sociological science

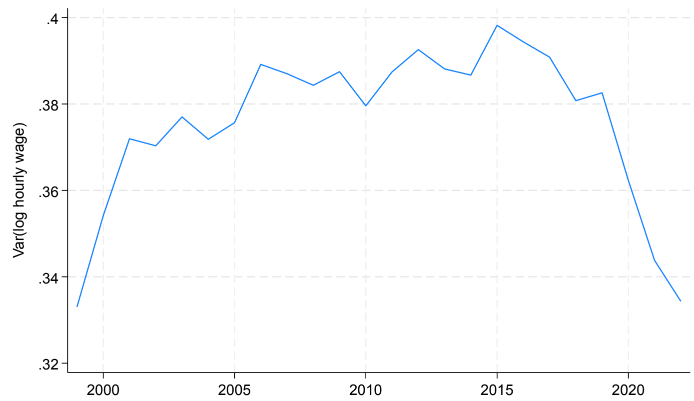
Supplement to:

Roh, Soohyun, Nathan Wilmers. 2026. "Declining Inequality and Persistent Inequality Structures" Sociological Science 13: 614-644.

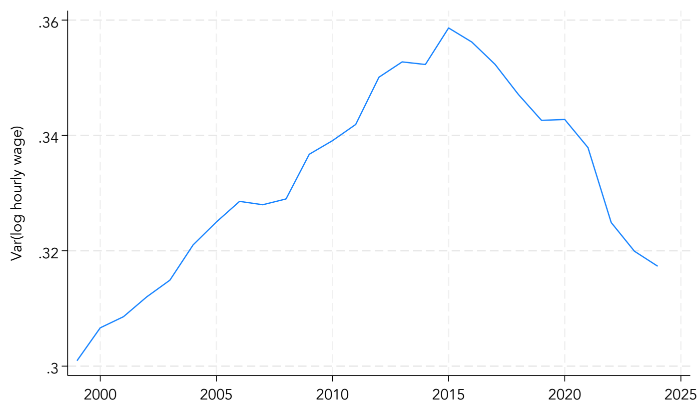
Online Appendix to

Declining Inequality and Persistent Inequality Structures

Appendix A. Comparisons to Publicly Available CPS Data



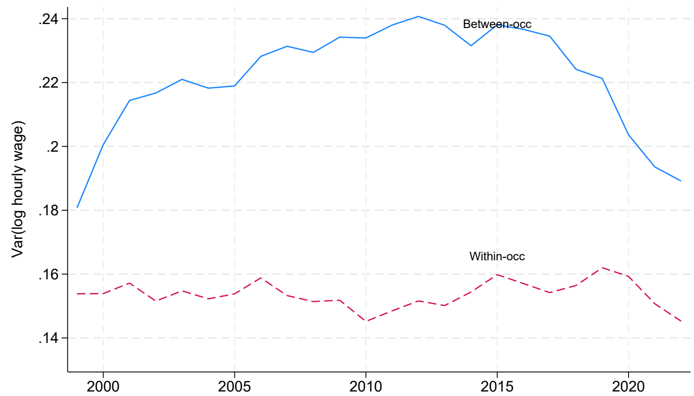
A. OEWS



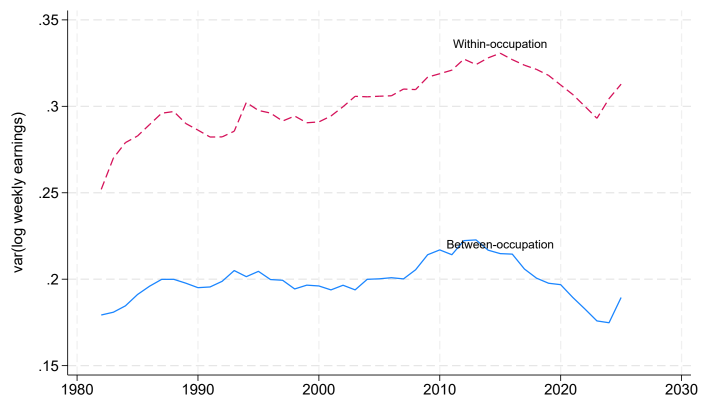
B. CPS

FIGURE A1. Rising and Declining Inequality.

Data is CPS-ORG and OEWS. CPS-ORG sample is workers 16 to 65 with at least 20 hours per week and 40 weeks worked per year.



A. OEWS: Occupation



B. CPS: Occupation

FIGURE A2. Between- and within-occupation variance declines.

CPS-ORG with IPUMS detailed occupation categories. OEWS with 5-digit SOC codes.

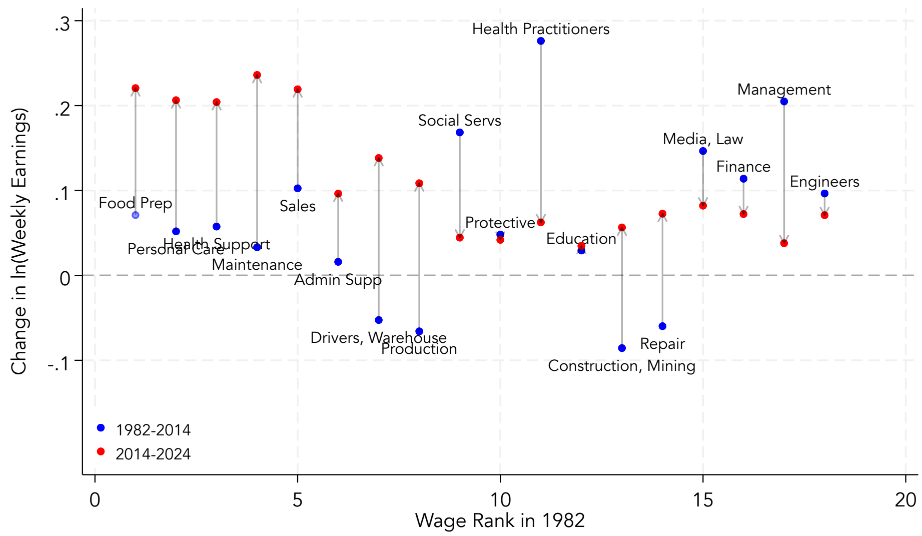


FIGURE A3. Changes in weekly earnings across major occupational categories

CPS-ORG with IPUMS detailed occupation categories.

Appendix B. Occupational Employment Statistics Data Details

B.1. Changes in the OEWS Data Structure

Beginning May 2021, OEWS shifted to a new model-based estimation strategy. It affected the final sample and, more importantly for our analysis, the survey weights offered from the BLS. On the other hand, survey methodology stayed mostly the same, except that now individual wage rates, in addition to wage intervals, are also collected from the majority of the employers.

Previously, as similar to most other surveys, the OEWS estimates have been produced directly from a stratified random sample and corresponding survey weights. Under the new estimation scheme, in contrast, the survey responses are used to model occupations and wages of all unobserved establishments in the population. In doing so, respondent establishments that do not match to their Quarterly Census of Employment and Wages (QCEW) records in terms of 6-digit NAICS, ownership, MSA, and total employment are considered to have failed to meet the stability criteria and are grouped together with the unobserved units. The wages and occupations of each unobserved unit, which includes non-sampled units, non-respondents, and respondents that do not pass stability criteria, are predicted from those of 10 nearest neighbors in terms of industry, size, ownership, and location. The final dataset for estimation thus represents the universe of establishments within the US, where all survey respondents get equal weights.

The change in the dataset poses issues for comparability of our analysis over the years we study, mostly coming from three dimensions. We ensure consistent estimation by addressing those three dimensions as follows. First, while individual wage rates became available for most units after 2020, we keep using the wage interval information for our analysis across the years. Second, we restrict our sample to survey respondents, dropping imputed observations following previous studies (Wilmers and Aeppli 2021). Third, we weight our analysis from 2019 with our emulated BLS weights.

Before the 2021 change in OEWS methodology, BLS offered combined panel weights to be applied to their stratified sample for estimation. However, since now the new model-based estimation assigns all respondents equal weights, combined panel weights are not offered anymore. We therefore emulate how BLS has constructed their weights in the past and apply that from 2019 estimates, while using official BLS weights for those until 2018.

OEWS has a unique survey design as it combines six panel surveys to construct a single cross-sectional estimate for a three-year period. Each panel is designed to represent the population, and each stratum, defined with 4 or 5-digit NAICS, state, and MSA, is assigned with survey weights that is inverse of its probability of being sampled. In combining six panels, therefore, the weights get divided by the number out of six that a particular stratum got sampled in the three-year period. Exceptions to this general rule are certainty and virtual certainty units, which are always assigned the combined weights of 1. These are establishments that are always expected to be sampled once during the three-year period.

We emulate this calculation process by dividing single panel survey weights by the total number a cell defined by 4-digit NAICS, state, and MSA is sampled across six panels. Also, we follow BLS documentation to identify certainty and virtual certainty units. The resulting weight allows us to consistently estimate the parameters of our interest across the years.

B.2. SOC Occupation Recodes

OEWS dataset provides wage and employment information for 6-digit SOC occupation codes. However, the codes are not consistent across the years, due to SOC reclassifications in 2000, 2010, and 2018. OEWS also have used some internal occupation codes as a response to these reclassifications. To ensure consistency in our occupation-based analysis across different years, we standardized the occupation classifications using 5-digit occupation codes. Specifically, for codes that do not have a one-to-one match across various classification systems, we introduced a new overarching code that encompasses all their subdivisions. For example, in the 2000 SOC classification, “Computer programmers” (15-102) and “Computer software engineers” (15-103) were distinct at the 5-digit level. However, in the 2010 SOC, they were combined under the category “Software developers and programmers” (15-113). We have designated 15-125 as the umbrella code that encompasses all these variations. The resulting adjusted 5-digit SOC codes comprises of 393 unique broad occupations, as compared to 459 in the 2018 SOC system.

Appendix C. Local Unemployment Regressions

To examine the relationship between local unemployment rates and the components of wage inequality at the local level, we first replace the occupation fixed-effects term ($\theta_{o,t}$) in Equation 3 with commuting-zone–specific occupation fixed effects ($\tilde{\theta}_{c,o,t}$), as shown in Equation A1.

$$y_{c,j,o,t} = \psi_{c,j,t} + \tilde{\theta}_{c,o,t} + \tilde{u}_{c,j,o,t} \tag{A1}$$

Note that we do not have to change the establishment fixed effects term as establishments are already nested within CZs. Then, we follow the same variance decomposition procedure from Equation 4 but *within* each CZ in Equation A2.

$$Var(\widehat{y}_{c,j,o,t}) = Var(\psi_{c,j,t}) + Var(\tilde{\theta}_{c,o,t}) + 2Cov(\psi_{c,j,t}, \tilde{\theta}_{c,o,t}) + Var(\tilde{u}_{c,j,o,t}). \tag{A2}$$

That is, we decompose the total CZ-level wage variance ($Var(\widehat{y}_{c,j,o,t})$) into its component terms: between-workplace wage variance ($Var(\psi_{c,j,t})$), between-occupation wage variance ($Var(\tilde{\theta}_{c,o,t})$), the covariance ($Cov(\psi_{c,j,t}, \tilde{\theta}_{c,o,t})$), and the residual wage variance ($Var(\tilde{u}_{c,j,o,t})$), each calculated at the CZ level.

We then estimate a series of CZ-level panel regressions in which the dependent variable is one of these variance components. Equation A3 illustrates the specification for the case where the dependent variable is the between-occupation pay variance term.

$$Var(\tilde{\theta})_{c,t} = U_{c,t} + \mu_c + \tau_t + \varepsilon_{c,t} \tag{A3}$$

Here, the CZ-by-year level between-occupation wage variance term ($Var(\tilde{\theta})_{c,t}$) is regressed on the local unemployment rate ($U_{c,t}$), controlling for CZ fixed effects (μ_c) and year fixed effects (τ_t). This specification captures how variation in local unemployment rates predicts variation in wage inequality between occupations, net of common time shocks and time-invariant CZ characteristics. In Table 3, we replicate Equation A3 using alternative dependent variables: total CZ-level wage variance ($Var(\widehat{y}_{c,j,o,t})$), between-establishment wage variance ($Var(\psi_{c,j,t})$), and the covariance between establishment and occupation premiums ($Cov(\psi_{c,j,t}, \tilde{\theta}_{c,o,t})$).