WORKING PAPER · NO. 2024-96

Disemployment Effects of Unemployment Insurance: A Meta-Analysis

Jonathan Cohen and Peter Ganong

AUGUST 2024

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Disemployment Effects of Unemployment Insurance: A Meta-Analysis *

Jonathan Cohen
Amazon

Peter Ganong
University of Chicago and NBER

August 5, 2024

Abstract

We systematically review studies of how unemployment benefits affect unemployment duration. Statistically significant findings are eleven times more likely to be published. Correcting for publication bias reduces average elasticity by factor of two. Meta-analysis provides a principled way for sufficient statistics methods to aggregate estimates across policy contexts and speak to the optimality of large reforms. Although existing consumption drop-based approaches typically imply an optimal replacement rate near zero, our corrected estimates imply an optimal replacement rate of 28%. The "micro" elasticity is equal to the "macro" elasticity, suggesting that general equilibrium effects are unimportant or cancel out.

Keywords: Unemployment Benefits, Publication Bias, Meta-analysis, Baily-Chetty

JEL Codes: C13, E24, E64, J64, J65

*Cohen: jonpcohen@gmail.com. Ganong: ganong@uchicago.edu (corresponding author). This research was conducted prior to the author's employment by Amazon. This paper is not sponsored or endorsed by, or associated with Amazon or any of its subsidiaries or affiliates. The views, opinions and positions included in this paper are the author's own and do not reflect the views, opinions and positions of Amazon. We thank Miguel Acosta, David Autor, Isaiah Andrews, Manasi Deshpande, Amy Finkelstein, Tomáš Havránek, Nathan Hendren, Simon Jaeger, Jonas Jessen, Max Kasy, Henrik Kleven, Bruce Meyer, Emi Nakamura, Pascal Noel, Matt Notowidigdo, and Charlie Rafkin for helpful conversations. We thank the staff of the Congressional Budget Office for their interest in the employment effects of unemployment insurace as well as seminar participants at the OECD and the Chicago Fed for useful suggestions. We thank Avik Garg, John Spence, Nicolas Wuthenow, Fatima Yusuf, and Madeline Zuckerman for excellent research assistance.

1 Introduction

How much do more generous unemployment benefits prolong the duration of unemployment? Although there are many published microeconomic studies estimating this relationship, there are some research questions which are difficult or impossible to answer in a single such study. For example, research questions which depend on the policy environment such as "How does the elasticity vary with the replacement rate?" are best answered by aggregating insights from multiple studies. Further, even a researcher interested in "just" the average elasticity may reach misleading conclusions if publication bias makes the set of published estimates unrepresentative of the distribution of latent estimates. We therefore re-examine the effects of unemployment benefits, applying recent advances in meta-analysis methods to 55 prior studies.

We have three major findings. First, a publication bias toward statistically significant findings is pervasive. If researchers and journals apply more scrutiny to statistically insignificant findings, then estimates of small elasticities with large standard errors will be censored from the published record. We document evidence of just such a pattern: estimates in the middle and top tercile of standard errors are about twice as large as estimates in the bottom tercile of standard errors.

The published record therefore overstates the mean of the latent distribution of duration elasticities. Using a correction method from Andrews and Kasy (2019), we find that statistically significant positive findings are about 11 times more likely to be published. In addition, because many factors—not just the standard error—may influence a study's estimated elasticity, we conduct a meta-analysis using Bayesian model averaging (Irsova et al., 2023). After correcting for publication bias, we find that the duration elasticity to replacement rates is 55% smaller using the first method and 37% smaller using the second method. The corrections to estimates of the duration elasticity to potential benefit duration are even more drastic.

Second, we show that meta-analysis is a complement to sufficient statistics methods by re-evaluating the optimal replacement rate formula of Baily (1978) and Chetty (2006). Using this formula and parameter estimates in the literature, a number of prior papers have found that it is optimal to have no benefits at all.[1]

Meta-analysis improves on the duration elasticity parameter input by correcting for publication bias and by enabling predictions for large reforms. In the canonical characterization of sufficient statistics, evaluating whether a large policy reform is optimal is either impossible or requires unpalatable structural assumptions, as discussed in Kleven (2021). Meta-analysis provides a way to preserve the transparency of the sufficient statistics approach while expanding its remit to characterize the optimality of large policy changes. Using the prior literature's consumption-based estimates for the gains together with these methodological improvements in characterizing the costs, we estimate an optimal replacement rate of 28%.

Third, we find no evidence of a difference between the "micro" and "macro" duration elasticity. An active theoretical and empirical literature seeks to understand and compare the causal effect of increasing benefits for a single worker (the "micro" effect) to the causal effect of increasing benefits for all workers (the "macro" effect).[2] In principle, these two effects could differ because of a number of different general equilibrium channels. Within our meta-analysis, we identify 14 studies where the treatment constitutes a change in both the worker's own benefits and the benefits of other workers participating in the same labor market. We find no systematic evidence of differences between micro and macro elasticities using cross-study heterogeneity in the share of workers affected. Our estimates therefore suggest that either these general equilibrium channels have limited aggregate importance, or that they cancel each other out.

Our analysis builds on several prior excellent literature reviews (Krueger and Meyer, 2002; Meyer, 2002; Schmieder and von Wachter, 2016; Lopes, 2022) by applying recent methods from the meta-analysis literature. Section 2 contains the study collection procedure; we follow the meta-analysis best practices laid out in Havránek et al. (2020). Section 3 presents evidence of publication bias. Section 4 documents heterogeneity in elasticities based on study characteristics. Section 5 draws out implications for the optimal level of benefits. Section 6 discusses the micro vs. macro elasticity. Section 7 concludes.

[1] For example, Gruber's (1997) classic paper "The Consumption-Smoothing Benefits of Unemployment Insurance" found that the optimal replacement rate is 2% (essentially zero) for a relative risk aversion coefficient of 2. Focusing on the social insurance motive for benefits, similar conclusions appear in Setty and Yedid-Levi (2021) and Krusell et al. (2010).

[2] See, e.g., Hagedorn et al. (2016), Michaillat (2012), Landais et al. (2018b), and Jessen et al. (2023).

2 Data

We aim to survey all microeconomic papers that estimate the causal effect of unemployment insurance (UI) potential benefit duration (PBD) or replacement rate (RR) on unemployment duration. We collect journal publications from Google Scholar following the guidelines compiled by the Meta-Analysis in Economics Research Network (Havránek et al., 2020). We limit the survey to papers published by the time the search was conducted on August 15, 2022, and we limit to the first 1,000 results that Google Scholar returns.[3] Of these 1,000 search results, 407 are journal publications, and 55 papers use microeconomic methods while reporting enough information to calculate a UI duration elasticity and standard error. Details on how we construct a comparable duration elasticity across different study methods are in Appendix C.

The 55 sample studies are of high quality based on the available metrics. 69% of the studies are identified using a quasi-experimental research design. Figure A.1 shows it is much more common for recently published papers to use quasi-experimental identification strategies. They are published in influential economics journals. The 25$^{\text{th}}$ percentile and median impact factors correspond to Oxford Economic Papers and the Journal of Public Economics, respectively. Half of the estimates are from such field journals, one-fifth of the estimates are from one of the "Top-5" general interest journals, 7% are from econometric methods journals, and the rest are from other general interest journals. Additional information on study characteristics and journal classifications is reported in Appendix C.

We attempt to restrict attention to the author's preferred specification. We rely on the paper's discussion to identify what constitutes the main estimate. In the absence of a discussion highlighting a single main estimate, we choose the estimate in the earliest table with the maximal set of controls. If the paper studies both RR and PBD variation or studies both covered duration and nonemployment, then we collect more than one estimate. Some papers do not include an overall elasticity, instead displaying only group-specific elasticities.[4] Be-

[3] Specifically, we use the software Publish or Perish with the following query: `duration "standard error" OR "standard errors" OR PBD OR "benefit duration" OR WBA OR "weekly benefit amount" OR "replacement rate" "unemployment insurance"`. In words, this requires the paper's text to contain the word "duration", the exact phrase "unemployment insurance", and at least one of the other phrases.

[4] For example, all statistical tests in Bennmarker et al. (2007) are gender-specific. The paper's elasticities

cause this disaggregation choice is potentially germane to publication bias, we collect each group-specific main estimate from these papers. Altogether we identify 89 estimates and standard errors. Table B.1 and Table B.2 contain the full list of included estimates and their sources for PBD elasticities and RR elasticities, respectively. We use a slightly more disaggregated dataset when testing for publication bias than we do when analyzing the systematic determinants of heterogeneity in elasticity estimates; the exact level of aggregation we use is described in Appendix C. The unadjusted mean PBD elasticity is 0.48, while the unadjusted mean RR elasticity is 0.44.

In addition to the main duration elasticities and their standard errors, we collect economic and methodological characteristics associated with each study. Section 4.1 describes these characteristics. Here, we highlight a few key patterns. In terms of economic characteristics, there is substantial variation across studies in the pre-reform level of the average replacement rate, which varies from 32% to 90% with an interquartile range of 25%. In terms of methodological characteristics, Table B.3 shows that quasi-experimental methods are more common for PBD elasticities (90%) compared to RR elasticities (47%).[5]

3 Publication Bias Evidence and Corrections

3.1 Graphical Evidence of Publication Bias

The key assumption underlying our preferred test is that a study's elasticity estimate and its standard error should be orthogonal. We argue this assumption is plausible in our setting with purely observational data. Although in principle the size of an estimate and the size of the standard error can be related because of power calculations, this concern is unlikely to apply to our meta-analysis. All UI studies in our review use observational data where sample size is fixed by data constraints (rather than explicitly designed experiments). Absent publication bias, a study's elasticity estimate and its standard error should be orthogonal;

for men are consistently positive and statistically significant while elasticities for women are consistently negative, of the same absolute magnitude, and marginally statistically significant.

[5]RR formulas typically depend on prior earnings with a minimum and maximum amount. Prior to the proliferation of regression kink designs, cross-sectional identification strategies for the RR elasticity would include parametric controls for the benefit formula's running variable.

both small and large estimated elasticities should come with small and large estimated standard errors. Alternatively, publication bias will generate a correlation between published elasticities and their standard errors by truncating the distribution of latent effects.

Figure 1 illustrates evidence of publication bias with increasing degrees of parametric structure: as a scatter plot, as a non-parametric density by standard error tercile, and finally as the mean elasticity by standard error tercile. In the absence of publication bias, because standard errors and elasticities are uncorrelated, the mean elasticities on the bottom panel should fall on the same vertical line (up to sampling variability).[6]

We find evidence of publication bias because estimates and standard errors are correlated. The dashed vertical line on the figure shows the mean unconditional elasticity. However, the figure shows that the mean elasticity is 0.27 for estimates in the bottom tercile, 0.50 in the middle tercile, and 0.56 for estimates in the top tercile. The estimates for each group are sufficiently precise that for two out of three terciles we can reject hypothesis that the group-specific means are equal to the unconditional mean.[7] This suggests that researchers and journals are indeed censoring small elasticity estimates with large standard errors.

In addition to the positive correlation, the figure also shows two other types of evidence for publication bias. First, 93% of elasticities are positive.[8] Second, Figure 1a shows that there are many studies just on the statistically significant side (below the diagonal gray line) but few just on the statistically insignificant side (above the line). Figure A.2 emphasizes this by plotting the t-statistic, or the ratio between the elasticity (x-coordinate in Figure 1a) and standard error (y-coordinate in Figure 1a). A higher-than-average density is readily apparent to the right of 1.96.

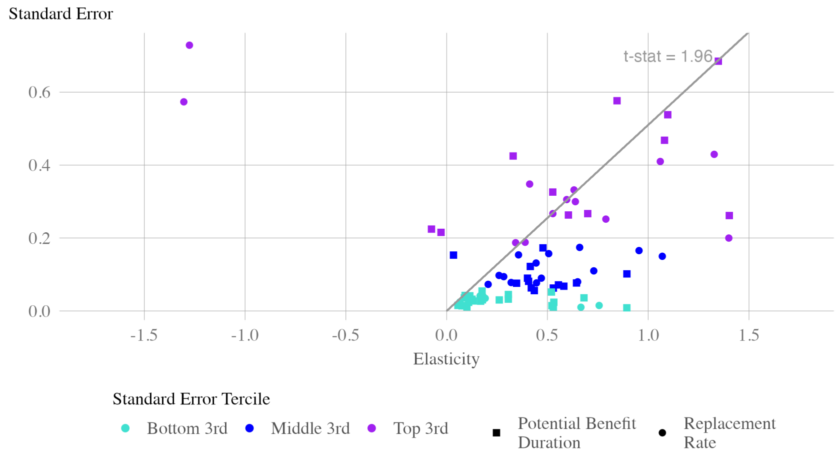
[6]The location of the line is not important; what matters is that the estimates grouped by standard error tercile should have estimates which fall on this line.

[7]If we drop the two noisy estimates with elasticity $<$ -1, then we reject the hypothesis for all three groups.

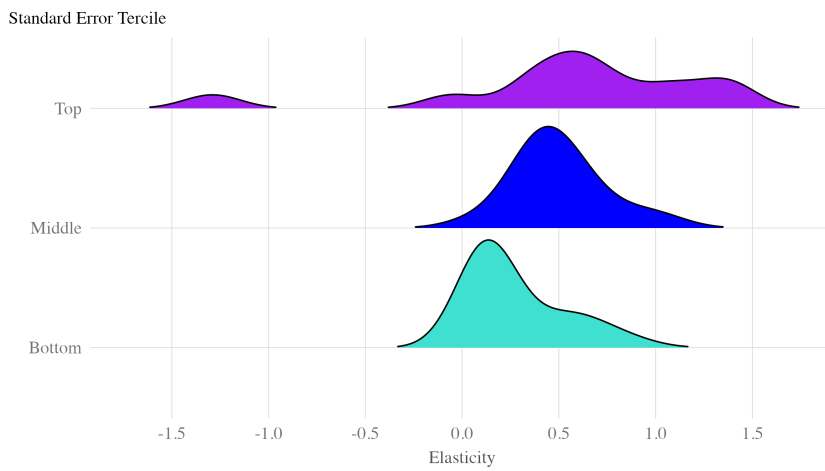
[8]This 93% estimate includes some negative estimates which are omitted from the figure for legibility. In the figure itself, 95% of the elasticities are positive.

Figure 1: Descriptive Evidence of Publication Bias

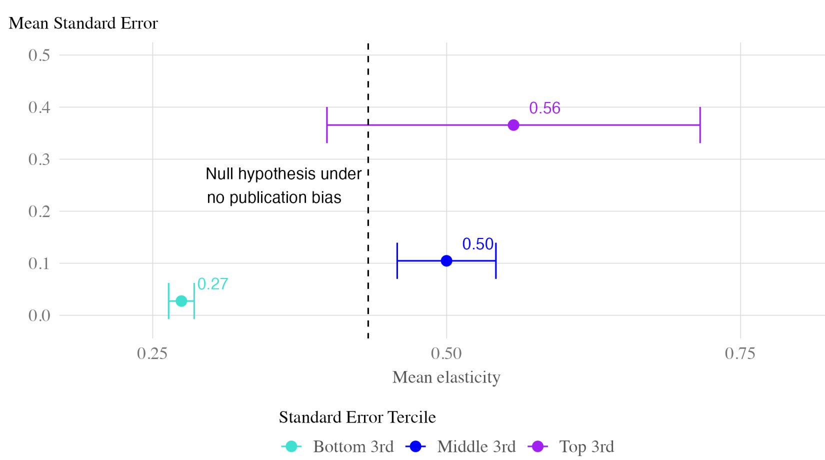
(a) Individual Estimates of Elasticities and Standard Errors



(b) Non-parametric Distribution of Elasticity Estimates



(c) Mean Elasticity Estimates



Notes: This figure describes the joint distribution of estimates of the elasticity and the standard error of unemployment duration with respect to unemployment benefit generosity. For visual clarity, six estimates with elasticities < -1.5 or > 1.5 are excluded here, but described in Appendix C. In panel (c), a 95 percent confidence interval is constructed for each tercile using the delta method.

3.2 Structural Bias Correction: Andrews and Kasy (2019)

We estimate the latent distribution of elasticities absent publication bias following Andrews and Kasy (2019).

The main idea of the approach is to jointly model the latent distribution of elasticities and the publication probability for different realizations. Latent estimates are defined as realizations of the estimated elasticity prior to the publication process. As discussed in Section 3.1, the key identification assumption for this approach is that the latent distribution of elasticities is independent of the latent distribution of standard errors of the elasticity.

Following Definition 1 in Section I of Andrews and Kasy (2019), we define the distribution of latent elasticities and their standard errors to be $(\Theta^*, \Sigma^*) \sim \mu_{\Theta,\Sigma}$. This distribution describes heterogeneity across studies and the noisiness of estimates. For a given study with distribution parameters (Θ, Σ), a noisy realization of the latent elasticity X^* is drawn: $X^* \mid \Theta^*, \Sigma^* \sim N(\Theta^*, \Sigma^{*2})$. Finally, a publication decision $D \mid X^*, \Theta^*, \Sigma^* \sim Ber(p(t^*))$ is drawn where $t^* = X^*/\Sigma^*$. We observe X, Θ, Σ if $D = 1$.

We make two functional form assumptions. First, in our baseline analysis, we assume that the latent distribution is

$$\Theta^* \sim \bar{\theta} + t(\nu) \cdot \tau$$

This is the same functional form used in the minimum wage application in Andrews and Kasy (2019). It is useful to use a t-distribution (rather than say a normal distribution) because the latent distribution of estimates has fatter tails than would be implied by a normal distribution. In the robustness analysis (Table B.4), we consider alternative distributions (including a non-parametric analysis which leaves the latent distribution unrestricted). Second, motivated by the excess mass of t-statistics just above 1.96, we assume the following publication probabilities for each estimate as a function of its t-statistic:

$$p(t) = \begin{cases} \beta_p & \text{if } t < 1.96 \\ 1 & \text{if } t \geq 1.96 \end{cases} \tag{1}$$

Normalization of the publication probability for t-statistics above 1.96 is without loss of generality, as that pins down the total number of published estimates. Table 1 shows publication probability β_p for statistically insignificant elasticities is 7-9% of the publication probability for statistically significant positive elasticities. The publication probability estimates are statistically precise; even at the upper end of the largest 95% confidence interval, the publication probability for insignificant estimates is still 26% of that for significant estimates. This extent of publication bias is in line with other literatures examined in Andrews and Kasy (2019).

Publication bias leads the naive elasticity to overstate the latent elasticity. The mean of the latent elasticity distribution ($\bar{\theta}$) is 0.10 for PBD (naive mean: 0.48) and 0.20 for RR (naive mean: 0.44). The corrected estimates for PBD are significantly different from the average naive elasticities at a 5% significance level in the sense that the standard errors for $\bar{\theta}$ are sufficiently precise that the 95% confidence interval for $\bar{\theta}$ excludes the average naive elasticity. For RR, the difference is not significant.

Table 1 shows that there is not a single latent elasticity value for either RR or PBD which can rationalize the distribution of estimates. Instead, there is substantial dispersion in the latent elasticity distribution. The dispersion parameter (similar to the standard deviation) of the latent t-distribution (τ) is 0.28 for PBD and 0.32 for RR. Put otherwise, before publication bias, 90% of latent PBD estimates will fall between -0.37 and 0.56 while 90% of the latent RR estimates will fall between -0.33 and 0.73.

Table 1: Publication Bias Corrections

	Average Published Elasticity	Publication Prob if $t < 1.96$	Latent Dist. Parameters $\Theta^* \sim \bar{\theta} + t(\nu) \cdot \tau$		
	$\bar{\epsilon}$	β_p	$\bar{\theta}$	τ	ν
Potential Benefit Duration	0.48	0.07	0.10	0.28	3.56
	(0.06)	(0.04)	(0.10)	(0.04)	(1.66)
Replacement Rate	0.44	0.09	0.20	0.32	4.95
	(0.16)	(0.08)	(0.23)	(0.12)	(2.63)

Notes: The table reports statistics on the observed and latent distribution of elasticities for potential benefit duration and replacement rate. There are 49 potential benefit duration estimates and 40 replacement rate estimates. $\bar{\epsilon}$ is the sample mean of realized elasticities. The remaining columns report estimated parameters from the Andrews and Kasy (2019) publication bias correction. β_p is the publication probabilities for t-statistics below 1.96. The latent distribution of elasticities is assumed to be distributed with $\Theta^* \sim \bar{\theta} + t(\nu) \cdot \tau$ so the mean of the latent elasticity distribution is $\bar{\theta}$ and the dispersion parameter (similar to the standard deviation) is τ. Standard errors are reported below parameter estimates. In some cases, multiple papers study the same benefit variation in the same region. We refer to these as "contexts", and allow for serial correlation between estimates within the same context.

The quantitative pattern of publication bias (β_p) is robust across specifications, but the magnitude of the mean latent elasticity ($\bar{\theta}$) is sensitive to specification. We explore the robustness of the estimates to different functions for $p(t)$, different assumptions for the distribution of latent estimates, and different samples in Table B.4. Across all specifications, the publication probability for insignificant estimates relative to significant estimates remains far below one. However, the estimates for $\bar{\theta}$ vary meaningfully. Across the PBD specifications, the estimates vary from 0.01 to 0.40. Across the RR specifications, the estimates vary from -0.09 to 0.36. In addition, estimates with lower values of $\bar{\theta}$ tend to also find more dispersion in latent estimates (higher value of τ). The GMM procedure in particular is fragile in the sense that the inclusion or exclusion of Hunt (1995) produces widely differing estimates of $\bar{\theta}$.[9]

[9] Hunt (1995) estimates an elasticity of -3.32 and standard error of 2.25. The correction procedure where the latent estimates are assumed to follow a t-distribution gives little weight to this estimate (because it is so imprecise) so its inclusion has little effect on the estimates. However, the results from the correction procedure which uses GMM are much more fragile. The GMM procedure upweights insignificant estimates by a factor of $1/\beta_p$ to estimate $\bar{\theta}$. When insignificant estimates are rare, $1/\beta_p$ will be large and the procedure

Overall, the corrections for publication bias are always substantial in economic terms, but uncertainty remains about the appropriate magnitude of the correction. In addition, across all specifications in Table 1 and Table B.4, we find a high degree of dispersion in the latent elasticity distribution. This finding motivates a search for systematic heterogeneity in elasticities in Section 4.

4 Predictors of Elasticity Heterogeneity

4.1 Motivation for Included Predictors

We look for evidence of observable characteristics that statistically explain heterogeneity in elasticities. The collected study characteristics for predicting the elasticity can be separated into two categories: economic characteristics and methodological characteristics. Economic characteristics should be interpreted as the key dimensions of heterogeneity. These are factors that policymakers can account for when prospectively setting UI benefit parameters in their own economic context. In contrast, methodological characteristics should be interpreted as auxiliary controls. We include them to account for any estimation choices that could affect the estimated elasticity. Table B.5 contains the full set of variables.

Economic characteristics: Meta-analysis's comparative advantage when applied to UI benefits is in characterizing heterogeneity by baseline benefit parameters. When including the baseline PBD (in weeks) and baseline RR (fraction) as predictors, we additionally interact them with the policy margin. This is because a given increase in benefit PBD increases total benefit entitlement more with a higher RR and vice versa. One other economic dimension of theoretical interest is whether the benefit variation affects many people or only a limited set of people. We discuss the interpretation of this covariate in more detail in Section 6. Finally, we include several other economic characteristics in the analysis whose construction is discussed in Table B.5 and ex post prove to be unimportant for predicting heterogeneity.

Methodological characteristics: Most of the methodological characteristics are proxies for study quality: data source, research design, and publication bias susceptibility. First,

will dramatically upweight a small number of estimates.

administrative data is less prone to measurement error. Second, quasi-experimental research designs—particularly regression discontinuity designs—are less prone to selection bias due to policy endogeneity. Third, noisier estimates are more prone to publication bias. Following the meta-regression literature, we include the standard error as a control (Stanley, 2008).

Other methodological characteristics are less clearly related to study quality but are included to increase comparability across estimates. For example, it is unclear whether elasticities derived from hazard models are systematically different from those derived from duration regression. Additionally, elasticities with the outcome as either total nonemployment duration or covered unemployment are related but conceptually distinct estimands.

4.2 Documenting Heterogeneity Using Bayesian Model Averaging

When projecting elasticities onto study characteristics, the power for detecting the importance of a given predictor diminishes as the number of predictors increases. In particular, as more predictors are added, the probability that a given variable is highly correlated with another will increase. While the augmented model will better predict the final outcome, it becomes more difficult to interpret which factors drove the prediction.

Following recent developments in the meta-analysis literature, we use Bayesian model averaging (BMA) to highlight heterogeneity correlates (Havranek et al., 2022; Gechert et al., 2022; Zigraiova et al., 2021; Bajzik et al., 2020). The frequentist analogue to BMA is as follows: run separate regressions predicting the elasticity using all possible subsets of the study characteristics

$$\epsilon = \alpha + \beta X \tag{2}$$

and average coefficients across regressions. BMA instead imposes priors over included

Table 2: Study Characteristics Correlated with Elasticities

	Posterior Inclusion Probability	Posterior Mean
(Intercept)	1.000	0.203
Baseline benefits		
Baseline RR (fraction) x RR estimate	0.578	0.366
Baseline RR (fraction) x PBD estimate	0.178	-0.051
Baseline PBD (weeks) x RR estimate	0.398	0.001
Baseline PBD (weeks) x PBD estimate	1.000	0.007
Policy variation		
PBD estimate (vs. RR estimate)	0.337	-0.112
Aggregate variation	0.106	0.000
Study context		
Sample year (2023 = 0)	0.109	0.000
Relative unemployment (pp)	0.114	-0.001
Labor tax wedge (pp)	0.133	0.000
United States dummy	0.260	-0.040
Data and estimation		
Administrative data	0.115	0.002
Nonemployment as outcome	0.772	-0.159
Hazard model	0.114	-0.002
DID or RKD	0.221	0.022
RDD	0.137	0.009
Standard errors		
SE	1.000	1.488
Journal		
Impact Factor (z-score)	0.253	0.014

Notes: This table contains model output from Bayesian model averaging to predict elasticity estimates with study characteristics. The first column is the posterior inclusion probability and the second column is the posterior mean. Variables are described in more detail in the notes to Table B.5.

covariates and their coefficients, iteratively updates those priors based on the likelihood, and outputs Bayesian analogs to the regression coefficient (posterior mean) and p-value (posterior inclusion probability). We use priors common in the BMA literature and estimate the model using the `bms` package in R (Eicher et al., 2011; Zeugner and Feldkircher, 2015).[10] In our initial analysis we found that the results were incredibly sensitive to one study (Hunt, 1995) which estimates a RR elasticity of -3.32 and standard error of 2.25; we therefore omit that study in what follows.

We find strong evidence of systematic heterogeneity in the estimated elasticity. The model's posterior distribution for the variance of the elasticity based on observable study characteristics is 56%, meaning that about half of the variation in elasticities reflects systematic heterogeneity. Table 2 examines the relationship of each specific covariate with the predicted elasticity. The posterior inclusion probability in column 1 captures a Bayesian notion of statistical significance; a covariate based on random noise generally has a posterior inclusion probability of between 10 and 15%, so six of the covariates appear to have meaningful predictive power as measured by their posterior inclusion probability (defined as posterior inclusion probability > 30%). Column 2 captures the coefficient of the relationship in the linear model via the posterior mean.

We find strong evidence that the elasticity is higher when the replacement rate is higher and when potential benefit duration is longer. Among studies with replacement variation, a higher baseline replacement rate is associated with a posterior mean of 0.366. This means that the predicted elasticity is 0.366 higher when the baseline replacement rate is 100% than when it is 0%. We return to this particular finding below in Section 5. Along similar lines, we find that when baseline potential benefit duration is 50 weeks longer the elasticity is 0.05 larger for changes in the replacement rate and 0.35 larger for changes in number of potential weeks of benefits.[11]

The two key methodological characteristics are standard errors and the definition of

[10] In particular, we use Zellner's g-prior for the regression coefficients and a uniform model prior.

[11] The finding that the elasticity is larger when the baseline number of PBD weeks is larger is in part dependent on a single outlier context which studies a reform in Norway where the control group had PBD of 186 weeks (as compared to 100 weeks or less in all other studies) and the estimated elasticity is very high. Røed and Zhang (2003) estimate an elasticity of 1.85 w.r.t PBD and Røed and Westlie (2012a) estimate an elasticity for men of 0.95 w.r.t. RR. If we re-run BMA dropping these studies we get a posterior mean of 0.001, meaning that when baseline PBD is 50 weeks longer then the elasticity is 0.05 larger.

unemployment duration. First, a study with a standard error of 1 yields an elasticity estimate 1.5 higher than a study with a standard error of 0. This is consistent with substantial censoring of insignificant elasticity estimates. Second, we find that measuring unemployment duration as total nonemployment delivers a slightly lower elasticity. This pattern holds even within studies, which implicitly controls for all contextual characteristics. This finding is also in line with a recent study by Bell et al. (2024) which is not included in this meta-analysis because it appeared after we completed the meta-analysis.

Perhaps the most valuable output from the meta-analysis is aggregating studies across different contexts to estimate the elasticity for a much-studied policy context: state-financed UI benefits in the US. State-financed UI benefits are those benefits which are available outside of recessions, have an average replacement rate of 43.5% (U.S. Department of Labor, 2024), and in most states last 26 weeks. The meta-analysis predicts that the covered duration elasticity to the replacement rate in this context is 0.36.[12] This method accounts for publication bias by setting the standard error equal to zero in the prediction equation. Applying the same method to calculate the covered duration elasticity with respect to changes in PBD delivers an elasticity of 0.22.

The meta-analysis also implies that elasticities in a European context are an order of magnitude greater than in the US. To aggregate across a wide variety of policy environments, we compute a weighted median across OECD countries. We find a replacement rate of 62% and potential benefit duration of 80 weeks. Applying the method above predicts covered duration elasticity to replacement rate of 0.55. With respect to PBD, the predicted elasticity is 0.67.

[12] Specifically, allowing here for small rounding errors, the variables used are: 0.203 (intercept) + 0 (Aggregate variation) - 0.04 (United States dummy) - 0 * 28.4 (Labor tax wedge) + 0.002 (Administrative data) + 0.009 (RDD) + 0.001*26 (Baseline PBD) - 0.366*0.435 (Baseline RR) $\approx$ 0.36.

5 Application: Sufficient Statistics for Optimal UI Benefits

Sufficient statistics formulas have grown enormously in popularity in recent years in public economics (Kleven, 2018). We illustrate how meta-analysis can be used to solve two well-known implementation challenges in the context of the optimal replacement rate for benefits in the Baily (1978) - Chetty (2006) formula.

First, researchers are often interested not only in calculating the welfare gain from a small change in policy, but also studying whether a large change is optimal. Estimating a duration elasticity requires variation in benefits, but such an estimate typically comes from a single policy discontinuity or policy reform that in turn has a single baseline PBD or RR. When drawing on a single elasticity estimate, the sufficient statistics approach is not able to make predictions about the effects of large changes in policy without resorting to structural assumptions. For example, Kleven (2021) shows how applying a sufficient statistics approach which holds constant the elasticity of labor supply to analyze a large change in tax policy implicitly assumes an isoelastic cost function for labor supply. This may be one reason why researchers primarily use structural approaches rather than sufficient statistics approaches to analyze large changes in policy.

Second, most sufficient statistics formulas require multiple causal effects, but multiple natural experiments rarely occur in exactly the same policy context. For example, the most precise and highest-quality elasticity estimates often come from European contexts but Gruber (1997)'s consumption-smoothing estimates are for the US. Researchers thus face a trade-off between using higher-quality estimates and using estimates from the same policy environment.

Meta-analysis provides a solution to both challenges in the context of the sufficient statistics formula for unemployment benefits. First, because the replacement rate varies widely across studies, we can predict how the elasticity varies with the replacement rate. In our application, we assume this relationship is linear. If the meta-analysis has correctly specified the sources of heterogeneity in study elasticities, it enables researchers to substitute an assumption about how reduced-form elasticities vary with the policy environment (informed by data) in lieu of specifying a structural cost function. In this respect, our analysis builds on other recent work which substitutes reduced-form elasticities for structural assumptions (Chiang and Zoch, 2022; Auclert et al., 2024).

It is useful to emphasize that this argument still relies on the envelope theorem and is therefore unable to use sufficient statistics to quantify the welfare effects from a large policy change. Rather, we are using studies from a wide range of policy environments—in each of which we maintain the envelope theorem assumption that agents are choosing optimally—to answer the question "what are the welfare gains from a marginal change to policy in this environment?"

Second, because the method in Section 4 allows for systematic heterogeneity in the elasticity, we can predict the elasticity in the US context while drawing on all the prior high-quality estimates from other contexts (and allowing the estimates to differ between the contexts based on observable differences in policy across contexts).

The Baily-Chetty formula equates the consumption-smoothing gain from an additional dollar of UI benefits with welfare loss via fiscal externality from an additional dollar of UI benefits:

$$\gamma \frac{\Delta c(r)}{c} = \frac{\epsilon_{1-e,r}}{e}. \tag{3}$$

The right-hand side of the formula is the RR elasticity $\epsilon_{1-e,r}$ which is estimated for the US policy context (non-recession, covered benefit durations, purged of publication bias) using the Bayesian model averaging framework from the previous section divided by the employment rate e for which we assume 0.95. The posterior mean estimates in Table 2 imply that the elasticity ranges from 0.18 to 0.48 as the replacement rate varies from 0 to 84%.

To estimate the left-hand side we assume a coefficient of relative risk aversion of $\gamma = 2$ and we use estimates of the consumption drop $\frac{\Delta c(r)}{c}$ from Gruber (1997). The Gruber estimates are particularly useful for our purposes because they compute the consumption smoothing gain for every possible replacement rate.[13] Gruber estimates that $\frac{\Delta c(r)}{c} = 0.222 - 0.265r$ which means that if the replacement rate is 0 then consumption will fall by 22.2% and a

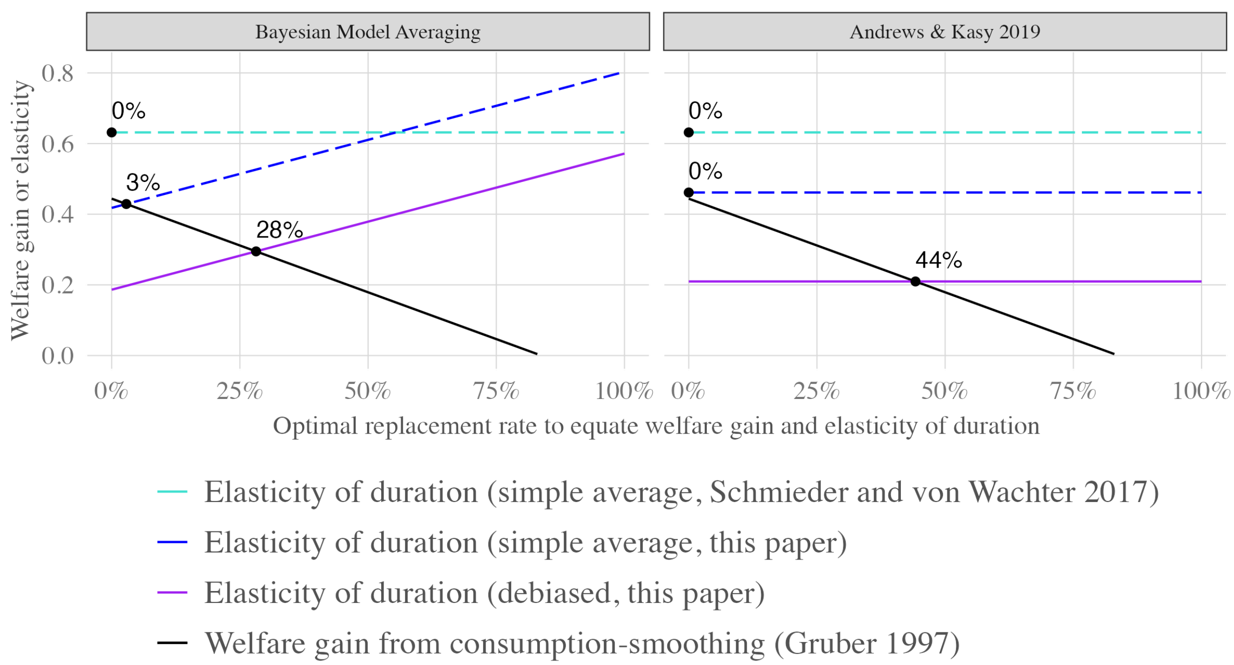
[13] There are several methods for estimating workers' willingness to pay for UI benefits (Landais and Spinnewijn 2021, Chetty 2008, Hendren 2017). However, they provide local estimates of the welfare gains from changing UI benefits and therefore cannot be used for estimating an optimal replacement rate.

replacement rate of $(0.222/0.265) = 84\%$ is sufficient to prevent any consumption drop during unemployment. We can therefore solve for the optimal replacement rate as

$$r = \frac{1}{0.265}(-0.222 + \frac{\epsilon_{1-e,r}}{e\gamma})$$

Figure 2 plots the welfare gain from increasing the replacement rate as a function of the existing replacement rate.

Figure 2: Optimal replacement rates for UI benefits



Notes: This figure shows the optimal replacement rates implied by the Baily-Chetty formula. We assume constant relative risk aversion with parameter $\gamma = 2$, a consumption drop during unemployment $\frac{\Delta c(r)}{c}$ from Gruber (1997), and present a range of estimates for $\epsilon_{1-e,r}$. The left panel predicts the elasticity using the Bayesian model averaging (BMA) approach from Section 4 with parameter values from footnote 12 and the right panel uses the Andrews-Kasy results from Section 3. The BMA predictions slope upward because we find in Table 2 that the elasticity is higher when the replacement rate is higher. The dashed horizontal light green line shows the mean nonemployment duration elasticity from Table 1 of Schmieder and von Wachter (2017). The dashed blue line shows the duration elasticity for each value of r without correcting for publication bias; in BMA, this is constructed by using the parameter values from footnote 12 plus multiplying the average standard error of 0.145 by the posterior mean of 1.488.

The optimal replacement rate solution of 28% is shown graphically in the left panel of Figure 2. In the right panel of the figure, we use the de-biased AK elasticity estimate for RR of 0.20 from Table 1. This estimate has none of the context-specific adjustments enumerated above but does use a more sophisticated methodology to account for publication bias.

Had we instead drawn on existing literature reviews or not accounted for publication bias, we would have found an optimal replacement rate close to zero. A recent literature review finds an RR elasticity of 0.59 (Schmieder and von Wachter, 2017) which in turn implies a an optimal replacement rate of 0%, which is to say that it is optimal not to have any UI benefits at all. The same conclusion holds if we use the naive average of the elasticities in our literature review. It also holds if "re"-bias our BMA estimate for the US by multiplying the posterior mean for the standard error times the average standard error. Correcting for publication bias is the crucial force which leads us to find higher optimal replacement rates.

6 Application: Micro vs Macro Elasticity

One recent area of active research is the general equilibrium effects of unemployment benefits. Interest in this topic surged in the wake of the Great Recession, when potential benefit durations were extended to be much longer than normal times. Researchers have typically summarized general equilibrium effects by comparing the causal effect of increasing benefits for a single worker (the "micro" effect) to the causal effect of increasing benefits for all workers (the "macro" effect).

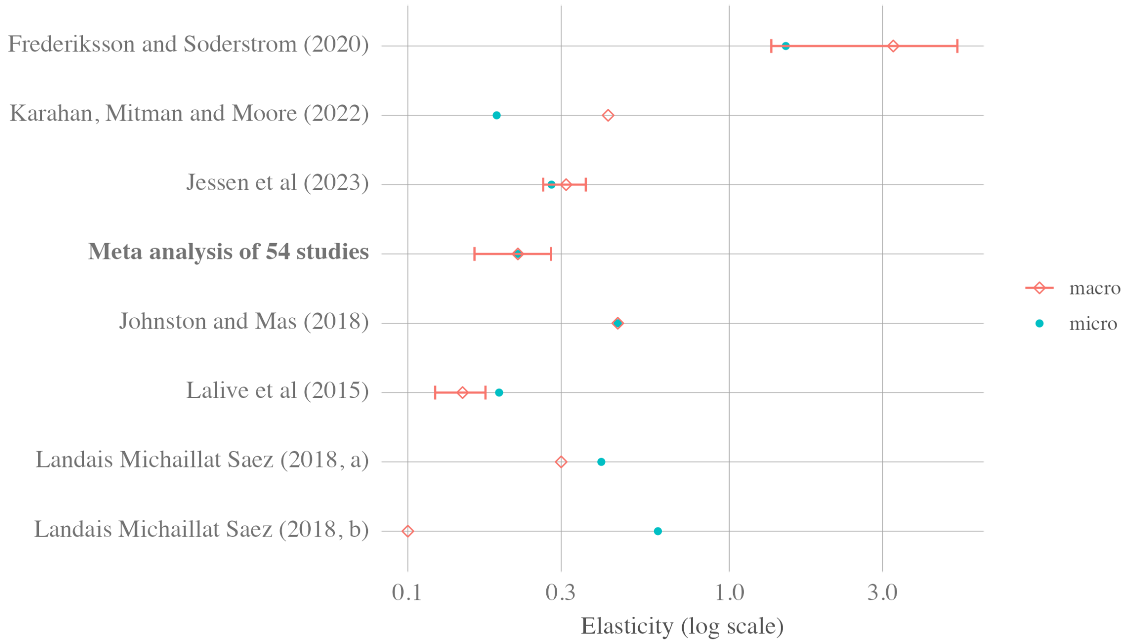
Different models of the labor market make different sign predictions about the general equilibrium effects of benefits. One line of research has studied the effects of benefits in a Diamond-Mortensen-Pissarides model where UI reduces search effort by jobseekers, leading employers to post fewer vacancies (Hagedorn et al., 2016). In such a model, the vacancy posting response amplifies the macro elasticity relative to the micro elasticity. A second line of research has studied the effects of benefits in a model with labor market congestion (Michaillat 2012 and Landais et al. 2018b). In such a model, UI reduces search effort for some unemployed jobseekers, but this is offset by a higher arrival rate of offers for other jobseekers, which Landais et al. call a "rat race" effect. In such a model, congestion mutes the macro elasticity relative to the micro elasticity.

Empirically measuring the difference between the micro elasticity and the macro elasticity has proved to be quite challenging. In order to estimate this pair of causal objects one needs a setting with "double randomization" (Lalive et al., 2015) where there is one experiment where some *workers* are treated with higher benefits (compared to control workers in the same labor market) and a second experiment where some *labor markets* are treated with higher benefits (compared to other control labor markets). To the best of our knowledge no such experiment has yet occurred; the closest available empirical counterpart comes from Jessen et al. (2023) studying a labor market policy in Poland.

Overall, substantial uncertainty remains about whether the micro elasticity is smaller than, larger than, or the same as the macro elasticity. This is because quasi-experimental variation to estimate both the micro and the macro elasticity is quite rare. Figure 3 summarizes the studies we are aware of which report both a micro and macro elasticity. Sufficient variation to also construct a confidence interval for the difference between the two elasticities is even rarer. The only study which finds a significant difference is (Lalive et al., 2015) and that study is limited to macro effects for workers over age 46.

We use meta-analysis to develop an alternative methodology for comparing the micro and the macro elasticity. Prior macro estimates rely on measuring the response of state- or region-level unemployment to a state- or region-level policy change (*e.g.* Chodorow-Reich et al. 2018 and Acosta et al. 2023). We adopt a different approach which is rooted in heterogeneity across microeconomic studies. Within our meta-analysis of 55 studies, we identified 14 studies where the treatment constitutes a change in both the worker's own benefits and the benefits of other workers participating in the same labor market. This type of variation arises in three ways in our meta-analysis: (i) when policy changes for all workers, and the study compares a cohort of treated workers to a prior cohort of control workers, (ii) when policy changes for all long-term unemployed workers, and (iii) when policy changes for all workers with the same level of pre-separation earnings. The remaining studies rely on variation that only affects a small share of workers who participate in the same labor market as the control workers.

Figure 3: Micro and Macro Unemployment Duration Elasticities



Notes: This figure compares several prior papers' estimates of the micro and macro elasticity of unemployment duration with respect to unemployment benefit generosity to our estimate. Studies are ordered by the difference between their macro and micro elasticities. The horizontal lines represent 95% confidence intervals for the *difference* between the micro and macro elasticities. Such standard errors are only available for Fredriksson and Söderström (2020), Jessen et al. (2023), Lalive et al. (2015), and our paper. We use suffixes "a" and "b" to denote what Landais et al. (2018a) calls their "lower bound" and "upper bound" estimates.

Using microeconometric studies to capture the difference between the micro and macro elasticity has distinct strengths and weaknesses relative to the existing approach in the literature. One strength of our approach is that it enables us to systematically draw on evidence from many different policy contexts rather than relying on case studies. A related strength is the ability to construct a confidence interval; this is impossible to construct in comparisons that rely on a single case study and can be quite imprecise when constructed based on estimates from a small number of regions. A limitation of the microeconomic approach is that it compares differences in outcomes among workers who are *already* unemployed. It is therefore silent on the question about how differences in UI affect the separation margin.[14]

We find no systematic evidence of differences between micro and macro elasticities using

[14]There is now a burgeoning literature on this topic. See Winter-Ebmer (2199) for one early example of such research and Jessen et al. (2023) for one recent example.

cross-study heterogeneity in the share of workers affected. More precisely, the "aggregate variation" row in Table 2 shows that the studies where a substantial share of workers are affected have an elasticity which is within 0.001 of the studies where a small share of workers are affected. The posterior inclusion probability is 0.107, which means the covariate is no different from random noise. Figure 3 shows a 95% confidence interval for macro PBD elasticity for the US that ranges from 0.16 to 0.28, which is substantially more precise than many of the prior estimates. This estimate and standard error together suggest that either these general equilibrium channels have limited importance or perhaps that they cancel each other out.

7 Conclusion

We comprehensively survey the literature and confirm prior reviews' finding that unemployment insurance benefit expansions increase unemployment duration. However, prior literature reviews take a naive average of existing estimates. The subset of published estimates tend to be larger than the (unobserved) latent elasticities. Publication bias inflates this naive average relative to the true average latent elasticity. After accounting for publication bias and study characteristics, a typical RR elasticity in terms of covered unemployment is 0.36, and a typical PBD elasticity in terms of covered unemployment duration is 0.22. Incorporating our estimates into an otherwise standard formula for the optimal replacement rate moves the optimal rate from close to zero to 28%. In doing so, we illustrate that meta-analysis can be used to adapt sufficient statistics for a wide variety of policy environments.

References

Acosta, Miguel, Andreas I. Mueller, Emi Nakamura, and Jón Steinsson. 2023. "Macroeconomic Effects of UI Extensions at Short and Long Durations." nber working paper 31784.

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review* 109 (8): 2766–2794. 10.1257/aer.20180310.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "The Intertemporal Keynesian Cross." *NBER Working Paper 25020*. 10.3386/w25020.

Baily, Martin Neil. 1978. "Some aspects of optimal unemployment insurance." *Journal of Public Economics* 10 379–402.

Bajzik, Josef, Tomas Havranek, Zuzana Irsova, and Jiri Schwarz. 2020. "Estimating the Armington elasticity: The importance of study design and publication bias." *Journal of International Economics* 127 103383. 10.1016/j.jinteco.2020.103383.

Bell, Alex, TJ Hedin, Geoffrey C. Schnorr, and Till M. von Wachter. 2024. "UI Benefit Generosity and Labor Supply from 2002-2020." working paper.

Bennmarker, Helge, Kenneth Carling, and Bertil Holmlund. 2007. "Do Benefit Hikes Damage Job Finding? Evidence from Swedish Unemployment Insurance Reforms." *LABOUR* 21 (1): 85–120, Publisher: CEIS.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83 (6): 2453–2483.

Carling, Kenneth, Bertil Holmlund, and Altin Vejsiu. 2001. "Do Benefit Cuts Boost Job Finding? Swedish Evidence from the 1990s." *The Economic Journal* 111 (474): 766–790. 10.1111/1468-0297.00659.

Chetty, Raj. 2006. "A General Formula for the Optimal Level of Social Insurance." *Journal of Public Economics* 1879–1901.

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy* 116 173–234.

Chiang, Yu-Ting, and Piotr Zoch. 2022. "Financial Intermediation and Aggregate Demand: A Sufficient Statistics Approach." *St. Louis Fed Working Paper 2022-038B*. 10.20955/wp.2022.038.

Chodorow-Reich, Gabriel, John Coglianese, and Loukas Karabarbounis. 2018. "The Macro Effects of Unemployment Benefit Extensions: a Measurement Error Approach*." *The Quarterly Journal of Economics* 134 (1): 227–279. 10.1093/qje/qjy018.

Eicher, Theo S., Chris Papageorgiou, and Adrian E. Raftery. 2011. "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* 26 (1): 30–55. 10.1002/jae.1112.

Fredriksson, Peter, and Martin Söderström. 2020. "The equilibrium impact of unemployment insurance on unemployment: Evidence from a non-linear policy rule." *Journal of Public Economics* 187 104199.

Gechert, Sebastian, Tomas Havranek, Zuzana Irsova, and Dominika Kolcunova. 2022. "Measuring capital-labor substitution: The importance of method choices and publication bias." *Review of Economic Dynamics* 45 55–82. 10.1016/j.red.2021.05.003.

Gerard, François, and Gustavo Gonzaga. 2021. "Informal Labor and the Efficiency Cost of Social Programs: Evidence from Unemployment Insurance in Brazil." *American Economic Journal: Economic Policy* 13 (3): 167–206. 10.1257/pol.20180072.

Gruber, Jonathan. 1997. "The Consumption Smoothing Benefits of Unemployment Insurance." *The American Economic Review* 87 (1): 192–205.

Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman. 2016. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects." Working Paper 19499, National Bureau of Economic Research.

Havranek, Tomas, Zuzana Irsova, Lubica Laslopova, and Olesia Zeynalova. 2022. "Publication and Attenuation Biases in Measuring Skill Substitution." *The Review of Economics and Statistics*. 10.1162/rest_a_01227.

Havránek, Tomáš, T. D. Stanley, Hristos Doucouliagos et al. 2020. "Reporting Guidelines for Meta-Analysis in Economics." *Journal of Economic Surveys* 34 (3): 469–475. 10.1111/joes.12363.

Hendren, Nathaniel. 2017. "Knowledge of Future Job Loss and Implications for Unemployment Insurance." *American Economic Review* 107 1778–1823.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *The Quarterly Journal of Economics* 135 1209–1318.

Hunt, Jennifer. 1995. "The Effect of Unemployment Compensation on Unemployment Duration in Germany." *Journal of Labor Economics* 13 (1): 88–120. 10.1086/298369, Publisher: The University of Chicago Press.

Irsova, Zuzana, Hristos Doucouliagos, Tomas Havranek, and TD Stanley. 2023. "Meta-analysis of social science research: A practitioner's guide." *Journal of Economic Surveys*.

Jessen, Jonas, Robin Jessen, Ewa Gałecka-Burdziak, Marek Góra, and Jochen Kluve. 2023. "The Micro and Macro Effects of Changes in the Potential Benefit Duration." `https://www.iza.org/publications/dp/15978/the-micro-and-macro-effects-of-changes-in-the-potential-benefit-duration`.

Kleven, Henrik. 2018. "Language Trends in Public Economics." https://www.henrikkleven.com/uploads/3/7/3/1/37310663/languagetrends_slides_kleven.pdf.

Kleven, Henrik J. 2021. "Sufficient statistics revisited." *Annual Review of Economics* 13 515–538.

Kolsrud, Jonas, Camille Landais, Peter Nilsson, and Johannes Spinnewijn. 2018. "The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden." *American Economic Review* 108 (4-5): 985–1033. 10.1257/aer.20160816.

Krueger, Alan B., and Bruce D. Meyer. 2002. "Chapter 33 Labor supply effects of social insurance." In *Handbook of Public Economics*, Volume 4. 2327–2392, Elsevier, . 10.1016/S1573-4420(02)80012-X.

Krusell, Per, Toshihiko Mukoyama, and Ayşegül Şahin. 2010. "Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations." *The Review of Economic Studies* 77 1477–1507.

Lalive, Rafael. 2007. "Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach." *American Economic Review* 97 (2): 108–112. 10.1257/aer.97.2.108.

Lalive, Rafael, Camille Landais, and Josef Zweimüller. 2015. "Market Externalities of Large Unemployment Insurance Extension Programs." *American Economic Review* 105 (12): 3564–3596. 10.1257/aer.20131273.

Landais, Camille, Pascal Michaillat, and Emmanuel Saez. 2018a. "A macroeconomic approach to optimal unemployment insurance: Applications." *American Economic Journal: Economic Policy* 10 (2): 182–216.

Landais, Camille, Pascal Michaillat, and Emmanuel Saez. 2018b. "A macroeconomic approach to optimal unemployment insurance: Theory." *American Economic Journal: Economic Policy* 10 (2): 152–81.

Landais, Camille, and Johannes Spinnewijn. 2021. "The Value of Unemployment Insurance." *The Review of Economic Studies* 88 3041–3085.

Lee, David S., Pauline Leung, Christopher J. O'Leary, Zhuan Pei, and Simon Quach. 2021. "Are Sufficient Statistics Necessary? Nonparametric Measurement of Deadweight Loss from Unemployment Insurance." *Journal of Labor Economics* 39 (S2): S455–S506. 10.1086/711594, Publisher: The University of Chicago Press.

Lopes, Marta C. 2022. "A review on the elasticity of unemployment duration to the potential duration of unemployment benefits." *Journal of Economic Surveys* 36 (4): 1212–1224. 10.1111/joes.12479.

Meyer, Bruce D. 2002. "Unemployment and Workers' Compensation Programmes: Rationale, Design, Labour Supply and Income Support." *Fiscal Studies* 23 (1): 1–49, Publisher: Wiley.

Michaillat, Pascal. 2012. "Do matching frictions explain unemployment? Not in bad times." *American Economic Review* 102 (4): 1721–1750.

Moffitt, Robert. 1985. "Unemployment insurance and the distribution of unemployment spells." *Journal of Econometrics* 28 (1): 85–101. 10.1016/0304-4076(85)90068-5.

Rebollo-Sanz, Yolanda F., and Núria Rodríguez-Planas. 2020. "When the Going Gets Tough... Financial Incentives, Duration of Unemployment, and Job-Match Quality." *Journal of Human Resources* 55 (1): 119–163. 10.3368/jhr.55.1.1015.7420R2, Publisher: University of Wisconsin Press.

Røed, Knut, Peter Jensen, and Anna Thoursie. 2008. "Unemployment duration and unemployment insurance: a comparative analysis based on Scandinavian micro data." *Oxford Economic Papers* 60 (2): 254–274. 10.1093/oep/gpm021.

Røed, Knut, and Lars Westlie. 2012a. "Unemployment Insurance in Welfare States: The Impacts of Soft Duration Constraints." *Journal of the European Economic Association* 10 518–554.

Røed, Knut, and Lars Westlie. 2012b. "Unemployment Insurance in Welfare States: The Impacts of Soft Duration Constraints." *Journal of the European Economic Association* 10 (3): 518–554. 10.1111/j.1542-4774.2011.01064.x.

Røed, Knut, and Tao Zhang. 2003. "Does Unemployment Compensation Affect Unemployment Duration?" *The Economic Journal* 113 190–206.

Sahin, Hasan, and A. Burca Kizilirmak. 2007. "Determinants of duration of unemployment insurance benefits in Turkey." *Applied Economics Letters* 14 (8): 611–615.

Schmieder, Johannes F., and Till von Wachter. 2016. "The Effects of Unemployment Insurance Benefits: New Evidence and Interpretation." *Annual Review of Economics* 8 (1): 547–581. 10.1146/annurev-economics-080614-115758.

Schmieder, Johannes F., and Till von Wachter. 2017. "A Context-Robust Measure of the Disincentive Cost of Unemployment Insurance." *American Economic Review* 107 (5): 343–348. 10.1257/aer.p20171026.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender. 2012. "The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years *." *The Quarterly Journal of Economics* 127 (2): 701–752. 10.1093/qje/qjs010.

Setty, Ofer, and Yaniv Yedid-Levi. 2021. "On the Provision of Unemployment Insurance when Workers are Ex-Ante Heterogeneous." *Journal of the European Economic Association* 19 664–706.

Stanley, Tom D. 2008. "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection." *Oxford Bulletin of Economics and statistics* 70 (1): 103–127.

U.S. Department of Labor. 2024. "UI Replacement Rates Report." `https://oui.doleta.gov/unemploy/ui_replacement_rates.asp`.

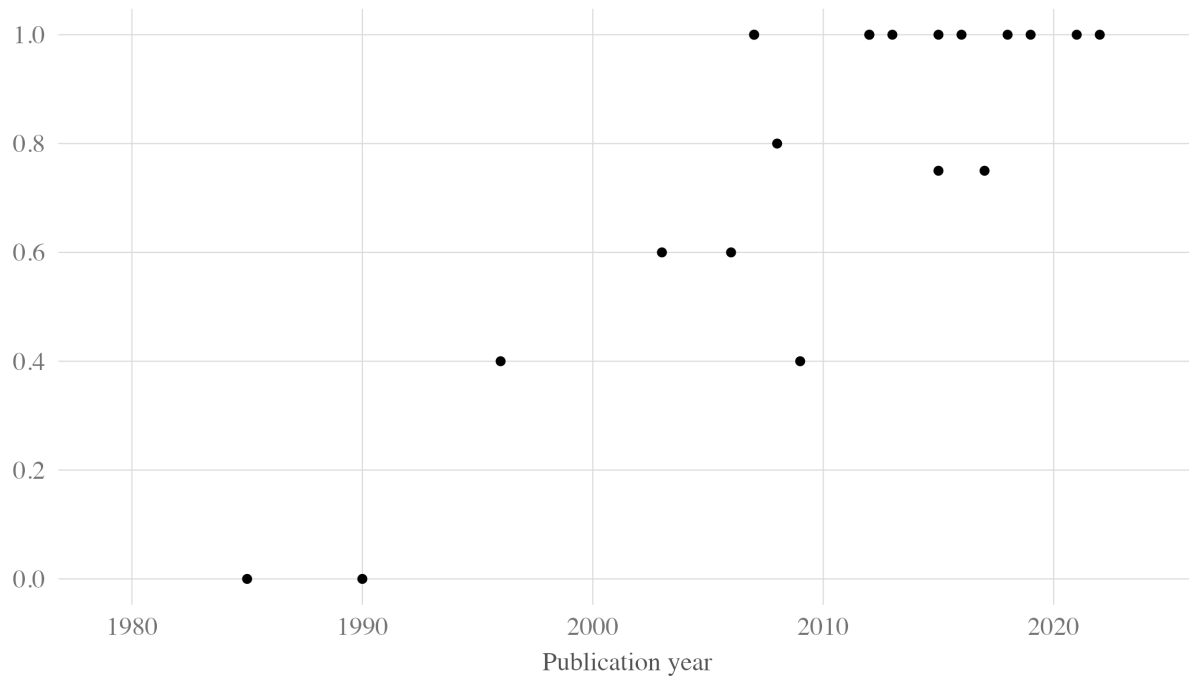
Winter-Ebmer, R., "Benefit duration and unemployment entry: A quasi- experiment in Austria." *European Economic Review* 47 259–273.

Zeugner, Stefan, and Martin Feldkircher. 2015. "Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R." *Journal of Statistical Software* 68 1–37. 10.18637/jss.v068.i04.

Zigraiova, Diana, Tomas Havranek, Zuzana Irsova, and Jiri Novak. 2021. "How puzzling is the forward premium puzzle? A meta-analysis." *European Economic Review* 134 103714. 10.1016/j.euroecorev.2021.103714.

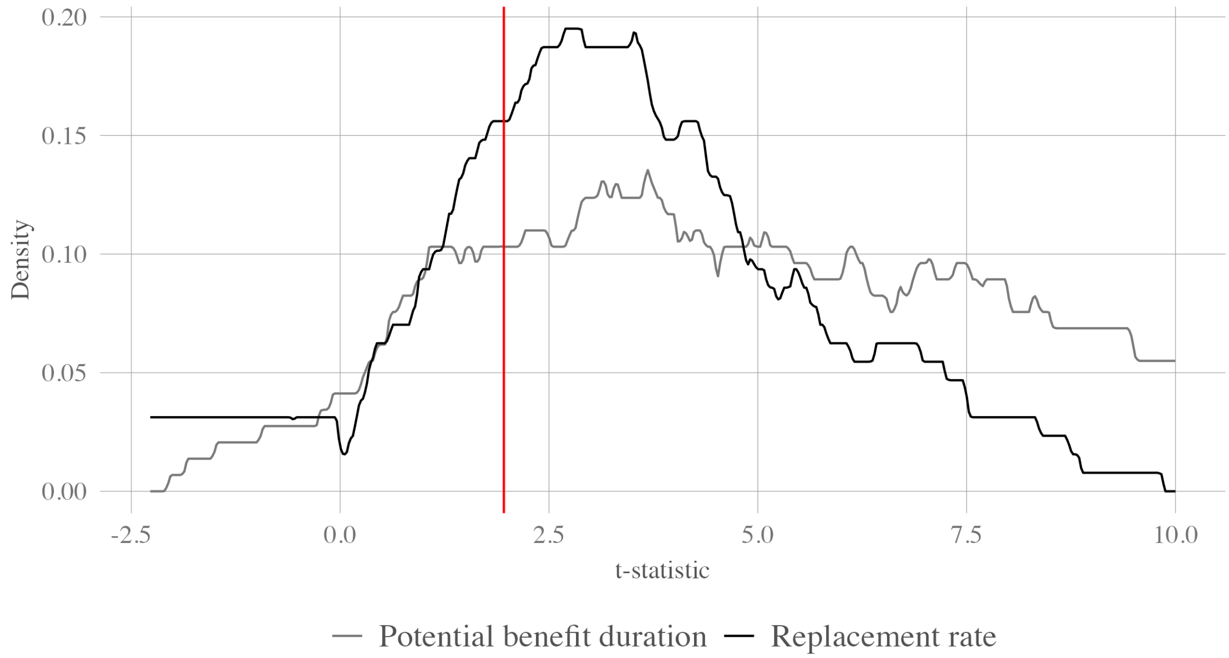
Appendix A Online Appendix – Supplemental Figures

Figure A.1: Quasi-Experimental Share of Studies by Publication Year



Notes: The figure is a binned scatterplot of the conditional mean by ventiles of year. Quasi-experimental studies are defined as those that identify the elasticity using a difference-in-differences design, regression discontinuity design, or regression kink design. The only other type of study in the sample is cross-sectional variation that uses a selection on observables assumption for identification.

Figure A.2: t-statistic bunching by UI policy margin



Notes: This plot shows the density of t-statistics, which are the ratios of the elasticity to its standard error. They include all main estimates. For visual clarity, they exclude t-statistics above 10 from Gerard and Gonzaga (2021); Rebollo-Sanz and Rodríguez-Planas (2020); Schmieder and von Wachter (2016); Schmieder et al. (2012); Røed and Westlie (2012b); Lalive (2007); Moffitt (1985).

Appendix B Online Appendix – Supplemental Tables

Table B.1: Included Studies: Disemployment Elasticities with Respect to Potential Benefit Duration

Paper	**Policy**	**Design**	**UE measure**	**Elasticity**	**SE**	**Source**
Addison & Portugal ’08 EL	PBD	DID	Total	1.08	0.47	Table 2
Caliendo & Tatsiramos ’13 JAE	PBD	RDD	Total	0.85	0.58	Text for elasticities (pg 624) and Table V for SEs
Caliendo & Tatsiramos ’13 JAE	PBD	RDD	Total	0.60	0.26	Text for elasticities (pg 624) and Table V for SEs
Card & Levine ’00 JPubEc	PBD	DID	Covered	0.31	0.05	Table 6 Column 2 Row 2
Card et al ’07 QJE	PBD	RDD	Total	0.18	0.03	Table II Column 3 Row 2
Centeno & Novo ’09 PEJ	PBD	RDD	Total	0.44	0.06	Figure 1, Table 2 Column 2
Centeno & Novo ’09 PEJ	PBD	RDD	Total	0.40	0.09	Figure 1, Table 2 Column 4
De Groot et al ’19 Labour Econ.	PBD	DID	Total	0.41	0.08	Text (pg 207) and Table 2
Fackler et al ’19 Labour Econ.	PBD	RDD	Total	-0.03	0.22	Table 1 Column 2
Fackler et al ’19 Labour Econ.	PBD	RDD	Covered	0.48	0.17	Table 1 Column 2

Farber et al '15 AEAPP	PBD	Cross-sectional	Total	0.03	0.15	Figure 2 and Column 2 Row 2 Table 1
Filiz '17 Labour	PBD	RDD	Covered	0.08	0.01	Table 3 Column 1
Gerad & Gonzaga '21 AER	PBD	RDD	Total	0.09	0.04	Figure 5 Panel (a) and Text (pg 192)
Gerad & Gonzaga '21 AER	PBD	RDD	Covered	0.89	0.01	Figure 5 Panel (a) and Text (pg 192)
Hunt '95 JOLE	PBD	DID	Total	1.10	0.54	Table 7 Column 2, Table 2, Table 5 Columns 3 and 5, Text (pg 91-92)
Johnston & Mas '18 JPE	PBD	RDD	Total	0.42	0.06	Footnote 15
Johnston & Mas '18 JPE	PBD	RDD	Covered	0.89	0.10	Footnote 15
Katz & Meyer '90 JPubEc	PBD	Cross-sectional	Covered	0.53	0.33	Table 3 Columns 1 and 2, Rows 2 and 5
Kyyra & Pesola '20 Labour Econ.	PBD	DID	Total	0.55	0.07	Table 2 Column 4 and text (footnote 12)

Lalive & Zweimuller ’04 JPubEc	PBD	DID	Total	0.09	0.04	Table 4 bottom row and Table 3
Lalive ’07 AEAPP	PBD	RDD	Total	0.26	0.03	Table 1 Row 1
Lalive ’07 AEAPP	PBD	RDD	Total	1.57	0.09	Table 1 Row 1
Lalive ’07 AEAPP	PBD	RDD	Total	-0.08	0.22	Table 1 Row 1
Lalive ’07 AEAPP	PBD	RDD	Total	0.70	0.27	Table 1 Row 1
Lalive ’08 JOE	PBD	RDD	Total	0.35	0.08	Table 2 Column 2 Panel B and Figures 3 and 8
Lalive ’08 JOE	PBD	RDD	Total	0.64	0.08	Table 2 Column 2 Panel B and Figures 3 and 8
Lalive et al ’06 RESTUD	PBD	DID	Total	0.17	0.03	Table 5 Rows 2-4 and Table 4 Rows 1-2
Lalive et al ’06 RESTUD	PBD	DID	Total	0.09	0.01	Table 5 Rows 2-4 and Table 4 Rows 1-2
Lalive et al ’15 AER	PBD	DID	Total	0.58	0.07	Table 1 Column 2 and Table 2 Columns 3 and 4

Lalive et al '15 AER	PBD	DID	Covered	1.40	0.26	Table 1 Column 2 and Table 2 Columns 3 and 4
Landais '15 AEJ:EP	PBD	RKD	Total	0.33	0.43	Table 4 Column 2
Landais '15 AEJ:EP	PBD	RKD	Covered	1.35	0.68	Table 4 Column 2
Le Barbanchon '16 Labour Econ.	PBD	RDD	Total	0.12	0.04	Table 7 Column 2 Rows 5
Le Barbanchon et al '19 JPubEc	PBD	DID	Covered	0.31	0.03	Table 4 Column 4 Row 2, Table A1
Lichter & Schiprowski '21 JPubEc	PBD	DID	Total	0.18	0.05	Table 2 Column 6
Lichter & Schiprowski '21 JPubEc	PBD	DID	Covered	0.53	0.06	Table 2 Column 6
Moffitt '85 JOE	PBD	Cross-sectional	Total	0.52	0.05	Table 3 Column 4
Nekoei & Weber '17 AER	PBD	RDD	Total	0.06	0.02	Table 2 Column 1
Petrunyk & Pfeifer '22 BER	PBD	DID	Total	0.09	0.02	Tabe 3 Column 4 bottom panel and Table 1 Column 2
Roed & Westlie '12 JEEA	PBD	DID	Total	1.85	0.07	Table 7 Column 2

Schmieder et al '12 QJE	PBD	RDD	Total	0.10	0.01	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 1
Schmieder et al '12 QJE	PBD	RDD	Covered	0.53	0.01	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 1
Schmieder et al '12 QJE	PBD	RDD	Total	0.11	0.02	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 2
Schmieder et al '12 QJE	PBD	RDD	Covered	0.53	0.02	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 2
Schmieder et al '12 QJE	PBD	RDD	Total	0.13	0.03	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 4
Schmieder et al '12 QJE	PBD	RDD	Covered	0.68	0.04	Table 2 Column 1 Panel B, Figure 1, Table W-1 Row 2 Column 4

Schmieder et al '16 AER	PBD	RDD	Total	0.13	0.03	Table 1 Columns 1 and 2
Schmieder et al '16 AER	PBD	RDD	Covered	0.52	0.01	Table 1 Columns 1 and 2
Van Ours & Vodopivec '06 JOLE	PBD	DID	Total	0.42	0.12	Table 5 Row 1

Notes: Each row is a separate main estimate. The first column is the authors, publication year, and journal abbreviation. The second column indicates that all of the estimates correspond to the elasticity with respect to potential benefit duration. The third column contains mutually exhaustive categories for regression discontinuity design (RDD), regression kink design (RKD), difference-in-differences (DID), and cross-sectional variation with controls (cross-sectional). The fourth column describes whether the unemployment outcome is total nonemployment or covered unemployment. The fifth column is the elasticity of the unemployment duration outcome with respect to the benefit generosity parameter and the sixth column is its standard error. The seventh describes the calculation sources from the published paper.

Table B.2: Included Studies: Disemployment Elasticities with Respect to Replacement Rate

Paper	**Policy**	**Design**	**UE measure**	**Elasticity**	**SE**	**Source**
Arranz et al '09 MyC	RR	Cross-sectional	Covered	0.44	0.13	Table 7 Column 4
Belzil '01 JAE	RR	Cross-sectional	Total	0.28	0.09	Text (pg 634)
Bennmarker et al '07 LABOUR	RR	DID	Total	2.01	0.68	Table 6 Row 1 Column 2, Table 3 Row 2 Columns 1 and 5, Table 7 Row 2 Column 4

Bennmarker et al '07 LABOUR	RR	DID	Total	2.32	0.60	Table 6 Row 2 Column 2, Table 3 Row 3 Columns 1 and 5, Table 7 Row 3 Column 4
Bennmarker et al '07 LABOUR	RR	DID	Total	1.33	0.43	Table 6 Row 2 Column 2, Table 3 Row 4 Columns 1 and 5, Table 7 Row 4 Column 4
Bennmarker et al '07 LABOUR	RR	DID	Total	-1.95	1.12	Table 6 Row 1 Column 4, Table 3 Row 6 Columns 1 and 5, Table 7 Row 7 Column 4
Bennmarker et al '07 LABOUR	RR	DID	Total	-1.28	0.73	Table 6 Row 2 Column 4, Table 3 Row 7 Columns 1 and 5, Table 7 Row 8 Column 4
Bennmarker et al '07 LABOUR	RR	DID	Total	-1.30	0.57	Table 6 Row 3 Column 4, Table 3 Row 8 Columns 1 and 5, Table 7 Row 9 Column 4
Blau & Robins '86 JPubEc	RR	Cross-sectional	Total	0.18	0.04	Table 4 Column 1 and text (pg 188)
Blau & Robins '86 JPubEc	RR	Cross-sectional	Total	0.26	0.10	Table 4 Column 1 and text (pg 188)

Card et al '15 AEAPP	RR	RKD	Covered	0.21	0.07	Table 1 Column 1 Row 2 of NBER WP
Card et al '15 ECMA	RR	RKD	Total	1.40	0.20	Table 1 Panel E Column 1
Carling et al '01 EJ	RR	DID	Total	1.97	0.97	Table 4 Column 4 Row DPOL and text (footnote 19)
Carling et al '96 JPubEc	RR	Cross-sectional	Total	0.06	0.02	Text (pg 327) for elasticity and Table 3 Column 1 row for SE
Chetty '08 JPE	RR	Cross-sectional	Total	0.53	0.27	Table 2 Column 1
Classen '77 ILRR	RR	Cross-sectional	Covered	0.45	0.08	Table 2 Column 1
Eugter '15 SJES	RR	DID	Total	0.39	0.19	Table 4 Column 1
Hunt '95 JOLE	RR	DID	Total	-3.32	2.25	Table 7 Column 2, Table 2, Table 5 Columns 3 and 5, Text (pg 91-92)
Katz & Meyer '90 JPubEc	RR	Cross-sectional	Covered	0.66	0.17	Table 3 Columns 1 and 2, Rows 2 and 5
Katz & Meyer '90 QJE	RR	Cross-sectional	Covered	0.41	0.35	Table 4 Column 1 Row 3 and Table 1
Kolsrud et al '18 AER	RR	RKD	Covered	1.53	0.13	Table 2 Panel 1 Column 1

Kroft & Notowidigdo '16 RESTUD	RR	Cross-sectional	Total	0.63	0.33	Table 2 Column 1 Row 1
Lalive et al '06 RESTUD	RR	DID	Total	0.17	0.04	Table 5 Rows 2-4 and Table 4 Rows 1-2
Landais '15 AEJ:EP	RR	RKD	Total	0.32	0.08	Table 4 Column 1
Landais '15 AEJ:EP	RR	RKD	Covered	0.73	0.11	Table 4 Column 1
Lee et al '21 JOLE	RR	RKD	Covered	1.06	0.41	Table 4 Column 5
Meyer & Mok '14 NTJ	RR	DID	Total	0.19	0.03	Table 3 Column 1
Moffitt '85 JOE	RR	Cross-sectional	Total	0.67	0.01	Table 3 Column 4
Portugal & Addison '90 ILRR	RR	Cross-sectional	Total	0.60	0.31	Table 3 Column 3 and Table 1
Poterba & Summers '95 RESTAT	RR	Cross-sectional	Total	0.36	0.15	Table 4 for elasticity Table 3 Column 2 Row 4 for SE
Rebollo-Sanz '20 JHR	RR	DID	Total	0.76	0.02	Table 3 Column 5;corrected simulation from text (pg 149)
Roed & Zhang '03 EJ	RR	Cross-sectional	Total	0.95	0.17	Table 4 Column 1 Row 1
Roed & Zhang '03 EJ	RR	Cross-sectional	Total	0.34	0.19	Table 4 Column 1 Row 1
Roed & Zhang '05 EER	RR	Cross-sectional	Total	0.65	0.08	Table 3 Column 1 and Text (pg 1823)

Roed et al ’08 OEP	RR	Cross-sectional	Total	1.07	0.15	Table 2 Column 1
Roed et al ’08 OEP	RR	Cross-sectional	Total	0.47	0.09	Table 2 Column 1
Topel ’84 JOLE	RR	Cross-sectional	Total	0.64	0.30	Table 3 Column 3 Row 6 and Table 2
Topel ’84 JOLE	RR	Cross-sectional	Total	0.51	0.16	Table 3 Column 3 Row 6 and Table 2
Uusitalo & Verho ’10 Labour Econ.	RR	DID	Total	0.79	0.25	Table 3 Column 1 Row 1 and Text (pg 650)
Winter-Ebmer ’98 OBES	RR	DID	Total	0.07	0.02	Table 1

Notes: Each row is a separate main estimate. The first column is the authors, publication year, and journal abbreviation. The second column indicates that all of the main estimates correspond to the elasticity with respect to replacement rate. The third column contains mutually exhaustive categories for regression discontinuity design (RDD), regression kink design (RKD), difference-in-differences (DID), and cross-sectional variation with controls (cross-sectional). The fourth column describes whether the unemployment outcome is total nonemployment or covered unemployment. The fifth column is the elasticity of the unemployment duration outcome with respect to the benefit generosity parameter and the sixth column is its standard error. The seventh describes the calculation sources from the published paper.

Table B.3: Distribution of Research Design by Policy Margin Among Included Studies

	DID	RDD	RKD	Other	Total
PBD	13 (43%)	13 (43%)	1 (3%)	3 (10%)	30 (100%)
RR	9 (30%)	0 (0%)	5 (17%)	16 (53%)	30 (100%)
Total	22 (37%)	13 (22%)	6 (10%)	19 (32%)	60 (100%)

Notes: The table includes 60 total observations from the 55 included papers because 5 papers estimate both an elasticity with respect to potential benefit duration and an elasticity with respect to replacement rate. The rows split policy parameters by whether the elasticity is with respect to potential benefit duration or replacement rate. The columns correspond to mutually exclusive research designs. The first three columns are quasi-experimental designs: DID is difference-in-differences, RDD is regression discontinuity design, and RKD is regression kink design. Other refers to papers using only cross-sectional variation, which implicitly relies on a selection on observables assumption for identification. The numbers correspond to the total number of estimates in that cell, and the percentages in parentheses refer to the fraction of the row's observations in that cell. Percentages may not add up to 100% due to rounding.

Table B.4: Robustness Analysis of Structural Bias Correction

	margin	Difference from baseline	β_p	$\bar{\theta}$	τ	$\bar{\epsilon}$	$\bar{\theta} - 1.64\tau$	$\bar{\theta} + 1.64\tau$
1	RR		0.09	0.20	0.32	0.44	-0.33	0.73
2	RR	normal distribution	0.05	-0.04	0.55	0.44	-0.94	0.86
3	RR	p(t) symmetric	0.16	0.36	0.26	0.44	-0.06	0.79
4	RR	drop Hunt (1995)	0.09	0.20	0.31	0.53	-0.31	0.71
5	RR	GMM non-parametric	0.20	-0.09	1.07	0.47	-1.85	1.67
6	RR	GMM non-parametric; p(t) symmetric	0.17	-0.07	1.08	0.47	-1.84	1.70
7	RR	GMM non-parametric; drop Hunt (1995)	0.19	0.14	0.88	0.49	-1.30	1.58
8	PBD		0.07	0.10	0.28	0.48	-0.36	0.56
9	PBD	normal distribution	0.06	0.01	0.49	0.48	-0.80	0.81
10	PBD	p(t) symmetric	0.22	0.29	0.23	0.48	-0.09	0.68
11	PBD	GMM non-parametric	0.16	0.40	0.30	0.47	-0.09	0.89
12	PBD	GMM non-parametric; p(t) symmetric	0.17	0.40	0.30	0.47	-0.09	0.89

Notes: The baseline specification in row (1) assumes a parametric t distribution, which is the distribution used in AK's application to the effect of the minimum wage on employment. Row (2) assumes the latent distribution is normal, $N(\bar{\theta}, \tau^2)$. Row (3) returns to the t distribution but assumes that p(t) has three regions: t < -1.96, $|t|$ < 1.96, and t > 1.96. In this row, β_p captures the estimate for $|t|$ < 1.96 which is where the bulk of estimates occur. Row (4) drops Hunt (1995) which is a very negative estimate. Rows (5), (6), and (7) repeat (1), (3), and (4) but use a non-parametric specification of AK. The remaining rows repeat the analysis above, except looking at the PBD margin.

Table B.5: Study Characteristics for Predicting the Elasticity

Category	Variables
Economic characteristics	
Policy	Potential benefit duration (vs. replacement rate), All affected by variation (vs. targeted variation)
Environment	Baseline potential benefit duration, Baseline replacement rate, Sample year, Relative unemployment rate, United States, Labor tax wedge
Methodological characteristics	
Data	Administrative data (vs. survey), total nonemployment (vs. covered unemployment) as the outcome
Estimation technique	RDD, DID/RKD, hazard model
Publication	Journal impact factor (z-score)

Notes: Dummy variables are coded up taking the value 0 with the category in parentheses.
Economic characteristics definitions: The baseline potential benefit duration is defined as the amount for the control group in quasi-experimental designs or the average sample amount in cross-sectional designs. The sample year is the year of the initial policy reform or the average sample year; it is coded up relative to the present-day (2023) such that all values are positive. The relative unemployment rate measures the time-specific macroeconomic environment. It comes from the World Bank's World Development Indicators database that is available since 1991. The difference comes from subtracting the average across all available years from the value in the sample year. The labor tax wedge summarizes the country's tax code. It is defined as the ratio between the amount of taxes paid by an average single worker without children and the corresponding total labor cost for the employer. It is the latest available value from the OECD: 2019 for Brazil and 2021 for all other countries. The estimation technique variables are all dummy variables. Difference-in-differences (DID) and regression kink designs (RKD) are pooled together, regression discontinuity designs (RDD) is its own category, and the omitted category is selection on observables designs relying on cross-sectional variation. The journal impact factor is the IDEAS/RePEc Simple Impact Factor as of April 10, 2023.

Appendix C Online Appendix – Data

Defining the elasticity

The elasticity of unemployment duration with respect to UI benefits ($\frac{\text{d(unemployment duration)}}{\text{d(UI benefits)}} \times \frac{\text{UI benefits}}{\text{unemployment duration}}$) is our effect size of interest. We distinguish between two UI benefit margins: potential benefit duration (PBD) and replacement rate (RR).[15] We distinguish between two measures of unemployment duration: weeks without any employment (total nonemployment) and weeks claiming UI benefits (covered unemployment).

For a given UI benefit margin and unemployment duration definition, there are four different decision points:

1. Whether the authors simulate the elasticity
2. Whether the outcome and regressor are in levels vs. logarithms
3. Whether the outcome is unemployment duration or hazard rate
4. Whether the regressor is continuous benefit generosity or discrete policy eligibility

First, if the authors simulate the elasticity, then we directly use the simulated elasticity. If they do not, then we follow subsequent steps to extract the elasticity from a regression.

Second, whether the outcome and regressor are in levels vs. logarithms determines whether we scale the UI benefit coefficient. If both the outcome and regressor are in logs, then we do not scale the coefficient. If the outcome is in logs but the regressor is in levels, then we scale the coefficient by avg. UI benefit in control group. If the outcome is in levels but the regressor is in logs, then we scale the coefficient by $\frac{1}{\text{avg. unemployment duration in control group}}$. If both the outcome and regressor are in levels, then we scale the coefficient by $\frac{\text{avg. UI benefit in control group}}{\text{avg. unemployment duration in control group}}$.

Third, if the outcome is a hazard rate, then we need to translate the hazard rate elasticity to an unemployment duration elasticity. In the special case of a constant hazard rate, the

[15] The potential benefit duration is the maximum number of benefit weeks a claimant can receive, and the replacement rate is the fraction of prior weekly earnings a claimant receives as benefits.

duration elasticity is the additive inverse of the hazard elasticity. Accordingly, we calculate the duration elasticity assuming a constant hazard rate.

Fourth, if the regressor is a discrete policy change and the outcome is a hazard rate, then we take final two steps. We divide by the benefit change to express the elasticity in terms of benefit generosity units. Additionally, we exponentiate the expression and subtract 1 to avoid the logarithm approximation of a discrete change in the regressor.

We note two useful methodological practices not fully adopted in our sample. First, among papers whose main specification is a hazard model, only one-fifth simulate the model-implied duration elasticity. Approximating a duration elasticity using the hazard model coefficient requires assuming a constant hazard rate, as discussed in Appendix C. However, this does not appear to hold in the data; there is strong evidence of duration dependence among papers that graphically report hazard rates by unemployment duration. Specifically, the typical range in weekly hazard rates (nine percentage points) is slightly larger than the average weekly hazard rate (7%); nearly every paper finds that it falls with unemployment duration. Second, the ratio of behavioral costs to mechanical costs is a sufficient statistic for the efficiency cost of UI benefit expansions. It facilitates comparisons with other tax and transfer policies because it is the fiscal externality term in the Marginal Value of Public Funds formula (Hendren and Sprung-Keyser, 2020). Accordingly, we recommend authors directly compute and report the UI reform's behavioral costs, mechanical costs, and the ratio of behavioral costs to mechanical costs (Lee et al., 2021).

Study characteristics

The vast majority of studies pertain to either Europe or the United States. Just under one-third (17 of 55) pertain to the United States, while only three of the rest pertain to non-European countries (Brazil, Canada, Turkey). The most commonly represented European countries are Austria (9), Germany (7), Sweden (5), and Norway (4); Finland, France, the Netherlands, Portugal, Slovenia, Spain, and Switzerland are the other European countries with 1 or 2 papers.

Journal classifications

Field journals: American Economic Journal: Economic Policy, ILR Review, Journal of

Human Resources, Journal of Labor Economics, Journal of Public Economics, LABOUR, Labour Economics, National Tax Journal, The Review of Economics and Statistics.

"Top-5" journals: American Economic Review, Journal of Political Economy, The Quarterly Journal of Economics, The Review of Economic Studies.

Econometric methods journals: Journal of Applied Econometrics and Journal of Econometrics.

Other general interest journals: American Economic Association Papers and Proceedings, Bulletin of Economic Research, Economics Letters, European Economic Review, Journal of The European Economic Association, Moneda Y Credito, Oxford Bulletin of Economics and Statistics, Oxford Economic Papers, Portuguese Economic Journal, Swiss Journal of Economics and Statistics, and The Economic Journal.

Levels of Aggregation

We use two different levels of aggregation of estimates at different places in the study.

1. *One Estimate Per Policy Margin-Outcome-Group-Paper* Through most of the analysis, we define the unit of observation as one estimate for each policy margin (RR versus PBD), one estimate for each outcome (unemployment vs nonemployment), and one estimate for each group when the paper does not indicate a preferred estimate (e.g. men and women are separate groups in Bennmarker et al. (2007), Schmieder et al. (2012) studies different age groups, and Røed et al. (2008) has different elasticity estimates for Norway and Sweden). This allows for the possibility that authors will report additional group-level estimates because of publication bias. However, in some cases, multiple papers study the same benefit variation in the same region. We refer to these as "contexts", and allow for serial correlation between estimates within the same context. This analysis has 89 observations.

2. *One Estimate Per Policy Margin-Outcome-Paper* For the Bayesian model averaging in Table 2, we aggregate estimates across groups such that there is one estimate for each policy margin and outcome. We aggregate across groups by averaging, weighting each group by the inverse variance of the group. One paper does not report sample sizes for older vs younger workers (Lalive et al., 2015). For this group, we weight the two

groups equally to aggregate. We do this to avoid putting excess emphasis on author reporting decisions. Because BMA itself relies on computing averages across studies, averaging from the group level to the study level has little impact on the results. This analysis has 68 observations.

Ad Hoc Inclusions and Exclusions

Inclusions There are two papers (Card et al., 2015; Kolsrud et al., 2018) we knew estimated the elasticity of unemployment duration with respect to benefit generosity that did not appear in our Publish and Perish search. Because of our goal of surveying the entire literature, we decided to include these studies in producing our estimates.

Exclusions We identified six studies which meet our meta-analysis study inclusion criteria but are excluded from some or all of our analysis.

1. We exclude Hunt (1995) (elasticity of -3.32 and standard error of 2.25) from the BMA analysis shown in Table 2, but include Hunt (1995) in the Andrews-Kasy estimates of Table 1. We made this choice because Hunt (1995) had an outsized effect on the BMA model, which is sensitive to outliers, but had negligible effect on the parametric Andrews-Kasy model (can compare Table B.4 rows 1 and 4).

2. Additionally, for visual clarity, in Figure 1, Hunt (1995), three estimates from Bennmarker et al. (2007) (elasticities are -1.95, 2.01, and 2.32; standard errors are 1.12, 0.68, and 0.60), Carling et al. (2001) (elasticity of 1.97 and standard error of 0.97), Kolsrud et al. (2018) (elasticity of 1.53 and standard error of 0.13), and Røed and Westlie (2012b) (elasticity of 1.85 and standard error of 0.07) are excluded. All of these estimates except Hunt (1995) are included in BMA.

3. We exclude throughout Sahin and Kizilirmak (2007). This study reports an elasticity of 10 with a t-statistic of 200. Obviously, if we had included this estimate it would have swamped all the other results because of the large elasticity and large t-statistic.