, since the CPS uses an address based sampling scheme. However, longitudinal measures of occupational mobility will be unaffected by recall bias, and can be related to wage and earnings changes over time, information that is not available from retrospective measures.

2.2 Retrospective and Longitudinal Measures of Occupational Mobility

Our primary data source is IPUMS-CPS (Flood et al. (2018)) for the period 1980-2018. We focus on the privately employed adult civilian population in the U.S., dropping individuals under the age of 18 and individuals working in government industries.[9] We also drop all observations where

[8]Because occupation is independently coded when the respondent reports changing employers, this is most salient when an individual remains with the same employer and is asked whether his or her current job tasks and duties have changed.

[9]We follow Kambourov and Manovskii (2008) in excluding government workers, however we find that adding them in to our sample does not substantially alter our findings.

occupation is imputed (including entirely imputed responses in the March CPS); we discuss the impact on our sample of these adjustments in Appendix A.[10] To account for changes in the occupational coding system over time, we use the time consistent occupational codes of Dorn (2009) and Autor and Dorn (2013). We report results for occupations defined at what we term the one, two and three digit level, which provide 6, 17, and 325 distinct occupations, respectively. Complete details of how we construct our data sample, including a list of one and two digit occupations and codes, are available in Appendix A.

As our ultimate interest is to directly compare retrospective and longitudinal responses, we longitudinally link responses from two consecutive March supplements to the CPS using identifiers developed in Rivera Drew et al. (2014) from 1989-2018, and using the algorithm set out in Madrian and Lefgren (2000) from 1980-1989. To guard against errors generated by spurious links, we follow the methodology of Madrian and Lefgren (2000) and drop all linked responses where sex or race disagree or where increases in education and age are greater than a year. As pointed out in both Madrian and Lefgren (2000) and Rivera Drew et al. (2014), because of changes in household identifiers, we are unable to link CPS records between 1985-1986 and 1995-1996.

Figure 1 plots occupational mobility measures from the CPS using retrospective (left panels) and longitudinal measures (right panels). In the left panel, we report three retrospective measures: (1) the "headline" measure from the March CPS, which includes individuals who recently moved; (2) a "restricted" measure from the March CPS, where sample restrictions are imposed to be consistent with the longitudinal measure (only those individuals who can be linked in consecutive Marches who are employed and not imputed both periods); and (3) an additional retrospective measure of occupational mobility obtained from the job tenure and occupational mobility supplement, which is available in select years. This last measure has the virtue of asking survey respondents about their occupation exactly one year ago, and thus more precisely captures annual occupational mobility.

For each of these cases, annual occupational mobility is generally low, 2-6%, 3-7%, and 4-9% a year at the one, two and three digit levels, respectively. Further, measured retrospectively, at all levels of aggregation occupational mobility has been trending downward, especially since the year 2000. When we impose the same sample restrictions needed for measuring longitudinal occupational mobility, the rates drop even lower and still show evidence of a downward trend. This level shift is largely due to the omission of individuals who move residences. Notably, the job tenure and occupational mobility supplement data generates point estimates for occupational switching which are only slightly above the March measures, which suggests a minimal bias in

[10] In the case where the entire set of responses to the March CPS is imputed, IPUMS data does not identify these. To properly identify these requires merging the FL-665 variable from the original CPS files (or the suprec variable from the Unicon produced CPS files).

retrospective measures from the March supplement due to the time horizon framing.

The right panel of Figure 1 shows occupational mobility rates measured using longitudinally linked data. In contrast to retrospective measures, longitudinal measures of occupational mobility are almost an order of magnitude higher, ranging from 19-24%, 25-32%, and 36-45% a year at the one, two and three digit levels. Further, at all levels of aggregation, longitudinal measures of occupational mobility are trending in the opposite direction, rising upward over time. The spikes in the years 1983 and 2003 are attributable to the change in occupational coding systems in those years, as small imperfections in coding consistency persist even with the time consistent occupational coding approach. The primary take away is that retrospective and longitudinal measures of occupational mobility in the CPS fundamentally disagree about both the levels and the trends in occupational mobility.

Given that we can observe both a longitudinal and a retrospective response for a subset of individuals in the data, we further decompose these reported switching rates by into the universe of possible reported outcomes for both occupational switching measures– the cases where both measures agree about whether an occupational switch has occurred or not and the cases where occupational switching is only reported in one measure. This decomposition is reported in Figure 2. The primary insight from this decomposition is that instances where a retrospective switch is reported, but not a longitudinal one, are infrequent. The primary divergence in the two measures of occupational switching lies in the many cases where longitudinal measures report a switch, but retrospective measures do not. We naturally expect a priori that many of these cases are spurious and due to coding error present in longitudinal responses; determining the exact magnitude of this error, and any underreporting error in retrospective responses is the quantitative question we are interested in.

These divergent patterns of occupational switching by retrospective and longitudinal measures hold even with further sample restrictions – focusing just on men, prime-aged workers (or other age groups), married or unmarried workers, etc. While we do not report these results, the strategy we use in identifying the true rate of occupational mobility explicitly accounts for the possibility that individual demographic characteristics influence the rate at which occupational switches occur and are reported. Thus, we defer a discussion of the relationship between reported occupational switching and individual characteristics to Section 4, when we report our primary findings.

3 Framework for Estimating Measurement Error and Actual Occupational Mobility

We now lay out a generalized method of moments approach to estimating the measurement error in retrospective and longitudinal measures of occupational mobility. Our approach is based on the approach used in Kane et al. (1999), and we use the relationship between occupational mobility and both labor market outcomes and individual characteristics to estimate the measurement error in each of the switching rates and the true rate of switching. The key data feature that allows us to do this is that we have retrospective and longitudinal answers for the same individual regarding their past year's occupation, providing multiple signals of occupational mobility.

3.1 Generalized Method of Moments Framework

Suppose that the conditional expectation of some labor market outcome Y_{it} for individual i in year t, given a binary indicator for occupational switching in the past year, SW_{it}, and a 1xK vector of individual characteristics, X_t, is given by:

$$\mathbb{E}\left[Y_{it} \mid SW_{it}, X_{it}\right] = \beta_{0,t} + \beta_{1,t} SW_{it} + X_{it}\beta_{2,t} \tag{1}$$

For convenience, assume that the vector of individual characteristics is normalized to have zero mean, $\mathbb{E}\left[X_{it}\right] = 0$. As a result, $\beta_{0,t} + \beta_{1,t}$ is the average value of Y_{it} for occupational switchers in year t and $\beta_{0,t}$ is the average value of Y_{it} for occupational stayers.

However, this relationship cannot be directly estimated because, instead of directly observing SW_{it} for each individual, we instead observe two noisy signals of switching, $\tilde{SW}_{it}^{R}$ (retrospective) and $\tilde{SW}_{it}^{L}$ (longitudinal). We allow for the errors in each of these signals to be correlated with individual characteristics, but assume that conditional on individual characteristics X_{it} and the value of SW_{it}, the errors in each of these signals are independent of each other and independent of the labor market outcome variable. We further assume a linear relationship between individual characteristics and these signals of occupational switching.[11] These assumptions collectively imply the

[11]The assumption of linearity is made for ease of interpreting the relationship between individual characteristics and occupational switching signals and ease of numerical optimization in the estimation.

following relationships, illustrated in Equations (2) and (3):[12]

$$P(\tilde{SW}^R_{it} = 1 \mid SW_{it}, \tilde{SW}^L_{it}, X_{it}, Y_{it}) = \alpha_{R,0,t} + \alpha_{R,1,t} SW_{it} + X_{it}\alpha_{R,X,t} \tag{2}$$
$$P(\tilde{SW}^L_{it} = 1 \mid SW_{it}, \tilde{SW}^R_{it}, X_{it}, Y_{it}) = \alpha_{L,0,t} + \alpha_{L,1,t} SW_{it} + X_{it}\alpha_{L,X,t} \tag{3}$$

We assume that, conditional on individual characteristics, the signals are positively correlated with actual occupational switching, implying that $\alpha_{R,1,t} > 0$ and $\alpha_{L,1,t} > 0$.[13]

Finally, we assume that the relationship between the probability of switching and the underlying controls is given by:

$$P(SW_{it} = 1 \mid X_{it}) = \delta_{0,t} + X_{it}\delta_{1,t} \tag{4}$$

Since X_{it} has been normalized to be mean zero, this implies that the average probability of an occupational switch is given by $\delta_{0,t}$.

The parameters to estimate in this model are $\beta_{0,t}$, $\beta_{1,t}$, $\alpha_{R,0,t}$, $\alpha_{R,1,t}$, $\alpha_{L,0,t}$, $\alpha_{L,1,t}$, and $\delta_{0,t}$, and the four Kx1 vectors $\beta_{2,t}$,$\delta_{1,t}$, $\alpha_{R,X,t}$, and $\alpha_{L,X,t}$. The objects of greatest interest, given our focus on occupational mobility, are the α_t parameters, which determine the measurement error in each of the occupational switching measures and the δ_t parameters, which determine actual occupational switching.

Before we discuss our estimation strategy, it is convenient to first define a sequence of indicator variables which indexes all possible values of both signals of occupational switching – switching in both retrospective and longitudinal measures, not switching in either measure, and switching in one but not the other. These are defined as the variables $\tilde{Z}_{i,1,t}$, $\tilde{Z}_{i,2,t}$, $\tilde{Z}_{i.3,t}$, and $\tilde{Z}_{i,4,t}$, where

$$\begin{aligned}
\tilde{Z}_{i,1,t} &= 1 \text{ if } \tilde{SW}^R_{it} = \tilde{SW}^L_{it} = 1 \\
\tilde{Z}_{i,2,t} &= 1 \text{ if } \tilde{SW}^R_{it} = \tilde{SW}^L_{it} = 0 \\
\tilde{Z}_{i,3,t} &= 1 \text{ if } \tilde{SW}^R_{it} = 1, \tilde{SW}^L_{it} = 0 \\
\tilde{Z}_{i,4,t} &= 1 \text{ if } \tilde{SW}^R_{it} = 0, \tilde{SW}^L_{it} = 1
\end{aligned}$$

All parameters can be estimated via GMM using the means of these indicator variables as well as the relationship of these variables with both the outcome variable Y_t and the vector of controls

[12]Alternatively, we could allow for a full set of interactions between actual occupational switching and individual characteristics in our model. We have explored this alternative specification and found that the inclusion of so many incidental parameters dramatically reduced the precision of our estimates without providing substantial additional insight and so we do not pursue this approach in the paper.

[13]This is a necessary condition for identification, equivalent to the monotonicity assumption of Hausman et al. (1998). The implication is simply that the probability of reporting an occupational switch rises if an actual occupational switch occurred.

X_t and also the relationship between X_t and Y_t. Defining the vector of all these indicator variables as $\tilde{Z}_{it} = [\tilde{Z}_{i,1,t}; \tilde{Z}_{i,2,t}; \tilde{Z}_{i,3,t}; \tilde{Z}_{i,4,t}]$, this amounts to identifying parameters using the moments $\mathbb{E}\left[\tilde{Z}_{it}\right]$, $\mathbb{E}\left[\tilde{Z}_{it} Y_{it}\right]$, $\mathbb{E}\left[\tilde{Z}_{it} X_{it}\right]$, $\mathbb{E}\left[X_{it}' Y_{it}\right]$. We provide a full listing of these moments as functions of model parameters in Appendix B.

3.2 Identification

We now sketch some intuition for how identification is obtained using these moments. The covariance between individual characteristics, X_{it} and the outcome variable, Y_{it}, $\mathbb{E}\left[X_{it}' Y_{it}\right]$, helps identify the parameters describing the linear relationship between X and Y $(\beta_{2,t})$, similar to standard OLS.[14] The covariance between the possible switching responses and the individual characteristics, $\mathbb{E}\left[\tilde{Z}_{it} X_{it}\right]$, helps identify the the relationship between individual characteristics and actual occupational switching $(\delta_{1,t})$, as well as the relationship between individual characteristics and the measurement error in both of the signals $(\alpha_{R,X,t}, \alpha_{L,X,t})$. Finally, the combination of the mean of $\tilde{Z}_{it}$ and the relationship between $\tilde{Z}_{it}$ and the labor market outcome Y_{it}, $\mathbb{E}\left[\tilde{Z}_{it}\right]$ and $\mathbb{E}\left[\tilde{Z}_{it} Y_{it}\right]$, identify the underlying occupational switching rate $(\delta_{0,t})$, the underlying error rates in the signals of switching, $(\alpha_{R,0,t}, \alpha_{R,1,t}, \alpha_{L,0,t}, \alpha_{L,1,t})$, and the relationship between occupational switching and the labor market outcome $(\beta_{0,t}, \beta_{1,t})$.

A natural concern here is that by using a linear relationship to model occupational switching, we may be unable to identify all the parameters of our model, an issue discussed in Hausman et al. (1998). Indeed, if we were to estimate the measurement error in each signal separately without using information in the other signal, we would be unable to recover the parameters of our model. This is not an issue in our framework because of two key things: (1) for each individual, we can observe two measures of occupational switching, and (2) we assume that the measurement error in the two measures is conditionally independent. Given these two assumptions, we can combine information from these two independent pieces of information to identify the actual rate of occupational switching.[15] This is analogous to identification with classical measurement error in a two stage least squares setting where there are two error-ridden measures of an independent variable and one is used as an instrument; the key requirement in that case is that the measurement error in the two signals is independent. Because we are dealing with misclassification error and the measurement error is thus correlated with the variable itself, our procedure uses a full GMM estimation instead of the simpler two stage least squares approach.[16]

[14]Note that since we have normalized X to mean zero, this is exactly the covariance between these two variables.

[15]Though we do not provide a formal proof of this result, we have validated this in our numerical estimation, finding that the Jacobian of the moment conditions always has full rank.

[16]Kane et al. (1999) make this point, showing that with misclassification error, using one noisy signal as an instrument for the other in a regression framework will not generate consistent estimates of parameters of interest even if the two errors are conditionally independent. This point is also made more recently in Bingley and Martinello (2017).

The intuition of identification in our model can be further described as follows. With our assumption of conditional independence, the probability of a worker actually switching occupations is highest when retrospective and longitudinal measures both indicate an occupational switch occurred. Correspondingly, the probability of a worker switching occupations is lowest when both measures indicate no switch occurred. Given our assumption of independence, we can then use the relationship between labor market outcomes and switching when only one of the measures indicates a switch to assess the relative accuracy of each measure, and ultimately the actual rate of occupational mobility.

For example, ignoring covariates for the moment, consider the data in the last row of Table 1, which provides summary statistics for the data used in our estimation. We see that the fraction of individuals whose part-time/full-time (PT/FT) work status changes between two years varies across each possible response for occupational switching. Conditional on both signals indicating a switch, 28% of workers switched PT/FT status; conditional on both signals indicate no switch, only 9% of workers switched PT/FT status. This suggests a substantial positive relationship between occupational switching and changing PT/FT status. Now, if we examine the rate at which workers switch PT/FT status for individuals who only switched according to the retrospective measure, 22% switch PT/FT status. In contrast, only 11% switch when only the longitudinal measure indicated a switch. This simple evidence suggests that a retrospective switch is closer to a true occupational switch than a longitudinal switch alone, because the correlation between switching and changing part-time/full-time status is closer to the case where both signals indicate an occupational switch has occurred. Abstracting from covariates, the combination of this data and the frequency with which each possible pair of occupational switching outcomes is observed alone would exactly identify the key parameters of the model $(\delta_{0,t},\alpha_{R,0,t},\alpha_{R,1,t},\alpha_{L,0,t},\alpha_{L,1,t})$, and the relationship between occupational switching and the labor market outcome $(\beta_{0,t},\beta_{1,t})$.

Critical to our identification strategy is our assumption of conditional independence. This assumption will violated if either (1) there is correlation between the measurement errors in the two signals even after controlling for observables or (2) there is correlation between the measurement error in either signal with the labor market outcome itself after controlling for observables. This latter potential violation is particularly concerning, as it seems plausible that the forces generating errors in occupational switching could be related to the value of the labor market outcome. For example, a worker might switch from part-time to full-time status within the same occupation, and naturally absorb more job duties with longer work hours. Given that added job duties could increase the likelihood of a spurious occupational switch, as a lengthier job description leads to a greater possibility of coding errors, this would violate our independence assumption. We discuss this concern further and how we attempt to mitigate it in the next subsection when deciding on appropriate labor market outcomes to use in our estimation.

Because our full estimation accounts for covariates, we have $5K+7$ moments for identifying $4K+7$ parameters, implying that the estimation is overidentified. The overidentification can be observed given that the moment condition $\mathbb{E}\left[\tilde{Z}_{it}X_{it}\right]$ provides $4K$ moments, but is intuitively only needed to identify $3K$ parameters $(\delta_{1,t},\alpha_{R,X,t},\alpha_{L,X,t})$, the parameters describing the relationship of actual or reported occupational switching with these individual characteristics. However, these $3K$ parameters are only sufficient because of our assumption that the measurement error in the two signals of occupational switching is independent. If the two measurement errors are not independent, the model will struggle to fit all the moments in estimation. Thus, a formal test for overidentification provides an implicit test of the independence assumption between the two measurement errors.

3.3 Empirical Implementation

We use CPS data for the occupational switching measures, control variables and labor market outcomes from 1980-2018 to estimate the above model; additional measurement details for all variables used in estimation in Appendix A. The data on the multiple signals for occupational switching comes from the linked March supplements to the CPS, with the same sample restrictions described in Section 2 for longitudinal measures. For control variables, we use age, age squared, and indicators for sex, race and ethnicity (white/nonwhite and Hispanic), educational attainment (below HS/HS degree/some college/college degree or more), marital status (married/unmarried), and two digit occupational fixed effects measured using the occupation in the second of the two years a worker is observed.[17] We use occupation measures after a possible switch as these are the same for both retrospective and longitudinal measures; using instead occupation fixed effects measured from the prior year, either from retrospective or longitudinal measures, does not meaningfully change our primary findings. Given that a key potential source of measurement error is how job duties are coded into occupations, we find it natural to account for a potential relationship between this error and a worker's occupation; however, we show in Appendix C that our results are robust to excluding these occupation fixed effects.

In principle, there are many possible labor market outcomes we could choose from in our estimation. As a filter, we only consider labor market outcome variables that: (1) have an a priori plausible relationship with occupational switching; (2) are available in each year we consider, 1980-2018; and (3) are ex ante unlikely to exhibit a mechanical correlation with measurement error. This last criterion is particularly important, as while it might be natural to consider whether or not a worker switched industries as a labor market outcome, we would naturally expect many

[17]Results using one digit occupational fixed effects are very similar. We do not consider three digit occupational fixed effects, as many three digit occupations are not observed year by year and with so many occupations, it is harder to interpret results regarding the relationship between occupational switching and actual occupations.

of the forces driving measurement error in occupational switching could also be driving measurement error in industry switching as well. Indeed, though we do not report it, we observe similar divergent patterns for industry switching over time when measured by retrospective and longitudinal methods; as this is not the focus of our paper and other datasets may provide compelling alternative for industry transitions, we do not explore this data further.

Given these criteria, we focus on the following set of labor market outcomes: whether or not a worker's hourly wage (prior year income divided by the product of usual weekly hours and weeks worked in the prior year) changed by more than 10%, whether or not a worker reported having more than one employer in the past year (not simultaneously), whether or not a worker's part-time/full-time work status changed across years, and whether or not a worker reported being employed for less than half the prior year.[18] Of these outcomes, the change in hourly wage and part-time/full-time work status are measured longitudinally and the number of employers and amount of time working in the past year are measured retrospectively. We intentionally select a balance of dependent variables measured retrospectively and longitudinally to mitigate the threat to identification that could come a correlation between measurement error and these outcomes on the basis of how the variables are measured (retrospectively or longitudinally).

With these four labor market outcomes, we stack the moments used in estimation and simultaneously estimate all model parameters. However, in reporting our results, to better understand how identification is occurring, we also report parameter estimates from estimating parameters from one labor market outcome at a time. In estimating the parameters from these moments via GMM, we follow the common practice of using a two-step procedure to obtain the optimal weighting of moment conditions.

By stacking moments from all four labor market outcomes and estimating parameters simultaneously, this further overidentifies our estimation, as with N dependent variables, as we have $(4+N)K+4N+3$ moments to estimate $(3+N)K+2N+5$ parameters, implying $K+2N-2$ additional moments. This additional overidentification is generated by the assumption that the errors in occupational switching are uncorrelated with the labor market outcomes conditional on observables and whether or not an actual occupational switch occurred. In other words, it should not matter which labor market outcome we use to estimate the parameters, as each should return the same results. Thus, a failure to reject the assumption of overidentification with multiple dependent variables additionally indicates a failure to reject the assumptions that the two signals of occupational switching are conditionally independent of both each other and the labor market outcomes. We argue this test thus provides additional evidence regarding one of our primary threats to identification.

[18] Importantly, the number of employers in the past year is asked *independently* of questions about occupation and industry, so there is not a mechanical correlation based on survey design.

We report the summary statistics of the variables used in estimation in Table 1. For illustrative purposes, we also report the mean of each variable conditional on each possible outcome from both occupational switching measures (measured at the one digit level).

4 Estimated Corrected Occupational Mobility and Measurement Error

4.1 Baseline Estimates of Measurement Error and Actual Occupational Mobility

We start by presenting estimates of the model parameters governing average actual occupational mobility and the average underlying measurement error for entire pooled sample 1981-2018.[19] This gives us summary averages of measurement error in both retrospective and longitudinal measures of occupational switching. We then report our estimates of occupational mobility and measurement error year by year.

Table 2 reports the estimates for δ_0, the actual occupational switching rate, and estimates of the measurement error in both retrospective and longitudinal measures of occupational mobility. Since we have normalized the vector of individual characteristics to be mean zero, these estimates capture the average across all individuals and not just for a subset of the population. In lieu of reporting the actual parameter estimates for $\alpha_{R,0},\alpha_{R,1}\alpha_{L,0},\alpha_{L,1}$, the parameters which govern the measurement error in each signal, for ease of interpretation, we report the false positive and the false negative rate for each measure. These are defined as the fraction of reported occupational switching and non-switching in each measure that is spurious. In the context of our model, the false positive rate for signal $j \in \{R,L\}$ is defined as $\frac{P(\tilde{SW}^j=1 \cap SW=0)}{P(\tilde{SW}^j=1)} = \frac{(1-\delta_0)\alpha_{j,0}}{(1-\delta_0)\alpha_{j,0}+\delta_0(\alpha_{j,0}+\alpha_{j,1})}$ and the false negative rate is defined as $\frac{P(\tilde{SW}^j=0 \cap SW=1)}{P(\tilde{SW}^j=0)} = \frac{\delta_0(1-\alpha_{j,0}-\alpha_{j,1})}{\delta_0(1-\alpha_{j,0}-\alpha_{j,1})+(1-\delta_0)(1-\alpha_{j,0})}$. For comparison's sake, we also report the average measured occupational switching rate in both longitudinal and retrospective measures as well.

Our pooled estimates suggest that the actual rate of occupational switching is much closer to the retrospective measure than the longitudinal measure. In the case of one-digit occupational switching, we estimate the average occupational switching rate from 1981-2018 to be 4.3%, higher than the retrospective measure of 2.5%, but much lower than the 21.2% from the longitudinal measure. Across all levels of aggregation, we find that the true switching rate is about two to three percentage points or roughly 70% higher than the retrospective measure.

[19]These pooled estimates do not include year fixed effects, but adding them does not have a significant impact on our estimates.

Examining the estimated false positive and false negative rates for each measure, we see that there is a large difference in the false positive rate across measures. We estimate that only 3-5% of reported retrospective switches are incorrect, but that 83-85% of longitudinally reported occupational switches are incorrect. On the other hand, the false negative rate is higher for retrospective switches, with about 2-3% of observations reporting no switch actually experiencing a switch, compared to only 1% from the longitudinal measure. While these numbers are small, since the retrospectively measured fraction of non-switchers is large (97% in the one digit case), the 2-3% false negative rate explains why the actual switching rate is 70% larger than what is reported retrospectively. These results are consistent with the idea that coding error generates a substantial upward in occupational switching measured longitudinally and that there is also non-trivial recall error generating a downward bias in occupational switching measured retrospectively.

Figure 3 presents our year by year estimates of occupational mobility and measurement error, which determine what the actual trend in occupational mobility has been over time. We report our estimates of actual occupational mobility opposite both retrospective (left panels) and longitudinal (right panels) measures. In Figure 4, we present estimates of the false positive (left panels) and false negative (right panels) rates over time.

Consistent with our findings from the pooled sample, we find that in each year, the actual level of occupational mobility is closer to the level of mobility measured retrospectively than measured longitudinally. Further, our estimates of actual occupational mobility trend downward over time, similar to retrospective measures. Given that longitudinal measures of occupational mobility have been trending up over this period, this implies that the false positive rate for longitudinal measures has been increasing over time, which we see in Figure 4. All other error rates appear approximately stable over time. Thus, we find that not only is measurement error in longitudinal measures of occupational switching large, but also that it has been worsening over time.

4.2 Identification and Overidentification

Given that we have used multiple labor market outcomes in our estimation, it is helpful to consider how our results would differ if we only considered a single labor market outcome. This is particularly important for understanding which labor market outcomes are driving our results. For example, given our findings that measurement error is most severe in longitudinal measures, we would be concerned if we saw that the only precise estimates of occupational mobility came from labor market outcomes which were measured retrospectively, as this might be driven by correlation with some unobserved factor specific to retrospective measurement methods.

We report our separate year by year estimates of actual one digit occupational mobility for each labor market outcome in Figure 5. We report results at the one digit level only for brevity's sake;

results for two digit and three digit occupational mobility reach similar conclusions about the relative strength of each labor market outcome for identifying occupational mobility. For comparison, we also plot actual occupational mobility estimated jointly from all labor market outcomes.

Estimates from each of the individual labor market outcomes agree with our primary conclusions about occupational mobility – the true rate is fairly low, trending down and closer to what is measured retrospectively. Of our four labor market outcomes, the one that has the tightest confidence intervals, is closest to the joint estimate, and thus is driving much of our identification, is whether or not a worker worked for multiple employers last year, which is measured retrospectively. However, if we were to only use the longitudinally computed measure of whether or not an individual changed part-time/full-time status, we also observe downward trend, albeit slightly noisier, and similar point estimates to our baseline, much closer to retrospective mobility measures than longitudinal measures. These similarities, along with the similar estimates obtained from the labor market outcome of whether a worker worked more than 26 weeks in the prior year, are encouraging that our identification is not sensitive to one particular labor market measure.

The notable exception here is the outcome of whether or not hourly wages rose by 10% or more in the past year, where there is no trend in occupational switching over time and point estimates are larger, albeit still much lower than longitudinal measures of occupational switching. However, as can be seen in Figure 5, the confidence intervals for this labor market outcome are also large, likely because of the large amount of noise in hourly wages. Given this much noise, there is little we can infer regarding the actual rate of occupational mobility from hourly wage changes. However, our jointly estimated levels of occupational mobility largely lie within estimated confidence intervals. Thus, given the similarity in the responses between these labor market outcomes measured in different ways, we take greater confidence that we have actually identified occupational mobility rates and not some idiosyncratic relationship between the measurement error in responses to occupational switching and the individual labor market outcomes.

Given that our model has more moments than parameters, and is thus potentially overidentified, we perform a Sargan-Hansen test of overidentification for our results. We find that in our year by year estimation, we consistently fail to reject the null hypothesis (at a threshold of 10%) that our model is overidentified. Thus, the data does not reject our model specification, particularly the assumptions of conditional independence of measurement error between the two signals and between the signals and the labor market outcomes. We provide further discussion and a full reporting of the p-values for this test for each year in Appendix C.

4.3 Correlates with Actual and Reported Occupational Switching

We now consider how individual characteristics are related to occupational switching and measurement error. Tables 3 (demographic characteristics) and 4 (occupational fixed effects) present the estimates from the pooled sample of the parameters δ_1, $\alpha_{L,X}$ and $\alpha_{R,X}$, which relate individual characteristics to the probability of an actual occupational switch or a reported retrospective or longitudinal switch. Because the parameters $\alpha_{L,X}$ and $\alpha_{R,X}$ indicate the marginal impact on reporting a switch, holding fixed whether or not an actual switch was reported, we also report the marginal impact of reporting either a retrospective or longitudinal switch without conditioning on whether an actual occupational switch occurred.[20] To provide context for the magnitudes on these coefficients, we also report the average rates at which an occupational switch is reported conditional on switching or not switching. We again report results only at the one digit level, but patterns are very similar at the two and three digit levels.

Examining first the relationship between demographic characteristics and occupational switching, the most striking relationship with actual occupational switching is age. Holding all other characteristics fixed, for every year older a worker is, the probability of making an occupational switch is roughly one percentage point lower, with some flattening in this slope in older age. This is similar to evidence on occupational switching over the life cycle as shown in Gervais et al. (2016). Other notable demographic characteristics associated with reduced occupational switching are being nonwhite, Hispanic, or married.

On the other hand, when it comes to the probability of reporting an occupational switching, we see notably different relationships with individual characteristics. For example, men, who are no more likely than women to switch occupations, are four percentage points more likely to *report* a longitudinal occupational switch. We also note that the probability of observing a longitudinal switch is higher for workers who are young, nonwhite, Hispanic, unmarried, or who are moderately educated (HS degree or some college). In contrast, although many point estimates are significant, we do not observe many strong relationships between demographic characteristics and reporting a retrospective occupational switch.

Turning to the relationship between one's occupation in the year after a possible a switch and actual and reported occupational switching, we observe substantial heterogeneity in both the probability of actually experiencing an occupational switch and the probability of reporting an occupational switch. For example, workers in professional specialty occupations (doctors, lawyers, teachers, etc.) are much less likely to have switched one digit occupations in the past year than

[20]This is done by computing the conditional expectation of actually switching conditional on individual characteristics. For example, since $P(\tilde{SW}_{it}^L = 1 \mid SW_{it}, X_{it}) = \alpha_{L,0,t} + \alpha_{L,1,t} SW_{it} + X_{it}\alpha_{L,X,t}$, we compute $P(\tilde{SW}_{it}^L = 1 \mid X_{it}) = \alpha_{L,0,t} + \alpha_{L,1,t}\mathbb{E}[SW_{it} \mid X_{it}] + X_{it}\alpha_{L,X,t}$ where $\mathbb{E}[SW_{it} \mid X_{it}] = \delta_0 + X_{it}\delta_{1,t}$. Thus, what we report as the marginal impact of an individual characteristic on reporting a longitudinal switch is $\alpha_{L,1,t}\delta_{1,t} + \alpha_{L,X,t}$.

workers in housekeeping, cleaning or protective service occupations. Further, some occupations have a very strong propensity to report occupational switches, especially longitudinally, such as executive, managerial, extractive, precision production and technical occupations. This is unsurprising given that much of the measurement error in longitudinal switching is tied to how occupation descriptions are coded into numerical values. However given the substantial magnitudes in misreporting errors and how much they differ from actual rates of mobility, this suggests that longitudinal measures of occupational switching are ill-suited to providing accurate information regarding heterogeneity across occupations in the propensity of making a switch. While we do not delve further into a detailed analysis of what occupations and tasks are associated with the errors in switching occupations, we note that this heterogeneity is potentially very important for studies that focus on occupational mobility across particular occupations. We consider an example of this in our application to wages and tradable/offshorable occupations in Section 5.2.

Given that we document that the average false positive rate in longitudinally measured occupational switches is rising, it is natural to consider how much of this is driven by increased propensity for measurement error across observable characteristics and changes in composition of those characteristics. For example, if mapping the job descriptions to numeric occupations is particularly difficult for some occupations, an increase in employment in these occupations over time could account for this upward trend. To answer this question, we consider a "residual" measure of the false positive rate which nets out changes in the both the coefficients on individual characteristics and the composition of these individual characteristics from the average false positive rate. To construct this measure, we residualize our estimates of δ_0 and $\alpha_{L,0}$, the constant terms in our expressions for actual and longitudinally reported occupational mobility, subtracting out the average "explained" portion of switching.[21] With these residualized measures, we then construct again the false positive rate described in Section 4.1. We also construct a residual false positive rate only using the residualized measure of $\alpha_{L,0}$, isolating just the residual changes in measurement error in longitudinal measures.

We plot our residual measures of the false positive rate alongside the actual false positive rate in Figure 6. In both cases, our residual measures have almost an identical trend over time as our actual estimated false positive rate. Thus, while we can explain some of the factors driving the observed level of measurement error in occupational mobility, the trend in errors over time is not well explained by these observable factors. One possibility for this residual upward trend could be that the set of job tasks performed within all occupations has been expanding over time, leading to greater coding errors in all occupations. Recent work in Atalay et al. (2018) suggests that there have been substantial changes in the set of tasks done even within narrowly defined occupational

[21]Namely, we construct $\hat{\hat{\delta}}_0 = \delta_0 - \bar{X}\delta_1$ and $\hat{\alpha}_{L,0} = \alpha_{L,0} - \bar{X}\alpha_{L,X}$. This is necessary because we have normalized all individual characteristics, X, to have zero mean, and thus the average levels are absorbed into these constants.

titles, so there is some support for this conjecture. We leave the exploration of this and other potential hypotheses to further work.

4.4 Estimates of the Relationship Between Occupational Switching and Labor Market Outcomes

Another outcome of our estimation is the estimated relationship between labor market outcomes and occupational switching, the parameter β_1 in our model. In particular, it is interesting to compare how our estimated true relationship between occupational switching and labor market outcomes, such as a change in hourly pay or switching employers, compares to estimates obtained from the noisily reported measures of either retrospective or longitudinal occupational switching. We report estimates of these relationships from our pooled sample from 1980-2018 opposite comparable linear specifications in Table 5. Again, we report results only at the one digit occupational level for brevity's sake, but the differences between point estimates are similar at the two and three digit levels as well.

Our estimates of the relationship between labor market outcomes and occupational mobility are consistent with our findings thus far. Individuals who switch occupations are more likely to have changed wages, switched part-time/full-time work status, had multiple employers in the past year, and worked less than half of the prior year. In every case, the point estimates we obtain for these outcomes are larger in absolute value than either estimate from retrospective or longitudinally linked data. However, we find that the bias in point estimates from retrospective occupational switches is small. This suggests that even though there is measurement error in retrospective measures of occupational mobility, using these measures in cross-sectional contexts where occupational mobility is a variable of interest may have only a small impact on results.

4.5 Total Occupational Mobility Estimates and Comparing to the PSID

Finally, we consider what our estimates imply for total occupational mobility and how this compares to estimates of corrected occupational mobility in data from the Panel Study of Income Dynamics (PSID). To estimate actual occupational mobility, we have necessarily restricted our focus to observations which can be linked across time, allowing us to observe both retrospective and longitudinal measures of mobility. However, this restriction substantially reduces the sample of all retrospective observations, in part because half of the sample cannot be matched back one year in time and because we have omitted occupational mobility patterns of individuals who have moved residences.

To construct an aggregate measure of occupational mobility, we have to take a stance on how to apply our corrected estimates of occupational mobility to the other occupational switches observed in the retrospective measure obtained from the March CPS. Given that the level of occupational switching can differ for the rest of the sample, primarily for movers, we correct the existing measures of occupational mobility using a scale factor obtained from our estimation. This scale factor is the estimated ratio of true occupational mobility to measured retrospective occupational mobility, and is given by: $\frac{P(SW=1)}{P(\tilde{SW}^R=1)} = \frac{\delta_0}{\alpha_{R,0}+\alpha_{R,1}\delta_0}$.

We consider two approaches – one where we correct the occupational switching rate of all individuals who did not report moving residences across time, and one where we correct the entire switching rate. It is possible that workers who report moving may more accurately report an occupational switch retrospectively given the accompanying geographic move. However, given that we do not find a strong relationship between observable worker characteristics and retrospective switching and thus have a limited understanding of what drives errors in reporting, there is limited support for this conjecture. Thus, we report results with adjustments to all workers and only non-movers.

We plot our two corrected time series of occupational mobility opposite the aggregate occupational mobility measured retrospectively in the March CPS in Figure 7. In the case where we only adjust the occupational switching patterns of non-movers, corrected occupational mobility is approximately 50% higher than the level directly measured from the CPS; when we adjust all occupational switches, the corrected level of occupational mobility is approximately 70% higher, consistent with our baseline estimates.[22] This implies that our final estimates for the average rate of annual occupational mobility from 1981-2018 at the one, two and three digit levels are, respectively, 6-7%, 7-9%, and 9-11%.

A natural comparison to our findings are the findings of Kambourov and Manovskii (2008), who estimate actual annual occupational mobility in the PSID from 1981-1997, correcting for measurement error using recoded occupational switching for the years 1968-1980.[23] They estimate these corrections taking as given a recoding of occupations for the years 1968-1980, which are done observing a worker's full history of occupations. By observing a worker's full history for occupation, the recoding approach reduces coding error attributable to sensitivity in how a worker's description of an occupation is mapped to an actual numeric code. Their methodology is different from ours, however, in that their estimated corrections to occupational mobility take as given recoded occupational mobility from 1968-1980, and then estimate the magnitude of the affine shift upwards in occupational mobility post-1980.

There are a few adjustments needed to make a comparison to their findings, as well a few

[22]In both 1985 and 1995, information about geographic moves is unavailable, and thus the two time series provide the exact same estimate of occupational mobility.

[23]We could also compare our findings to those from other datasets, such as the SIPP or NLSY. However, since we are not aware of corrected time series of occupational switching from these datasets, it is unclear what can be gleaned from the comparison of our corrected numbers to the raw numbers in these data.

significant caveats to making this comparison. We adjust our sample to correspond to theirs – head of household men between the ages of 23-61 who are self-employed – and re-estimate actual occupational mobility. This sample adjustment will reduce our estimated levels of occupational mobility, almost solely due to the exclusion of young workers. It is also important to observe that the measures of occupational mobility obtained in Kambourov and Manovskii (2008) differ in at least two critical ways. First, their definition of occupational mobility does not require that individuals actually be employed in the past year, but computes the rate at which employed workers today had a different occupation in their last measured employment spell (at least a year prior). As they point out, this tends to raise their estimates of occupational mobility by 2.5 percentage points. Second, the PSID uses a different occupational coding scheme than the CPS, and thus we should not ex ante expect to obtain the same results. In particular, the PSID has roughly 30-50% more distinct occupations (depending on the level of aggregation), and thus, this could generate significantly higher levels of occupational mobility than we would measure in the CPS.[24]

Figure 8 plots our re-estimated aggregate measure of occupational mobility, adjusting all switches including movers, opposite the measures of corrected occupational mobility from the PSID in Figure 3 of Kambourov and Manovskii (2008) from 1981-1997, the years for which our two samples overlap.[25] In each case, our point estimates for occupational mobility are lower than those obtained by Kambourov and Manovskii (2008), differing by 8-11 percentage points on average. Again, roughly 2.5 percentage points of these differences are attributable to differences in definitions of occupational mobility, suggesting final gaps of 5.5-8.5 percentage points. A sizable portion of this gap is likely attributable to the differences in coding schemes. However, although we find somewhat lower rates of occupational mobility over this time period, our estimated time series of corrected occupational mobility is closer to their findings than any other measure of occupational mobility from the CPS.

4.6 Robustness

We provide a variety of robustness checks in Appendix C. There are potentially other labor market outcomes or individual characteristics we could include in our estimation, some of which are not available for every year in our sample. In the Appendix, we consider using other possible labor market outcomes, such percent changes in income, percent changes in reported hours worked, and changes in inclusion in an employer health plan, and find comparable results. We also expand the

[24]The PSID codes reported in the Appendix to Kambourov and Manovskii (2009b) indicate 9 one digit codes, 26 two digit codes and 429 three digit codes. In contrast, the time consistent occupation codes we use from the CPS has 6 one digit codes, 17 two digit codes and 325 three digit codes.

[25]Because the PSID becomes biennial after 1997, there is not a natural comparison to our findings post-1997, and thus we do not report our estimates for this sample post-1997. However, the downward trend in occupational mobility over time in our estimates is still observed, albeit with greater noise due to the reduced sample size.

set of individual characteristics that we control for, expanding them to include the level of income in the past year, average hours worked in the past year, self employed status, veteran status, full-time/part-time status in the prior year, and state fixed effects. We also consider results where we exclude occupation fixed effects from the set of individual characteristics. For a shorter time window, we can consider accounting for whether or not a proxy was used in responding to the survey, whether that proxy status changed between surveys, school enrollment, whether a person was born outside the United States, disability status, and whether or not a worker held multiple jobs. We find very similar results in all of these specifications, including for the point estimates relating individual characteristics to occupational switching and measurement error. We also report results where we use the job tenure and occupational mobility supplement instead of the March supplement and find similar results.

5 Applications

We now consider two applications of our estimates of annual occupational mobility. First, we show how we can use our annual estimates of occupational mobility to generate corrections to the time series of monthly occupational mobility. Second, we discuss how our findings regarding occupational mobility impact conclusions from two papers in the literature on the impact of trade and offshoring on workers.

5.1 Corrections to Monthly Occupational Switching Rates

Thus far, our study of occupational mobility has been focused on annual measures of occupational mobility, as annual mobility has multiple signals regarding occupational switching that can be used to estimate the actual level of occupational mobility. However, occupational mobility at the monthly frequency is also of interest and has been commonly used in the literature (for two recent examples, see Cortes and Gallipoli (2017) and Robinson (2018)). Using our estimates from Section 4, we can construct predicted probabilities for occupational switching in monthly CPS data. We then compare these predicted probabilities to both the raw data and existing approaches to correct occupational mobility in the monthly data.

One appeal of studying occupational mobility at the monthly frequency is that since 1994, for each of the four month windows a household is interviewed for, occupations are coded dependently for months two through four, similar to the coding in annual retrospective measures of occupational mobility. In every month prior to 1994, and for the first and fifth months households are surveyed after 1994, an individual's occupation is coded independently, similar to the coding in longitudinal annual measures of occupational mobility. Post-1994, occupations are also coded independently if an individual doesn't provide any information about whether or not he or she remained with the same employer. This naturally implies a substantial discontinuity in monthly occupational mobility around the year 1994, as can be seen in Figure 9, which plots the raw time series for occupational mobility from the linked monthly CPS.[26]

The most common approach to addressing occupational mobility measurement error in the monthly CPS comes from Moscarini and Thomsson (2007) (henceforth MT), who aim to isolate legitimate occupational switches by identifying "suspicious" switches and then filtering these suspicious switches in a variety of ways. MT label all switches prior to 1994 as suspicious and any switch after 1994 is labeled as suspicious when an individual has changed employers or their activities of work for the same employer or that data is missing. Suspicious switches are then determined to be spurious either if the worker didn't switch industry, didn't switch class of work (self-employed, private, gov't) and didn't report recently looking for work or if the full worker history of occupational switching before and after the reported switch exhibits unusual patterns.[27] MT acknowledge that this approach is conservative and requires substantial judgment in assessing what occupation histories are unusual, but their method produces a time series beginning in 1979 which appears consistent with other data sources for occupational switching and exhibits no large discontinuities.

Although our estimates for correcting occupational mobility were generated at the annual frequency, given data on the type of switch reported, individual characteristics and labor market outcomes, we can use our annual estimates to construct a predicted switching rate for each response in the monthly CPS. For each individual response, we compute a predicted probability that this person switched, given the data provided: $P(SW = 1 \mid X, Y, \tilde{SW})$. We provide a full description how this predicted probability is computed in Appendix D and give a general sketch of the method here.

We first map reported monthly occupational switching into the types of occupational switching observed in the annual data as follows. For monthly responses prior to 1994 and for cases after 1994 where there is missing data on whether or not the worker changed employers, we treat these switches as corresponding to a longitudinal switch from our annual framework (as independent coding is used in measuring occupations). For all other occupational switching responses from

[26]For comparability with the corrected series for monthly occupational mobility we present in Figure 10, we construct the raw monthly occupational mobility rate using only data on individuals occupation transitions between months 2 and 3 in the first four months the household is sampled. As described in the notes to Figure 9, we impose the same sample criteria for computing monthly occupational mobility as for computing annual occupational mobility in the linked sample (18+, non-government workers, no imputed occupations).

[27]MT provide a full listing of the types of occupational switching histories that they consider legitimate and spurious, and allow these to vary pre- and post-1994. For example, in any year, a switch from months 2 to 3 in the sample would be ruled spurious if an individual reported three different occupations in months 1-3 and then was unemployed in month 4 (flag 10 in their terminology).

1994 onward, we treat responses as corresponding to a reported retrospective switch in our annual data (since dependent coding has been used). All of the individual worker characteristics (X) we consider are observable for each individual in each month of the monthly CPS. However, of our labor market outcomes, we can only observe a change in part-time/full-time status, and since 1994, whether or not a worker switched employers. Given this data and estimates from our model, we construct the predicted probability of switching occupations for each response in the monthly data. In the results we report here, we use estimates from the pooled data to minimize noise in the predicted probabilities, however we report our results using the year by year estimates in Appendix D.

For consistency with MT, we restrict our sample of potential occupational switches in each month to only individuals observed in the second and third months of the first four month stretch they are surveyed. Otherwise, we impose the same sample restrictions as in our annual data, focusing on workers 18 and older not employed in the public sector, dropping all imputed values for occupations and all linked observations with inconsistent information on sex, race or age. We construct the MT filtered data following the procedures described in their paper for this same sample of workers for comparison. To reduce noisiness and facilitate comparison between different series of occupational switching, we smooth all monthly time series of occupational mobility using a 12 month moving average. Acknowledging the break in survey procedures in 1994, we do not smooth the data across this year.

Figure 10 presents monthly occupational mobility as predicted using our annual estimates, as calculated using the MT approach, and for the raw monthly data from 1994 on. Our predicted probability is higher than the monthly MT predictions, given that we estimate retrospective measures understate true mobility. However, given that the degree of recall bias at the annual and monthly frequencies may be different, our corrections for recall bias may be too large in the monthly data. Regardless, the dynamics of these two series are remarkably similar. Our monthly estimate of occupational mobility still exhibits a discontinuity around the year 1994, but it is much less pronounced than in the raw data.[28] Notably, the raw monthly data tracks the MT data very well until about the year 2007. However, after 2007, the raw monthly data indicates a substantial rise in occupational mobility, in contrast to both our and MT's estimates. This discrepancy is driven by increased nonresponse to whether or not an individual switched employers between months, triggering an independent coding of occupations, where there is a high potential for coding error.

We thus conclude that our estimates regarding measurement error and occupational mobility

[28]This is in part due to the lack of information on employer switches prior to 1994. If we don't use data on employer switches at all, only using the labor market outcome of a full-time/part-time switch, our time series of monthly occupational mobility still has a discontinuity around 1994, although it is upward instead of downward. Only using full-time/part-time switch to infer occupational switching results in a slightly higher level of monthly switching, consistent with our estimates of annual occupational mobility using only this data in Section 4.2.

at the annual time horizon can potentially be fruitfully applied to obtain corrected estimates of monthly data. Since our estimates track closely the dynamics of the MT estimates, we would argue that either approach yields a reasonable estimate of the trends in occupational mobility over time. We prefer our estimates because they account for the possibility of measurement error even in dependently coded measures in occupational mobility and they do not rely on judgment calls regarding which types of occupational switches are spurious. However, the implied rates of mobility may be biased upward because our implied corrections for recall bias in retrospective measures are drawn from annual data, not the monthly data. Regardless of one's preferences between our estimates and those from MT, we argue that some form of correction to monthly data is important, otherwise the increase of nonresponse in monthly data generates a spurious increase in occupational mobility since 2007.

We also conclude that whether using our estimates of monthly occupational mobility or those from MT, great caution is needed when trying to aggregate monthly occupational mobility to longer time horizons. For example, if one assumes that the probability of switching occupations operates as a Poisson process, arriving each month independently, then one would conclude that the annual occupational switching rate, SW_t^A, could be obtained from the monthly switching rate, SW_t^M, as $SW_t^A = f(SW_t^M) = 1-(1-SW_t^M)^{12}$. If this is applied to our monthly estimates of occupational mobility, it would imply annual switching rates of 20-40%, which are far greater than our annual estimates for occupational mobility.

One reason that this approach to time aggregation may fail is pointed out by MT – the probability of making an occupational switch is likely not independent over time, as workers who switch occupations today may be more likely to switch occupations in the future. Another possibility is simply that there exists substantial heterogeneity in occupational switching across observable or unobservable groups which may confound the time aggregation procedure described above. In this case, monthly occupational switching is given by a weighted average of groups, $SW_t^M = \sum_i \omega_{it} SW_{it}^M$. However, if one aggregates total monthly switching instead of monthly switching for each group separately, one will overstate actual annual occupational mobility. Aggregating the monthly total, one obtains $SW_t^A = f(SW_t^M) > \sum \omega_{it} f(SW_{it}^M)$, where the relationship follows because of Jensen's inequality. Regardless of the reason, our estimates of monthly mobility suggest that simple time aggregation exercises will significantly overstate the level of occupational mobility at longer time horizons.

5.2 Understanding the Worker Level Impacts of Trade and Offshoring

To conclude, we consider the implications of our estimates for recent literature on the worker-level impacts of trade and offshoring. A sizable literature in trade has focused on how worker flows in and out of industries or occupations can mitigate or exacerbate the costs and benefits of increased trade and offshoring. This literature is often interested in the costs associated with switching, either those associated directly with the direct cost of making a switch or because of the specificity of a worker's human capital in a given occupation or industry, and often links these to the levels of worker flows. Thus, the levels of worker mobility play a potentially important role in understanding the impact on workers from increases in trade.

5.2.1 Revisiting Structural Estimates of the Impact of Trade Liberalization: Artuç et al. (2010)

We first consider what our findings imply for structural models of trade adjustment which use worker flows across industries or occupations to estimate switching costs. While the emphasis of this paper is on occupational mobility, our findings can be naturally generalized to industry mobility, as it is measured in the same way as occupational mobility in the CPS.[29] In particular, we consider an approach used by Artuç et al. (2010) (henceforth ACM) to address concerns about measurement error in retrospectively measured industry mobility in the March CPS, an approach used since by Artuç and McLaren (2015) and Caliendo et al. (2015). ACM argue that there may be bias in retrospectively measured industry mobility because the exact time window over which a switch is occurring is not precisely defined in the March CPS. As a result, they compare mobility rates in the CPS to those in the NLSY and argue that the March CPS mobility rates actually only capture switching over a five month time horizon. As a result, they then use a time aggregation procedure, similar to the one described in the prior section, to transform the retrospective mobility rates in the CPS into an annual time frequency, assuming that the measured rates only represent five months worth of mobility. In practice, this amounts to increasing retrospectively measured industry to industry mobility rates by, on average, 130% (see their Table 4).

We argue that this approach is too conservative. First, returning to our Figure 1, we observe that both the job tenure and occupational mobility supplement and the March CPS give very similar estimates of the occupational mobility rate even though the March CPS compares the current occupation to the longest job of the prior year. We are thus skeptical that the time horizon used to measure occupational mobility in the March CPS is of first order concern. Second, ACM use the NLSY to benchmark their worker flows. However, the NLSY is also subject to similar types of coding error (see, for example, Speer (2016)) and thus may not provide a reliable benchmark. Third, even if the retrospective switching rates in the March CPS actually only capture mobility rates across five months, the time aggregation procedure used to inflate mobility rates will overstate actual annual mobility for the reasons discussed in the prior section.

[29] Again, if we look at industry mobility measured retrospectively and longitudinally, it exhibits the same differential levels and trends as for occupational mobility.

While we agree with their conclusion that unadjusted retrospective measures of worker mobility understate actual worker flows, the corrections we estimate for occupational mobility suggest that actual mobility rates are only approximately 56% higher than what is measured for the years they study, 1976-2000.[30] Using lower levels for mobility rates in estimating structural models of trade adjustment, such as the model of ACM, will tend to imply higher estimates of moving costs for workers across sectors or occupations. To give a sense of the quantitative impact of these higher moving costs, we first re-estimate the worker switching cost parameters of their model, but instead of using their adjustment, inflate all industry-industry flows by 56%, based off our estimates from occupational mobility. We then re-simulate one of the trade experiments in ACM using these new parameter estimates. The experiment we replicate is the effect of a sudden trade liberalization where a 30% tariff on manufacturing is suddenly removed, exposing manufacturing workers to increased competition from imports. A key message of ACM is that in doing welfare analysis, one needs to account for the fact that even manufacturing workers exposed to increased imports can benefit from a trade liberalization. Although wages for these workers may fall, because workers can switch industries, a reduction in prices from the elimination of a tariff can benefit workers via improving their option value of work in other sectors.

For brevity's sake, we focus on changes in welfare (measured as discounted lifetime utility) for workers in the import-competing sector, manufacturing in this case. In Figure 11, we plot both the original outcomes reported in ACM and our outcomes with the different parametrization for the same simulations for workers in the manufacturing sector as reported in the first column of Figure 5 of ACM.[31] The outcomes in the two panels differ based on assumptions regarding the annual discount factor ($\beta = 0.97$ or $\beta = 0.90$); these differences matter because a worker's option value from future switching opportunities will depend on how much they discount the future.

Our findings are consistent with the primary qualitative message of ACM, that even import-competing workers can benefit from a trade liberalization in the long run, and in some cases, the short run. However, with these higher moving costs in the simulation, we find that the benefits (costs) to workers in the import competing sectors are much smaller (larger) in the short run and take much longer to converge to their steady state values. In simulations with the higher discount factor, the welfare gains in our simulations are roughly 45% lower on impact, and in the simulations with the lower discount factor, the welfare losses on impact are more than 50% larger. In the original simulations of ACM, they report convergence to the steady state taking approximately eight years; in our simulations, on average, it takes fourteen years, almost twice as long, to reach

[30] This number is the average of our estimated scaling factor, $\frac{P(SW=1)}{P(\tilde{SW}^R=1)} = \frac{\delta_0}{\alpha_{R,0}+\alpha_{R,1}\delta_0}$, across all levels of aggregation for the time period 1981-2000.

[31] This simulation assumes that all goods in the economy are tradable. We have also looked at the results in the case where they assume that some goods in the economy are non-tradable; the differences we find from these simulations are so similar to the baseline ones we have reported that we omit them for brevity.

that same level of convergence. The direction of these findings is not surprising, as with higher moving costs, worker adjustment to shocks will be much more sluggish.

We again emphasize that our message is not that the primary conclusions of ACM are incorrect. Even with a different parametrization of their model and simulations, we still find that trade liberalizations can be beneficial to even directly impacted workers and we share their conclusion that our understanding of these impacts depends heavily on understanding patterns of worker mobility. We simply emphasize that the exact quantitative predictions regarding worker welfare to a trade shock appear to be sensitive to the levels of worker flows used to discipline model parameters.[32] However, we observe that since worker flows have been trending down over time, not only is it possible that trade adjustment is more painful and sluggish for exposed workers than previous estimates imply, but that this may be worsening over time.

5.2.2 Revisiting Reduced Form Estimates of the Impact of Trade: Ebenstein et al. (2014)

We conclude with a final example illustrating the impact of our corrections to occupational mobility by revisiting results from Ebenstein et al. (2014) (henceforth, EHMP). EHMP construct empirical measures of exposure to trade and offshoring and find that there are significant wage effects for workers when trade and offshoring exposure are measured at the occupational level. As a final exercise in their paper, they estimate the wage losses to occupation switchers who were originally in occupations classified as "tradable." They do this by instrumenting a reported longitudinal occupational switch with whether or not the original occupation was tradable. According to their findings, occupational switching for workers in tradable jobs is associated with wage losses of 12-17%.

Our findings of a very high error rate in longitudinal occupational switching suggests that there is a strong potential that these estimates are contaminated by measurement error. Importantly, even though their empirical strategy uses an instrumental variables approach, because the measurement error is non-classical, even if the instrument satisfies the usual conditions for identification, the IV estimates will still be biased. Further, given our findings that the measurement error in longitudinal occupational switching is higher for certain occupations, it is plausible that the instrument of whether or not an occupation is tradable may be correlated with the measurement error, further confounding estimates.

To assess the impact of measurement error in occupational switching for these findings on the wage effects of occupational switching for workers in tradable occupations, we extend our estimation strategy to include the specifications of EHMP and estimate the same IV coefficients

[32]We also note that the sensitivity of these adjustments to worker flows will depend on the estimation strategy used to recover model parameters. More recent work, such as Artuç and McLaren (2015), has used alternative estimation strategies, and thus the quantitative impact of these flow adjustments in other papers may be different.

as in their exercise, albeit corrected for measurement error. We provide greater detail of all steps of this procedure in Appendix E and provide a sketch of the approach here. The conditional expectations to be estimated in the original two stage least squares specification of EHMP can be written as:

$$\mathbb{E}\left[\tilde{SW}^{L}_{i,o,t} \mid Tradable_o, C_{i,o,t}\right] = \eta_0 + \eta_1 Tradable_o + C_{i,o,t}\eta_2 \tag{5}$$

$$\mathbb{E}\left[\Delta ln(w_{i,o,t}) \mid \tilde{SW}^{L}_{i,o,t}, C_{i,o,t}\right] = \xi_0 + \xi_1 \tilde{SW}^{L}_{i,o,t} + C_{i,o,t}\xi_2 \tag{6}$$

where $\tilde{SW}^{L}_{i,ot}$ is 0/1 variable measuring whether or not an individual i in occupation o last year reported a longitudinal switch; $Tradable_o$ is the instrument for occupational switching, measuring whether or not an individual's prior year occupation is exposed to globalization and offshoring (as constructed in EHMP); $\Delta ln(w_{it})$ is the one year log change in wages for a worker i who was in occupation o last year; and $C_{i,o,t}$ is a vector of individual and occupational characteristics used as controls – age, female, nonwhite, union status, a set of educational dummy variables, state fixed effects, year fixed effects, and finally a pre-switching occupation wage premium for occupational switchers.

We extend our generalized method of moments estimation to include the moment conditions for the first stage and IV parameters of their estimation and estimate the entire system of moments simultaneously. We adjust our sample to match theirs, focusing on non-self-employed workers between the ages 16-64. We also use the individual characteristics from their specification, along with the instrument of whether or not an occupation is tradable, as the covariates in our estimation of measurement error (setting $X_{it} = [Tradable_o; C_{i,o,t}]$). Finally, the CPS data used in EHMP comes from the outgoing rotation group sample from the monthly CPS, linked to responses one year prior. To access certain variables only available in this sample (union status, in particular), we need to merge outgoing rotation group data into our sample of linked responses. We do this using the linking keys constructed in Flood and Pacas (2016), which link data in March basic files (including individuals in the outgoing rotation group) and the March supplement to the CPS used in our baseline analysis. Because these linking keys extend back only to 1989, even though the original results of EHMP study the years 1984-2002, we generate results only for the time period 1989-2002. Even with this different sample and set of covariates in estimation, we show in Appendix E that our estimates for occupational switching and measurement error are very similar to those reported in Table 2.

To maximize comparability to the original findings of EHMP, we present baseline GMM estimates where we simultaneously estimate all parameters, but estimate measurement error and occupational mobility parameters using just the linked March sample (with the above adjustments), and estimate the parameters of the EHMP specifications on their original data sample of outgoing rotation groups. With this approach, our estimation of their specification is almost identical to their original work. The only adjustments made are to restrict the time frame to 1989-2002 and convert the occupational definitions in their data to the time-consistent occupational codes that we use from Autor and Dorn (2013). In estimating parameters from two distinct samples, we are treating the two samples as disjoint when they are actually not, which could bias our estimates, particularly our standard errors. As a result, in the Appendix, we also report results where we either exclude all potentially overlapping observations, to ensure the two samples are disjoint, and results where we link as much of their data as possible to our core estimating sample of linked March supplement responses and estimate all parameters jointly. Our findings in these two alternate estimations are very similar to the baseline we report.

Before reporting the results of our estimation, we first illustrate how our measurement error corrections could impact the IV estimates. Given the framework for measurement error and occupational switching described in equations (1)-(4) and the two stage least squares specification from EHMP in (5)-(6), we can represent the IV coefficient corrected for measurement error, $\hat{\xi}_1^{IV}$, relative to the uncorrected coefficient, $\tilde{\xi}_1^{IV}$, in the following way:

$$\hat{\xi}_1^{IV} = \frac{\tilde{\eta}_1}{\hat{\eta}_1}\tilde{\xi}_1^{IV} = \frac{\alpha_{L,1}\delta_1(Trad.) + \alpha_{L,X}(Trad)}{\delta_1(Trad.)}\tilde{\xi}_1^{IV} \tag{7}$$

where $\tilde{\eta}_1$ is the uncorrected first stage estimate of the relationship between tradable occupations and occupational switching, $\hat{\eta}_1$ is the first stage estimate corrected for measurement error, $\alpha_{L,1}$ is the parameter that tells how much more likely a reported longitudinal occupational switch is when there is an actual occupational switch, $\delta_1(Trad.)$ is the coefficient relating the tradable instrument to actual occupational switching, and $\alpha_{L,X}(Trad.)$ is the coefficient relating the tradable instrument to the measurement error in longitudinal occupational switching.

The key insight here is that the corrections to measurement error amount to correcting the first stage estimates. If there are no errors in the first stage estimates, $\hat{\eta}_1 = \tilde{\eta}_1$, then the IV estimate will not be affected by measurement error in occupational switching. The correctly estimated first stage relationship between occupational switching and the tradable instrument is $\hat{\eta}_1 = \delta_1(Trad.)$ and is estimated from our core set of moments, whereas the first stage estimated from the data without correction is captured by $\tilde{\eta}_1 = \alpha_{L,1}\delta_1(Trad.) + \alpha_{L,X}(Trad)$.

There are two key sources of bias here – the fact that $\alpha_{L,1}$ may be less than one, representing the fact that there are some false positive longitudinal switches (since $\alpha_{L,0} + \alpha_{L,1} \leq 1$), and the fact that $\alpha_{L,X}(Trad.)$ may not be zero, which would be the case if the tradable instrument is correlated with the error in longitudinal occupational switching. In the case where being in a tradable occupation is orthogonal to the measurement error in occupational switching ($\alpha_{L,X}(Trad.) = 0$), then the true IV coefficient will be smaller in absolute value than the estimated one in the data, as the first stage coefficient is attenuated by the measurement error in occupational switching (because $\alpha_{L,1} < 1$). On the other hand, if there is some correlation between being in a tradable occupation and measurement error in occupational switching, then the nature of the bias can't be signed ex ante, as it will depend on the signs and magnitudes of all the parameters. In fact, if $\delta_1(Trad.)$ and $\alpha_{L,X}(Trad.)$ are of opposite sign, the true IV coefficient could be of the opposite sign of that estimated in the data.

We report in Table 6 our estimates of the both the uncorrected and corrected IV coefficients and first stage coefficients, at one, two and three digit levels of occupational switching. Similar to the original findings of EHMP, we find that the uncorrected IV coefficient implies wage losses for workers in tradable occupations who switch occupations ranging between 10-17%. However, when we correct for the measurement error in occupational switching, the IV coefficient flips sign and increases in magnitude. Our results imply that occupational switchers who were originally in tradable occupations see wage *gains* of 40-60 log points. The differences between these two point estimates are statistically significant.

The difference in these two estimates is driven by the different estimates of the first stage relationship between occupational switching and working in a tradable occupation. This can be seen in the bottom panel of Table 6 , which reports the estimated first stage coefficients. Without correcting for measurement error, the first stage is positive – workers in tradable occupations are more likely to switch occupations. However, when measurement error is corrected for, the sign of the first stage coefficient flips and becomes negative – workers in tradable occupations are *less* likely to switch occupations. That is, workers originally in tradable occupations are more likely to *report* an occupational switch longitudinally, but less likely to have *actually* switched.

Our robustness exercises in Appendix E show that this same result obtains when we vary the samples used in estimation to make them perfectly disjoint or perfectly overlapping. Though unreported, we also find similar results if we extend the sample window through the year 2017, if we assign whether or not an occupation is tradable based on the retrospectively reported prior occupation, and if estimate all specifications without covariates.[33] We also find similar results in linked data where we use the retrospective occupational switching measure instead of the longitudinal occupational switch. This is particularly appealing since using linked retrospective responses is less computationally intensive and does not require structural estimation, yet recovers the same result. In each case, we find that workers in tradable occupations were less likely to switch occupations and the IV estimates of wage changes with occupational switching imply wage gains to occupational switchers in those occupations.

[33] An additional possible concern is that wage changes are being used in both the final results specification of EHMP and as part of our estimation of measurement error. Since we estimate everything jointly, there is no circular contamination of our estimates from using hourly wage changes in this way. Nevertheless, our results are robust to not using hourly wage changes in estimating the measurement error in occupational switching.

How should we interpret these results? Even with the corrections to measurement error, we are skeptical that these IV estimates actually reflect causal wage impacts of displacement by trade or offshoring, as there are many potential confounding factors that could hinder proper identification here, including selection and unobservable worker differences. The finding we place greater emphasis on, however, is that correcting for measurement error, workers in tradable jobs are less likely to switch occupations than other workers. This implies sluggish worker-level adjustment to trade and offshoring shocks. Thus, a potential concern is that in response to a trade shock, workers may be "stuck" in these tradable occupations and face substantial welfare losses from their relative immobility.

This provides but one example of how longitudinally measured occupational switching can lead to misleading inference about the impacts of occupational switching. In particular, this concern will inevitably arise when using the CPS to study occupational switching and wage changes, since data on changes in wages and income can only be constructed when measuring occupational mobility longitudinally.[34] As a result, a proper understanding of the relationship between wage changes and occupational switching from CPS data necessitates addressing this measurement error.

6 Conclusion

Our paper starts from the observation that measuring occupational mobility within the Current Population Survey retrospectively and longitudinally yield very different answers about the levels and trends of occupational mobility over the last 40 years. Using linked data on individuals for whom we can observe both measures of an occupational switch, we estimate the true level of occupational switching and the nature of the measurement error in both measures of occupational mobility. We find that occupational mobility has indeed been trending downwards, and that measurement error in longitudinal measures has been increasing over time. Further, this measurement error trend cannot be explained by observable worker characteristics, including two digit occupation fixed effects. While retrospective measures of occupational mobility are closer to our estimated levels of occupational mobility, we find that they are still significantly lower than the actual rate of occupational mobility. While we naturally conjecture that these measurement errors are driven by coding errors in the longitudinal responses and recall bias in the retrospective responses, further work is needed to understand better the sources of error and how they have changed over time.

We also show how our estimates can be used to construct a corrected time series of monthly

[34] A notable exception to this is for displaced workers, as the displaced worker survey both retrospectively asks about a worker's prior occupation and the wages or income received at that job. That said, depending on the estimation strategy used, there could still be significant bias arising from using even retrospectively measured occupational switching to understand changes in wages.

occupational mobility. Importantly, consistent with previous work, we find that standard time aggregation methods that might be used to convert monthly switching rates to annual switching rates are likely to substantially overstate the actual level of occupational mobility. Further, failure to correct for measurement error in monthly occupational mobility will lead to a spurious increase in occupational mobility since the year 2007.

We apply our corrections for occupational mobility to literature on the impact of trade on workers, namely Artuç et al. (2010) and Ebenstein et al. (2014). We find that the response of workers to trade shocks may be even more sluggish than expected, with lower worker flows implying higher moving costs in structural models of trade and greater losses along the transition for workers exposed to import competition. We also find that workers in tradable occupations are less likely to switch occupations, suggesting the possibility that workers may be "stuck" in tradable jobs and have less mobility prospects than previously believed.

We provide online all our estimated series for occupational mobility as well as the code used to estimate and construct them. Our hope is that providing these improved estimates of occupational mobility will be helpful in further research on the subject.

References

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

AIGNER, D. J. (1973): "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, 1, 49–59.

ARTUÇ, E., S. CHAUDHURI, AND J. MCLAREN (2010): "Trade shocks and labor adjustment: A structural empirical approach," *American economic review*, 100, 1008–45.

ARTUÇ, E. AND J. MCLAREN (2015): "Trade policy and wage inequality: A structural analysis with occupational and sectoral mobility," *Journal of International Economics*, 97, 278–294.

ATALAY, E., P. PHONGTHIENGTHAM, S. SOTELO, AND D. TANNENBAUM (2018): "The Evolving US Occupational Structure," .

AUTOR, D. H. AND D. DORN (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review*, 103, 1553–97.

BINGLEY, P. AND A. MARTINELLO (2017): "Measurement error in income and schooling and the bias of linear estimators," *Journal of Labor Economics*, 35, 1117–1148.

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement error in survey data," in *Handbook of econometrics*, Elsevier, vol. 5, 3705–3843.

CALIENDO, L., M. DVORKIN, AND F. PARRO (2015): "The impact of trade on labor market dynamics," Tech. rep., National Bureau of Economic Research.

CARRILLO-TUDELA, C., L. VISSCHERS, AND D. WICZER (2019): "Cyclical Earning and Employment Transitions," .

CORTES, G. M. (2016): "Where have the middle-wage workers gone? a study of polarization using panel data," *Journal of Labor Economics*, 34, 63–105.

CORTES, G. M. AND G. GALLIPOLI (2017): "The costs of occupational mobility: An aggregate analysis," *Journal of the European Economic Association*, 16, 275–315.

CORTES, G. M., N. JAIMOVICH, AND H. E. SIU (2018): "The" End of Men" and Rise of Women in the High-Skilled Labor Market," Tech. rep., National Bureau of Economic Research.

CUBAS, G. AND P. SILOS (2017): "Social Insurance and Occupational Mobility," .

DORN, D. (2009): "Essays on Inequality, Spatial Interaction, and the Demand for Skills," Ph.D. thesis, University of St. Gallen.

EBENSTEIN, A., A. HARRISON, M. MCMILLAN, AND S. PHILLIPS (2014): "ESTIMATING THE IMPACT OF TRADE AND OFFSHORING ON AMERICAN WORKERS USING THE CURRENT POPULATION SURVEYS," *The Review of Economics and Statistics*, 96, 581–595.

EROSA, A., L. FUSTER, G. KAMBOUROV, AND R. ROGERSON (2017): "Hours, occupations, and gender differences in labor market outcomes," .

FLOOD, S., M. KING, S. RUGGLES, AND J. R. WARREN (2018): "Integrated Public Use Microdata Series, Current Population Survey: Version 6.0," [Machine-readable database]. Minneapolis: University of Minnesota.

FLOOD, S. AND J. PACAS (2016): "Using the Annual Social and Economic Supplement with Current Population Survey Panels," *Minnesota Population Center WP*, 4.

FORSYTHE, E. (2018): "Careers within Firms: Occupational Mobility over the Life Cycle," .

GERVAIS, M., N. JAIMOVICH, H. E. SIU, AND Y. YEDID-LEVI (2016): "What should I be when I grow up? Occupations and unemployment over the life cycle," *Journal of Monetary Economics*, 83, 54–70.

GORRY, A., D. GORRY, AND N. TRACHTER (2014): "Learning and life cycle patterns of occupational transitions," .

GROES, F., P. KIRCHER, AND I. MANOVSKII (2014): "The U-shapes of occupational mobility," *The Review of Economic Studies*, 82, 659–692.

GUVENEN, F., B. KURUSCU, S. TANAKA, AND D. WICZER (2015): "Multidimensional skill mismatch," Tech. rep., National Bureau of Economic Research.

HAUSMAN, J. A., J. ABREVAYA, AND F. M. SCOTT-MORTON (1998): "Misclassification of the dependent variable in a discrete-response setting," *Journal of econometrics*, 87, 239–269.

HUCKFELDT, C. (2016): "Understanding the scarring effect of recessions," *Report, Economics Department*.

KAMBOUROV, G. AND I. MANOVSKII (2008): "Rising occupational and industry mobility in the United States: 1968–97," *International Economic Review*, 49, 41–79.

——— (2009a): "Occupational mobility and wage inequality," *The Review of Economic Studies*, 76, 731–759.

——— (2009b): "Occupational specificity of human capital," *International Economic Review*, 50, 63–115.

——— (2013): "A cautionary note on using (march) current population survey and panel study of income dynamics data to study worker mobility," *Macroeconomic Dynamics*, 17, 172–194.

KANE, T. J., C. E. ROUSE, AND D. STAIGER (1999): "Estimating returns to schooling when schooling is misreported," Tech. rep., National Bureau of Economic Research.

LALÉ, E. (2012): "Trends in occupational mobility in France: 1982–2009," *Labour Economics*, 19, 373–387.

MADRIAN, B. C. AND L. J. LEFGREN (2000): "An approach to longitudinally matching Current Population Survey (CPS) respondents," *Journal of Economic and Social Measurement*, 26, 31–62.

MATHIOWETZ, N. A. (1992): "Errors in reports of occupation," *The Public Opinion Quarterly*, 56, 352–355.

MATHIOWETZ, N. A. AND G. J. OUNCAN (1988): "Out of work, out of mind: Response errors in retrospective reports of unemployment," *Journal of Business & Economic Statistics*, 6, 221–229.

MOLLOY, R., C. L. SMITH, AND A. K. WOZNIAK (2014): "Declining migration within the US: The role of the labor market," Tech. rep., National Bureau of Economic Research.

MOSCARINI, G. AND K. THOMSSON (2007): "Occupational and Job Mobility in the US," *The Scandinavian Journal of Economics*, 109, 807–836.

PAPAGEORGIOU, T. (2014): "Learning your comparative advantages," *Review of Economic Studies*, 81, 1263–1295.

POLETAEV, M. AND C. ROBINSON (2008): "Human capital specificity: evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000," *Journal of Labor Economics*, 26, 387–420.

RIVERA DREW, J. A., S. FLOOD, AND J. R. WARREN (2014): "Making full use of the longitudinal design of the Current Population Survey: Methods for linking records across 16 months\m {1}," *Journal of economic and social measurement*, 39, 121–144.

ROBINSON, C. (2018): "Occupational Mobility, Occupation Distance, and Specific Human Capital," *Journal of Human Resources*, 53, 513–551.

SPEER, J. D. (2016): "How bad is occupational coding error? A task-based approach," *Economics Letters*, 141, 166–168.

TRAIBERMAN, S. ET AL. (2017): "Occupations and Import Competition," in *2017 Meeting Papers*, Society for Economic Dynamics, 1237.

WICZER, D. (2015): "Long-term unemployment: Attached and mismatched?" .

XU, M. (2019): "Understanding the Decline in Occupational Mobility," .

Table 1: Summary Statistics

Variable	Mean	Mean $(L=Y,R=Y)$	Mean $(L=N,R=N)$	Mean $(L=N,R=Y)$	Mean $(L=Y,R=N)$	Std. Dev.
Longitudinal Occupational Switch (one digit)	0.212	1.000	0.000	0.000	1.000	0.409
Longitudinal Occupational Switch (two digit)	0.283	1.000	0.089	0.264	1.000	0.450
Longitudinal Occupational Switch (three digit)	0.414	1.000	0.254	0.559	1.000	0.493
Retrospective Occupational Switch (one digit)	0.025	1.000	0.000	1.000	0.000	0.157
Retrospective Occupational Switch (two digit)	0.032	1.000	0.007	1.000	0.006	0.176
Retrospective Occupational Switch (three digit)	0.042	1.000	0.018	1.000	0.015	0.200
Male	0.540	0.567	0.525	0.533	0.599	0.498
Age	43.44	35.14	43.93	37.94	42.44	12.90
Non-white	0.112	0.099	0.111	0.120	0.117	0.315
Hispanic	0.050	0.046	0.047	0.050	0.060	0.217
Marital Status (=1 if married)	0.667	0.521	0.676	0.559	0.646	0.471
Has Less than HS education	0.079	0.080	0.079	0.084	0.079	0.270
Completed HS degree	0.325	0.358	0.319	0.297	0.346	0.468
Completed Some College	0.266	0.337	0.251	0.343	0.315	0.442
Has College Degree (and/or additional education)	0.331	0.226	0.351	0.275	0.260	0.470
% ΔHourly Wage > 10%	0.738	0.822	0.731	0.810	0.757	0.440
Number of employers in past year >1 (measured year t)	0.094	0.537	0.077	0.506	0.105	0.291
Weeks employed between year $t-1$ and $t > 26$	0.976	0.885	0.981	0.891	0.971	0.152
Switched full-time/part-time status	0.096	0.280	0.087	0.220	0.108	0.294
N	526,718	10,738	413,453	2,818	99,709	526,718

The above table presents summary statistics for the labor market outcomes, individual characteristics and switching measures used in estimation. Data is from linked March supplements to the CPS, 1981-2018. Sample is restricted to individuals employed in both samples, 18 or older, not employed in government industries, and with no imputations for any variables used in the estimation. All moments are computed using survey weights. With a linked sample, each observation corresponds to a pair of responses in the March CPS, one in year $t-1$ and in year t, and changes over time are measured as the differences across these time periods. Columns 3-6 of the table present the means of each variable conditional on each possible combination of outcomes in both occupational switching measures, where occupational switching is defined at the one digit level. See Section 2 for more details on data construction and sample.

Table 2: Pooled Estimates of Actual Occupational Mobility and Measurement Error in Reported Occupational Mobility, 1981-2018

	One Digit	Two Digit	Three Digit
Actual Avg. Occupational Switching, δ_0	0.043	0.054	0.068
	(0.001)	(0.001)	(0.001)
Reported Ret. Occ. Mobility	0.025	0.032	0.042
	(0.000)	(0.000)	(0.000)
Reported Long. Occ. Mobility	0.212	0.283	0.414
	(0.001)	(0.001)	(0.001)
False Positive Rate, Retrospective	0.027	0.032	0.047
	(0.006)	(0.004)	(0.004)
False Positive Rate, Longitudinal	0.837	0.836	0.850
	(0.002)	(0.002)	(0.002)
False Negative Rate, Retrospective	0.020	0.024	0.030
	(0.001)	(0.001)	(0.001)
False Negative Rate, Longitudinal	0.011	0.011	0.010
	(0.000)	(0.000)	(0.000)
N	526,718	526,718	526,718

These results report occupational mobility and measurement error estimates from the model described in Section 3. Data used is longitudinally linked March CPS responses ranging from 1980-2018, providing occupational mobility estimates from 1981-2018; details of the linking and sample selection are described in Section 2. Point estimates and standard errors (reported below in parentheses) are obtained from standard two stage GMM estimates. Standard errors for false positive and negative rates (defined in the text in Section 4.1) are constructed via the delta method. Reported standard errors for measured longitudinal and retrospective occupational mobility are the standard deviation of occupational mobility divided by $\sqrt{N}$. All estimates are from the model where the vector of individual characteristics is defined as de-meaned age, age squared, and indicators for sex, race and ethnicity (white, nonwhite and Hispanic), educational attainment (below HS, HS degree, some college, college degree or more) marital status (married and unmarried), and two digit occupational fixed effects measured using the present occupation.

Table 3: Estimated Relationships Between Individual Characteristics and Actual and Reported Occupational Mobility, 1981-2018

Individual Characteristic	Actual	Longitudinal		Retrospective	
		Control for Switch	Unconditional	Control for Switch	Unconditional
Male	0.000	0.038	0.038	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Hispanic	-0.022	0.035	0.022	0.003	-0.009
	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
Age/1000	-8.902	-1.308	-6.764	0.282	-4.719
	(0.316)	(0.312)	(0.284)	(0.097)	(0.129)
Age squared/1000	0.077	0.009	0.056	-0.002	0.041
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Nonwhite	-0.016	0.016	0.007	0.003	-0.006
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Married	-0.007	-0.011	-0.015	-0.001	-0.005
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Below HS degree	0.001	0.016	0.017	0.004	0.004
	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
Has only HS degree	0.001	0.030	0.031	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Completed some college	0.006	0.054	0.058	0.004	0.008
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Avg rate | switch = 0	-	0.185	0.185	0.001	0.001
		(0.001)	(0.001)	(0.000)	(0.000)
Avg rate | switch = 1	0.043	0.798	0.798	0.563	0.563
	(0.001)	(0.005)	(0.005)	(0.008)	(0.008)
N	526,718	526,718	526,718	526,718	526,718

These results report the estimated relationships between individual demographic characteristics and actual and reported occupational mobility rates. For reported occupational mobility rates, we report both the coefficient on individual characteristics holding constant whether or not a switch occurred as well as the relationship with measured switching unconditional of an actual switch. The bottom rows reporting the average rate report the probability of observing a switch conditional on a switch occurring or not occurring. For education, the reference category which has been omitted is having a college degree or more education. Data used is longitudinally linked March CPS responses ranging from 1980-2018, providing occupational mobility estimates from 1981-2018; details of the linking and sample selection are described in Section 2. Point estimates and standard errors (reported below in parentheses) are obtained from standard two stage GMM estimates. Coefficients and standard errors for age squared are scaled to reduce the number of leading zeros. These results are obtained in a specification also including two digit occupation fixed effects; results for these are reported in Table 4.

Table 4: Estimated Relationships Between Occupation and Actual and Reported Occupational Mobility, 1981-2018

Occupation in Year *t*	Actual	Longitudinal		Retrospective	
		Control for Switch	Unconditional	Control for Switch	Unconditional
Executive, Administrative and Managerial Occ.	-0.029	0.066	0.048	0.005	-0.012
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Management Related Occ.	-0.024	0.051	0.036	0.007	-0.006
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Professional Specialty Occ.	-0.033	-0.069	-0.090	0.005	-0.014
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Technical and Related Support Occ.	-0.034	0.071	0.051	0.004	-0.015
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Sales Occ.	-0.018	0.003	-0.009	0.002	-0.008
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Administrative Support Occ.	-0.025	0.004	-0.012	0.002	-0.012
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Housekeeping and Cleaning Occ.	0.007	-0.079	-0.075	-0.007	-0.003
	(0.008)	(0.007)	(0.007)	(0.002)	(0.003)
Protective Service Occ.	0.024	-0.056	-0.041	-0.008	0.006
	(0.009)	(0.008)	(0.008)	(0.003)	(0.004)
Other Service Occ.	0.004	-0.012	-0.009	-0.003	-0.001
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Farm Operators and Managers	-0.011	-0.104	-0.110	-0.005	-0.012
	(0.005)	(0.005)	(0.005)	(0.001)	(0.002)
Other Agricultural and Related Occ.	0.003	0.002	0.003	-0.001	0.000
	(0.008)	(0.009)	(0.008)	(0.003)	(0.004)
Mechanics and Repairers	-0.029	0.025	0.007	0.002	-0.014
	(0.003)	(0.004)	(0.004)	(0.001)	(0.001)
Construction Trades	-0.021	0.000	-0.013	0.003	-0.009
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Extractive Occ.	-0.006	0.139	0.135	0.015	0.012
	(0.017)	(0.022)	(0.020)	(0.007)	(0.009)
Precision Production Occ.	-0.023	0.123	0.108	0.003	-0.010
	(0.003)	(0.005)	(0.004)	(0.001)	(0.002)
Machine Operators, Assemblers and Inspectors	-0.008	-0.003	-0.009	-0.001	-0.006
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)
Avg rate | switch = 0	-	0.185	0.185	0.001	0.001
		(0.001)	(0.001)	(0.000)	(0.000)
Avg rate | switch = 1	0.043	0.798	0.798	0.563	0.563
	(0.001)	(0.005)	(0.005)	(0.008)	(0.008)
N	526,718	526,718	526,718	526,718	526,718

These results report the estimated relationships between reported occupation in year *t* (the year following a switch) and actual and reported occupational mobility rates. For reported occupational mobility rates, we report both the coefficient on individual characteristics holding constant whether or not a switch occurred as well as the relationship with measured switching unconditional of an actual switch. The bottom rows reporting the average rate report the probability of observing a switch conditional on a switch occurring or not occurring. The reference occupation omitted is transportation and material moving occupations. A full list of the mapping of occupation codes into these sixteen occupations is provided in Appendix A. Data used is longitudinally linked March CPS responses ranging from 1980-2018, providing occupational mobility estimates from 1981-2018; details of the linking and sample selection are described in Section 2. Point estimates and standard errors (reported below in parentheses) are obtained from standard two stage GMM estimates. Coefficients and standard errors for age and age squared are scaled to reduce the number of leading zeros. These results are obtained in a specification also including individual demographic characteristics; results for these are reported in Table 3.

Table 5: Estimates of the Relationship Between Occupational Mobility and Labor Market Outcomes, 1981-2018

Labor Market Outcome	Actual	Retrospective	Longitudinal
Number of employers in past year > 1	0.437	0.421	0.065
	(0.000)	(0.002)	(0.001)
Switched full-time/part-time work status	0.178	0.159	0.039
	(0.000)	(0.003)	(0.001)
Weeks employed between year $t-1$ and t > 26	-0.097	-0.085	-0.018
	(0.000)	(0.001)	(0.001)
% Δhourly wage > 10%	0.083	0.071	0.033
	(0.000)	(0.004)	(0.002)
N	526,718	526,718	526,718

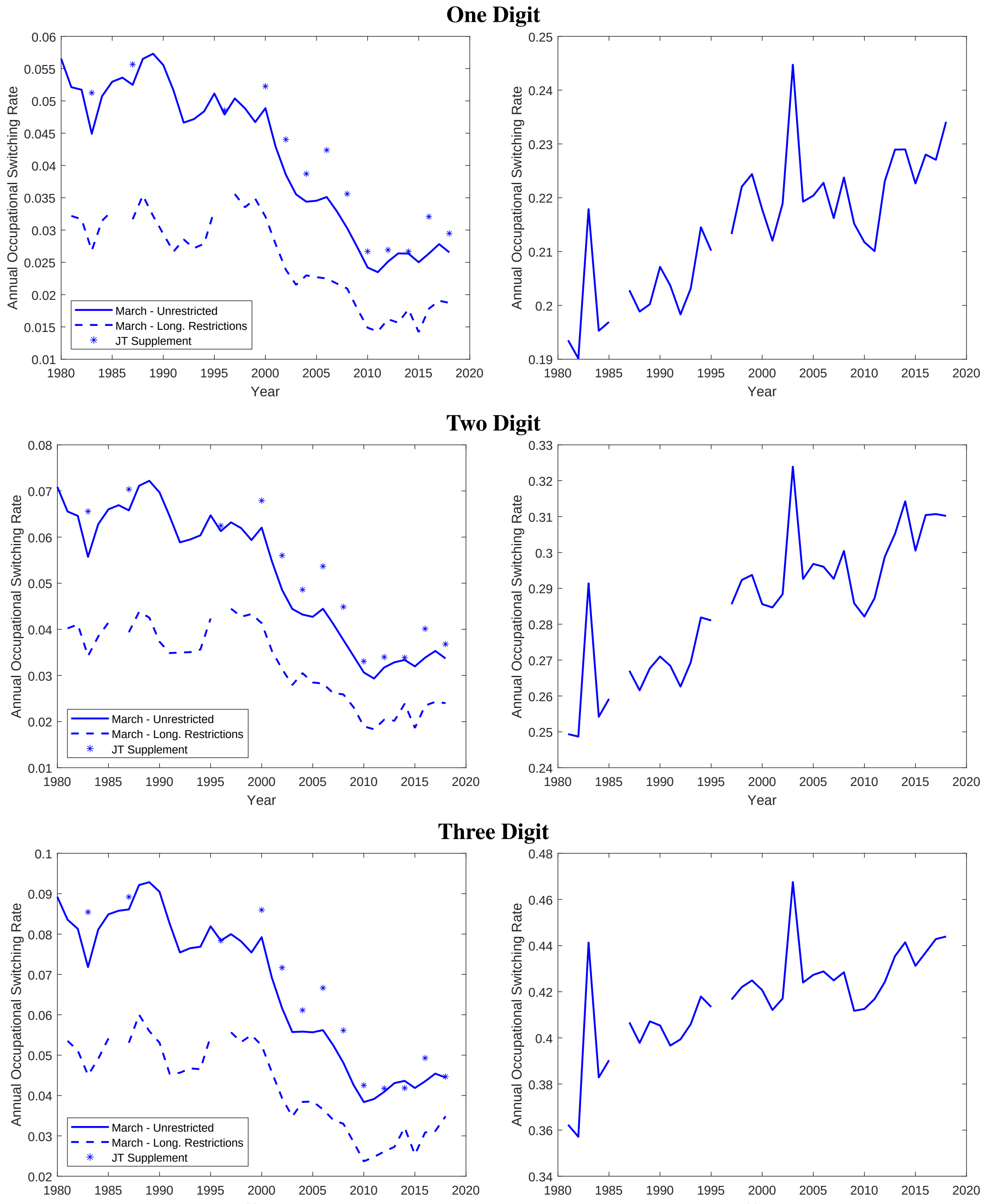
These results report the estimated relationship between actual occupational mobility and the labor market outcomes used in the estimation of the model described in Section 5. Parameter estimates correspond to the parameter β_1. Data used is longitudinally linked March CPS responses ranging from 1980-2018, providing occupational mobility estimates from 1981-2018; details of the linking and sample selection are described in Section 2. All estimates are obtained from a postulated linear relationship between the outcome and occupational switching, with controls for age, age squared, and indicators for sex, race and ethnicity (white, nonwhite and Hispanic), educational attainment (below HS, HS degree, some college, college degree or more) marital status (married and unmarried), and two digit occupational fixed effects measured using the present occupation. Point estimates and standard errors (reported below in parentheses) for actual occupational mobility are obtained from standard two stage GMM estimates. Point estimates and standard errors for the relationship between outcomes for retrospective and longitudinal measures of occupational mobility are obtained from a standard OLS regression.

Table 6: Estimates of Wage Changes with Occupational Switching for Workers in Tradable Occupations, 1989-2002

Parameter	One Digit	Two Digit	Three Digit
IV estimate of occ. switching on wages with tradable	-0.119	-0.173	-0.106
instrument, uncorrected ($\tilde{\xi}_1^{IV}$)	(0.033)	(0.022)	(0.016)
IV estimate of occ. switching on wages with tradable	0.403	0.559	0.348
instrument, corrected ($\hat{\xi}_1^{IV}$)	(0.117)	(0.075)	(0.055)
Difference	-0.522	-0.732	-0.454
	(0.149)	(0.094)	(0.071)
First stage estimate of tradable occupation on	0.045	0.070	0.091
occupational switching, uncorrected ($\tilde{\eta}_1^{IV}$)	(0.002)	(0.002)	(0.002)
First stage estimate of tradable occupation on	-0.013	-0.022	-0.028
occupational switching, corrected ($\hat{\eta}_1^{IV}$)	(0.002)	(0.002)	(0.003)
Difference	0.058	0.091	0.119
	(0.003)	(0.003)	(0.003)

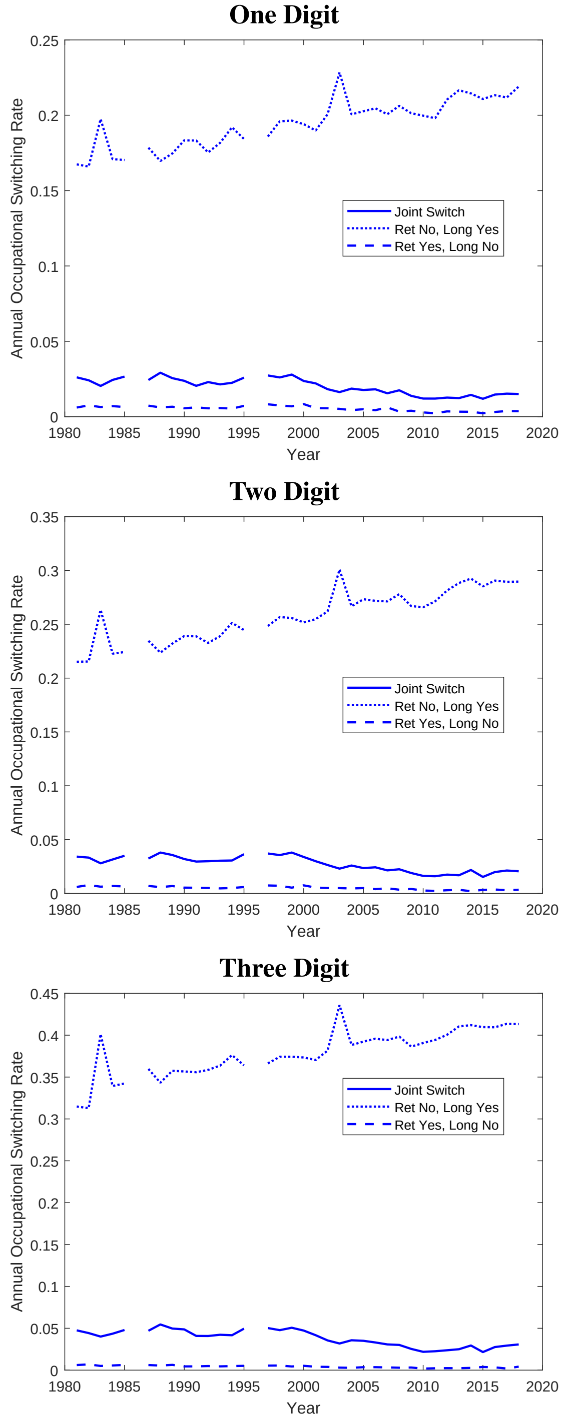
These results report the estimated IV relationship between changes in log wages and changes in occupational switching, where occupational switching is instrumented for using whether or not an occupation is tradable. These coefficients are estimated from a sample of non-self-employed workers ages 16-64 over the years 1989-2002 who report employment and wages in two consecutive years in the CPS outgoing rotation group data. These specifications are drawn from Ebenstein et al. (2014), and the first row of the top and bottom panels reports estimates comparable to the results in Table 7 of their paper. Differences between these point estimates and their results primarily result from reducing the time window from 1984-2002 to 1989-2002 and recoding occupations and occupational switches to use time-consistent occupational codes from Autor and Dorn (2013). The top panel reports the IV coefficients; the bottom panel reports the first stage coefficients. All specifications include controls for age, female, nonwhite, union status, a set of educational dummy variables, state fixed effects, year fixed effects, and a pre-switching occupation wage premium for occupational switchers (details of construction are in Appendix E). Each specification is estimated jointly with parameters determining the actual occupational mobility rate and measurement error in occupational mobility via two stage GMM; the sample for these estimates is of linked responses to the March CPS supplement for the same worker population and time frame. In estimation, the two samples are treated as disjoint, however, Appendix E reports results with this assumption relaxed. Specifications for the IV regressions reported above can be found in equations 5 and 6 in the text; correction terms for IV and first stage coefficients can be found in equation 7. Standard errors for IV correction estimates are constructed using the delta method.

Figure 1: Annual Occupational Mobility Rates in the CPS, Retrospective vs. Longitudinal



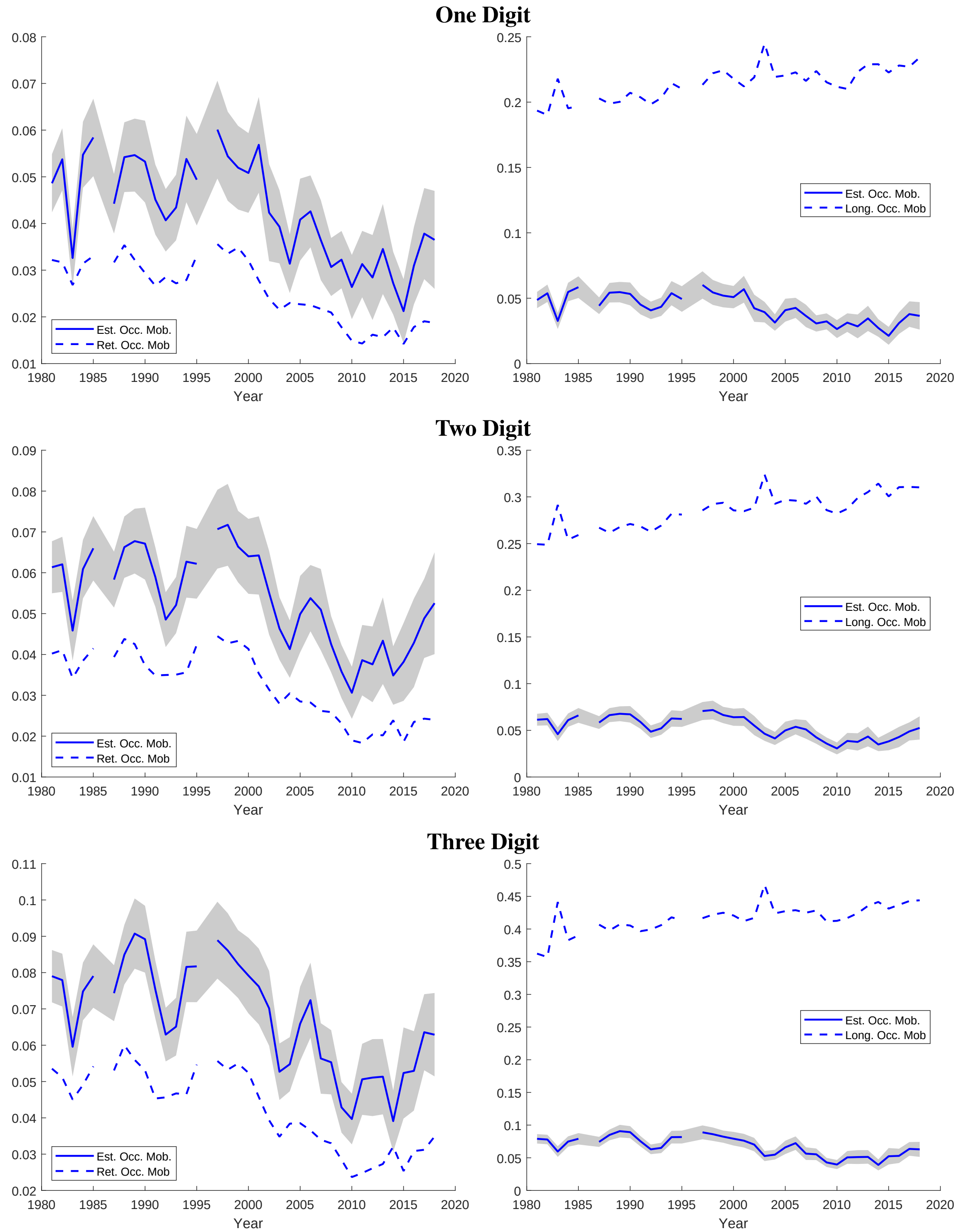
The above figures plot the fraction of employed individuals in each year who reported working in a different occupation the prior year. Left panels: Retrospective measures, computed using the March supplement and the Job Tenure and Occupational Mobility supplement; Right panels: longitudinal measures constructed using linked responses in March CPS. The dashed line in the left panel represents the retrospective switching rate for the same sample as the longitudinal measures – able to link across consecutive Marches (meaning no movers) and privately employed without occupation imputation in either response. The stars in the left panel represent data from the Job Tenure and Occupational Mobility supplement in the years in which that data is available from IPUMS-CPS. Spikes in the longitudinal data in 1983 and 2003 are attributable to changes in the occupational coding systems between 1982-1983 and 2002-2003. Consecutive March surveys cannot be linked between 1985-1986 and 1995-1996, and hence data for longitudinal mobility rates are missing for these years.

Figure 2: Decomposition of Occupational Switching Responses in the CPS, Retrospective and Longitudinal



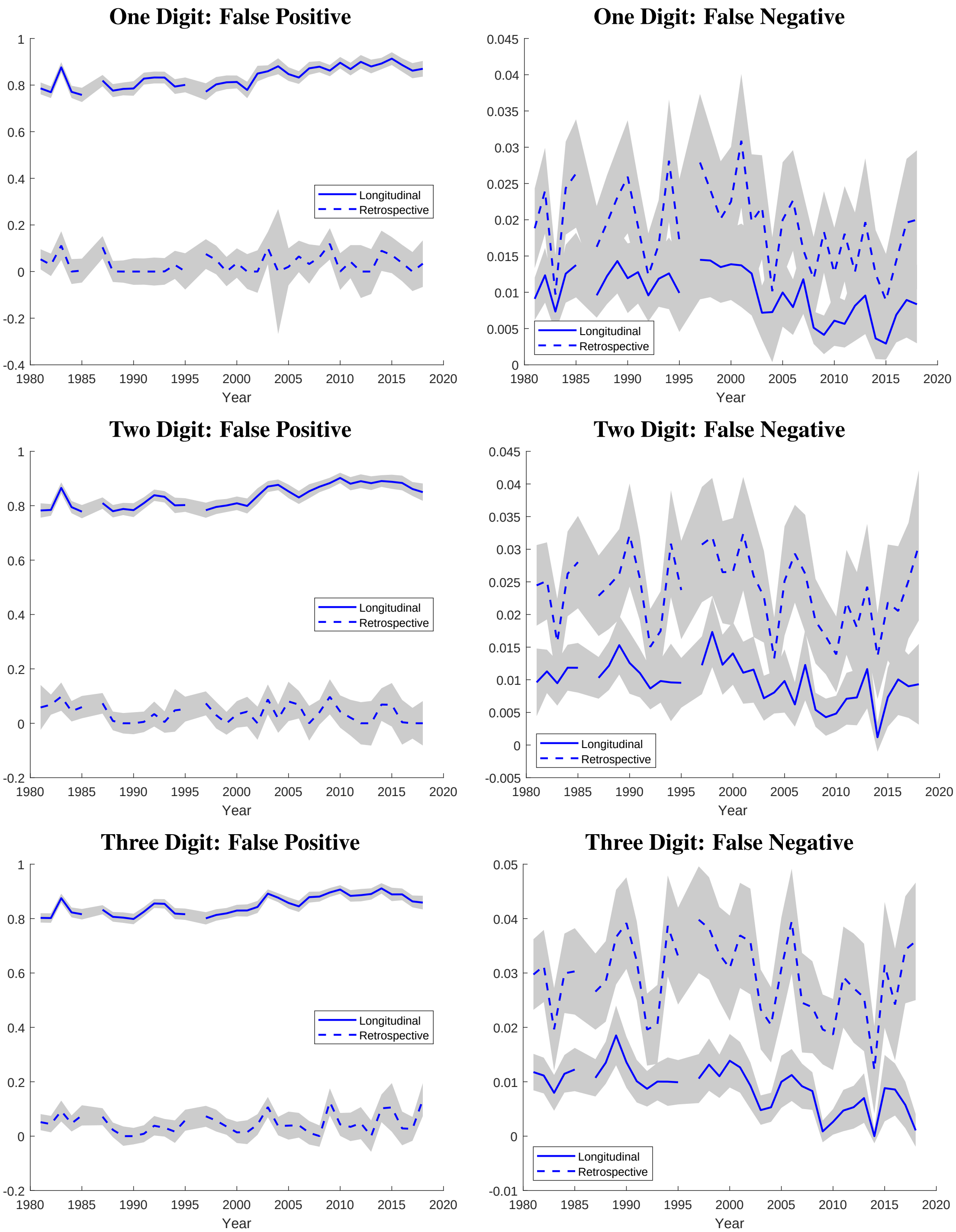
The above figures decompose reported occupational switching into the different cases where measures of occupational switching agree or disagree. The solid line represents the fraction of all responses where both occupational measures report a switch, the dashed line reports the case where only the retrospective measures report a switch, and the dotted line reports the case where only the longitudinal measure reports a switch. See notes to Figure (1) for more details.

Figure 3: Estimated Actual Occupational Mobility, Year by Year, 1981-2018



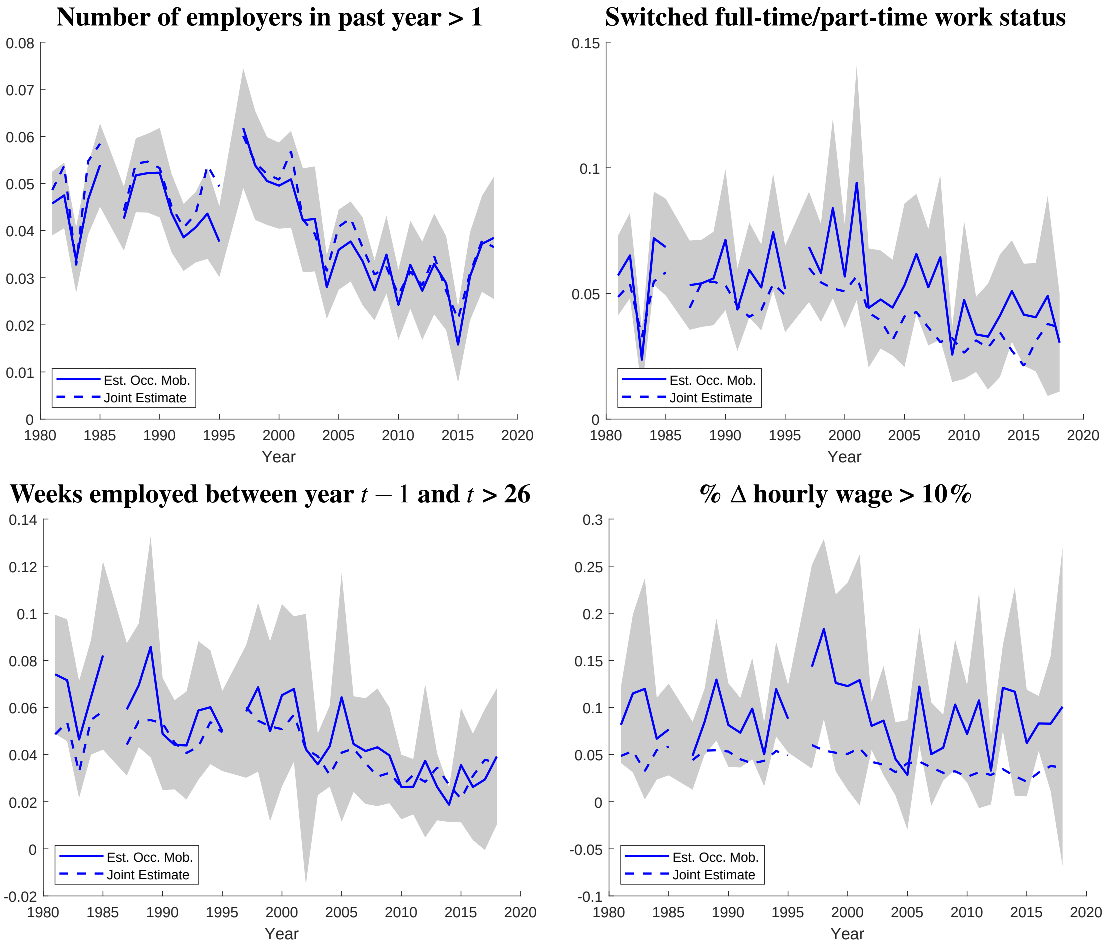
The above figures plot the estimates of actual annual occupational mobility, the fraction of employed individuals in each year who reported working in a different occupation the prior year. Left panels: Estimates compared with retrospective measures with longitudinal sample restrictions; Right panels: estimates compared with longitudinal measures. Dashed lines represent measured switching, solid lines represent estimated switching. The shaded bars represent 95% confidence intervals each year for the estimated occupational mobility rate. Spikes in the longitudinal data in 1983 and 2003 are attributable to changes in the occupational coding systems between 1982-1983 and 2002-2003. Consecutive March surveys cannot be linked between 1985-1986 and 1995-1996, and hence data for longitudinal mobility rates is missing for these years. See Figure 1 for more details regarding the data.

Figure 4: Estimated False Positive (left panels) and False Negative (right panels) Rates in Retrospective and Longitudinal Measures of Occupational Mobility, 1981-2018



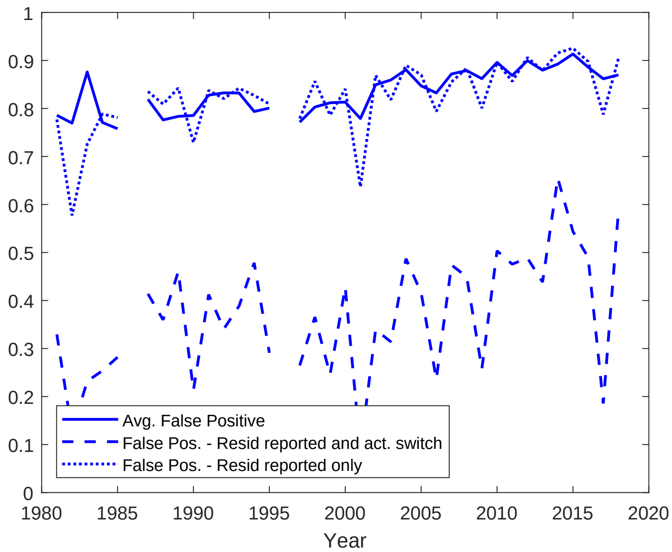
The above figures plot estimates of false positive and false negative rates over time in both retrospective and longitudinal measures of occupational mobility. Left panels: False positive rates; right panels: false negative rates. Dashed lines represent retrospective measures, solid lines represent longitudinal measures. False positive and false negative rates are defined in the text in Section 4.1. The shaded bars represent 95% confidence intervals in each year for the estimated error rates.

Figure 5: Estimated Actual Occupational Mobility, Separate Labor Market Outcomes



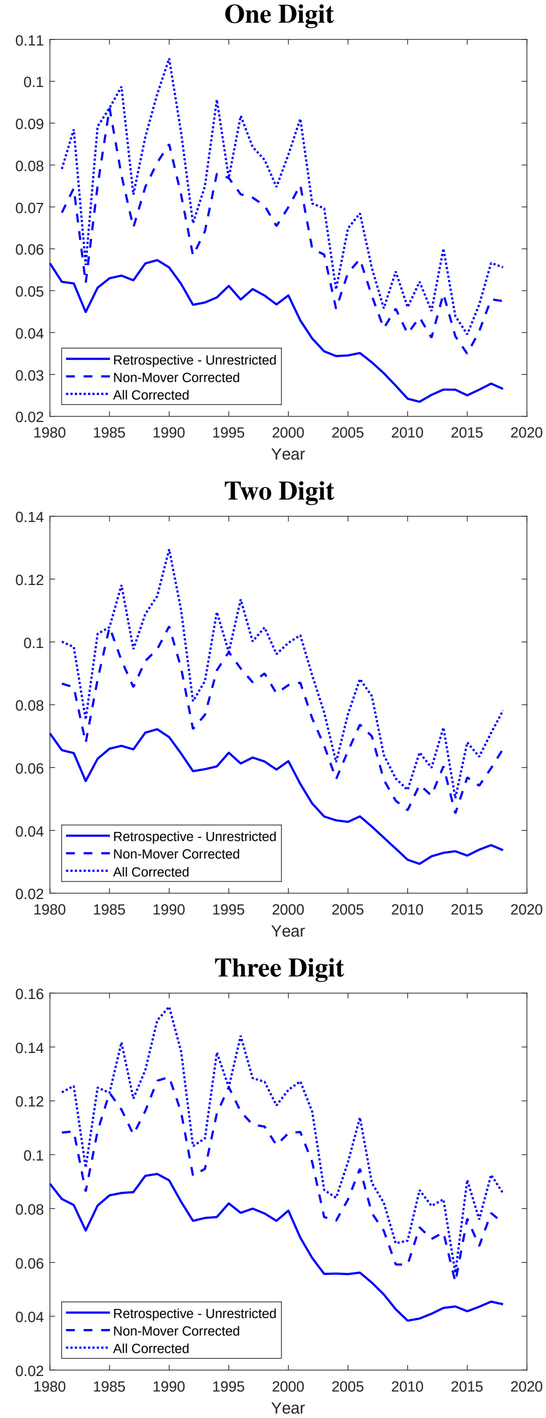
The above figures plot estimates of actual occupational mobility using each labor market outcome variable separately. Solid lines denotes the estimates from the individual labor market outcome case; dashed lines plot the jointly estimated rate for comparison. The shaded bars represent 95% confidence intervals in each year for annual occupational mobility rate estimates using just one labor market outcome variable; error bars for the joint estimates are not plotted.

Figure 6: Longitudinal Occupational Mobility False Positive Rate over Time, Average and Residual Measure



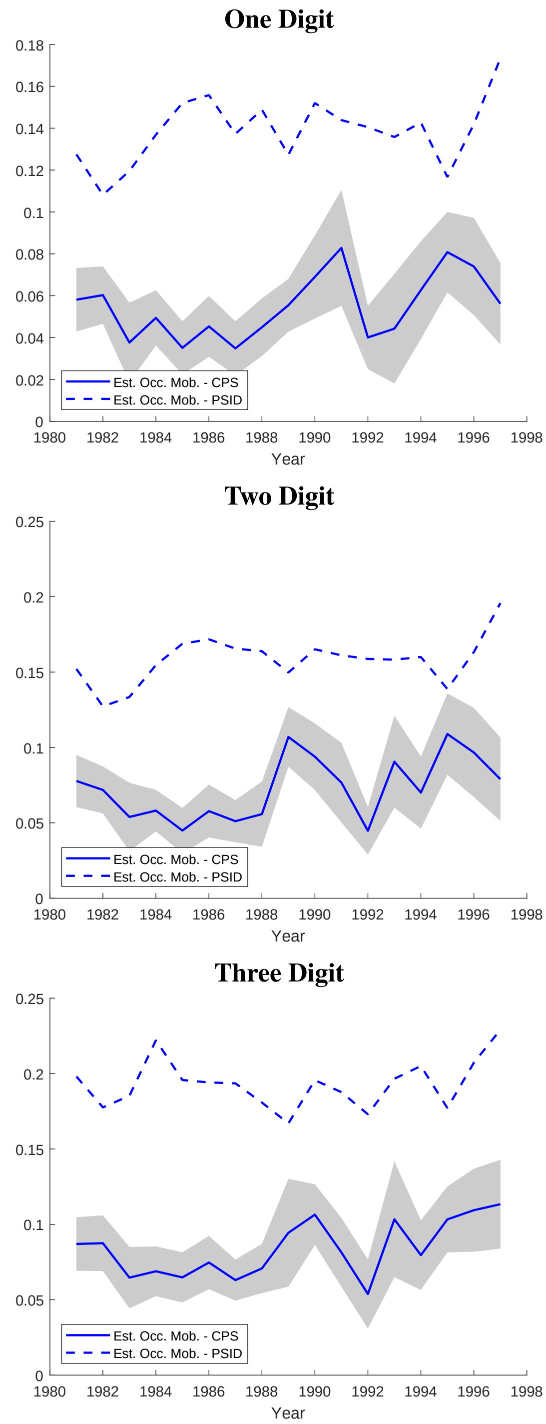
The above figures plot estimates of the false positive rate in longitudinal occupational mobility, both the average value and the value after residualizing out average observable characteristics. The solid line represents the estimated average false positive rate. The dashed line represents the false positive rate where both occupational mobility and longitudinal reports of occupational mobility have been residualized prior to constructing the false positive rate; the dotted line shows the results when only residualizing the longitudinal measure. Residualized occupational mobility is given by $\hat{\delta}_0 = \delta_0 - \bar{X}\delta_1$ and the residualized constant term in longitudinal reports of occupational mobility is given by $\hat{\alpha}_{L,0} = \alpha_{L,0} - \bar{X}\alpha_{L,X}$. These measures are then used in the false positive rate formula in place of δ_0 and $\alpha_{L,0}$, which formula is given by $\frac{P(\tilde{SW}^L=1 \cap SW=0)}{P(\tilde{SW}^L=1)} = \frac{(1-\delta_0)\alpha_{L,0}}{(1-\delta_0)\alpha_{L,0}+\delta_0(\alpha_{L,0}+\alpha_{L,1})}$. All estimates are obtained from the pooled sample from 1981-2018.

Figure 7: Aggregate Estimates of Actual Occupational Mobility, Year by Year, 1981-2018



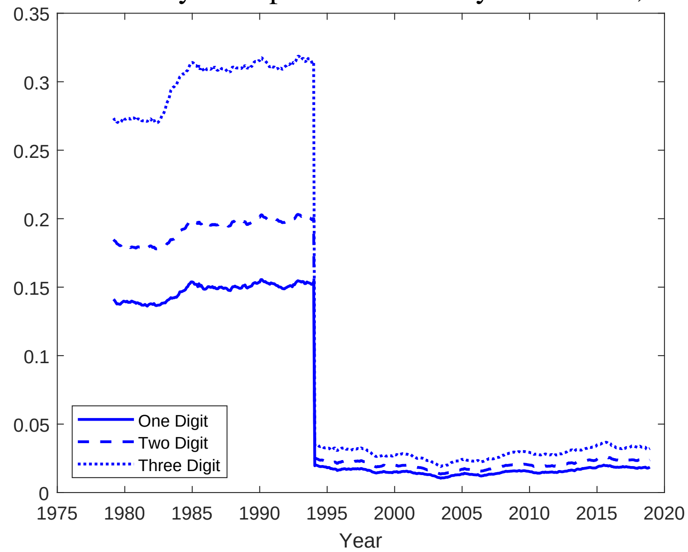
The above figures plot estimates of aggregate annual occupational mobility, adjusting occupational switching patterns for individuals who can't be matched across consecutive years. Dashed lines represent corrections to occupational switching applied only to switching rates of individuals who did not report a geographic move in the past year; dotted lines represent corrections to occupational mobility for all workers. Corrections are applied by multiplying the underlying switching rates by the estimated ratio $\frac{P(SW=1)}{P(\tilde{SW}^R=1)}$. The solid line is the uncorrected, unadjusted retrospective occupational mobility rate, the same as the solid lines in the left panels of Figure 1. Shaded error bars are not presented because of the close overlap between the two corrected time series. In 1985 and 1995, the two time series are equal to each other because there is not data available on geographic movers. In the years when it is not possible to estimate the correction term (1986 and 1996), we use the five year moving average of the correction terms instead.

Figure 8: Comparison of Estimated Annual Occupational Mobility in the CPS and the PSID, Year by Year, 1981-1997



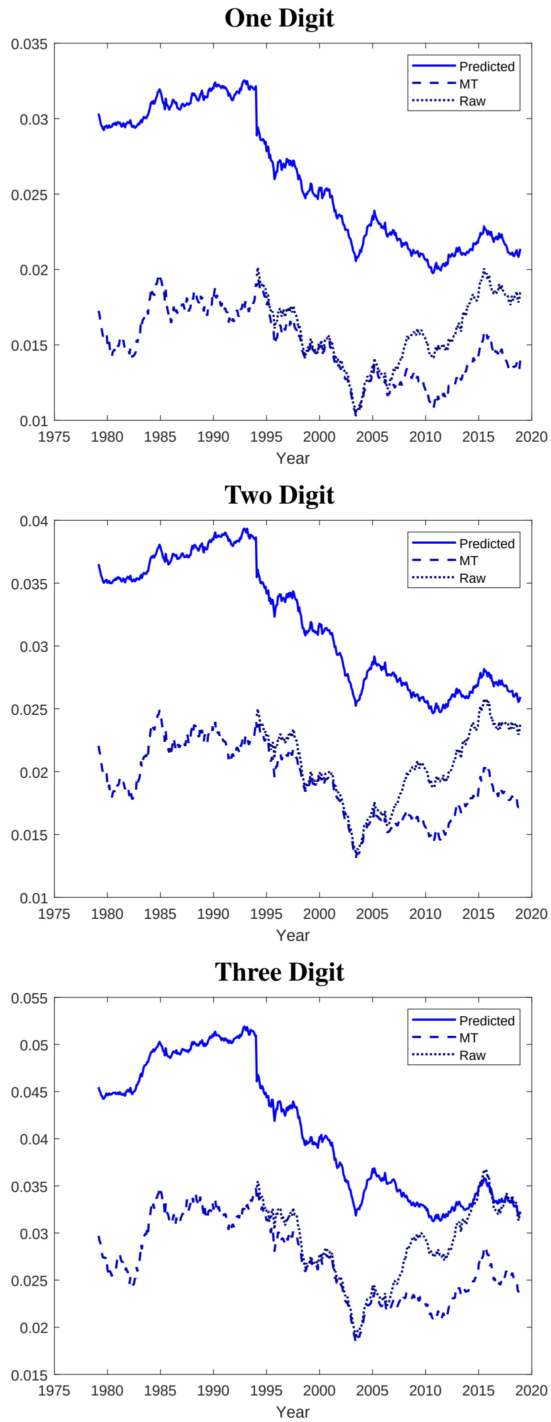
The above figures plot estimates of aggregate annual occupational mobility from the CPS (including movers) against annual estimates of occupational mobility from Kambourov and Manovskii (2008). Both samples are for male heads of household who are not self-employed and between the ages of 23 and 61. adjusting occupational switching patterns for individuals who can't be matched across consecutive years. Dashed lines are the estimates of occupational mobility from, based on a probit model, presented in Figure 3 of Kambourov and Manovskii (2008). Shaded error bars represent 95% confidence intervals for the scaling correction term multiplied by the actual estimated corrected level of occupational mobility. In the years when it is not possible to estimate the correction term (1986 and 1996), we use the five year moving average of the correction terms instead.

Figure 9: Raw Monthly Occupational Mobility in the CPS, 1979-2018



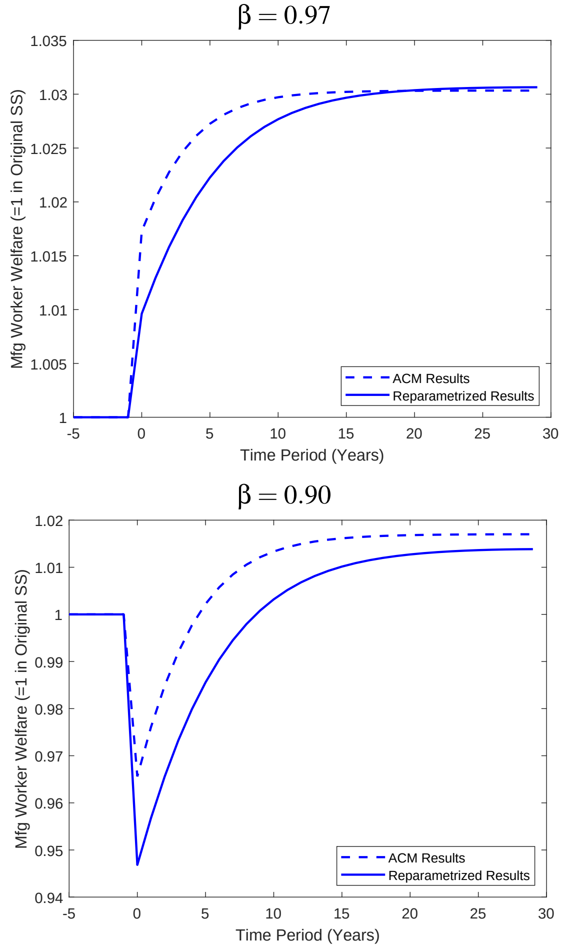
The above figure plots the fraction of employed individuals in each year who reported working in a different occupation the prior year. Data is from the monthly CPS constructed using linked responses for consecutive months. Sample is restricted to individuals 18+ who are in their second or third month in the survey, and who are privately employed and for whom occupation is not imputed or missing. For readability, data prior to 1994 and after 1994 are a 12 month moving average.

Figure 10: Comparison of Corrections to Monthly Occupational Mobility to Raw Data and Moscarini and Thomsson (2007), 1979-2018



The above figure plots the fraction of employed individuals in each year who reported working in a different occupation the prior year. Data is from the monthly CPS constructed using linked responses for consecutive months. Sample is restricted to individuals 18+ who are in their second or third month in the survey, and who are privately employed and for whom occupation is not imputed or missing. For readability, data prior to 1994 and after 1994 are a 12 month moving average. MT time series for monthly occupational mobility is obtained applying the filters of Moscarini and Thomsson (2007) to the monthly time series. Predicted monthly occupational mobility is obtained by applying our estimates for occupational mobility and measurement error to monthly data; for details, see Appendix D.

Figure 11: Simulations of Worker Welfare in the Model of Artuç et al. (2010) in Response to Trade Liberalization



The above figure plots welfare simulations from the model of Artuç et al. (2010) where workers face moving costs to switching industries and the economy is exposed to a trade liberalization in manufacturing where a 30% tariff is suddenly removed. These plots correspond to the left column of Figure 5 in their paper, with the dashed lines being an exact replica of their findings in that figure. The upper figure corresponds to their model with a discount factor of $\beta = 0.97$, the lower panel corresponds to their model with a discount factor of $\beta = 0.90$; the value of the discount factor is particularly important in these simulations as it strongly influences the importance of the option value to manufacturing workers of switching to other sectors in their overall utility. In the above figures, the surprise liberalization occurs at $t = 0$; each period corresponds to a full year. Welfare is normalized to 1 in the initial steady state to allow for direct comparisons between the two simulations, which have different initial and terminal steady states. For additional simulation details, as well as the details of the model, see Artuç et al. (2010). The values for their model parameters we estimate using our flow corrections are, in the $\beta = 0.97$ case, $C = 8.99$ and $\nu = 2.21$; in the $\beta = 0.90$ case, $C = 5.88$ and $\nu = 1.36$.

Appendix A: Additional Data Details and Results

A.1 Description of Sample Construction

We obtain Current Population Survey data from IPUMS-CPS (Flood et al. (2018)) for the period 1980-2018, focusing on data for the March supplement to the CPS (Annual Social and Economic Supplement, or ASEC). In addition to the raw data from IPUMS, we also merge on data from Unicon samples of the CPS (which are no longer publicly available, having been merged with IPUMS) and the NBER extracts of the CPS. The reason we do this is because the raw IPUMS data does not include an indicator variable for whether or not an individual's responses to all questions in the CPS were imputed (whole case imputation). Unicon data uniquely has this information for the years 1988-1990 (in the variable "suprec") and this data is available in the raw NBER CPS samples (in the variable FL-665) beginning in the year 1991. The code and linking procedures to obtain this variable are made available with our code and data posted online.

We focus on the privately employed adult civilian population in the U.S., dropping individuals under the age of 18 and individuals working in government industries, including armed forces.[35] We also drop all observations with imputed values for occupation. This includes the observations where all survey questions have been imputed, which can only be identified using the Unicon and NBER data.[36]

To account for changes in the occupational coding system over time, we use the time consistent occupational codes of Dorn (2009) and Autor and Dorn (2013). We further adjust and extend these codes using Census crosswalks to account for coding adjustments subsequent to the year 2010 and to account for some occupation codes which are only used in certain years (for example, the code for CEOs only begins being used in the year 2003).

These sample restrictions are used to compute what we refer to as the "headline" measure of retrospective occupational mobility. For all reported results, we use individual-level supplement weights provided in the CPS.

To construct longitudinally measured occupational mobility, we link observation across two consecutive March supplements. For the period 1989-2018, this is accomplished using individual identifiers developed in Rivera Drew et al. (2014); for 1980-1989, this is accomplished using the algorithm set out in Madrian and Lefgren (2000). To guard against errors generated by spurious links, we follow the methodology of Madrian and Lefgren (2000) and drop all linked responses where sex or race disagree or where increases in education and age are greater than a year. As

[35]We follow Kambourov and Manovskii (2008) in excluding government workers, however we find that adding them in to our sample does not substantially alter our findings.

[36]The coding of imputations in the CPS changed substantially in the year 1988, the first year when these whole-case imputation identifiers are available. Prior to the year 1988, the imputation flags for occupation suffice for identifying imputed observations.

pointed out in both Madrian and Lefgren (2000) and Rivera Drew et al. (2014), because of changes in household identifiers, we are unable to link CPS records between 1985-1986 and 1995-1996.

Consistent with how we are measuring occupational mobility, for the longitudinal sample, we require that an individual is employed in both of the linked years and has an non-imputed occupation response. Given this restriction and limits to the observations hat can be matched across years, this generates a significantly smaller subset of the sample than the one used to compute the headline measure of retrospective mobility. Thus, when we measure "restricted" retrospective mobility, we obtain a lower rate of occupational mobility. This is largely due to the omission of individuals who have moved their residence in the past year.

Finally, when we construct our sample used for estimation of actual occupational mobility and measurement error, we make a few additional adjustments. We set age equal to 80 for all individuals 80 years old or older, to make consistent the data with top codes imposed on age in some years. We also make several adjustments to the labor market outcomes used in the estimation. We convert the part-time/full-time status indicator into a binary variable – full-time or part-time. We construct hourly wages as the past year's income divided by the product of the number of weeks worked in the past year and the usual weekly hours worked. When hourly pay measured this way is zero, we replace it with 0.00001. Since we measure whether or not hourly pay has changed by 10% or more, this simply ensures that these observations are not dropped, but are coded as having experienced a large wage change. Finally, we also drop all observations for which there has been imputation or where there is missing data for any of the variables used in the estimation (including the inputs to constructing hourly wages).

A.2 List of One and Two Digit Occupational Codes

To construct what we refer to as one and two digit occupational codes from the harmonized occupation codes used in Autor and Dorn (2013), we rely on occupational groupings defined by the Census and reported in Dorn (2009). The listing of these occupation codes is provided in Table A.1, as well as the detailed three digit 1990 occupation codes that they correspond to.

Table A.1: List of One and Two Digit Occupations

One Digit Occupation	Two Digit Occupation	Three Digit Codes
Managerial and Professional Specialty Occ.	Executive, Administrative and Managerial Occ.	4-22
	Management Related Occ.	23-37
	Professional Specialty Occ.	43-199
Technical, Sales and Admin. Support Occ.	Technicians and Related Support Occ.	203-235
	Sales Occ.	243-283
	Administrative Support Occ.	303-389
Service Occ.	Housekeeping and Cleaning Occ.	405-408
	Protective Service Occ.	415-427
	Other Service Occ.	433-472
Farming, Forestry and Fishing Occ.	Farm Operators and Managers	473-475
	Other Agricultural and Related Occ.	479-498
Precision Production, Craft and Repair Occ.	Mechanics and Repairers	503-549
	Construction Trades	558-599
	Extractive Occ.	614-617
	Precision Production Occ.	628-699
Operators, Fabricators and Laborers	Machine Operators, Assemblers and Inspectors	703-799
	Transportation and Material Moving Occ.	803-889

Alternatively, we could take literally the notion of digit and define one and two digit occupations using the leading digits in each occupation code. We have experimented with this approach and found that it while it does impact the level of occupational mobility for one and two digit codes, it does not change the general punchlines of our analysis. We prefer the definitions for occupation codes from Table A.1 based on occupational groupings as these have more informational content than simply using the leading digits, as the coding system implemented in the 1990 Census was not designed to take the digit structure so literally. We also note that the Standard Occupational Classification System (SOC) does produce occupation codes where digits literally correspond to levels of aggregation, and the 1998 SOC codes are very similar to the one and two digit groupings we employ.

A.3 Sensitivity of Occupational Switching Rates to Imputation

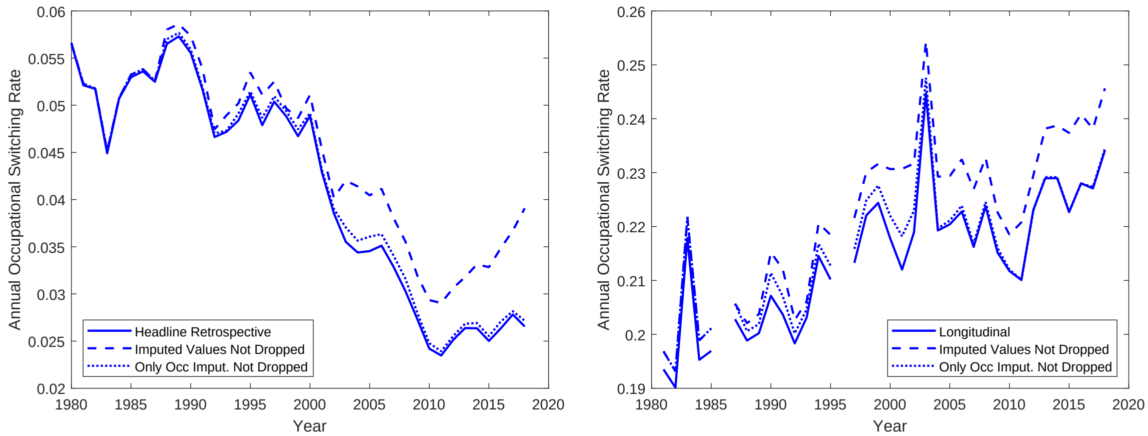
Throughout the paper, we drop all observations where the value for a worker's occupation is imputed. This imputation can occur through two ways – an imputation of a worker's response to the question(s) about occupations or through imputing all worker responses (whole-case imputation). Here we illustrate the impact of accounting for imputation on occupational switching rates.

Figure A.1 plots our baseline observed occupational switching rates, where all imputed responses are dropped, opposite occupational switching rates calculated where either all imputed observations are left in and switching rates where only observations with imputations for just occupation are kept in. The left panels plot "headline" or unrestricted retrospective occupational switching rates at the one, two and three digit levels; the right panels plot longitudinal occupational switching rates.

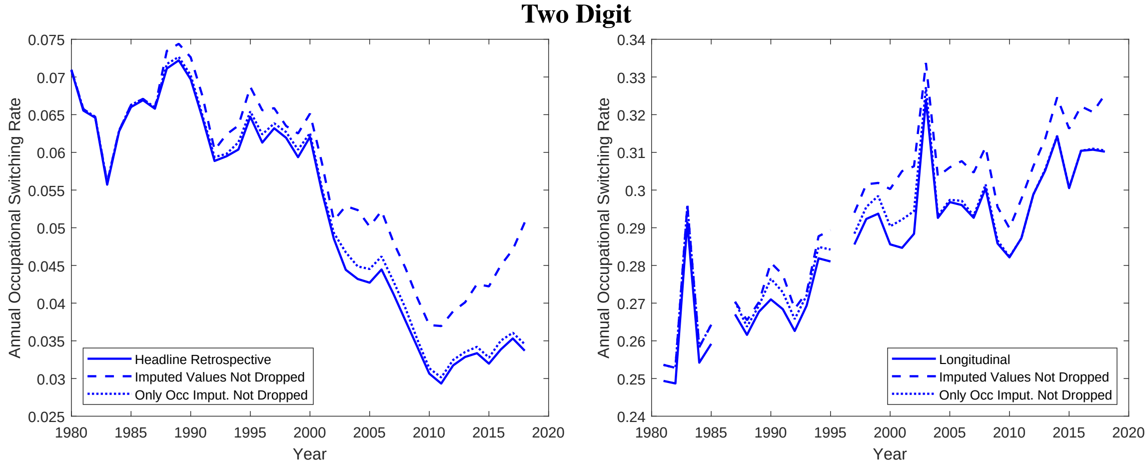
We observe that failing to account for imputations can have a significant impact on occupational switching rates. Perhaps unsurprisingly, this impact is largest at the three digit level, where occupational switching is most sensitive to small differences in occupations. We also observe that when we only drop whole-case imputations, leaving in imputation for just occupational questions, we get occupational switching rates that look very similar to our baseline measures. Thus, accounting for these whole-case imputations is critical to getting right the occupational switching rates. We note that this is particularly true when examining retrospective occupational switching, as, for example, when failing to account for these whole-case imputations, occupational mobility appears to be rising substantially since the year 2010.

Figure A.1: Sensitivity of Occupational Switching Rates to Keeping Imputed Values

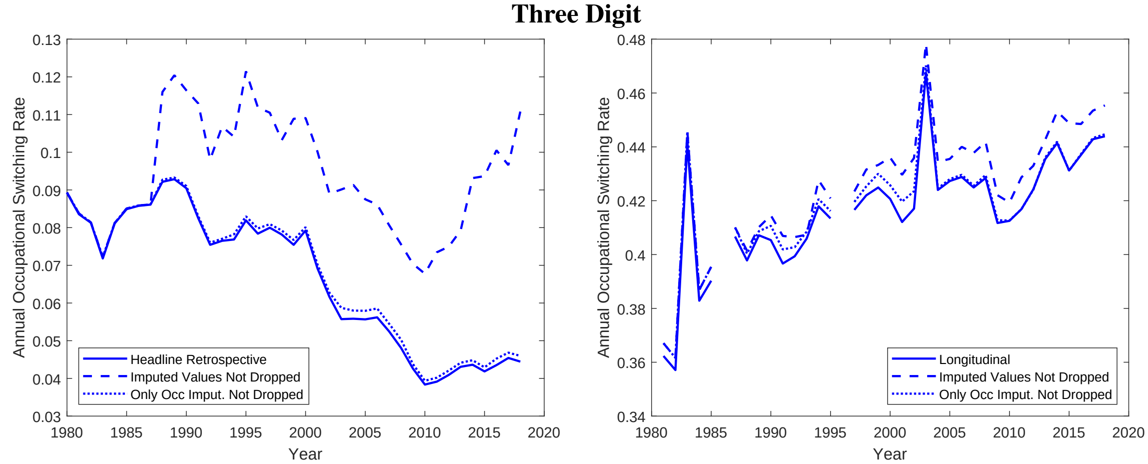
One Digit



Two Digit



Three Digit



The above figures plot the fraction of employed individuals in each year who reported working in a different occupation the prior year. Left panels: Retrospective measures, computed using the March supplement; Right panels: longitudinal measures constructed using linked responses in March CPS. Greater details on the construction of these rates and the sample used to do so can be found in the notes to Table 1. The dashed lines indicate the occupational switching rates when we leave in all imputed values. The dotted lines indicate the occupational switching rate where we leave in only the imputations for occupation questions, but omit the imputations where responses to all questions have been imputed.

Appendix B: GMM Moments for Estimation

In our baseline results, we can apply the assumptions of our framework and write the moments used in estimation as the following:

$$
\begin{aligned}
\mathbb{E}\left[\tilde{Z}_t\right] &= T_t[(1-\delta_{0,t});\delta_{0,t}] + [1;1;-1;-1]\alpha'_{R,X,t}\Sigma_{XX,t}\alpha_{L,X,t} \\
&\quad + [1;1;-1;-1]\delta'_{1,t}\Sigma_{XX,t}\left(\alpha_{L,X,t}\alpha_{R,1,t} + \alpha_{R,X,t}\alpha_{L,1,t}\right) \\
\mathbb{E}\left[\tilde{Z}_t Y_t\right] &= T_t[(1-\delta_{0,t})\beta_{0,t};\delta_{0,t}\left(\beta_{0,t}+\beta_{1,t}\right)] + T_t[-1;1]\delta'_{1,t}\Sigma_{XX,t}\beta_{2,t} \\
&\quad + [1;1;-1;-1]\left(\alpha_{R,1,t}\alpha'_{L,X,t} + \alpha_{L,1,t}\alpha'_{R,X,t}\right)\mathbb{E}\left[X_t'X_t\delta_{1,t}X_t\right]\beta_{2,t} \\
&\quad + [1;1;-1;-1]\alpha'_{R,X,t}\mathbb{E}\left[X_t'X_t\alpha_{L,X,t}X_t\right][\beta_{2,t} + \delta_{1,t}\beta_{1,t}] \\
&\quad + \left(T_{L,t}\left[-\beta_{0,t};\beta_{0,t}+\beta_{1,t}\right]\alpha'_{R,X,t} + T_{R,t}\left[-\beta_{0,t};\beta_{0,t}+\beta_{1,t}\right]\alpha'_{L,X,t}\right)\Sigma_{XX,t}\delta_{1,t} \\
&\quad + \left(T_{L,t}\left[(1-\delta_{0,t});\delta_{0,t}\right]\alpha'_{R,X,t} + T_{R,t}\left[(1-\delta_{0,t});\delta_{0,t}\right]\alpha'_{L,X,t}\right)\Sigma_{XX,t}\beta_{2,t} \\
&\quad + [1;1;-1;-1]\left[(1-\delta_{0,t})\beta_{0,t} + \delta_{0,t}\left(\beta_{0,t}+\beta_{1,t}\right)\right]\alpha'_{R,X,t}\Sigma_{XX,t}\alpha_{L,X,t} \\
\mathbb{E}\left[\tilde{Z}_t X_t\right] &= T_t[-1;1]\delta'_{1,t}\Sigma_{XX,t} + [1;1;-1;-1]\left(\alpha_{R,1,t}\alpha'_{L,X,t} + \alpha_{L,1,t}\alpha'_{R,X,t}\right)\mathbb{E}\left[X_t'X_t\delta_{1,t}X_t\right] \\
&\quad + [1;1;-1;-1]\alpha'_{R,X,t}\mathbb{E}\left[X_t'X_t\alpha_{L,X,t}X_t\right] \\
&\quad + \left(T_{L,t}\left[(1-\delta_{0,t});\delta_{0,t}\right]\alpha'_{R,X,t} + T_{R,t}\left[(1-\delta_{0,t});\delta_{0,t}\right]\alpha'_{L,X,t}\right)\Sigma_{XX,t} \\
\mathbb{E}\left[X_t'Y_t\right] &= \Sigma_{XX,t}\left([-\delta_{1,t} \quad \delta_{1,t}][\beta_{0,t};\beta_{0,t}+\beta_{1,t}] + \beta_{2,t}\right)
\end{aligned}
$$

where $\Sigma_{XX,t}$ is the variance-covariance matrix of X in year t and T_t is given by:

$$
\begin{aligned}
T_t &= \begin{bmatrix}
P(\tilde{Z}_{1,t}=1 \mid SW_t=0) & P(\tilde{Z}_{1,t}=1 \mid SW_t=1) \\
P(\tilde{Z}_{2,t}=1 \mid SW_t=0) & P(\tilde{Z}_{2,t}=1 \mid SW_t=1) \\
P(\tilde{Z}_{3,t}=1 \mid SW_t=0) & P(\tilde{Z}_{3,t}=1 \mid SW_t=1) \\
P(\tilde{Z}_{4,t}=1 \mid SW_t=0) & P(\tilde{Z}_{4,t}=1 \mid SW_t=1)
\end{bmatrix} \\
&= \begin{bmatrix}
(\alpha_{R,0,t})(\alpha_{L,0,t}) & (\alpha_{R,0,t}+\alpha_{R,1,t})(\alpha_{L,0,t}+\alpha_{L,1,t}) \\
(1-\alpha_{R,0,t})(1-\alpha_{L,0,t}) & (1-\alpha_{R,0,t}-\alpha_{R,1,t})(1-\alpha_{L,0,t}-\alpha_{L,1,t}) \\
(\alpha_{R,0,t})(1-\alpha_{L,0,t}) & (\alpha_{R,0,t}+\alpha_{R,1,t})(1-\alpha_{L,0,t}-\alpha_{L,1,t}) \\
(1-\alpha_{R,0,t})(\alpha_{L,0,t}) & (1-\alpha_{R,0,t}-\alpha_{R,1,t})(\alpha_{L,0,t}+\alpha_{L,1,t})
\end{bmatrix}
\end{aligned}
$$

and where $T_{L,t}$ and $T_{R,t}$ are given by:

$$T_{L,t} = \begin{bmatrix} (\alpha_{L,0,t}) & (\alpha_{L,0,t} + \alpha_{L,1,t}) \\ -(1-\alpha_{L,0,t}) & -(1-\alpha_{L,0,t}-\alpha_{L,1,t}) \\ (1-\alpha_{L,0,t}) & (1-\alpha_{L,0,t}-\alpha_{L,1,t}) \\ -(\alpha_{L,0,t}) & -(\alpha_{L,0,t}+\alpha_{L,1,t}) \end{bmatrix}$$

$$T_{R,t} = \begin{bmatrix} (\alpha_{R,0,t}) & (\alpha_{R,0,t} + \alpha_{R,1,t}) \\ -(1-\alpha_{R,0,t}) & -(1-\alpha_{R,0,t}-\alpha_{R,1,t}) \\ -(\alpha_{R,0,t}) & -(\alpha_{R,0,t}+\alpha_{R,1,t}) \\ (1-\alpha_{R,0,t}) & (1-\alpha_{R,0,t}-\alpha_{R,1,t}) \end{bmatrix}$$

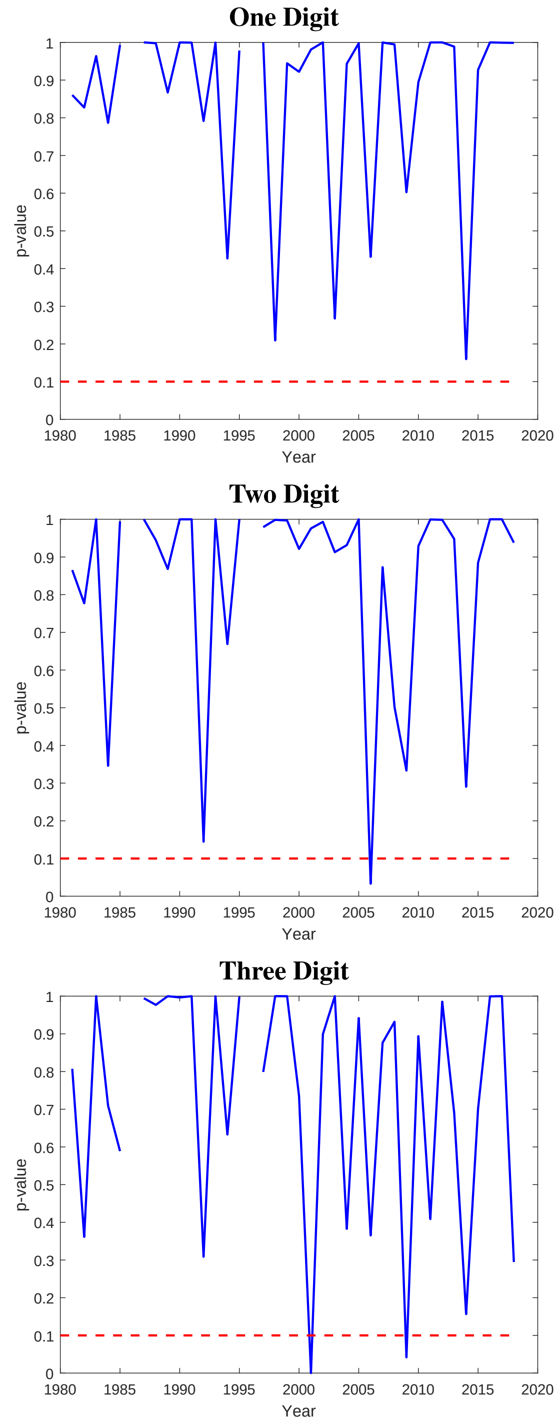
These moments are naturally similar to those reported in Kane et al. (1999), although they do not allow for correlation between the measurement error in the signals and individual characteristics, and thus have a significantly simpler set of moment conditions.

Appendix C: Additional Results and Robustness

C.1 Overidentification Tests

In this section, we report the results of overidentification tests in our model and discuss their interpretation. In Figure C.1, we report the p-value from the overidentification test for estimating the model year by year for one, two and three digit results.

Figure C.1: P-values from Overidentification Tests



The above figures report the p-values from the Sargan-Hansen test of overidentification, based off the χ^2 distribution. The red dashed line indicates the 10% threshold used for determining rejection of the null hypothesis.

As mentioned in the body of the paper, we fail to reject at the 10% level the null hypothesis that the model is overidentified in almost every case, the exceptions being the year 2006 for 2-digit switching and the years 2001 and 2009 for 3-digit switching. Given that overidentification in our estimation comes from the assumption that the measurement error in each signal of occupational switching is conditionally independent of the other signal and labor market outcomes, we view this evidence as supportive for these assumptions.

We do acknowledge, however, that the overidentification test does indeed reject the null hypothesis in our pooled estimates of occupational mobility. We are not concerned by this, however, for a few reasons. First, if we run our pooled year estimation for each labor market outcome separately, we fail to reject overidentification. In terms of our model, this suggests that the concern is about whether or not the labor market outcomes are conditionally independent of the measurement error and not whether the two types of error are independent of each other. Second, we note that running our estimation in the pooled case allows for much less flexibility than the year by year case, as we do not allow the relationships between measurement error, switching, labor market outcomes and individual characteristics to vary over time. This restriction is potentially important as changes in business cycle conditions and long run secular trends could cause at least the relationships between individual characteristics and labor market outcomes to change over time, if not also the relationships with the measurement error. Alternatively, a natural conjecture is that the rejection of overidentification is occurring because we fail to account for some omitted variable which is correlated with the measurement error and occupational switching.

In the subsequent robustness checks in this Appendix, we find that the inclusion of additional individual characteristics in our estimation increases the p-value substantially in the pooled estimation case. When we include this substantial set of additional covariates, we again fail to reject overidentification. Importantly, accounting for these additional covariates has a minimal impact on our estimates, as we show later in this Appendix. Though unreported, what we do find is that the noisier labor market outcomes become less noisy with the addition of these covariates and the point estimates on occupational switching converge towards the joint estimate obtained in our baseline. Thus, we do not find it concerning that we fail to reject the null of overidentification in our baseline pooled estimates.

As a note, a key reason that our estimates are so insensitive to the variation in the controls included in our estimation is because the weighting matrix obtained by two-stage GMM estimation will place greater weight on the least noisy moments. In our case, those moments are the ones involving the labor market outcome of whether or not a worker had more than one employer in the past year. Thus, although some of the other moments are noisier, since the estimation places greater weight on these moments, this leads to fairly consistent estimation of the levels of occupational mobility and measurement error regardless of the individual characteristics included.

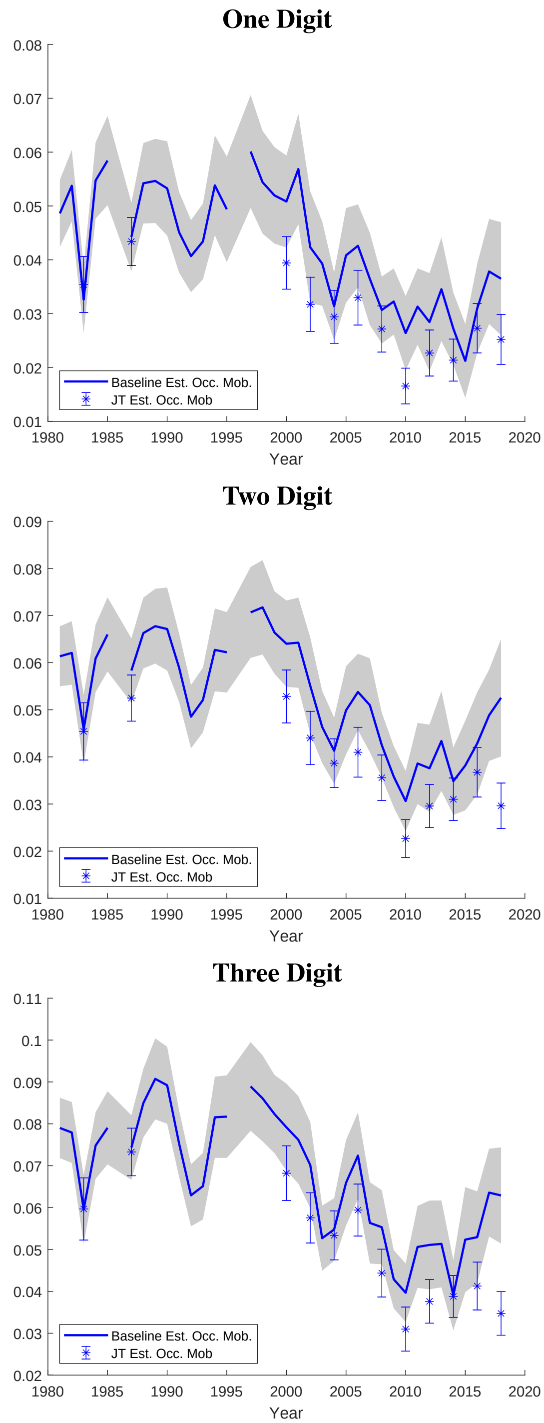
C.2 Using the Job Tenure and Occupational Mobility Supplement

As seen in Section 2, an alternative retrospective measure of occupational switching can be obtained from the job tenure supplement to the Current Population Survey. Here, we show that similar results are obtained when we estimate actual occupational mobility using linked job tenure supplement data instead of linked March CPS data.

The primary differences with using the job tenure supplement data have to do with the labor market outcomes. Two of the four labor market outcomes used in our baseline estimation are unobservable in the job tenure supplement, as we neither observe changes in wages nor the weeks worked in the past year. We still observe whether or not an individual changed part-time or full-time status, however, and we can construct a measure of the number of employers in the past year that is similar to the measure in the March supplement. The job tenure supplement asks for the current employer tenure with a worker's current employer, and we code workers who report being with their current employer for less than a year as having had more than one employer in the past year (since all workers must have been employed a year ago to appear in our sample). Thus we use these two observable labor market outcomes in our estimation. All individual characteristics are measured in the same way, and define our sample for estimation in the same way as in our baseline estimation.

Figure C.2 presents the year by year estimates of occupational mobility from our baseline estimation as well as from the job tenure supplement, both with 95% confidence error bands. Estimates of occupational mobility from the job tenure supplement track fairly closely the patterns observed in our baseline results. On average, the job tenure supplement generates estimates of occupational mobility that are slightly lower than those in our baseline. However, with the exception of results from the year 2018, in every case, the confidence intervals of the estimates overlap, and in quite a few cases, the point estimates are nearly identical. Given that these two sets of estimates are computed from different samples, different months, and slightly different labor market outcomes, we view these results as strongly supportive of our baseline findings using the March supplement to the CPS.

Figure C.2: Estimated Actual Occupational Mobility Rate, March Supplement vs. Job Tenure Supplement



The above figures plot the estimates of actual annual occupational mobility from the March CPS (originally shown in Figure 3) opposite estimates obtained using the job tenure supplement. The shaded bars and vertical error bars represent 95% confidence intervals each year for the estimated occupational mobility rate. See Figures 1 and 3 for more details regarding the data and the baseline estimated results.

C.3 Using Alternate Sets of Outcomes or Individual Characteristics

In this section, we report robustness to using alternate labor market characteristics in estimation and also using different individual characteristics as controls in the estimation. For brevity's sake, we report robustness checks for just the one digit level.

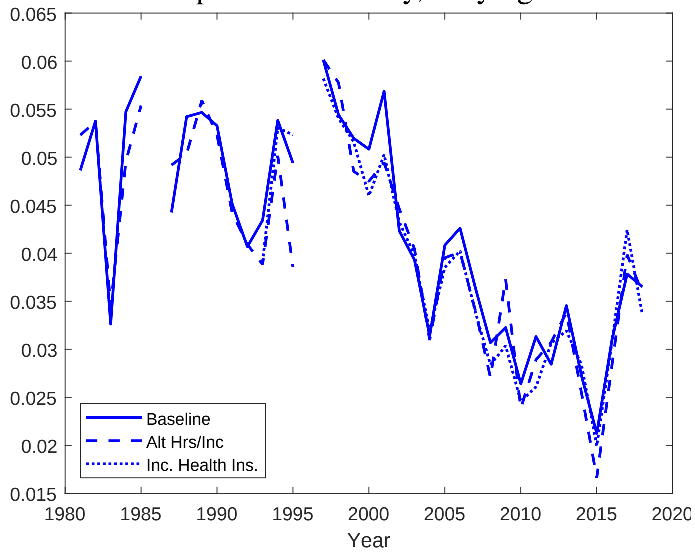
We first consider how our estimation would change if we varied the labor market characteristics used in estimation. The criteria we require for labor market characteristics – available for all years in the sample, unlikely to be mechanically correlated with measurement error, and having an a priori plausible relationship with occupational switching – limit the set of variables we can consider. However, we use both measures of hours and wages in our estimation, and we first report robustness to varying how we measure these. Instead of using whether or not a worker switched part-time/full-time status, we use the absolute value of the percent change (in logs) of actual hours worked between survey years, and instead of using whether or not hourly pay increases (in absolute value) by 10% or more, we consider the absolute value of the percent change (in logs) of total income. As a result, in our first robustness exercise, we use these two measures instead of our original measures, and estimate them alongside our original outcomes of the number of employers in the past year and whether or not the worker worked half of the weeks in the prior year.

We also consider a labor market outcome which satisfies all of our criteria except being available in every year – whether or not a worker reported a change in whether or not their employer contributed to health insurance. This data is only available from 1992 onward, so we can use the data on a switch beginning in 1993. With this labor market outcome, we consider a robustness exercise where we re-estimate the parameters of the model adding this labor market outcome to our original four labor market outcomes.

Figure C.3 reports the annual occupational switching rate at the 1 digit level we estimate in our baseline alongside the two robustness checks described above. We do not report standard error bands because of difficulty in visualizing these for all series simultaneously. However, as can be seen in the figure, in both of these robustness exercises, we find that the annual switching rate is very close to the original switching rate we estimate. Thus, we do conclude that our results are not sensitive to varying how we measure changes in hours worked or wages, or the inclusion of whether or not employer contributions to health insurance changed across the two years.[37] We also fail to reject overidentification at the 10% level in all but one year in the first exercise varying the measures of hours and income, and we fail to reject in all but two years where we add employer contributions to health insurance.

[37] Though unreported, if one estimates occupational switching simply using data on whether or not there was a change in whether or not an employer contributed to health insurance, we get very noisy estimates with very large confidence intervals, that always intersect with the confidence intervals of our baseline results.

Figure C.3: Estimated Occupational Mobility, Varying Labor Market Outcomes

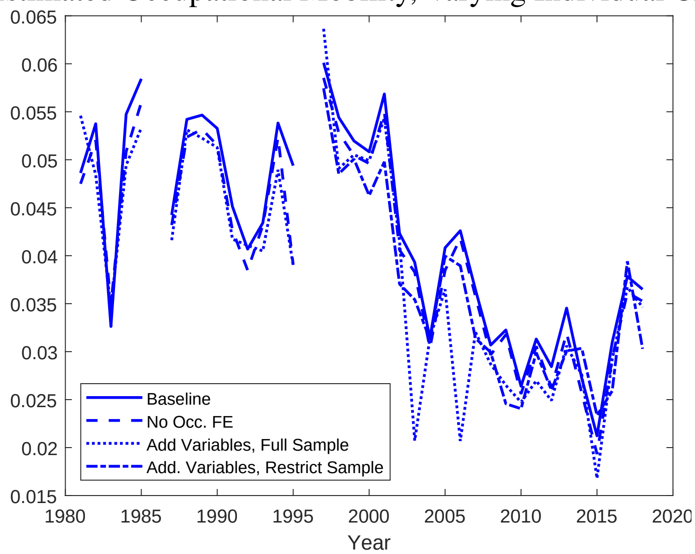


We also consider how our estimates change if we vary the set of individual characteristics used in estimation. We first consider what happens to our estimates if we omit the two digit occupation fixed effects from estimation, and just use the baseline set of worker characteristics. We then add additional controls that are available in all years – whether a worker worked full-time last year, the usual hours worked last year, veteran status, whether a worker was self employed last year, the log of reported income last year, and state fixed effects. Finally, there are some variables which cannot be observed in all years, and we consider exercises beginning in 1995 where all of these variables can be observed, as well as the other variables we have added thus far. These variables are whether or not an individual is presently enrolled in school, whether or not an individual was born outside the United States, whether or not a worker has a disability, whether or not the response for a worker was provided by a proxy (or whether the proxy reporting status changed across two years), and whether or not a worker simultaneously held two jobs.[38]

Figure C.4 reports the annual occupational switching rate at the 1 digit level for each of these checks. We see that in each case, reducing or increasing the set of individual characteristics in the ways described above does not have a significant impact on the level we estimate for occupational switching.

[38]This last variable is identified by workers who report a different level of actual total hours worked at all jobs from what they report for actual total hours worked at their primary job.

Figure C.4: Estimated Occupational Mobility, Varying Individual Characteristics



However, it is possible that our estimates relating individual characteristics and occupational switching, actual and reported, may change with the inclusion of these variables. We report in Tables C.1 and C.2 the estimated coefficients on occupational switching for our original worker characteristics in the pooled sample when we estimate beginning in 1995 to allow for all these worker characteristics to be included. These results can naturally be compared to those reported in Tables 3 and 4 in the paper, and we see that while the coefficient magnitudes change by modest amounts, the patterns we identified in the body of the paper are fairly similar. We continue to find a small negative correlation between the actual occupational switching rates and longitudinally reported occupational switching rates across individual occupations.

We also report in Table C.3 the relationships between these additional worker characteristics and occupational switching, with the exception of state fixed effects, which are usually small and insignificant. We see that non-full-time workers, not-self-employed workers, lower income workers, workers with multiple jobs and disabled workers are more likely to switch occupations. Similar to our baseline findings, we do not find strong relationships between any observables and whether or not a retrospective switch is reported. However, we do find that workers working fewer hours, who are self employed, who earn higher incomes, and who are in school are more likely to report an occupational switch, holding constant whether or not a switch actually occurred.

Table C.1: Estimated Relationships Between Individual Characteristics and Actual and Reported Occupational Mobility, 1995-2018, Controlling for Additional Characteristics

Individual Characteristic	Actual	Longitudinal		Retrospective	
		Control for Switch	Unconditional	Control for Switch	Unconditional
Male	0.009	0.038	0.044	0.001	0.006
	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Hispanic	-0.019	0.028	0.017	0.003	-0.008
	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)
Age/1000	-6.639	-1.225	-5.307	0.172	-3.638
	(0.311)	(0.324)	(0.296)	(0.101)	(0.131)
Age squared/1000	0.055	0.006	0.040	-0.002	0.030
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Nonwhite	-0.014	0.015	0.007	0.003	-0.005
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Married	-0.007	-0.012	-0.016	-0.001	-0.005
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Below HS degree	-0.015	0.020	0.011	0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Has only HS degree	-0.008	0.033	0.028	0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Completed some college	-0.002	0.055	0.054	0.005	0.004
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Avg rate | switch = 0	-	0.194	0.194	0.001	0.001
		(0.001)	(0.001)	(0.000)	(0.000)
Avg rate | switch = 1	0.038	0.799	0.799	0.562	0.562
	(0.001)	(0.007)	(0.007)	(0.012)	(0.012)
N	301,187	301,187	301,187	301,187	301,187

See notes to Tables 3 and 4.

Table C.2: Estimated Relationships Between Occupation and Actual and Reported Occupational Mobility, 1995-2018, Controlling for Additional Characteristics

Occupation in Year *t*	Actual	Longitudinal		Retrospective	
		Control for Switch	Unconditional	Control for Switch	Unconditional
Executive, Administrative and Managerial Occ.	-0.018	0.059	0.048	0.004	-0.007
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Management Related Occ.	-0.014	0.047	0.038	0.006	-0.002
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Professional Specialty Occ.	-0.028	-0.072	-0.089	0.004	-0.012
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Technical and Related Support Occ.	-0.027	0.072	0.055	0.003	-0.012
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Sales Occ.	-0.015	-0.003	-0.012	0.002	-0.007
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Administrative Support Occ.	-0.024	0.003	-0.011	0.002	-0.012
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Housekeeping and Cleaning Occ.	0.000	-0.078	-0.078	-0.006	-0.006
	(0.008)	(0.007)	(0.007)	(0.002)	(0.003)
Protective Service Occ.	0.018	-0.052	-0.041	-0.007	0.003
	(0.009)	(0.008)	(0.008)	(0.003)	(0.004)
Other Service Occ.	0.000	-0.015	-0.014	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Farm Operators and Managers	-0.004	-0.127	-0.130	-0.008	-0.011
	(0.006)	(0.006)	(0.005)	(0.002)	(0.002)
Other Agricultural and Related Occ.	-0.003	-0.001	-0.003	-0.002	-0.004
	(0.009)	(0.009)	(0.008)	(0.003)	(0.004)
Mechanics and Repairers	-0.024	0.023	0.009	0.002	-0.012
	(0.003)	(0.004)	(0.004)	(0.001)	(0.001)
Construction Trades	-0.015	-0.010	-0.019	0.002	-0.007
	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)
Extractive Occ.	-0.001	0.143	0.143	0.016	0.015
	(0.016)	(0.022)	(0.020)	(0.007)	(0.009)
Precision Production Occ.	-0.019	0.123	0.112	0.003	-0.008
	(0.003)	(0.005)	(0.004)	(0.001)	(0.002)
Machine Operators, Assemblers and Inspectors	-0.006	-0.001	-0.009	-0.001	-0.004
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)
Avg rate | switch = 0	-	0.194	0.194	0.001	0.001
		(0.001)	(0.001)	(0.000)	(0.000)
Avg rate | switch = 1	0.038	0.799	0.799	0.562	0.562
	(0.001)	(0.007)	(0.007)	(0.012)	(0.012)
N	301,187	301,187	301,187	301,187	301,187

See notes to Tables 3 and 4.

Table C.3: Estimated Relationships Between Additional Individual Characteristics and Actual and Reported Occupational Mobility, 1995-2018

Individual Characteristic	Actual	Longitudinal		Retrospective	
		Control for Switch	Unconditional	Control for Switch	Unconditional
Full-time Last year	-0.024	0.005	-0.010	0.001	-0.013
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Usual Hours Worked Last Year/100	-0.008	-0.0435	-0.039	-0.001	0.003
	(0.007)	(0.007)	(0.007)	(0.002)	(0.003)
Veteran	0.008	-0.009	-0.004	-0.000	0.005
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Self-employed Last Year	-0.016	0.045	0.035	0.003	-0.006
	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)
Log Income Last Year/100	-1.482	0.279	-0.632	0.101	-0.749
	(0.146)	(0.132)	(0.070)	(0.051)	(0.043)
In School	0.035	0.025	0.047	0.005	0.025
	(0.007)	(0.006)	(0.005)	(0.002)	(0.003)
Proxy Response (Either Year)	-0.001	0.007	0.006	0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)
Switch in Proxy Response	0.002	0.008	0.009	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Held Multiple Jobs Simultaneously	0.025	0.009	0.024	0.002	0.016
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Disabled	0.014	-0.003	0.005	-0.001	0.007
	(0.004)	(0.004)	(0.004)	(0.001)	(0.002)
Born Outside the United States	-0.005	0.006	0.003	-0.001	-0.003
	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)
Avg rate | switch = 0	-	0.194	0.194	0.001	0.001
		(0.001)	(0.001)	(0.000)	(0.000)
Avg rate | switch = 1	0.038	0.799	0.799	0.563	0.563
	(0.001)	(0.007)	(0.007)	(0.008)	(0.008)
N	301,187	301,187	301,187	301,187	301,187

See notes to Tables 3 and 4.

Appendix D: Predicted Occupational Switching in Monthly Data

In this Appendix, we describe how we construct predicted probabilities for occupational switching in monthly data.

The probability of actually experiencing an occupational switching given observable data is given by $P(SW = 1 \mid X, Y, \tilde{SW})$, where SW is a dichotomous variable for whether or not a switch actually occurred, X is the set of individual characteristics, Y is data on labor market outcomes, and $\tilde{SW}$ is some signal of switching. Given the assumptions of our model, it is straightforward to apply Bayes' rule and obtain the following expression for the probability of actually experiencing an occupational switch:

$$P(SW=1 \mid X,Y,\tilde{SW}) = \frac{P(\tilde{SW},Y \mid SW=1,X)P(SW=1,X)}{P(\tilde{SW},Y \mid SW=1,X)P(SW=1,X)+P(\tilde{SW},Y \mid SW=0,X)P(SW=0,X)}$$
$$= \frac{P(\tilde{SW} \mid X,Y,SW=1)P(Y \mid X,SW=1)P(SW=1 \mid X)}{P(\tilde{SW} \mid X,Y,SW=1)P(Y \mid X,SW=1)P(SW=1 \mid X)+P(\tilde{SW} \mid X,Y,SW=0)P(Y \mid X,SW=0)P(SW=0 \mid X)}$$

Because the errors in reported occupational switching are uncorrelated with labor market outcomes conditional on actually switching and because the labor market outcomes we study are themselves all dichotomous, it is straightforward to express this object as a function of model parameters. For example, if we observe an individual who reports switching ($\tilde{SW}=1$) and also reports changing part-time/full-time work status ($Y=1$), we would write the predicted probability in terms of model parameters as:

$$P(SW=1 \mid X,Y=1,\tilde{SW}=1)$$
$$= \frac{(\alpha_0+\alpha_1+X\alpha_X)(\beta_0+\beta_1+X\beta_2)(\delta_0+X\delta_1)}{(\alpha_0+\alpha_1+X\alpha_X)(\beta_0+\beta_1+X\beta_2)(\delta_0+X\delta_1)+(\alpha_0+X\alpha_X)(\beta_0+X\beta_2)(1-\delta_0-X\delta_1)}$$

We compute a similar expression for all possible realizations of worker outcomes, and thus assign a predicted probability for occupational switching to each observation in the monthly data. Because our underlying model models dichotomous variables linearly, it is possible to get predicted probabilities for switching greater than 1 and less than 0. While this occurs very infrequently in the data, in some cases, the predicted probabilities for just a very few observations can be so large as to bias the monthly total. As a result, we censor predicted probabilities and require that they be within the interval $[0,1]$. If we do not do this, we end up with very similar results with the exception of a few months where individual outliers bias the monthly estimate to be unusually large or small. In practice, we also omit results for January 1983, January 1994, January 2003, January 2011, as these are months in which substantial changes in either occupational coding or CPS survey procedures generate notable outliers for monthly switching (in all series, not just ours).

To compute this expression numerically, we need data on X, Y, and $\tilde{SW}$, and we need to take a stance on how the reported switching rates in the monthly data correspond to the different types of reported switching rates in the annual data. Fortunately, the same individual characteristics, X, we use in the annual data are observed in every period in the monthly data. For the labor market outcomes, we only consistently observe whether or not a full-time/part-time switch occurred. From 1994, with the introduction of dependent coding in the monthly CPS, we have data on whether or not there was a change in employer. However, we have no data on intervening unemployment between the months an individual is observed, and we only observe income for the outgoing rotation groups (the fourth and eighth months an individual is surveyed) and thus do not have the ability to construct the change in wages. Thus, in our baseline estimates, we use solely data on a full-time/part-time switch prior to 1994 and data on both a full-time/part-time switch and an employer switch after 1994.

As described in the text, the way in which occupations are coded in the monthly CPS changed in 1994, with a shift to dependent coding. Given that the retrospective measures of annual mobility in the March CPS also use dependent coding, when we estimate the predicted probability of switching, we treat most occupational switches after the year 1994 as retrospective switches and thus use $\alpha_{R,0}, \alpha_{R,1}$, and $\alpha_{R,X}$ in our computation. The exceptions post-1994 are the cases in which an individual did not respond to the employer switch question. In this instance, occupation coding reverts to independent coding, and we treat the switch as corresponding to the longitudinal switches we observed in the linked annual data. Since monthly switching can only be observed in the second, third or fourth months, we are unaffected by the independent coding of occupations that occurs in the first and fourth months an individual is surveyed.

In constructing the predicted probabilities, we continue to normalize the vector of individual characteristics, X, to zero and thus do not require any adjustments to the parameters estimated. The key effect of this is that it imposes the same sample means for the individual characteristics in the monthly samples as in the annual samples. We do this to reduce variability in the predicted switching rates due to changes in sample composition from month to month and to maximize comparability with the annual data.

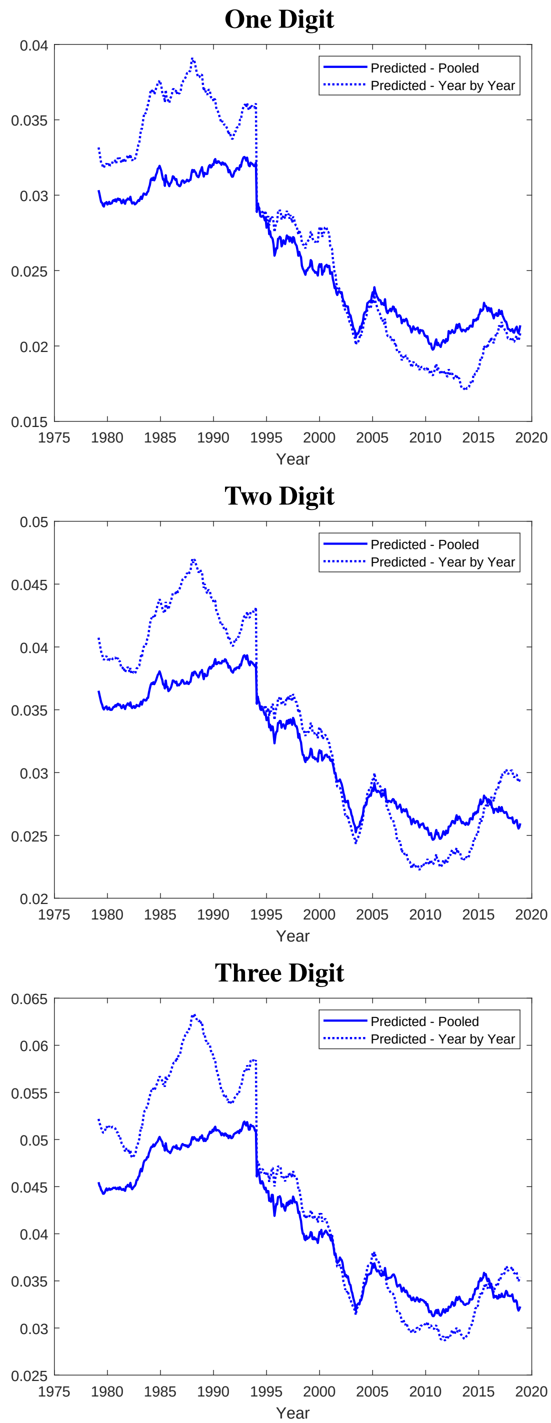
We also must decide whether to use our pooled estimates of parameters or our year by year estimates. The trade-offs are straightforward. Using our parameters from the pooled sample, we avoid the substantial variance and noise that would come from using parameters that change from year to year and we do not need to take a stance on how to map each year's parameter estimates to each month. On the other hand, using these pooled estimates will fail to capture trends in the error rates for longitudinal measures. Since the monthly CPS switches to (mostly) dependent coding beginning in 1984, we consider this abstraction reasonable and thus, in the paper, we report estimates using the pooled estimates.

However, as a robustness check, we present below the predicted monthly switching rate using year by year parameter estimates. To minimize additional noise generated from using these parameters, we do the following things. First, for the two years in which we are unable to obtain estimates because of linking limitations, 1986 and 1996, we average the parameters estimated in the two adjacent years. Further, we also use a five year moving average of the year by year parameters obtained from the model. This allows us to capture trends in occupational mobility and measurement error without the full variability present in the year by year estimates. Finally, since annual occupational mobility is computed in March, we construct our monthly estimates using a 13 month moving average, where we assume that annual occupational mobility estimates for year t map to monthly occupational mobility estimates for April of the prior year to March of the present year.As with our baseline results, we then take the 13 month moving average of the total monthly switching rates to minimize noise and allow for easier interpretation of our results.

We report the predicted monthly switching rates using year by year parameters in Figure D.1 below, alongside our estimates obtained using the pooled sample parameters. The primary differences between the two estimates are observed prior to the year 2000, as the year by year estimates generate higher levels of occupational mobility prior to 1994 and a larger discontinuity in the year 1994. However, aside from these differences, the dynamics of the series are very comparable, primarily in the last couple of decades of the sample.[39]

[39]For some of our results with these dynamically computed predicted switching rates, we observe a notable uptick in occupational switching in the year 2017, which persists to the end of the sample. However, this uptick is in part driven by the fact that we do not have estimates for the year 2019, and thus the moving average of parameters used to construct occupational switching places particularly high weight on the parameter estimates from the year 2017, which similarly estimate an uptick. We would expect with further data for 2019 and beyond that this increase would be smoothed out.

Figure D.1: Comparison of Corrections to Monthly Occupational Mobility using Pooled and Year by Year Parameter Estimates, 1979-2018



Appendix E: Details and Robustness to Exercises Revisiting Trade, Occupational Switching, and Wage Changes in Ebenstein et al. (2014)

In this Appendix, we provide added detail regarding our replication and extension of the results of Ebenstein et al. (2014) (henceforth, EHMP) as well as robustness checks where we vary the samples used in estimation.

E.1 Implementation Details

The conditional expectations of the original two stage least squares specification of EHMP can be written as:

$$\begin{aligned}\mathbb{E}\left[\tilde{SW}^L_{i,o,t} \mid Tradable_o, C_{i,o,t}\right] &= \eta_0 + \eta_1 Tradable_o + C_{i,o,t}\eta_2 \\ \mathbb{E}\left[\Delta ln(w_{i,o,t}) \mid \tilde{SW}^L_{i,o,t}, C_{i,o,t}\right] &= \xi_0 + \xi_1 \tilde{SW}^L_{i,o,t} + C_{i,o,t}\xi_2\end{aligned} \tag{8}$$

where $\tilde{SW}^L_{i,ot}$ is 0/1 variable measuring whether or not an individual i in occupation o last year reported a longitudinal switch; $Tradable_o$ is the instrument for occupational switching, measuring whether or not an individual's prior year occupation is exposed to globalization and offshoring (as constructed in EHMP); $\Delta ln(w_{it})$ is the one year log change in wages for a worker i who was in occupation o last year; and $C_{i,o,t}$ is a vector of individual and occupational characteristics used as controls – age, female, nonwhite, union status, a set of educational dummy variables, state fixed effects, year fixed effects, and finally a pre-switching occupation wage premium for occupational switchers. This last term is constructed by running a regression of log wages on the other individual characteristics listed above and a set of occupational fixed effects in the year prior to when individuals may have switched. The regression is run separately for all workers and workers who report working in a different occupation in the following year, and the premium for each occupation is constructed as the difference in the fixed effect coefficients in the two regressions for each occupation.

The moment conditions used to estimate the first stage regression for the above specification can be written as:

$$\begin{aligned}\mathbb{E}\left[X'_{i,o,t}\tilde{SW}^L_{i,o,t}\right] &= \mathbb{E}\left[X'_{i,o,t}\eta_0\right] + \mathbb{E}\left[X'_{i,o,t}\eta_1 Tradable_o\right] + \mathbb{E}\left[X'_{i,o,t}C_{i,t}\right]\eta_2 \\ &= \Sigma_{XX,t}\vec{\eta}\end{aligned} \tag{9}$$

where $X_{i,o,t} = [Tradable_o, C_{i,o,t}]$, $\Sigma_{XX,t}$ is the covariance matrix of $X_{i,o,t}$, $\vec{\eta} = [\eta_1; \eta_2]$ and we have normalized $X_{i,o,t}$ such that $\mathbb{E}[X_{i,o,t}] = 0$.

The moment conditions used to estimate the instrumental variables based on the above specification can be written as:

$$\begin{aligned}\mathbb{E}\left[X'_{i,o,t}\Delta ln(w_{i,o,t})\right] &= \mathbb{E}\left[X'_{i,o,t}\xi_0\right] + \mathbb{E}\left[X'_{i,o,t}S\tilde{W}^{L}_{i,o,t}\right]\xi_1 + \mathbb{E}\left[X'_{i,o,t}C_{i,o,t}\right]\xi_2 \\ &= \mathbb{E}\left[X'_{i,o,t}\tilde{X}_{i,o,t}\right]\vec{\xi}\end{aligned} \tag{10}$$

where $\tilde{X}_{i,o,t} = [S\tilde{W}^{L}_{i,o,t}, C_{i,o,t}]$ and $\vec{\xi} = [\xi_1; \xi_2]$. Manipulation of these two sets of moments conditions yields standard formulas for least squares and IV estimation in the exactly identified cases (which these are).

The estimation we consider estimates simultaneously the parameters $\vec{\xi}$ and $\vec{\eta}$ from the above moment conditions with the estimation of the parameters β_0, β_1, $\alpha_{R,0}$, $\alpha_{R,1}$, $\alpha_{L,0}$, $\alpha_{L,1}$, δ_0,β_2,δ_1, $\alpha_{R,X}$, and $\alpha_{L,X}$ from the original model, using the moment conditions spelled out in Appendix B. We assume that the set of individual characteristics used to estimate our original set of parameters is the same set of individual characteristics used to estimate the above specification as well as the instrument indicating whether or not an occupation is tradable. The key question for estimation is then whether or not the samples used in estimating both sets of moments are the same, which will impact the optimal weighting matrix in the GMM estimation and thus the standard errors for all parameters. We first describe the data samples used and then return to this question.

The original study of EHMP focuses on workers between the ages 16-64 who are not self-employed with two wage observations in the outgoing rotation groups of the CPS. Their data are obtained originally from CEPR and cover the years 1984-2002. Their original data files and code are available from the Data Archive site hosted at the Review of Economics and Statistics.

For our baseline exercises, we aim to adjust the original data of EHMP as minimally as possible, however, there are a few necessary adjustments we must make to estimate their specifications jointly with ours. First, we convert their occupation codes into the time-consistent occupation codes of Autor and Dorn (2013). Having converted occupational codes in their original data, we then merge the measures of tradability for each occupation from their data to our linked March sample, so that the definition of whether or not an occupation is tradable or not is consistently applied in all samples. Second, if we want to use the same covariates in estimating measurement error as they use in their specification, we need to merge the outgoing rotation group data on to our linked March supplement sample to have the data for whether or not an individual is in a union. The linking keys we use to map outgoing rotation group data to our linked March supplement sample come from Flood and Pacas (2016), but they only go back to the year 1989. Thus, the second adjustment we make is to only use data from 1989-2002. On top of that, since we are unable to link workers between 1995-1996 and because there is no information available on state for the year 1995, we drop workers whose second observation occurs in the years 1995 and 1996. The final adjustment we make relative to their original code is to adjust the computation of the occupation premium, as there is a slight error in the original code. However, all results are robust to using their original measure of this control or excluding it altogether.[40]

Our baseline estimates leave their sample further untouched and use that sample to estimate the parameters of their IV specification. We then use our original sample, adjusted to match theirs (all non-self-employed workers ages 16-64), to estimate the parameters governing occupational mobility and measurement error. We estimate these two specifications simultaneously, but construct the standard errors and weighting matrix under the assumption that the two samples are disjoint. While this assumption is false, given the overlap of the March supplement and outgoing rotation group sample, it allows for the least disruption to the original findings of EHMP.

As robustness checks, we report later in this Appendix two alternate sample structures. First, we remove all workers in their original outgoing rotation group sample who are observed in the months of March – June. We drop the months of Apr-June because workers who complete the outgoing rotation group questions could have been observed in the prior three months and may have completed the March supplement questions. In this specification, the two samples are truly disjoint, and thus our baseline assumptions for estimation are most justified in this case. As a second robustness check, we merge the original data from EHMP to a complete set of the outgoing rotation group data obtained from IPUMS, then use the linking keys of Flood and Pacas (2016) to merge this data to our March supplement data.[41] In this case, we literally have the same samples for both sets of moments, and estimate everything jointly, allowing for covariance between all the moments in estimation.

Before reporting the results of our estimation, we first illustrate the nature of our measurement error corrections. The primary emphases of our corrections has to do with the first stage of the instrumental variables strategy. The estimated parameter η_1 in equation (8) aims to capture the relationship between being in a tradable job and occupational switching, but is estimated using the error-laden longitudinal measure of occupational switching. Because the measurement error is non-classical, using the longitudinal measure of switching will result in a biased estimate of η_1. The correct estimate will be estimated from the parameter vector δ_1 from equation (4), which estimates the relationship between individual characteristics, including whether or not an occupation

[40]We also find that if we construct this premium measure using retrospective measures of occupations when we link the two samples together, this has no impact on the results.

[41]To do this, we link the outgoing rotation group observations to their observation in the month of March and then link this to the March supplement data using the linking keys from Flood and Pacas (2016). Thus, in our full linked March sample, we have outgoing rotation group observations included from April-June. We do this to increase our sample size and power in estimation. This multi-step linking procedure leads to some loss of observations in the cases where we fail to match, however, we have not found this to have a significant quantitative impact on our estimates.

is tradable, and actual occupational switching.

To illustrate the relationship between the corrected and uncorrected IV coefficients, observe that in our model, the proper moment condition for the IV estimation is:

$$\begin{aligned}\mathbb{E}\left[X'_{i,o,t}\Delta ln(w_{i,o,t})\right] &= \mathbb{E}\left[X'_{i,o,t}\xi_0\right] + \mathbb{E}\left[X'_{i,o,t}SW_{i,o,t}\right]\xi_1 + \mathbb{E}\left[X'_{i,o,t}C_{i,o,t}\right]\xi_2 \\ &= \Sigma_{XX,t}\delta_1\xi_1 + \mathbb{E}\left[X'_{i,o,t}C_{i,o,t}\right]\xi_2 \\ &= \Sigma_{XX,t}\left(\delta_1\xi_1 + [0;\xi_2]\right)\end{aligned}$$

The corrected IV coefficient, $\hat{\xi}_1^{IV}$, will be obtained by estimating ξ_1 using the above moment condition. To relate this directly to the uncorrected IV coefficient, $\tilde{\xi}_1^{IV}$, estimated from the moment condition in (10), we first manipulate (10) to be written as:

$$\begin{aligned}\mathbb{E}\left[X'_{i,o,t}\Delta ln(w_{i,o,t})\right] &= \mathbb{E}\left[X'_{i,o,t}\xi_0\right] + \mathbb{E}\left[X'_{i,o,t}S\tilde{W}^L_{i,o,t}\right]\xi_1 + \mathbb{E}\left[X'_{i,o,t}C_{i,o,t}\right]\xi_2 \\ &= \Sigma_{XX,t}\vec{\eta}\xi_1 + \Sigma_{XX,t}[0;\xi_2] \\ &= \Sigma_{XX,t}\left(\vec{\eta}\xi_1 + [0;\xi_2]\right)\end{aligned}$$

where the second step is obtained by plugging in from (9). The combination of these two moment conditions implies that:

$$\delta_1\hat{\xi}_1^{IV} + [0;\hat{\xi}_2^{IV}] = \vec{\eta}\tilde{\xi}_1^{IV} + [0;\tilde{\xi}_2^{IV}]$$

and thus that $\hat{\xi}_1^{IV} = \frac{\tilde{\eta}_1}{\hat{\eta}_1}\tilde{\xi}_1^{IV}$ where $\tilde{\eta}_1$ is the first stage estimate of the relationship between tradable occupations and a reported longitudinal occupational switch and$\hat{\eta}_1$ (given by the first element of δ_1, or rather $\delta_1(Trad.)$, the element of the δ_1 vector corresponding to the tradable indicator) is the first stage estimate corrected for measurement error.

If we go a step further, our moments used in estimating measurement error can specify the nature of the bias in the first stage estimate. The moment used in estimating the first stage, $\mathbb{E}\left[X'_{i,o,t}S\tilde{W}^L_{i,o,t}\right]$ is also used in estimating the measurement error since we are using the same set of individual characteristics and since our indicators for each possible switching outcome observed can be related to a longitudinal switch as $S\tilde{W}^L_{i,o,t} = \tilde{Z}_{i,1,t} + \tilde{Z}_{i,4,t}$ (for definitions of those indicators, see Section 3.1 in the body of the paper). Drawing from the moments in Appendix B, the implied moment for $\mathbb{E}\left[X'_{i,o,t}S\tilde{W}^L_{i,o,t}\right]$ in estimation of the measurement error is given by:

$$\mathbb{E}\left[X'_{i,o,t}S\tilde{W}^L_{i,o,t}\right] = \Sigma_{XX,t}\left(\alpha_{L,1}\delta_1 + \alpha_{L,X}\right)$$

which implies that $\vec{\eta} = \alpha_{L,1}\delta_1 + \alpha_{L,X}$, where again, δ_1 represent the true first stage coefficients. This allows us to write the expression given in the body of the paper, that

$$\hat{\xi}_1^{IV} = \frac{\tilde{\eta}_1}{\hat{\eta}_1}\tilde{\xi}_1^{IV} = \frac{\alpha_{L,1}\delta_1(Trad.) + \alpha_{L,X}(Trad)}{\delta_1(Trad.)}\tilde{\xi}_1^{IV}$$

where $\alpha_{L,X}(Trad.)$ is the coefficient relating the tradable instrument to reported longitudinal occupational switching. In practice, since our GMM system is overidentified, it may not be that $\tilde{\eta}_1 = \alpha_{L,1}\delta_1(Trad.) + \alpha_{L,X}(Trad)$ exactly. As a result, in all our results, we independently estimate the first stage parameters (using the moment conditions in (9)). However, we have checked and find that the implied first stage coefficient is in practice very similar.

There are two key sources of bias here, the fact that $\alpha_{L,1}$ may be less than one, representing the fact that there are some false positive longitudinal switches (since $\alpha_{L,0} + \alpha_{L,1} \leq 1$), and the fact that $\alpha_{L,X}(Trad.)$ may not be zero, which would be the case if the tradable instrument is correlated with the error in longitudinal occupational switching. In the case where being in a tradable occupation is orthogonal to the measurement error in occupational switching ($\alpha_{L,X}(Trad.) = 0$), then the true IV coefficient will be smaller in absolute value than the estimated one in the data, as the first stage coefficient is attenuated by the measurement error in occupational switching (because $\alpha_{L,1} < 1$). On the other hand, if there is some correlation between being in a tradable occupation and measurement error in occupational switching, then the nature of the bias can't be signed ex ante, as it will depend on the signs and magnitudes of all the parameters. In the case where $\delta_1(Trad.)$ and $\alpha_{L,X}(Trad.)$ are of opposite sign, the true IV coefficient could be of the opposite sign of that estimated in the data.

E.2 Additional Results and Robustness Checks

A natural concern with the results we present is whether or not the different sample and different set of individual characteristics implies different estimates of occupational mobility and measurement error than our baseline results reported in Table 2. We show that this is not a concern in Table E.1, where we report the estimated occupational switching rates and error rates obtained from the sample used in this extension. The estimated parameters look very similar to those reported in Table 2.

Table E.1: Pooled Estimates of Actual Occupational Mobility and Measurement Error using Sample for Ebenstein et al. (2014) Extensions, 1989-2002

	One Digit	Two Digit	Three Digit
Actual Avg. Occupational Switching, δ_0	0.046	0.056	0.072
	(0.001)	(0.001)	(0.002)
Reported Ret. Occ. Mobility	0.029	0.037	0.048
	(0.001)	(0.001)	(0.001)
Reported Long. Occ. Mobility	0.203	0.272	0.408
	(0.001)	(0.001)	(0.001)
False Positive Rate, Retrospective	0.018	0.036	0.047
	(0.011)	(0.008)	(0.006)
False Positive Rate, Longitudinal	0.823	0.826	0.841
	(0.005)	(0.004)	(0.003)
False Negative Rate, Retrospective	0.019	0.021	0.027
	(0.001)	(0.001)	(0.001)
False Negative Rate, Longitudinal	0.013	0.012	0.012
	(0.001)	(0.001)	(0.001)
N	136,522	136,522	136,522

These results report occupational mobility and measurement error estimates from the model described in Section 3. Data used is longitudinally linked March CPS responses ranging from 1989-2002 for non-self-employed workers ages 16-64. Point estimates and standard errors (reported below in parentheses) are obtained from standard two stage GMM estimates. Standard errors for false positive and negative rates (defined in the text in Section 4.1) are constructed via the delta method. Reported standard errors for measured longitudinal and retrospective occupational mobility are the standard deviation of occupational mobility divided by $\sqrt{N}$. All estimates are from the model where the vector of individual characteristics is defined as de-meaned age, female, nonwhite, union status, a set of educational dummy variables, state fixed effects, year fixed effects, a pre-switching occupation wage premium for occupational switchers, and an indicator for whether or not the original occupation was tradable (measured using the longitudinally).

We now present our two robustness checks using different samples in estimation. Table E.2 presents our results for the first stage and IV coefficients when we omit all possibly overlapping observations from the data used to estimate the IV and first stage specification and Table E.3 presents our results for the case where we estimate all parameters jointly from a single linked sample. In both cases, we see that the corrections for measurement error flip the sign of the first stage coefficients and thus flip the sign of the IV estimates, while increasing them in absolute value.

Table E.2: Estimates of Wage Changes with Occupational Switching for Workers in Tradable Occupations, Disjoint Estimation Samples, 1989-2002

Parameter	One Digit	Two Digit	Three Digit
IV estimate of occ. switching on wages with tradable	-0.134	-0.176	-0.112
instrument, uncorrected ($\tilde{\xi}_1^{IV}$)	(0.041)	(0.026)	(0.020)
IV estimate of occ. switching on wages with tradable	0.459	0.578	0.370
instrument, corrected ($\hat{\xi}_1^{IV}$)	(0.146)	(0.092)	(0.068)
Difference	-0.593	-0.753	-0.482
	(0.184)	(0.116)	(0.087)
First stage estimate of tradable occupation on	0.045	0.071	0.091
occupational switching, uncorrected ($\tilde{\eta}_1^{IV}$)	(0.002)	(0.002)	(0.002)
First stage estimate of tradable occupation on	-0.013	-0.022	-0.028
occupational switching, corrected ($\hat{\eta}_1^{IV}$)	(0.002)	(0.002)	(0.003)
Difference	0.058	0.092	0.119
	(0.003)	(0.003)	(0.004)

See notes to Table 6.

Table E.3: Estimates of Wage Changes with Occupational Switching for Workers in Tradable Occupations, Jointly Estimated from Same Sample used to Estimate Measurement Error, 1989-2002

Parameter	One Digit	Two Digit	Three Digit
IV estimate of occ. switching on wages with tradable	-0.093	-0.125	-0.075
instrument, uncorrected ($\tilde{\xi}_1^{IV}$)	(0.053)	(0.039)	(0.028)
IV estimate of occ. switching on wages with tradable	0.425	0.460	0.289
instrument, corrected ($\hat{\xi}_1^{IV}$)	(0.256)	(0.152)	(0.112)
Difference	-0.518	-0.585	-0.363
	(0.306)	(0.188)	(0.140)
First stage estimate of tradable occupation on	0.054	0.074	0.101
occupational switching, uncorrected ($\tilde{\eta}_1^{IV}$)	(0.003)	(0.004)	(0.004)
First stage estimate of tradable occupation on	-0.012	-0.020	-0.026
occupational switching, corrected ($\hat{\eta}_1^{IV}$)	(0.002)	(0.002)	(0.003)
Difference	0.066	0.094	0.127
	(0.004)	(0.004)	(0.005)

See notes to Table 6.