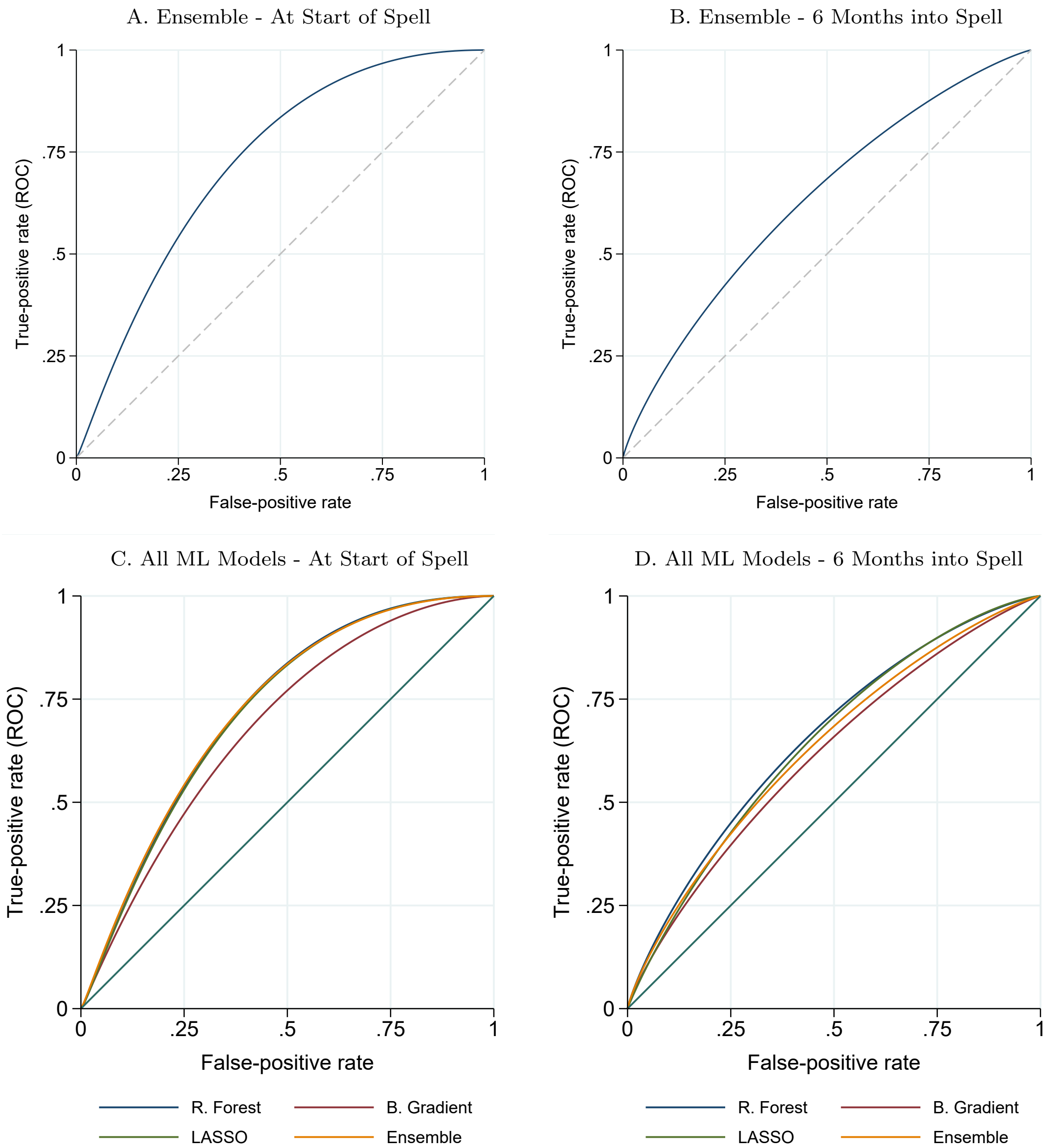
Appendix

A Additional Figures and Tables

A.1 Predictive Power: ROC Curves

Figure A1: Predictive Power of Ensemble Model



Notes: The Receiver Operating Characteristic (ROC) curves plot the combinations of true-positive and false-positive rates attained by binary classifiers based on various thresholds of our predicted job-finding probabilities. Panels A and B focus on the ensemble model, while panels C and D also show the three underlying ML models (random forest, gradient-boosted decision trees and lasso). All curves shown correspond to the hold-out sample for the year 2006.

A.2 Predictive Power: Different Horizons and Robustness

Table A1: ROBUSTNESS: JOB FINDING OVER DIFFERENT HORIZONS

		$E(\cdot)$		$Var(\cdot)$		$Cov(\cdot)$		$R^2(\cdot)$	
Job Finding Horizon	N	F	$\hat{F}$	F	$\hat{F}$	$\hat{F}, F$	$\hat{F}, F_{6m}$	$\hat{F}, F$	$\hat{F}, F_{6m}$
3 Months	122,590	0.483	0.484	0.250	0.030	0.027	0.028	0.097	0.122
6 Months	122,590	0.700	0.699	0.210	0.033	0.031	0.031	0.136	0.136
12 Months	122,590	0.860	0.861	0.120	0.017	0.016	0.021	0.132	0.131

Notes: The table reports summary statistics about observed and predicted job-finding probabilities at the start of the spell over different horizons. We consider job finding over three horizons: three months, six months (the baseline) and twelve months since the beginning of the spell.

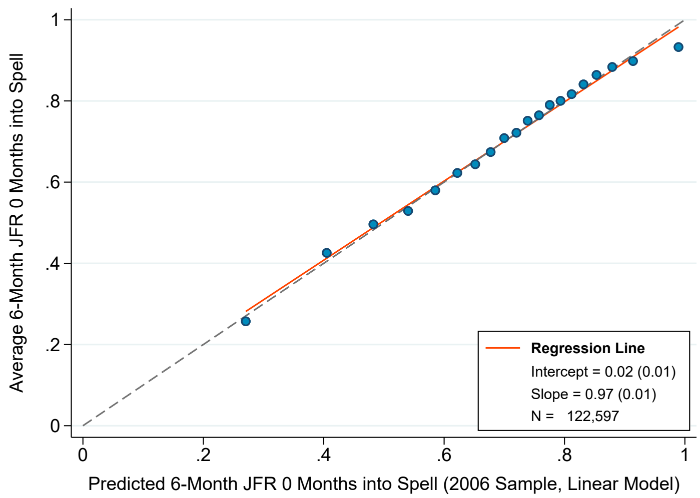
Table A2: ROBUSTNESS: ALMPS AND JOB FINDING DEFINITION

Model	Sample	N	$E(F_{i,0})$	$E(\hat{F}_{i,0})$	$Var(F_{i,0})$	$Var(\hat{F}_{i,0})$	$Cov(\hat{F}_{i,0}, F_{i,0})$	$R^2(\hat{F}_{i,0}, F_{i,0})$
A. Baseline								
Baseline	All	122,590	0.700	0.699	0.210	0.033	0.031	0.136
B. Robustness to ALMPs								
Baseline	No ALMPs	113,334	0.730	0.706	0.197	0.032	0.029	0.130
No ALMPs	All	122,590	0.700	0.726	0.210	0.030	0.029	0.134
No ALMPs	No ALMPs	113,334	0.730	0.731	0.197	0.030	0.028	0.133
C. Robustness to job finding definition								
Baseline	No AvOrs 7-8	105,300	0.683	0.698	0.216	0.033	0.031	0.133
Baseline	No AvOrs 5-8	65,590	0.742	0.743	0.192	0.026	0.024	0.117
D. Robustness to functional form								
Linear	All	122,590	0.700	0.700	0.210	0.030	0.030	0.138

Notes: The table reports summary statistics about observed and predicted job-finding probabilities at the start of the spell for different combinations of models and samples. Models considered include the baseline 2006 model, a model trained on the subset of spells that did not include ALMPs (in the narrow sense) during the first six months of unemployment ("No ALMPs"), and the linear regression model ("Linear"). The samples on which we evaluate the predictions are the full 2006 hold-out sample ("All") and several subsets thereof: excluding spells that include ALMPs during the first six months ("No ALMPs"), excluding spells that ended because the job seeker entered education other than training or died ("No AvOrs 7-8") and excluding spells that ended because the job seeker terminated contact with PES for unspecified or unknown reasons, entered education other than training or died ("No AvOrs 5-8").

A.3 Predictive Power: Linear Model

Figure A2: Comparing Predictions to Outcomes: Linear Model



Notes: The figure shows a binned scatter plot of observed job finding and the predictions of the linear model. That is to say, we split the hold-out sample into 20 vigintiles of predicted 6-month job-finding probability and report, for each bin, mean observed and predicted 6-month job-finding rates at the start of the spell. The red line shows the results of a linear regression at the individual level of a dummy for finding a job within 6 months on the predicted 6-month job-finding probability.

Table A3: R^2 for various submodels in the year 2006: ML Model vs Linear Model

	A. ML Model							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.057	0.086	0.094	0.128	0.132	0.136	0.139	0.136
Change (j) vs $(j-1)$	-	+51.0%	+8.9%	+37.2%	+2.7%	+3.2%	+2.3%	-2.4%
	B. Linear model							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.062	0.086	0.092	0.123	0.125	0.131	0.133	0.138
Change (j) vs $(j-1)$	-	+37.5%	+7.8%	+33.4%	+1.7%	+4.2%	+2.1%	+3.4%
Socio-demographics	X	X	X	X	X	X	X	X
Labour Income		X	X	X	X	X	X	X
Other Income			X	X	X	X	X	X
Employment History				X	X	X	X	X
Income History					X	X	X	X
Migration History						X	X	X
Industry							X	X
Municipality								X

Notes: The table shows the R^2 of the predicted 6-month job-finding probability and a dummy for actual job finding in the hold-out sample for the year 2006 for various models. Panel A reproduces Panel A in Table 3 for convenience. Panel B shows results from linear regression models that use the same variable groups, starting from the basic model in (1) and adding variable groups sequentially until all of the groups included in the baseline model are incorporated in (8).

A.4 Predictive Power: The Role of Pre-Unemployment History Variables

Table A4: R^2 DEPENDING ON PRE-UNEMPLOYMENT HISTORY VARIABLES

A. Groups of variables: sequential sub-models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.058	0.108	0.114	0.117	0.118	0.120	0.125
Change (j) vs $(j-1)$	-	+87.0%	+5.7%	+2.5%	+0.5%	+2.3%	+4.1%
Basic Socio-demographics	X	X	X	X	X	X	X
Individual History in $t-1$		X	X	X	X	X	X
Individual History in $t-2$			X	X	X	X	X
Individual History in $t-3$				X	X	X	X
Individual History in $t-4$					X	X	X
Individual History in $t-5$						X	X
Firm Characteristics							X

B. Groups of variables: marginal sub-models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.058	0.108	0.087	0.070	0.076	0.071	0.087
Change (j) vs (1)	-	+87.0%	+51.4%	+21.2%	+32.0%	+23.1%	+50.0%
Variables:							
Basic Socio-demographics	X	X	X	X	X	X	X
Individual History in $t-1$		X					
Individual History in $t-2$			X				
Individual History in $t-3$				X			
Individual History in $t-4$					X		
Individual History in $t-5$						X	
Firm Characteristics							X

C. Individual variables: sequential sub-models

	(1)	(2)	(3)	(4)	(5)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.058	0.070	0.076	0.097	0.110
Change (j) vs $(j-1)$	-	+20.8%	+9.7%	+26.4%	+13.9%
Basic Socio-demographics	X	X	X	X	X
Days on UI (2y)		X	X	X	X
Unemp. Spells (2y)			X	X	X
Number of Firms (2y)				X	X
Days on DI (2y)					X

D. Individual variables: marginal sub-models

	(1)	(2)	(3)	(4)	(5)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.058	0.070	0.074	0.079	0.075
Change (j) vs (1)	-	+20.8%	+28.3%	+37.1%	+30.8%
Basic Socio-demographics	X	X	X	X	X
Days on UI (2y)		X			
Unemp. Spells (2y)			X		
Number of Firms (2y)				X	
Days on DI (2y)					X

Notes: The table shows the R^2 of the predicted 6-month job-finding probability and a dummy for actual job finding in the hold-out sample for the year 2006 for various models. Panel A starts from the basic model in (1) and adds years of pre-unemployment history variables (including days on UI, days on DI, number of unemployment spells and number of firms) and firm characteristics (tenure, size, change in size and layoff rate)) sequentially, while Panel B adds the same groups one at a time. Panels C and D do the same for a selection of individual variables.

A.5 Predictive Power: SILC Survey Data

Table A5: Regressions with SILC survey data

	General Health (GH)			Mental Health (MH)			GH for MH sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pred. JFR	1.082	1.054		0.992	0.890		0.957	0.973	
	(0.098)	(0.101)		(0.347)	(0.353)		(0.348)	(0.360)	
Health PC1		0.014	0.040					-0.008	0.019
		(0.011)	(0.012)					(0.043)	(0.043)
Mental Health PC1					0.046	0.064			
					(0.033)	(0.034)			
R^2	0.153	0.155	0.016	0.095	0.117	0.044	0.089	0.090	0.003
Adj. R^2	0.142	0.142	0.014	0.083	0.094	0.031	0.078	0.066	-0.010
N	735	735	735	80	80	80	79	79	79

Notes: This table presents output from linear regressions of observed job finding on the predicted job-finding rate and measures of general and mental health obtained from the EU-SILC survey. Our measure of general health ("Health PC1") is constructed from three survey questions: general health (PH010), suffering from any chronic illness (PH020) and limitation in activities because of health problems (PH030). The mental health index ("Mental Health PC1") is constructed from five questions: overall life satisfaction (PW010), meaning of life (PW020), being very nervous (PW050), feeling "down in the dumps" (PW060) and feeling downhearted or depressed (PW080). In both cases, the index used in the regressions is the first principal component of the matrix of relevant survey answers. For the regressions, we match individual spells in our hold-out samples from 1992 to 2016 with responses to the survey, with Columns (1)-(3) including spells matched with general health answers, (4)-(6) with mental health answers and (7)-(9) with both. Note that the mental health module was only included in the 2013 version of the survey, hence the lower number of matches in Columns (4)-(9). The results show that general health does not add much explanatory power, potentially because our prediction model already incorporates this information via the number of days spent on DI in the years before the unemployment spell. In contrast, adding mental health increases the R^2 by 23% and adjusted R^2 by 13%, although the small sample size raises concerns about overfitting. A simple placebo exercise, where we perform the general health regressions on the mental health sample and find virtually no change in the R^2 and a decrease in the adjusted R^2, suggests that the added explanatory power of the mental health variables is not simply due to the small sample size.

A.6 Predictive Power: Extended Models

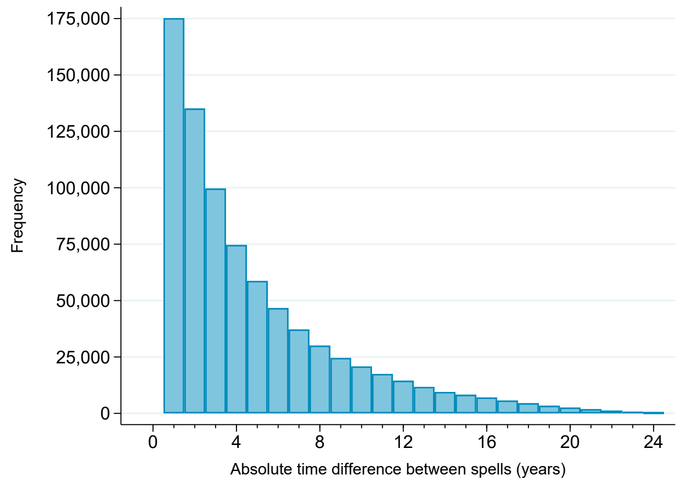
Table A6: R^2 FOR EXTENDED MODELS: STARTING FROM BASIC

	Extensions of Basic							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_{i,0}, F_{i,0})$	0.057	0.082	0.067	0.068	0.063	0.068	0.068	0.091
Change (j) vs (1)	-	+43.5%	+18.2%	+19.7%	+9.8%	+18.9%	+19.3%	+60.6%
Socio-demographics	X	X	X	X	X	X	X	X
Occupation		X						X
Union member			X					X
Wealth				X				X
IQ					X			X
UI choice						X		X
UI benefits							X	X

Notes: The table shows the R^2 of the predicted 6-month job-finding probability and a dummy for actual job finding in the hold-out sample for the year 2006 for various models. We start from the basic model using only socio-demographic information in column (1) and add additional information from other administrative data sets, first one at a time and then all at once in column (8).

A.7 Multiple Spell Data and Results

Figure A3: Two-Spell Sample: Time Difference Between Spells



Notes: The figure shows the distribution of the calendar year difference, in absolute terms, between the start of the two unemployment spells for individuals in our two-spell sample, as described in Table 4. The resulting sample consists of 791,524 individuals.

Table A7: Identifying heterogeneity using repeated spells vs. observables

	Two-spell sample			Control Sample
	$F^1_{i,0}, F^2_{i,0}$	$\hat{F}^1_{i,0}, F^2_{i,0}$	$\hat{F}^2_{i,0}, F^2_{i,0}$	$\hat{F}^2_{i,0}, F^2_{i,0}$
	(1)	(2)	(3)	(4)
	A. Spells in different calendar years			
Cov(·)	0.031	0.018	0.025	0.029
$R^2(\cdot)$	0.019	0.053	0.102	0.122
N	791,524	791,524	791,524	791,524
	B. Spells more than 2 years apart			
Cov(·)	0.026	0.014	0.026	0.029
$R^2(\cdot)$	0.012	0.032	0.105	0.123
N	481,205	481,205	481,205	481,205
	C. Spells more than 5 year apart			
Cov(·)	0.019	0.009	0.025	0.030
$R^2(\cdot)$	0.007	0.013	0.097	0.124
N	248,173	248,173	248,173	248,173

Notes: This table reports key statistics for the sample of individuals with multiple unemployment spells between 1992 and 2016. For a description of the samples and statistics, see the note to Table 4. Panel A reproduces Table 4, while panels B and C simply restrict the samples in A to spells more than 2 and 5 years apart, respectively.

A.8 Dynamics over Spell: Robustness

Table A8: Models Trained of Different Samples: Pooled 2006-2007 Data

Sample	**Model**	N	$E(F)$	$E(\hat{F})$	$Var(\hat{F})$	$Cov(\hat{F}, F)$	$R^2(\hat{F}, F)$
At Start of Spell	0M Model	220,439	0.704	0.702	0.035	0.034	0.156
	6M Model	220,439	0.704	0.594	0.023	0.022	0.100
	12M Model	220,439	0.704	0.525	0.021	0.013	0.041
6M into Spell	6M Model	72,347	0.538	0.537	0.024	0.020	0.063
	12M Model	72,347	0.538	0.489	0.020	0.012	0.031
12M into Spell	12M Model	37,242	0.481	0.474	0.020	0.009	0.018

Notes: The table reports summary statistics about models trained on different unemployment durations in 2006 and 2007. The first three rows correspond to the hold-out sample at the start of the unemployment spell. We generate three different predictions for this sample, with models trained at the start of the spell (the contemporaneous predictions), 6 months into the spell and 12 months into the spell. Rows 4 and 5 deal with the hold-out sample 6 months into the spell; for this sample, we generate predictions using the models trained contemporaneously and 12 months into the spell. Row 6 presents results for the hold-out sample 12 months into the spell, using the contemporaneous model.

Table A9: Models Trained of Different Samples: Pooled 2009-2010 Data

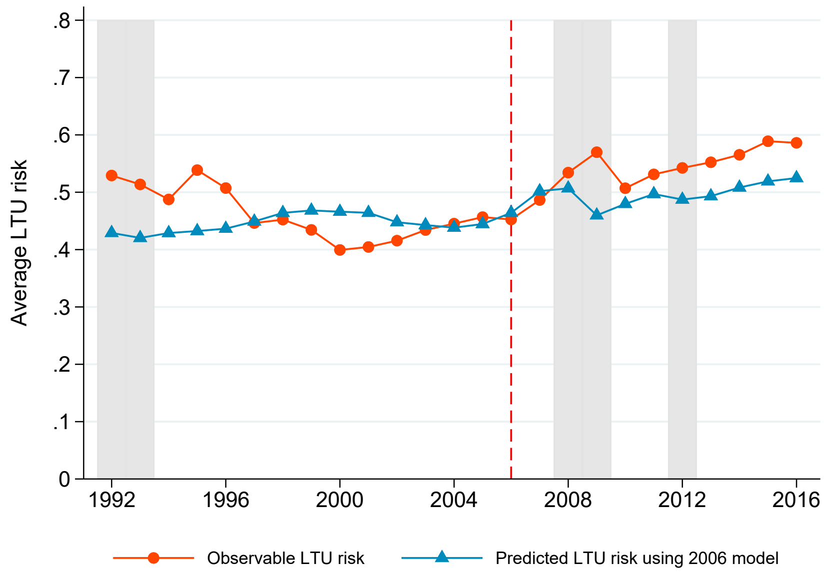
Sample	**Model**	N	$E(F)$	$E(\hat{F})$	$Var(\hat{F})$	$Cov(\hat{F}, F)$	$R^2(\hat{F}, F)$
At Start of Spell	0M Model	229,444	0.577	0.572	0.031	0.031	0.123
	6M Model	229,444	0.577	0.522	0.020	0.020	0.079
	12M Model	229,444	0.577	0.482	0.017	0.011	0.029
6M into Spell	6M Model	96,336	0.469	0.473	0.018	0.014	0.044
	12M Model	96,336	0.469	0.450	0.017	0.009	0.019
12M into Spell	12M Model	47,504	0.426	0.426	0.016	0.009	0.022

Notes: The table reports summary statistics about models trained on different unemployment durations in 2009 and 2010. The first three rows correspond to the hold-out sample at the start of the unemployment spell. We generate three different predictions for this sample, with models trained at the start of the spell (the contemporaneous predictions), 6 months into the spell and 12 months into the spell. Rows 4 and 5 deal with the hold-out sample 6 months into the spell; for this sample, we generate predictions using the models trained contemporaneously and 12 months into the spell. Row 6 presents results for the hold-out sample 12 months into the spell, using the contemporaneous model.

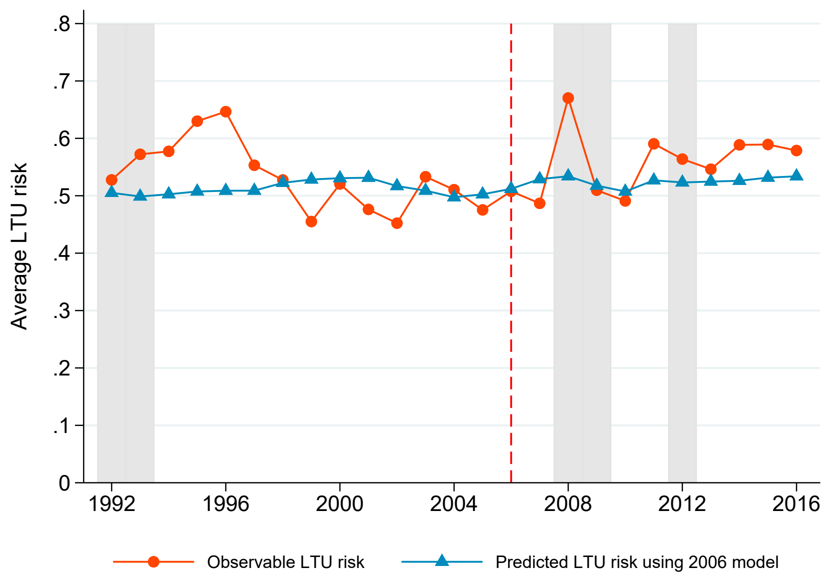
A.9 Selection over Business Cycle: Robustness

Figure A4: Average Risk and Selection into Long-Term Unemployment

A. 6 Months into Spell

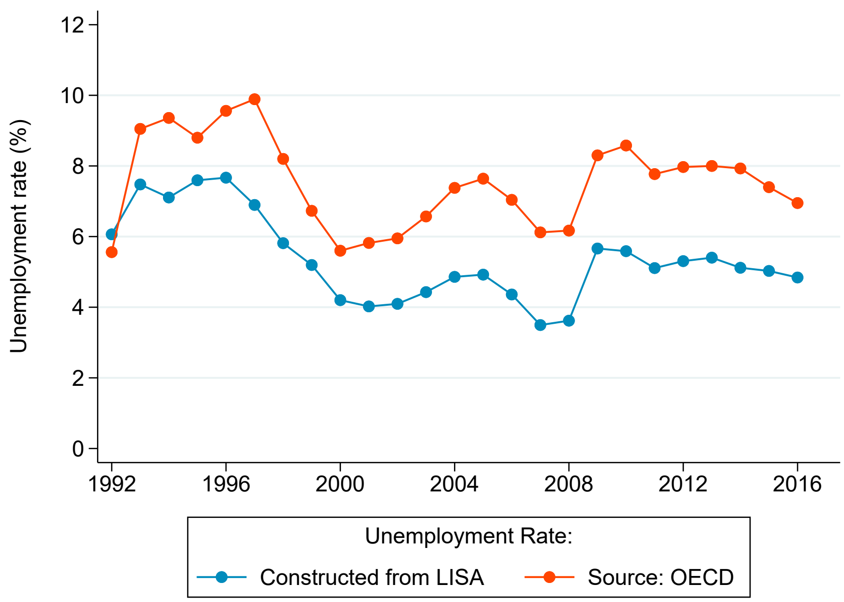


B. 12 Months into Spell



Notes: The figure shows the averages of 1 minus observed and predicted 6-month job-finding rates at different unemployment durations for the hold-out sample for the years 1992-2016. Panel A shows unemployment risk between the 6th and 12th months of unemployment, while Panel B shows unemployment risk between the 12th and 18th months. Predictions are obtained using the corresponding model trained on 2006 data. The grey shaded areas correspond to periods with two consecutive quarters of negative growth in Gross Domestic Product.

Figure A5: Unemployment Rate: LISA vs OECD



Notes: The figure compares the unemployment rate computed from the LISA panel and the official OECD statistics between 1992 and 2016. We use the LISA series throughout the analysis.

Table A10: Relationship between Unemployment and Long-Term Unemployment Risk

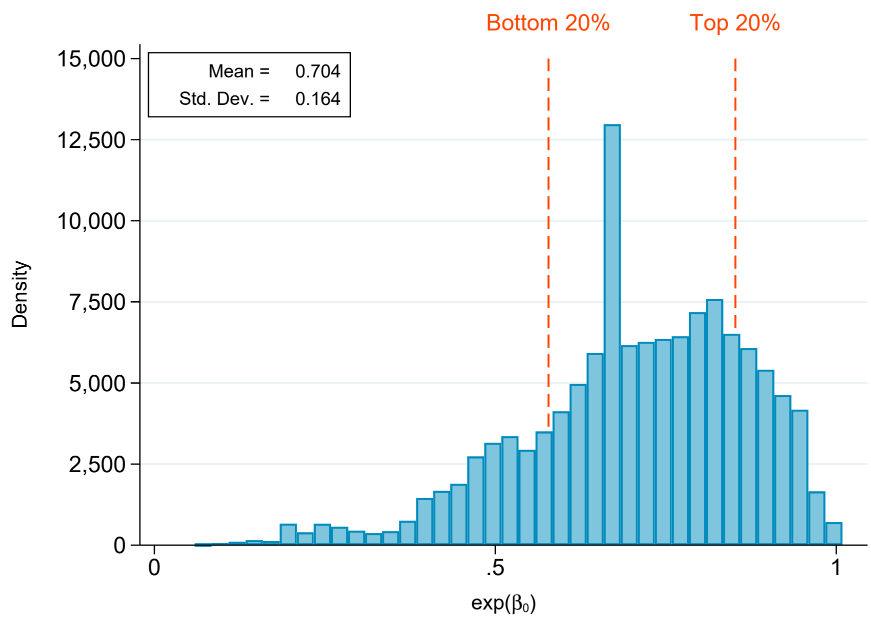
	Predicted log LTU risk (2006)		Observed log LTU risk	
	(1)	(2)	(3)	(4)
	A. 1992-2016			
Log unemployment rate	-0.072	0.119	0.384	0.782
	(0.088)	(0.060)	(0.189)	(0.135)
Time trend		0.012		0.025
		(0.002)		(0.004)
R^2	0.028	0.670	0.152	0.682
Adj. R^2	-0.014	0.640	0.116	0.653
Observations	25	25	25	25
	B. 1995-2016			
Log unemployment rate	-0.068	0.079	0.366	0.688
	(0.108)	(0.050)	(0.229)	(0.088)
Time trend		0.015		0.032
		(0.002)		(0.003)
R^2	0.019	0.817	0.113	0.889
Adj. R^2	-0.030	0.798	0.069	0.877
Observations	22	22	22	22

Notes: The table shows the results of linear regressions of the log of predicted and observed long-term unemployment risk on the log of the aggregate unemployment rate (1-4) and a linear time trend (2 and 4). Panel A uses every year in our sample period, while Panel B restricts to 1995-2016 to avoid early censoring of income and employment histories.

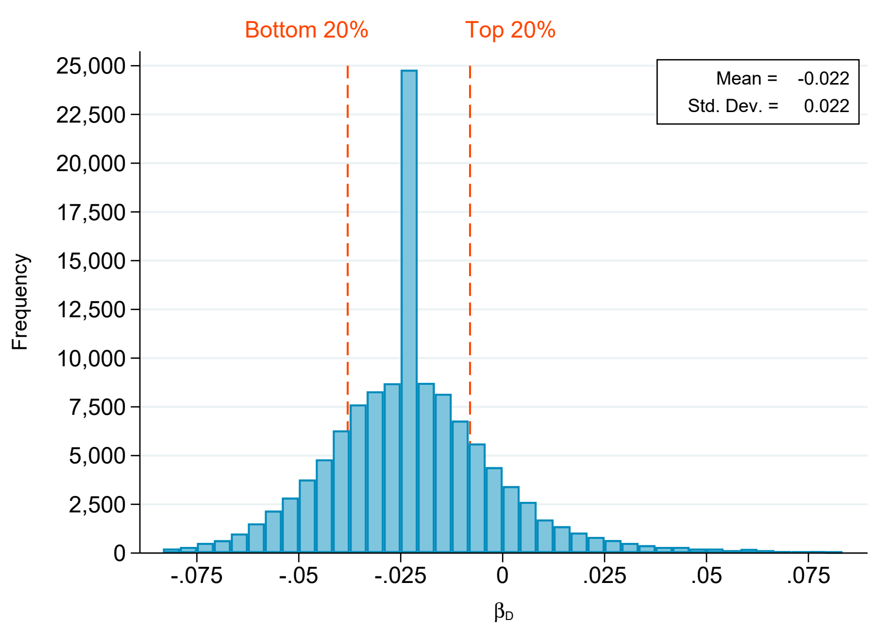
A.10 Heterogeneity in Dynamics over Spell: Additional Results

Figure A6: Distribution of Permanent and Duration-Dependent Component of Job-Finding Risk

A. Histogram of $\exp(\hat{\beta}_0)$



B. Histogram of $\hat{\beta}_D$

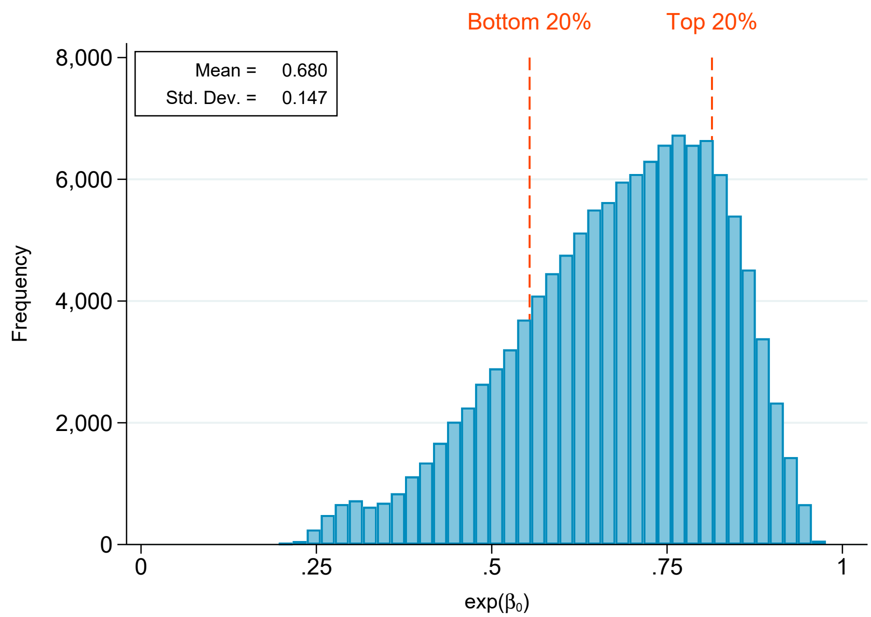


Notes: The figure shows the distribution of the coefficients from the individual-level regressions outlined in equation 7, after applying the shrinkage in equation 8. Panel A shows the histogram of the exponential of the intercept $\exp(\hat{\beta}_0)$, while Panel B shows the histogram of the duration-dependence coefficient $\hat{\beta}_D$.

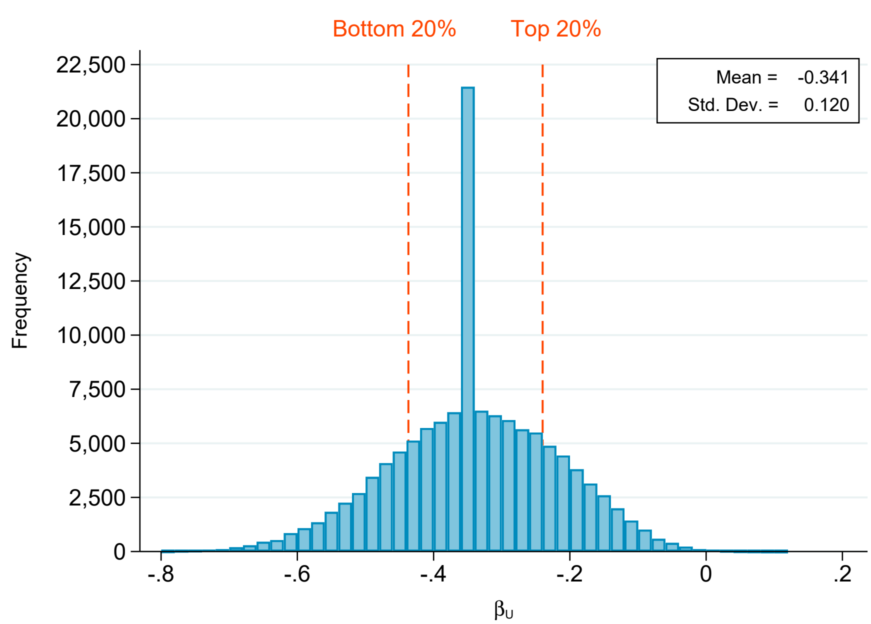
A.11 Heterogeneity in Dynamics over Business Cycle: Additional Results

Figure A7: Distribution of Permanent and Cyclical Component of Job-Finding Risk

A. Histogram of $\exp(\hat{\beta}_0)$

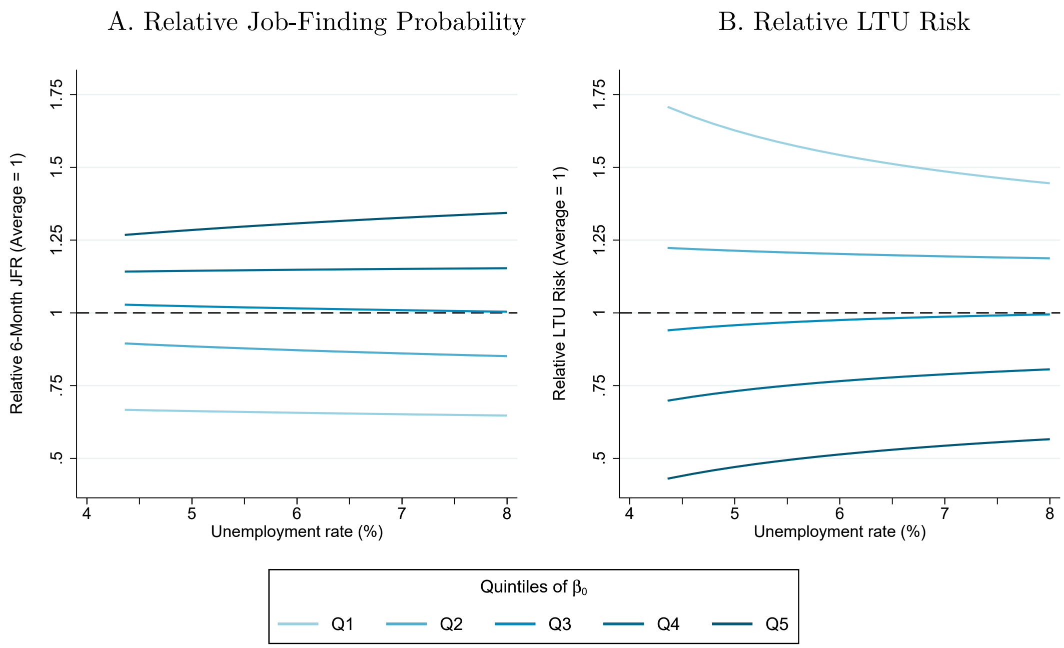


B. Histogram of $\hat{\beta}_U$



Notes: The figure shows the distribution of the coefficients from the individual-level regressions outlined in equation 9, after applying the shrinkage in equation 8. Panel A shows the histogram of the exponential of the intercept $\exp(\hat{\beta}_0)$, while Panel B shows the histogram of the cyclicality coefficient $\hat{\beta}_U$.

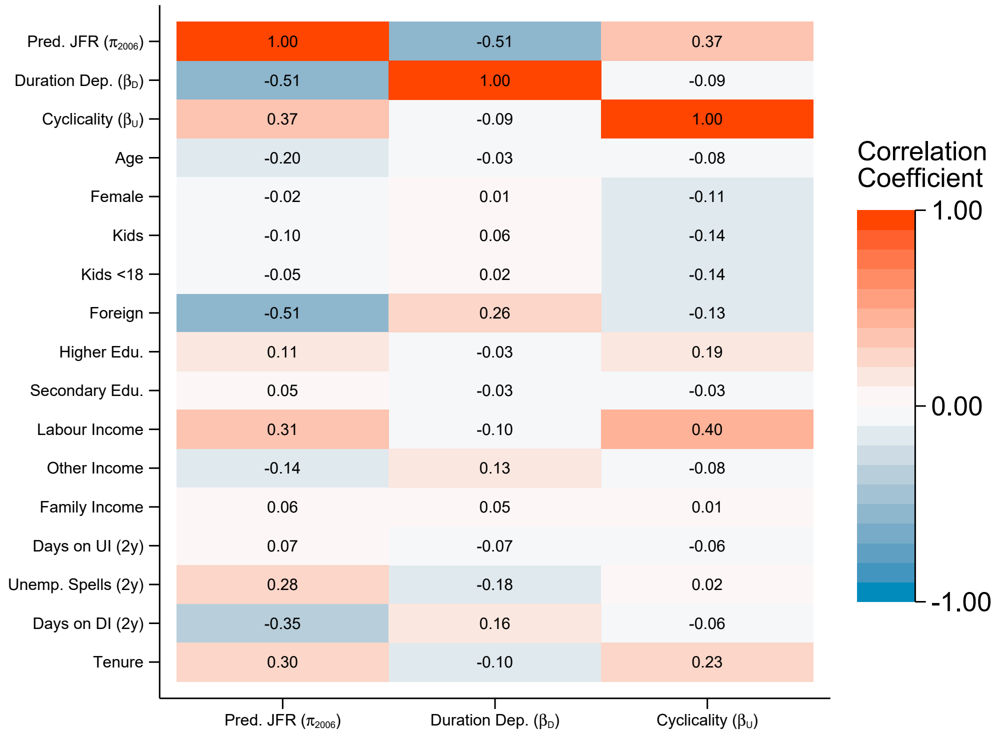
Figure A8: Heterogeneity in Individual Cyclicality: Relative To Average



Notes: Panel A shows the mean predicted individual job-finding rate for the five quintiles of the distribution of the intercept β_0, normalizing the job-finding rate to the mean in 2006 for each unemployment rate. Panel B shows the predicted change in individual LTU risk (defined as the complementary probability), relative to the mean profile, for the five quintiles of the distribution of β_0.

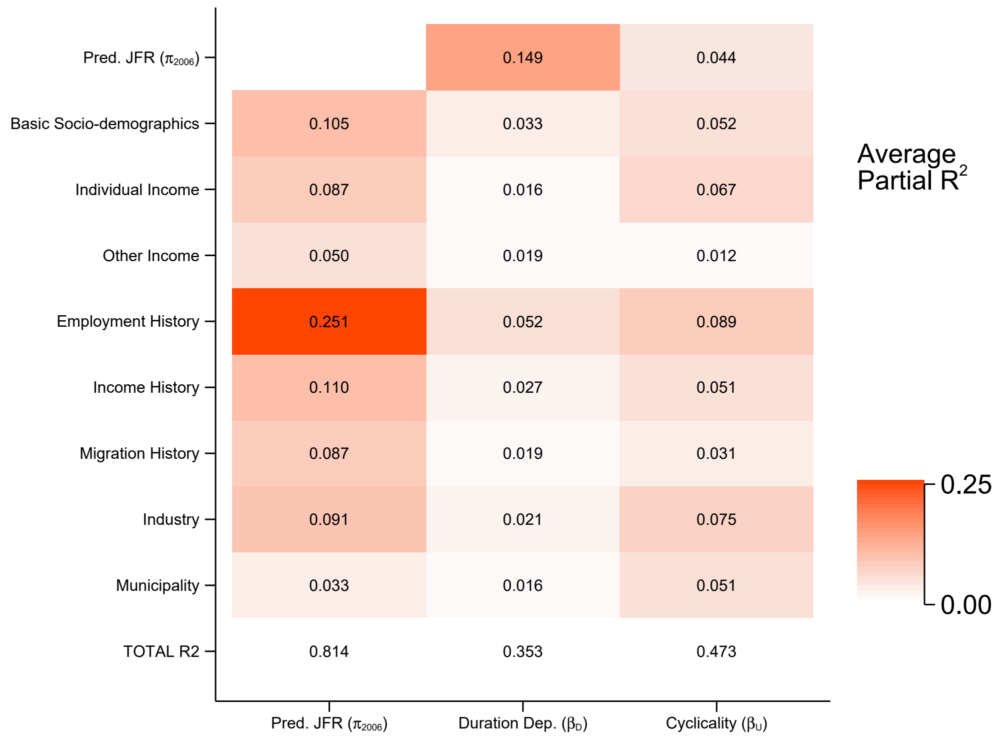
A.12 Heterogeneity and Relationship to Observables: Additional Results

Figure A9: Heterogeneity in Job Finding, Cyclicality and Duration Dependence: Correlation



Notes: This figure reports bivariate correlation coefficients between the predictions and a subset of the variables included in the baseline model. The first column shows correlations between the predicted 6-month job-finding probability at the start of the spell, from the baseline model in 2006, and the variables listed on the y-axis. Columns 2 and 3 show the coefficients for the duration dependence parameter β_D (see Section 4) and the cyclicality parameter β_U (see Section 5), respectively. All coefficients are computed on the 2006 hold-out sample.

Figure A10: Heterogeneity in Job Finding, Cyclicality and Duration Dependence: R^2



Notes: The figure reports Shapley-Owen decompositions of the total R^2 from a linear regression of the predictions on the variable groups included in the baseline model. See Grömping [2007]; Huettner and Sunder [2012] for a full description of the decomposition. The first column shows the Shapley-Owen values for the predicted 6-month job-finding probability at the start of the spell, from the baseline model in 2006. Columns 2 and 3 show the same decomposition for the duration dependence parameter β_D (see Section 4) and the cyclicality parameter β_U (see Section 5), respectively, also including the predicted job-finding probability as a separate variable group in the regressions. All values are computed on the 2006 hold-out sample.

A.13 Targeting: Descriptive Statistics and Robustness

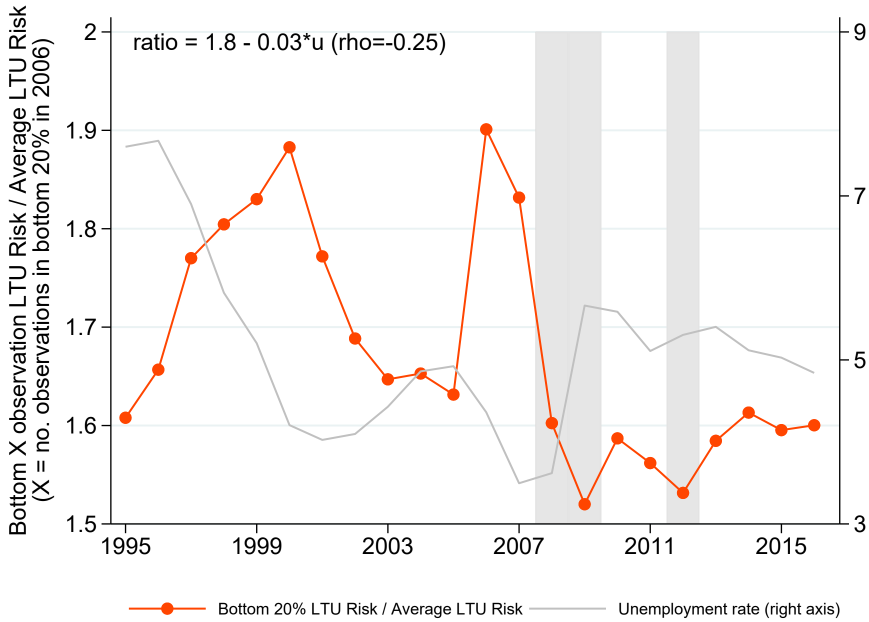
Table A11: ALMP Statistics

Category	% of sample	Prior spell dur.			Length	First 6M	Pred. JFR
		Mean	P25	P75	Mean		Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ⓐ Vocational training	9.1%	318	89	412	121	46.5%	0.686
Ⓑ Non-vocational training or job-search assistance	9.1%	327	48	391	91	54.1%	0.594
Ⓐ + Ⓑ **ALMPs (narrow)**	15.7%	273	57	343	102	54.9%	0.612
Ⓒ Work experience (may include some training)	12.9%	421	97	588	146	36.1%	0.650
Ⓓ Workfare	2.1%	800	190	1187	262	24.2%	0.697
Ⓒ + Ⓓ Work programs	13.7%	397	93	547	148	37.5%	0.654
Ⓔ Subsidized work for the non-disabled	6.2%	601	190	798	245	24.1%	0.663
Ⓕ Subsidized work for the disabled	1.8%	620	146	830	640	29.6%	0.583
Ⓖ Start-up incentive	1.5%	437	117	533	204	38.5%	0.716
Ⓔ + Ⓕ + Ⓖ Subsidies	9.1%	527	153	714	307	28.4%	0.648
Ⓐ + Ⓑ + Ⓒ + Ⓓ + Ⓔ + Ⓕ + Ⓖ **ALMPs (broad)**	25.6%	226	56	323	139	56.1%	0.627

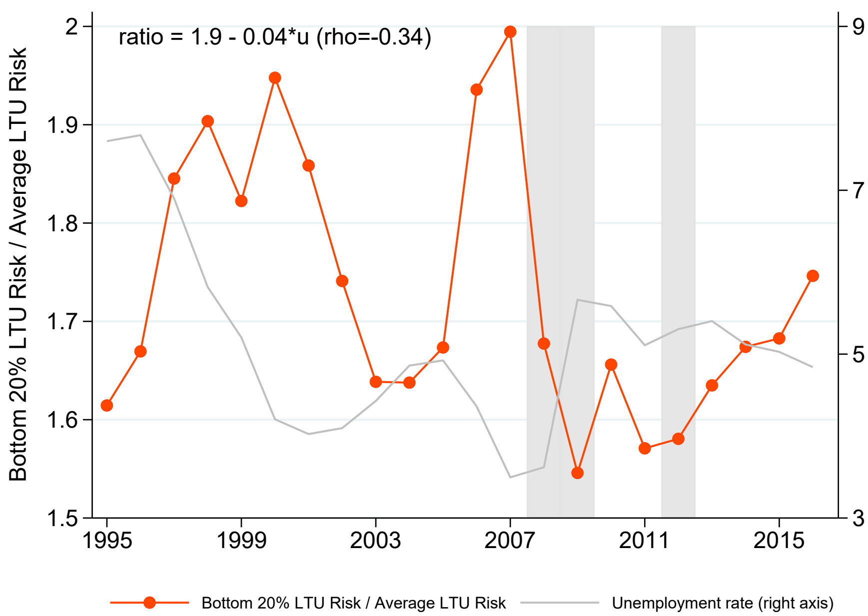
Notes: The table reports descriptive statistics about active labour market policies (ALMPs) in our full 1992-2016 sample. Column 1 shows the percentage of unemployment spells affected by a given policy. Columns 3 to 5 inform about the duration (in days) of the spell before entering the policy, while column 6 reports the length of the policy itself (in days). Column 6 shows the percentage of spells affected by the policy during the first six months of unemployment, relative to the total number of spells affected by the policy. Finally, Column 7 reports the mean predicted job-finding probability at the start of the spell for the fraction of the 2006 hold-out sample that included the policy during the first 6 months of unemployment.

Figure A11: The Value of Targeting over the Business Cycle (Robustness)

A. Fixed Number of Unemployed

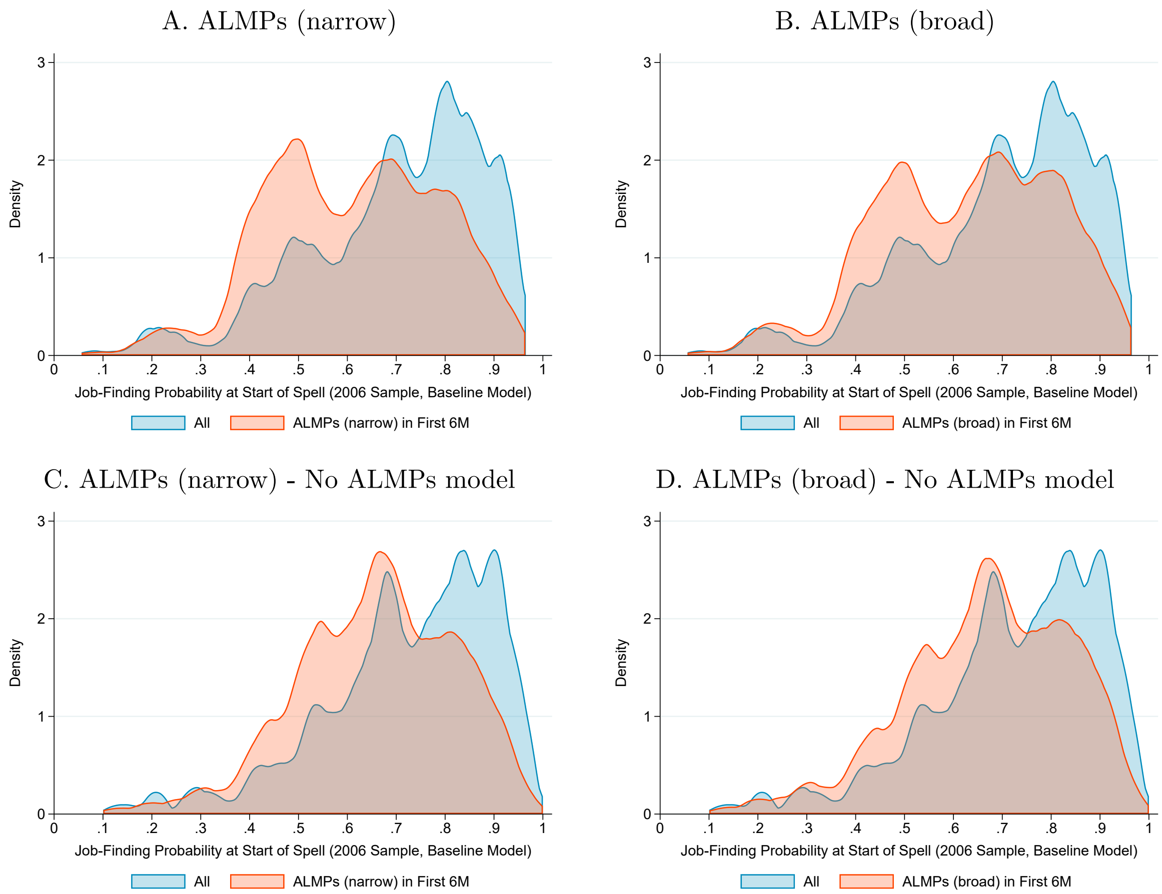


B. Excluding ALMPs



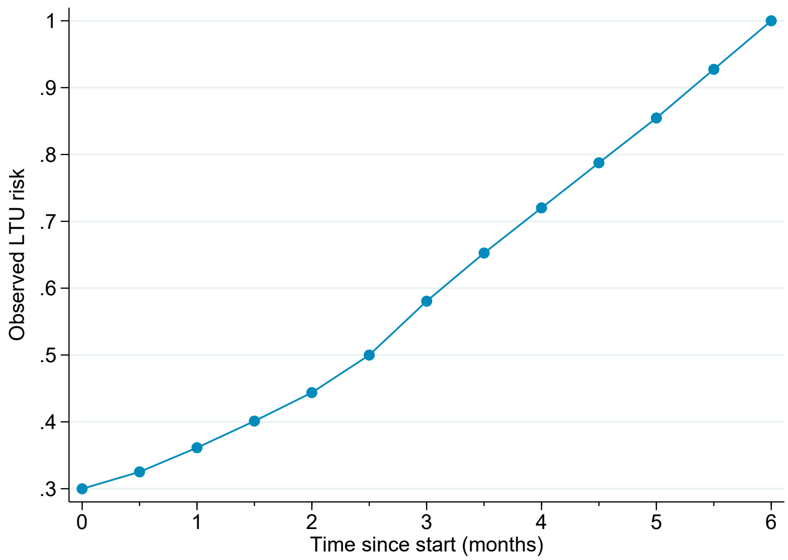
Notes: Panel A shows the value of targeting if we target a fixed number of unemployed (corresponding to the number of unemployed in the bottom 20% in 2006) instead of the bottom 20% of unemployed each year. The lower panel shows that value of targeting if we exclude all those who entered into any ALMP (in the narrow sense) during the first 6 months of their unemployment spell.

Figure A12: Targeting of Training Spells



Notes: This figure shows the distribution of predicted job-finding probabilities as in Figure 2, but separating out the spells that enter ALMPs in the narrow (panels A and C) or broad (panels B and D) sense during the first 6 months of the unemployment spell. Panels A and B show the distribution of predicted job-finding rates from the baseline model, while Panels C and D use a ML model that was trained on a sample that excludes any unemployment spell where the unemployed worker entered ALMPs (in the narrow sense) during the first 6 months of the unemployment spell.

Figure A13: Observed LTU Risk over the First 6 Months of the Spell



Notes: The graph shows observed long-term unemployment risk, defined at the probability of still being unemployed six months after the start of the unemployment spell, conditional on surviving a given amount of time since the start of unemployment. Conditional probabilities have been computed over 15-day increments for the 2006 hold-out sample.

B Proofs and Additional Propositions

B.1 Additional Propositions and Proofs

Proof of Proposition 1. We first note that

$$\begin{aligned}
cov\left(\hat{F}_i, F_i\right) &= E\left(\hat{F}_i F_i\right) - E\left(\hat{F}_i\right) E\left(F_i\right) \\
&= E\left(E\left(\hat{F}_i F_i | T_i\right)\right) - E\left(\hat{F}_i\right) E\left(T_i\right) \\
&= E\left(E\left(\hat{F}_i \times 1 | T_i\right) \Pr\left(F_i = 1 | T_i\right) + E\left(\hat{F}_i \times 0 | T_i\right) \Pr\left(F_i = 0 | T_i\right)\right) - E\left(\hat{F}_i\right) E\left(T_i\right) \\
&= E\left(E\left(\hat{F}_i T_i | T_i\right)\right) - E\left(\hat{F}_i\right) E\left(T_i\right) \\
&= E\left(\hat{F}_i T_i\right) - E\left(\hat{F}_i\right) E\left(T_i\right) \\
&= cov\left(\hat{F}_i, T_i\right).
\end{aligned}$$

Next, we use the assumption that $E\left(\varepsilon_i | X_i\right) = 0$ to show that

$$\begin{aligned}
cov\left(\hat{F}_i, T_i\right) &= cov\left(\hat{F}_i, T(X_i) + \varepsilon_i\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right) + cov\left(\hat{F}_i, \varepsilon_i\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right) + E\left(\hat{F}_i \varepsilon_i\right) - E\left(\hat{F}_i\right) E\left(E\left(\varepsilon_i | X_i\right)\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right) + E\left(\hat{F}_i \varepsilon_i\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right) + E\left(E\left(\hat{F}_i \varepsilon_i | X_i\right)\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right) + E\left(E\left(\hat{F}_i | X_i\right) E\left(\varepsilon_i | X_i\right)\right) \\
&= cov\left(\hat{F}_i, T(X_i)\right).
\end{aligned}$$

Combining the fact that $cov\left(\hat{F}_i, T_i\right) = cov\left(\hat{F}_i, T(X_i)\right)$ and that $cov\left(\hat{F}_i, F_i\right) = cov\left(\hat{F}_i, T_i\right)$, we get that:

$$cov\left(\hat{F}_i, F_i\right) = cov\left(\hat{F}_i, T(X_i)\right).$$

Now we can use the Cauchy-Schwarz inequality,

$$\begin{aligned}
var\left(T(X_i)\right) var\left(\hat{F}_i\right) &\geq cov\left(\hat{F}_i, T(X_i)\right)^2 \\
&= cov\left(\hat{F}_i, F_i\right)^2.
\end{aligned}$$

Hence, we have derived the first lower bound on the variance in job-finding rates,

$$\frac{var\left(T(X_i)\right)}{var\left(F_i\right)} \geq \frac{cov\left(\hat{F}_i, F_i\right)^2}{var\left(\hat{F}_i\right) var\left(F_i\right)} = R^2\left(\hat{F}_i, F_i\right).$$

Given our assumption that $E(\varepsilon_i|X_i) = 0$, we also have that $var(T_i) \geq var(T(X_i))$ and thus

$$R^2\left(\hat{F}_i, F_i\right) \leq \frac{var(T(X_i))}{var(F_i)} \leq \frac{var(T_i)}{var(F_i)}.$$

QED.

Variance of Types with Unbiased Predictors. We can prove the following proposition:

Proposition 3. *If the predictor is unbiased, i.e.* $E(\hat{F}_i|X_i) = T(X_i)$*, then the hold-out sample covariance of the observed realization and the prediction model is an estimate of the variance in observable types as follows:*

$$cov(F_i, \hat{F}_i) = cov(T(X_i), T(X_i)). \tag{A1}$$

Proof. We take from the proof of Proposition 1 that $cov(\hat{F}_i, F_i) = cov(\hat{F}_i, T(X_i))$ and then use the fact that $E(\hat{F}_i|X_i) = T(X_i)$ as follows

$$\begin{aligned}
cov(\hat{F}_i, F_i) &= cov(\hat{F}_i, T(X_i)) \\
&= E\left(\hat{F}_i T(X_i)\right) - E\left(\hat{F}_i\right) E\left(T(X_i)\right) \\
&= E\left(E(\hat{F}_i T(X_i)|X_i)\right) - E\left(E(\hat{F}_i|X_i)\right) E\left(T(X_i)\right) \\
&= E\left(E(\hat{F}_i|X_i)T(X_i)\right) - E\left(E(\hat{F}_i|X_i)\right) E\left(T(X_i)\right) \\
&= E\left(T(X_i)^2\right) - E\left(T(X_i)\right)^2 \\
&= var(T(X_i)).
\end{aligned}$$

QED.

Proof of Proposition 2. We take from the proof of Proposition 1 that, for any δ, $cov_\delta(\hat{F}_i^\delta, F_i^\delta) = cov_\delta(\hat{F}_i^\delta, T^\delta(X_i))$ and by extension $cov_\delta(\hat{F}_i^{\delta'}, F_i^\delta) = cov_\delta(\hat{F}_i^{\delta'}, T^\delta(X_i))$. We then use the fact that $E_\delta(\hat{F}_i^{\delta'}|X_i) = T^{\delta'}(X_i)$ and proceed as follows,

$$\begin{aligned}
cov_\delta(\hat{F}_i^{\delta'}, F_i^\delta) &= cov_\delta(\hat{F}_i^{\delta'}, T^\delta(X_i)) \\
&= E_\delta\left(\hat{F}_i^{\delta'} T^\delta(X_i)\right) - E_\delta\left(\hat{F}_i^{\delta'}\right) E_\delta\left(T^\delta(X_i)\right) \\
&= E_\delta\left(E_\delta(\hat{F}_i^{\delta'} T^\delta(X_i)|X_i)\right) - E_\delta\left(E_\delta(\hat{F}_i^{\delta'}|X_i)\right) E_\delta\left(T^\delta(X_i)\right) \\
&= E_\delta\left(E_\delta(\hat{F}_i^{\delta'}|X_i)T^\delta(X_i)\right) - E_\delta\left(E_\delta(\hat{F}_i^{\delta'}|X_i)\right) E_\delta\left(T^\delta(X_i)\right) \\
&= E_\delta\left(T^{\delta'}(X_i)T^\delta(X_i)\right) - E_\delta\left(T^{\delta'}(X_i)\right) E_\delta\left(T^\delta(X_i)\right) \\
&= cov_\delta(T^{\delta'}(X_i), T^\delta(X_i)).
\end{aligned}$$

QED.

Proof of Corollary 1. First, we follow Mueller, Spinnewijn and Topa [2021] and decompose the observation decline in job finding between two adjacent periods δ and δ' as follows

$$
\begin{aligned}
E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) &= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + E_\delta\left(T_i^{\delta'}\right) - E_{\delta'}\left(T_i^{\delta'}\right) \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \int T_i^{\delta'} dF^\delta(T_i^\delta) - \int T_i^{\delta'} dF^{\delta'}(T_i^{\delta'}) \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \int T_i^{\delta'} dF^\delta(T_i^\delta) - \frac{\int T^{\delta'}(1-T_i^\delta)dF^\delta(T_i^\delta)}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \frac{(1-E_\delta(T_i^\delta))E_\delta(T_i^{\delta'})}{1-E_\delta(T_i^\delta)} - \frac{\int T^{\delta'}(1-T_i^\delta)dF^\delta(T_i^\delta)}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \frac{(1-E_\delta(T_i^\delta))E_\delta(T_i^{\delta'})}{1-E_\delta(T_i^\delta)} - \frac{E_\delta(T^{\delta'}) - E_\delta(T^{\delta'}T_i^\delta)}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \frac{E_\delta(T^{\delta'}T_i^\delta) - E_\delta(T^{\delta'})E_\delta(T^\delta)}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left[T_i^\delta - T_i^{\delta'}\right] + \frac{cov_\delta(T_i^\delta, T_i^{\delta'})}{1-E_\delta(T_i^\delta)},
\end{aligned}
$$

where we used the fact that $dF^{\delta'}(T_i^{\delta'}) = \frac{(1-T_i^\delta)dF^\delta(T_i^\delta)}{\int(1-T_i^\delta)dF^\delta(T_i^\delta)}$. The equation above can be re-arranged to

$$
\begin{aligned}
E_\delta\left(T_i^\delta - T_i^{\delta'}\right) &= E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T_i^\delta, T_i^{\delta'})}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T^\delta(X_i)+\varepsilon_i^\delta, T^{\delta'}(X_i)+\varepsilon_i^{\delta'})}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T^\delta(X_i)+\varepsilon_i^\delta, T^{\delta'}(X_i))}{1-E_\delta(T_i^\delta)} - \frac{cov_\delta(T^\delta(X_i)+\varepsilon_i^\delta, \varepsilon_i^{\delta'})}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T^\delta(X_i), T^{\delta'}(X_i))}{1-E_\delta(T_i^\delta)} - \frac{cov_\delta(T^\delta(X_i)+\varepsilon_i^\delta, \varepsilon_i^{\delta'})}{1-E_\delta(T_i^\delta)} \\
&= E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T^\delta(X_i), T^{\delta'}(X_i))}{1-E_\delta(T_i^\delta)} - \frac{cov_\delta(\varepsilon_i^\delta, \varepsilon_i^{\delta'})}{1-E_\delta(T_i^\delta)} - \frac{cov_\delta(T^\delta(X_i), \varepsilon_i^{\delta'})}{1-E_\delta(T_i^\delta)}
\end{aligned}
$$

The last covariance term $cov_\delta(T^\delta(X_i), \varepsilon_i^{\delta'})$ equals 0 by the assumption that any unobserved heterogeneity is orthogonal to the obervables, $E(\varepsilon_i|X_i) = 0$. This indeed implies that unobserved heterogeneity and observable heterogeneity across adjacent periods are orthogonal for a given set of individuals.

If $cov_\delta(\varepsilon_i^\delta, \varepsilon_i^{\delta'}) \geq 0$, then the equation above

$$
E_\delta\left(T_i^\delta - T_i^{\delta'}\right) \leq E_\delta\left(T_i^\delta\right) - E_{\delta'}\left(T_i^{\delta'}\right) - \frac{cov_\delta(T^\delta(X_i), T^{\delta'}(X_i))}{1-E_\delta(T_i^\delta)}.
$$

Next, we use the fact that, in the hold-out sample, $E_\delta\left(F_i^\delta\right) = E_\delta\left(T_i^\delta\right)$, $E_{\delta'}\left(F_i^{\delta'}\right) = E_{\delta'}\left(T_i^{\delta'}\right)$, and $cov_\delta(\hat{F}_i^{\delta'}, F_i^\delta) = cov_\delta(T^{\delta'}(X_i), T^\delta(X_i))$ (from Proposition 2) and thus get

$$E_\delta\left(T_i^\delta - T_i^{\delta'}\right) \leq E_\delta\left(F_i^\delta\right) - E_{\delta'}\left(F_i^{\delta'}\right) - \frac{cov_\delta\left(F_i^\delta, \hat{F}_i^{\delta'}\right)}{1 - E_\delta\left(F_i^\delta\right)}.$$

QED.

Identification with Data on Multiple Unemployment Spells Per Person. We prove the following proposition:

Proposition 4. *The covariance of actual job finding across two unemployment spells of the same individual at time δ and δ' is equal to the covariance in the underlying probabilities across the two unemployment spells:*

$$cov_{\delta,\delta'}\left(F_i^\delta, F_i^{\delta'}\right) = cov_{\delta,\delta'}\left(T_i^\delta, T_i^{\delta'}\right).$$

Proof.

$$\begin{aligned}
cov_{\delta,\delta'}\left(F_i^\delta, F_i^{\delta'}\right) &= E_{\delta,\delta'}\left[F_i^\delta F_i^{\delta'}\right] - E_{\delta,\delta'}\left(F_i^\delta\right) E_{\delta,\delta'}\left(F_i^{\delta'}\right) \\
&= E_{\delta,\delta'}\left[E_{\delta,\delta'}\left(F_i^\delta F_i^{\delta'} | T_i^\delta, T_i^{\delta'}\right)\right] - E_{\delta,\delta'}\left(T_i^\delta\right) E_{\delta,\delta'}\left(T_i^{\delta'}\right) \\
&= E_{\delta,\delta'}\Big[E_{\delta,\delta'}\left(1 \times 1 | T_i^\delta, T_i^{\delta'}\right) \Pr\left(F_i^\delta = 1 \ \& \ F_i^{\delta'} = 1 | T_i^\delta, T_i^{\delta'}\right) \\
&\quad + E\left(1 \times 0 + 0 \times 1 + 0 \times 0 | T_i^\delta, T_i^{\delta'}\right)\left(1 - \Pr\left(F_i^\delta = 1 \ \& \ F_i^{\delta'} = 1 | T_i^\delta, T_i^{\delta'}\right)\right)\Big] \\
&\quad - E_{\delta,\delta'}\left(T_i^\delta\right) E_{\delta,\delta'}\left(T_i^{\delta'}\right) \\
&= E_{\delta,\delta'}\left[\Pr\left(F_i^\delta = 1 \ \& \ F_i^{\delta'} = 1 | T_i^\delta, T_i^{\delta'}\right)\right] - E_{\delta,\delta'}\left(T_i^\delta\right) E_{\delta,\delta'}\left(T_i^{\delta'}\right) \\
&= E_{\delta,\delta'}\left[T_i^\delta T_i^{\delta'}\right] - E_{\delta,\delta'}\left(T_i^\delta\right) E_{\delta,\delta'}\left(T_i^{\delta'}\right) \\
&= cov_{\delta,\delta'}\left(T_i^\delta, T_i^{\delta'}\right).
\end{aligned}$$

QED.

Corollary 2. *In the stylized model in Section 2, where $T_i^\delta = T_i + h(\delta) + \nu_i^\delta$, the covariance of actual job finding across two unemployment spells of the same individual at time δ and δ' is equal to the variance in the fixed heterogeneity across the two spells, T_i, but a lower bound for the total heterogeneity, i.e.:*

$$cov_{\delta,\delta'}\left(F_i^\delta, F_i^{\delta'}\right) = var_{\delta,\delta'}\left(T_i\right) \leq var_{\delta,\delta'}(T_i^\delta) = var_{\delta,\delta'}(T_i^{\delta'}).$$

Proof. We take from the proof of Proposition 4 above the fact that $cov_{\delta,\delta'}\left(F_i^\delta, F_i^{\delta'}\right) = cov_{\delta,\delta'}\left(T_i^\delta, T_i^{\delta'}\right)$. Then:

$$\begin{aligned} cov_{\delta,\delta'}\left(F_i^{\delta},F_i^{\delta'}\right) &= cov_{\delta,\delta'}\left(T_i^{\delta},T_i^{\delta'}\right) \\ &= cov_{\delta,\delta'}\left(T_i+h\left(\delta\right)+\nu_i^{\delta},T_i+h\left(\delta'\right)+\nu_i^{\delta'}\right) \\ &= cov_{\delta,\delta'}\left(T_i,T_i\right) \\ &= var_{\delta,\delta'}\left(T_i\right). \end{aligned}$$

It is obvious that $var_{\delta,\delta'}\left(T_i\right) \leq var_{\delta,\delta'}\left(T_i\right) + var_{\delta,\delta'}\left(\nu_i^{\delta}\right) = var_{\delta,\delta'}(T_i^{\delta})$, and that $var_{\delta,\delta'}\left(T_i\right) \leq var_{\delta,\delta'}\left(T_i\right) + var_{\delta,\delta'}\left(\nu_i^{\delta'}\right) = var_{\delta,\delta'}(T_i^{\delta'})$. QED.

C Additional Details on Prediction Model

In this Appendix, we describe the binary prediction algorithm that we use to obtain the job-finding probabilities, and report its accuracy across different subgroups.

C.1 Prediction Algorithm

The algorithm we use to predict the probability that an individual finds a job in the next 6 months is a standard machine learning method for binary classification, an ensemble learner that consists in our case of a random forest model, gradient boosted regression trees and LASSO model. To avoid overfitting, we train and calibrate the prediction algorithm on a training sample, for which we use 51.4% of the overall sample. We then use this trained prediction algorithm to obtain predictions for a hold-out sample, which consists of the remaining unemployment spells. All the analyses and statistics in the paper are developed use only this hold-out sample.

The prediction method we use follows four steps, which closely resemble the steps used in Einav et al. [2018]. First, we follow standard practice in machine learning by tuning key parameters that govern the prediction models by 3-fold cross-validation. Second, we train the three resulting prediction models separately. Third, we combine the three obtained predictions into one using a linear combination that we calibrate in the data. Finally, we calibrate the resulting final ensemble predictions using a linear spline. We describe each of the four steps in more detail here.

Parameter Tuning As the three machine learning models that we use have parameters that are at the discretion of the researcher, we follow standard practice and tune these parameters using 3-fold cross validation. More specifically, we tune the following parameters using 10 percent of the sample: minimal node size (mid.node.size), number of variables used at each node (mtry) for the random forest model, learning rate (eta) for the boosted regression trees, and the shrinkage parameter (lambda) for the LASSO.[1] For each of these parameters, we optimize among 5 to 7 alternatives. We tune these parameters using 3-fold cross validation, where we are optimizing the area under the receiver operating characteristic curve (AUC).[2] Thus, for each of the parameter values we want to test, the model is trained on 2 folds (subsets of the training sample), and then the performance is measured in the 3rd fold. The parameter values for which the AUC in the 2006 hold-out sample is highest for each prediction algorithm are: mtry = 50, min.node.size = 12, eta = 0.5, lambda = 0.01.

Estimating the Models Using these tuned parameter values, all models are estimated using 30% of the sample.

Obtaining Ensemble Predictor We combine the predictions from the random forest, gradient boosting regression trees, and LASSO into one ensemble prediction. Following Einav et al. [2018], we construct the ensemble prediction to be the linear combination $p_{ensemble} = \hat{\beta}_{rf}\hat{p}_{rf} + \hat{\beta}_{gb}\hat{p}_{gb} + \hat{\beta}_{lasso}\hat{p}_{lasso}$, where $\hat{p}_x$ is the prediction from algorithm x and $\hat{\beta}_x$ is the associated weight.

[1] We use the package CARET in R that provides a standardized way to tune parameters. The prediction models we use are RANGER (random forest), XGBLINEAR (boosted regression trees), and GLMNET (LASSO).

[2] This is a common metric used in the machine learning literature to measure the performance of a prediction model.

We obtain estimates for the weights from a constrained linear regression (with no constant and the weights summing to one) of the dummy for job finding on the three individual predicted probabilities. For this step, we use 6% of the sample. We find associated weights for the baseline model in 2006 that are $\hat{\beta}_{rf} = 1.03$, $\hat{\beta}_{gb} = 0.04$ and $\hat{\beta}_{lasso} = -0.07$. The gradient-boosted regression trees seems to perform less well than the other prediction models.

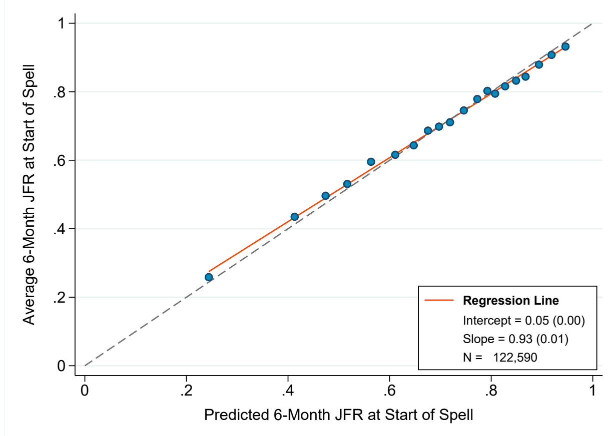
Calibrating Probabilities Finally, the raw probability predictions we get from the ensemble step are calibrated to the actual observed probabilities by estimating a linear spline. This calibration is done using 5.4% of the sample, again not used in any of the previous steps. 250 equal sized bins are created based on the ranked predicted probability. In every bin the mean probability is calibrated to the observed mean probability for these observations. The piece-wise linear spline that follows from linearly interpolating all intermediate points serves as the last step in the prediction mechanism.

C.2 Additional Discussion of Prediction Model

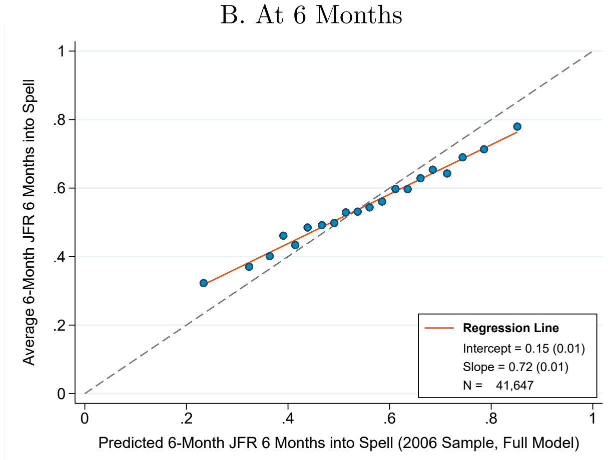
While Panel A of Figure 1 shows a calibration plot for the entire sample, Figures A15 to A19 show a calibration plot for certain subgroups of the sample. If we construct 144 groups by income decile, gender, citizenship, days on DI and days on UI, we see from Panel A, B and C of Figure A20 that average predicted probabilities within the groups remain well calibrated. This makes us comfortable that the observed differences in predicted long-term unemployment risks across different groups are not due to differential prediction accuracy of our ensemble predictor.

Figure A14: Comparing Predictions to Outcomes

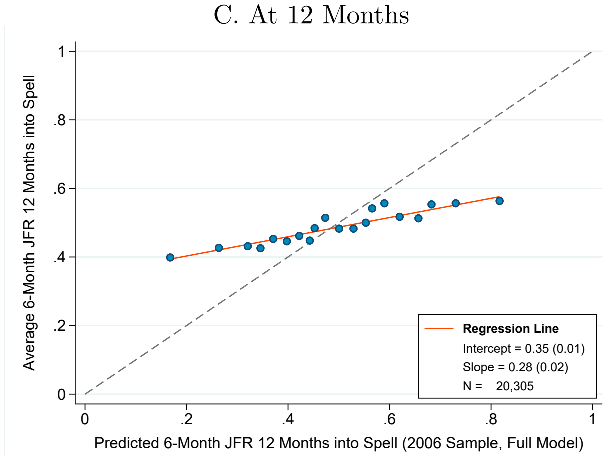
A. At 0 Months



B. At 6 Months

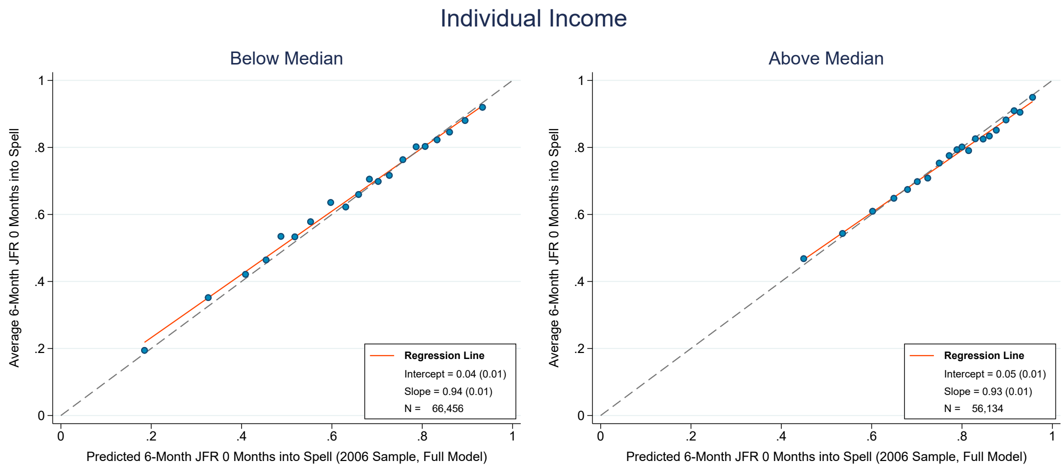


C. At 12 Months



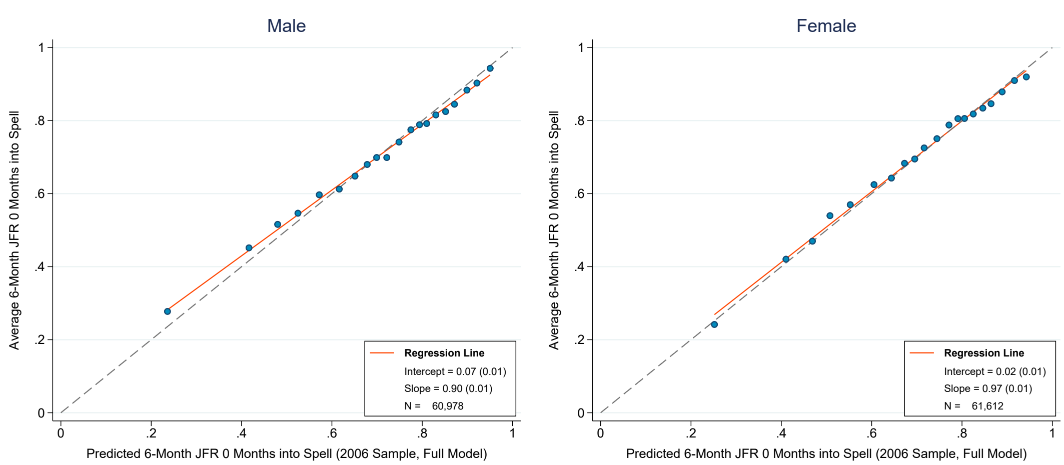
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates, as in Figure 1, for various unemployment durations. Panel A simply reproduces Panel A of 1, while Panels B and C show the predictions 6 and 12 months into the unemployment spell, respectively. All results correspond to the 2006 hold-out sample.

Figure A15: Comparing Predictions to Outcomes: by Income



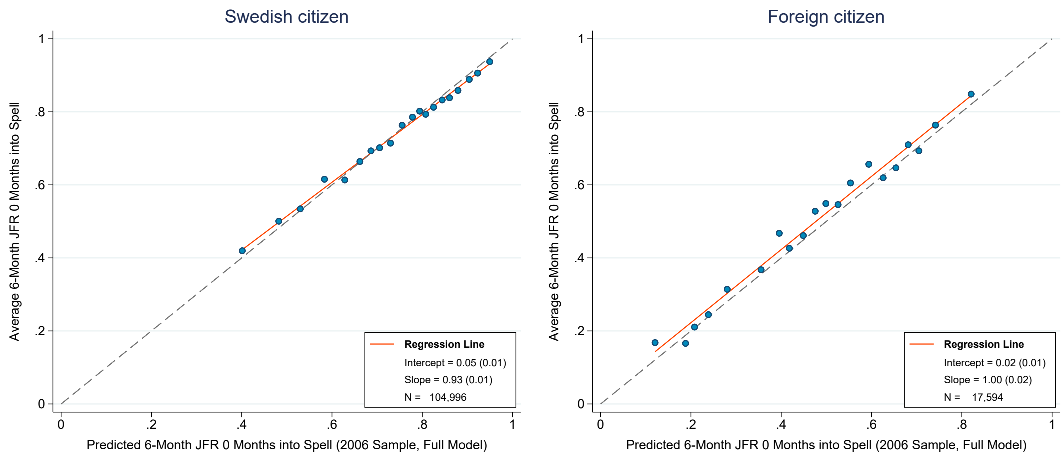
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by individual labour income.

Figure A16: Comparing Predictions to Outcomes: by Gender



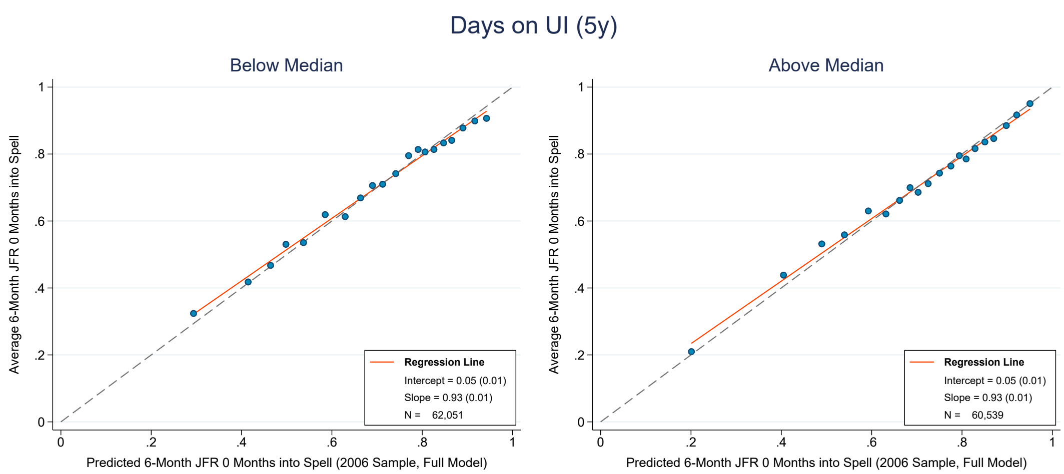
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by gender.

Figure A17: Comparing Predictions to Outcomes: by Citizenship



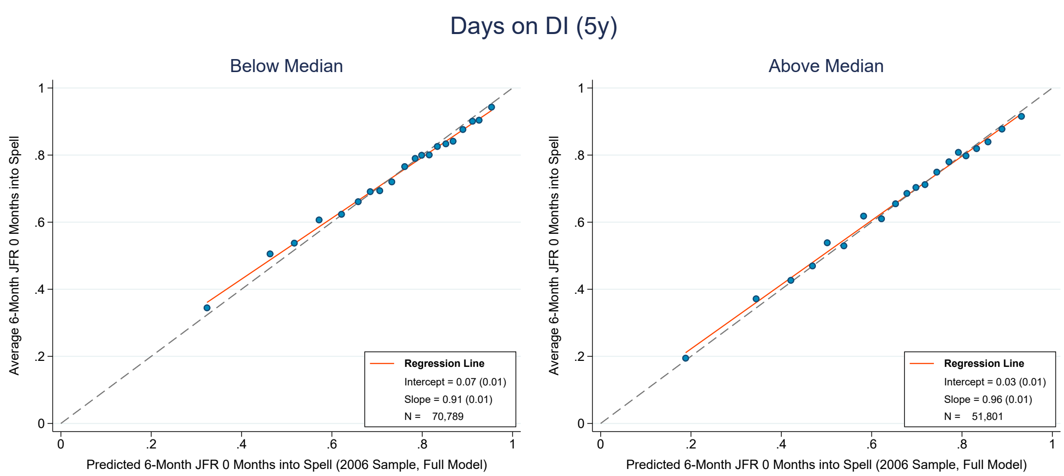
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by citizenship.

Figure A18: Comparing Predictions to Outcomes: by Days on UI



Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by days on UI during the 5 years preceding the unemployment spell.

Figure A19: Comparing Predictions to Outcomes: by Days on DI



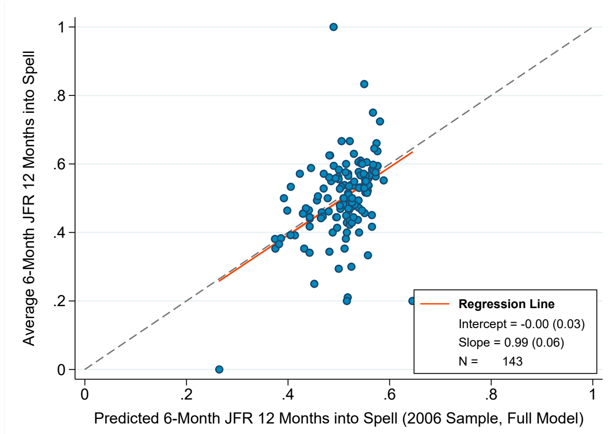
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by days on DI during the 5 years preceding the unemployment spell.

Figure A20: Comparing Predictions to Outcomes: by 144 groups

A. At 0 Months

B. At 6 Months

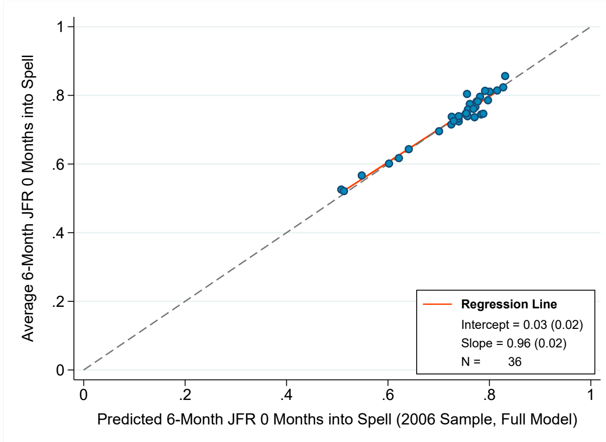
C. At 12 Months



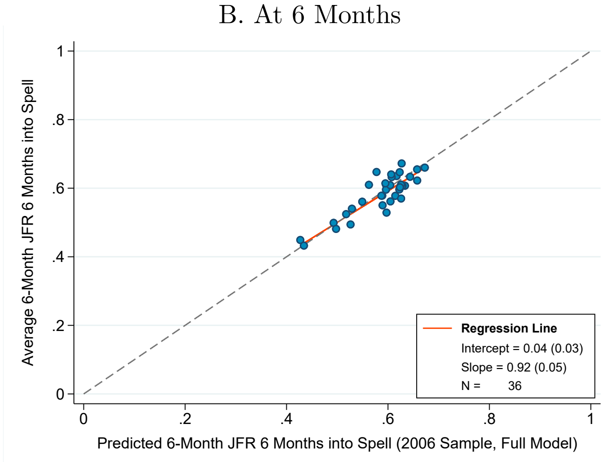
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates for various unemployment durations. Here we construct 144 bins by deciles of labour income, gender, citizenship, days on UI and days on DI, and we report average observed and predicted job-finding rates for each bin. The regression output corresponds to a regression of bin averages. Panel A uses the baseline predictions at the start of the spell, while Panels B and C show the predictions 6 and 12 months into the unemployment spell, respectively. All results correspond to the 2006 hold-out sample.

Figure A21: Comparing Predictions to Outcomes: by 36 groups

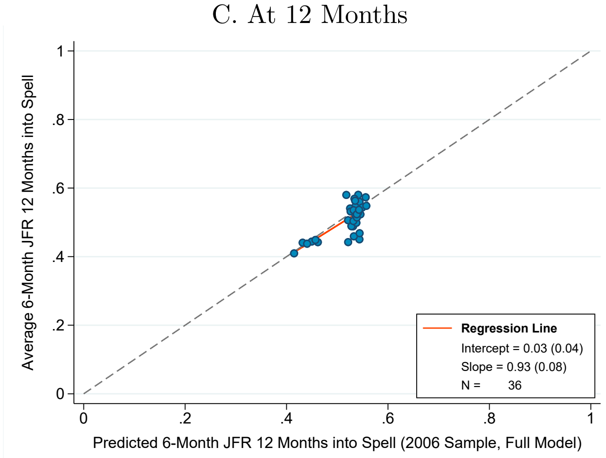
A. At 0 Months



B. At 6 Months



C. At 12 Months



Notes: The figure presents binned scatter plots of observed and predicted job-finding rates for various unemployment durations. Here we construct 36 bins by deciles of labour income, gender and citizenship, and we report average observed and predicted job-finding rates for each bin. The regression output corresponds to a regression of bin averages. Panel A uses the baseline predictions at the start of the spell, while Panels B and C show the predictions 6 and 12 months into the unemployment spell, respectively. All results correspond to the 2006 hold-out sample.