NBER WORKING PAPER SERIES

THE NATURE OF LONG-TERM UNEMPLOYMENT: PREDICTABILITY, HETEROGENEITY AND SELECTION

Andreas I. Mueller
Johannes Spinnewijn

Working Paper 30979
http://www.nber.org/papers/w30979

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2023, Revised June 2024

We thank Bart Cockx, Peter Fredrikson, Francois Gerard, Gregor Jarosch, Maximilian Kasy, Philipp Kircher, Jonas Kolsrud, Kory Kroft, Camille Landais, Alan Manning, Rachel Meager, Peter Nilsson, Oskar Nordstrom Skans, Torsten Persson, Jean-Marc Robin, Je˛ Smith, Johan Vikstrom, Josef Zweimüller and participants at seminars and workshops at Barcelona School of Economics, Carnegie Mellon University, Cornell University, Dallas Fed, HEC Montreal, IIES, LSE, NBER Summer Institute, Norwegian Workshop on Labor Research, Princeton, SITE, UCL, University of Essex, University of Zurich, Uppsala University, and UT Austin for their comments and suggestions and Sebastian Ernst, Miguel Fajardo-Steinhauser, Diego Ferreras Garrucho and Ada Zakrzewska for their excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Nature of Long-Term Unemployment: Predictability, Heterogeneity and Selection
Andreas I. Mueller and Johannes Spinnewijn
NBER Working Paper No. 30979
February 2023, Revised June 2024
JEL No. E24,J64,J68

ABSTRACT

This paper studies the predictability of long-term unemployment (LTU) using rich administrative data from Sweden. We establish substantial heterogeneity in LTU risk across individuals, accounting for both observed and unobserved heterogeneity using a wide range of observable predictors and multiple spell outcomes respectively. We apply our prediction algorithm to study the dynamics of job finding over the unemployment spell and the business cycle. Selection effects can explain most of the decline in average job finding over the unemployment spell, but little of its cyclicality. We also find sizeable heterogeneity in the profiles of job finding over the unemployment spell, but not so over the business cycle.

Andreas I. Mueller
The University of Texas at Austin
Department of Economics
2225 Speedway
Austin, TX 78712
and NBER
andimueller@gmail.com

Johannes Spinnewijn
Department of Economics
London School of Economics, Houghton Street
London, WC2A 2AE
J.Spinnewijn@lse.ac.uk

1 Introduction

Tackling the issue of long-term unemployment (LTU) is a long-standing challenge for labor market policy (Machin and Manning [1999]). Spells of prolonged joblessness are associated with worse economic outcomes, including lower job-finding rates, lower re-employment wages and less stable jobs. These concerns are heightened during bad economic times when the incidence of LTU rises dramatically, such as for example during the Great Recession in 2007-09 (Kroft et al. [2016]). The long-term unemployed are central in the policy debate on how to set unemployment benefits (Kolsrud et al. [2018] and are typically the target for active labor market programs (OECD [2019]).

Despite the policy-relevance of the issue, the sources of LTU are still subject to debate. A first source of contention centers around the observation that long-term unemployed workers typically exhibit lower job-finding rates. One interpretation is that likelihood of finding a job declines at the individual level of the unemployment spell and thus long-term unemployment is a trap, which is difficult to escape. An alternative interpretation, however, is that long-term unemployed workers are inherently different and generally less likely to find jobs throughout their spell. This dynamic selection thus could account for the lower observed job-finding among the long-term unemployed. A second unsettled debate concerns the rise of long-term unemployment in recessions. This rise may be due to a change in the composition of the pool of unemployed workers or be driven by a decrease in the within-individual job-finding chances during recessions for all or at least some unemployed workers.

Traditionally, research has found little role played by observable heterogeneity in explaining the dynamics of job finding over the spell of unemployment or over the business cycle, using observable characteristics in survey data (e.g., Baker [1992]; Kroft et al. [2016]). This in principle leaves an important potential role for within-individual dynamics, either over the unemployment spell or over the business cycle. A long-standing literature has been contesting the limited role of heterogeneity across unemployed job seekers and integrated unobservable heterogeneity in structural models (see Lancaster [1979], Heckman and Singer [1984] and Machin and Manning [1999]). Most recently, a series of papers find evidence for substantial heterogeneity across unemployed job seekers, using a range of different methods (e.g., Hall and Kudlyak [2019], Gregory et al. [2021], Mueller, Spinnewijn and Topa [2021], Alvarez et al. [forthcoming], Ahn et al. [2023]). This naturally raises the question whether studies, which found no or little role for observable heterogeneity, are limited by the range of observable characteristics typically available in survey data.

This paper uses administrative data on the universe of unemployment spells for all unemployed individuals in Sweden for the years 1992-2016 and combines it with rich and detailed information on these individuals' characteristics, which includes – in addition to standard socio-demographics – information on income, employment and benefit histories, as well as information on prior employers, occupation, asset portfolios and IQ scores. We leverage the rich observables to study the predictability of long-term unemployment risk and combine it with multiple spell data to estimate the heterogeneity in job-finding rates. We exploit the predictive power of our prediction model to study the dynamics of job finding over the unemployment spell and the business cycle. We first revisit the importance of heterogeneity in job finding across job seekers and the role of the resulting selection for the dynamics of the average job finding over the unemployment spell and the business cycle. In addition, the large-scale nature of our data both in terms of number of individuals and years allows us to study in detail the potential *heterogeneity in dynamics* over both dimensions.

The paper provides a conceptual framework that characterizes differences sources of heterogeneity in job-finding rates and demonstrates how these can be identified using both measures of predictive power from a prediction model and using the correlation in job finding across repeated spells. First, we show that the R-squared of the predicted job-finding rate with the actual job finding in a hold-out sample provides a lower bound for the heterogeneity in job-finding *probabilities*. The heterogeneity identified in this way includes both persistent and transitory heterogeneity. In a second step, we show that heterogeneity that persists – for example over the spell of unemployment or across cohorts of workers who become unemployed at different stages of the business cycle – can be separated using the correlation between the predicted job-finding rate with actual job finding at different times (e.g., in the spell or over the business cycle). Our measure of persistent heterogeneity over the unemployment spell provides a lower bound for the contribution of dynamic selection to the observed decline in job finding. Finally, we highlight the complementary value between our approach of using observable characteristics vs. the approach of using multiple unemployment spells. While the first approach uncovers all *observable* heterogeneity and allows separating its persistent and transitory components, the second approach uncovers also *unobserved* heterogeneity but at the same time remains limited to the heterogeneity that is persistent across spells (Honoré [1993], Alvarez et al. [forthcoming]). We show how we can combine the two methods to provide a tighter lower bound on the overall heterogeneity of job finding probabilities, which accounts for the unobserved heterogeneity missed by the prediction approach as well as the transitory heterogeneity missed by the approach using multiple spell data. Moving beyond the lower bound, we also show how to scale estimates of observable heterogeneity to gauge the importance of unobservable heterogeneity for dynamic selection.

For the empirical analysis we employ standard machine learning (ML) techniques that leverage the data-rich environment to predict job seekers' job-finding. Our baseline prediction model is an Ensemble Model, which uses a weighted average of the predictions from the LASSO, Gradient-Boosted Decision Trees and Random Forest algorithms (e.g., Einav et al. [2018]). These algorithms vary in their selection of variables and treatment of non-linearities and variable interactions. To deal with potential over-fitting the prediction model is trained on a training sample, but its predictive power is evaluated in a hold-out sample. We focus on predicting the probability of finding a job within 6 months from the start of the unemployment spell. We define this job-finding probability as one minus the probability to be still unemployed six months into the spell, which is a standard measure of long-term unemployment risk in labor-force statistics and a typical target in risk profiling models used by Public Employment Services. Our main analysis predicts the job-finding rates at the start of the spell in the year 2006, when we have all the different data sets available. We then extend our prediction exercise to other unemployment durations and to the years between 1992 and 2016, using a baseline set of characteristics, which are available consistently for each year in our sample and for the universe of unemployed workers in Sweden. Our prediction analyses provide three sets of empirical results shedding new light on the nature of long-term unemployment risk and its determinants.

First, we find substantial predictable heterogeneity in long-term unemployment risk. The predictive power, as measured by the hold-out sample R-squared between the predicted job-finding rate and actual job finding, equals 15% and is more than twice as large when adding income, employment and benefit histories relative to a prediction model that only uses basic socio-demographic variables such as education, age, gender, marital status, citizenship and the number and age of children. The prior employment history, even if only available for one or a few years, is particularly predictive, potentially serving as a proxy for workers' heterogeneity that is otherwise unobservable. The further gains in predictive power when using information on occupation, assets, IQ scores, UI benefits, etc., which are only available for limited samples or years, are modest, suggestive of the saturation of our baseline model. Compared to our Ensemble Model we find that the predictive power of a linear model is nearly as high. This shows both that the additional predictive power comes from the data-rich environment rather than non-linearities or interactions exploited by the ML algorithms and that the risk of over-fitting using the universe of unemployment spells is limited.

We complement our approach using observables by leveraging repeated unemployment spells for a subsample of job seekers and find a substantial role for *unobserved* heterogeneity. The unobserved heterogeneity estimated using the repeated spells corresponds to about half of the estimated *observable* heterogeneity. At the same, our approach using observables shows that an important part of the heterogeneity is not persistent across unemployment spells and that the sample of job seekers with repeated spells is less heterogeneous than the overall sample. Combining the two methods allows us to establish a tighter lower bound on the overall heterogeneity than using either method separately. That is, we find that at least 19% of the variation in job finding outcomes is *ex ante* determined for a cohort of job seekers at the start of their unemployment spell.

Second, we study the implications of the estimated heterogeneity in job finding for the dynamics of the average job finding over the unemployment spell. Similar to other countries, the observed job-finding rate in Sweden declines strongly with the duration of an unemployment spell. In 2006, unemployed job seekers' 6-month job-finding rate was 70% at the start of the unemployment spell but then declined to 55% at 6 months. We repeat the prediction exercise at different durations for ongoing unemployment spells and infer the persistent heterogeneity in job finding by computing the covariance between predicted job finding from 6 to 12 months in the unemployment spell with actual job finding from 0 to 6 months. We find that nearly three quarters of the predictable heterogeneity in job finding is persistent over the spell of unemployment. Applying the decomposition in our conceptual framework, the persistent heterogeneity accounts for a decline of job finding from 70% to 62.7% at 6 months, implying that dynamic selection accounts for 49% of the observed decline in job finding, or more, given the lower bound nature of our prediction exercise. Assuming proportionality in the selection on unobservables relative to the selection on observables into long-term unemployment, we can explain as much as 88% of the observed decline in job finding.

Our analysis thus suggests that the within-individual decline in job finding over the unemployment spell is relatively small on average. Our prediction exercises allow us to go beyond this and investigate the potential heterogeneity in the declines by relating the predicted probabilities of job finding at different durations. This also allows for an empirical test of the key assumption in proportional hazard models, which is that job-finding rates decline at the same rate across job seekers. We uncover substantial heterogeneity in the decline of job finding over the unemployment spell across individuals with different observable characteristics, even when adjusting for the sampling error inherent in this exercise, or when relying on non-parametric tests. Our results thus reject the key assumption in the most commonly used model of job search. We do find, however, that the individual-level declines are strongly negatively correlated with the job finding probability at the start of the spell. This convergence in within-individual job-finding rates further compresses the heterogeneity in job finding over the unemployment spell, above and beyond the dynamic selection.

Third, we turn to the time series of our sample and estimate our prediction model for each year from 1992 to 2016. While we find that some heterogeneity is transitory, the persistence in predictive power is very strong across years. In particular, the R-squared of actual job finding in a given cohort of unemployed workers with the predicted job-finding rate from the model estimated on a different cohort remains high even if distant years are used. The distribution of predicted job-finding risk, however, changes over the business cycle. Prior research has found that compositional changes in the pool of unemployed cannot account for the increased LTU risk in recessions [Baker, 1992; Kroft et al., 2016]. Using richer data, we still reject this so-called heterogeneity hypothesis, as we find that the compositional changes in the pool of unemployed do not translate into higher LTU risk in recessions. Unemployed workers are thus exposed to substantial changes in LTU risk over the business cycle. In parallel to our analysis on duration dependence, we also assess the heterogeneity in the cyclicality of LTU risk across job seekers with different observable characteristics, by relating their predicted job-finding rate to the unemployment rate in each year. In comparison with the duration-dependence analysis, we find only modest differences in the cyclicality in job finding. Recessions affect the employment prospects of both job seekers with higher and lower job-finding chances, but disproportionally hurt the job-finding prospects of workers with lower education and income.

Related Literature. A long and distinguished literature has studied long-term unemployment and its causes and consequences. Our analyses aim to contribute to three important strands in this literature.

First, the duration dependence in job finding and the importance of heterogeneous job finding therein has received wide attention, starting with the seminal work of Lancaster [1979] and Heckman and Singer [1984]. Several papers have argued that negative duration dependence is largely spurious, once both observed and unobserved heterogeneity are accounted for (e.g., Cockx and Dejemeppe [2005]). However, an important takeaway from this early literature is that identification is sensitive to functional and distributional form assumptions. A few recent papers have tried to overcome these challenges. Building on Honoré [1993] and his identification argument using multiple spell data, Alvarez, Borovičková and Shimer [forthcoming] have developed and implemented a novel approach that estimates heterogeneity in job finding without relying on proportional hazards. In a similar spirit, Güell and Lafuente [2022] decompose the observed variance in unemployment durations into a between- and within-spell component.[1] Mueller, Spinnewijn and Topa [2021] instead focus on predictable heterogeneity from job seekers' reported beliefs about their own job-finding risk.[2] All of these papers find substantial heterogeneity in job-finding rates, implying substantial dynamic selection over the unemployment spell, just like we do.[3] In comparison with Alvarez et al. [forthcoming]

[1] Güell and Lafuente's paper builds on an earlier paper by Alvarez, Borovičková and Shimer [2014].

[2] See also Arni et al. [2014] who look at the role of personality traits, beliefs and other behavior variables for predicting long-term unemployment.

[3] There is also direct evidence that call-back rates decline with unemployment duration (Eriksson and Rooth [2014]; Kroft et al. [2013]), though Farber et al. [2016] find no effects for older workers. Jarosch and Pilossoph [2018] also show

and Güell and Lafuente [2022], our paper does not rely on identifying assumptions regarding the persistence of job-finding probabilities across unemployment spells, but shows how to evaluate this empirically. In our data, we show that the persistent heterogeneity across spells identified by the multiple spell approach is about equally as large as the observable heterogeneity uncovered by our prediction exercise. The nature of the heterogeneity, however, is different across the two approaches (observable vs. unobservable; transitory vs. persistent) and, for this reason, the two approaches are highly complementary. In comparison with Mueller et al. [2021], we use rich observable characteristics from administrative data, including income, employment and benefit histories, rather than predictive beliefs. This gives us the statistical power to study heterogeneity in the dynamics of job-finding rates over the unemployment spell. Moreover, to adjust for unobservable heterogeneity, they rely on a model of beliefs, whereas in this paper we show conceptually how to combine our approach using observables with the approach using multiple spell data to estimate unobserved heterogeneity and implement it empirically. For all these reasons, our paper also differs starkly from the prior literature that found a small role of observables for dynamic selection but using data limited to socio-demographic characteristics (e.g., Kroft et al. [2016]).

Second, we apply our approach to a long time series of over 20 years of data, allowing us to speak to issues related to the business cycle. As already mentioned, prior work has found little role for observable heterogeneity for explaining the cyclicality of job finding or unemployment duration (e.g., Baker [1992]; Cockx and Dejemeppe [2005]; Krueger et al. [2014]; Kroft et al. [2016]). Still, a number of papers have found compositional changes in the pool of unemployed over the business cycle. For example, Mueller [2017] documents changes in the composition of the unemployed in terms of prior wage but these compositional shifts have no bearance for the heterogeneity hypothesis because high- and low-wage workers have similar job-finding rates. Elsby et al. [2015] show that the heterogeneity hypothesis matters for transition rates between unemployment and out of the labor force, but not for U-E transition rates. In contrast, Ahn and Hamilton [2020] estimated the heterogeneity in the context of a mixed proportional hazard model and find an important role of heterogeneity for cyclical movements in job finding. Hall and Schulhofer-Wohl [2018] find some evidence in favor of the heterogeneity hypothesis based on the reason for unemployment (layoff, quit, etc.) and show that the pool of job seekers sorts toward low-job-finding types in recessions. Overall, these papers typically rely on a small number of observable characteristics available in labor force survey data. Given the richness of our data, we believe it is important to re-evaluate the heterogeneity hypothesis of the cyclicality of long-term unemployment risk. Moreover, while some earlier work has studied the heterogeneity in the cyclicality job finding by groups of workers characterized by their average wages and hours worked (Bils et al. [2012], Mueller [2017]), we go beyond this by studying heterogeneity in the cyclicality of job finding across a rich set observable characteristics.

Our paper more generally relates to recent papers that use machine learning and related techniques to classify workers into types based on labor force histories. Gregory et al. [2021] use a k-means algorithm to cluster workers based on the similarity of their employment histories in administrative data. Hall and Kudlyak [2019] and Ahn et al. [2023] infer worker types from their labor market

that this may not necessarily translate into declining job-finding rates. Another related paper is Morchio [2020] who finds in data from the NLSY79 that workers with different unemployment histories in their 20s face different job-finding rates later in life.

transitions in labor force survey data. These papers generally find large amounts of heterogeneity in job finding, though they do not separately identify ex-ante heterogeneity and true duration dependence in job finding. Our paper is also complementary to the work studying the impact of job loss on earnings and wages in particular, focusing on the duration dependence [Schmieder et al., 2016], the cyclicality [Schmieder et al., 2022] and predictable sources of heterogeneity more generally [Bertheau et al., 2022].

Finally, many countries either use long-term unemployment or a measure of the risk of long-term unemployment as a criterion for the assignment of active labor market policies (OECD [2019]). This has received relatively little attention in the literature, with only a few empirical (e.g., Berger et al. [2001], Black et al. [2003], Ernst et al. [2024]) and theoretical exceptions (e.g., Pavoni and Violante [2007], Spinnewijn [2013]). This contrasts with the wide attention given to the duration dependence of unemployment benefits specifically (e.g., Shimer and Werning [2008], Schmieder et al. [2012], Kolsrud et al. [2018]) and the evaluation of active labor market programs more generally (see Card et al. [2017] for a meta-analysis).

This paper proceeds as follows. Section 2 provides a conceptual framework that characterizes different sources of heterogeneity and shows how they can be identified using observables and multiple spells. Section 3 presents the data and prediction model and analyses the heterogeneity in job finding. Section 4 analyses the predictability of job finding over the spell of unemployment and quantifies its role for dynamic selection. Section 5 analyses the predictability of job finding over the business cycle. Section 6 concludes.

2 Conceptual Framework

We present a conceptual framework of unemployment to show how to identify heterogeneity in job finding probabilities and how this heterogeneity is crucial for understanding the dynamics in job finding and the role of selection or sorting effects in particular.

2.1 Statistical Model

Setup. We model a continuum of types defined by their likelihood of starting an unemployment spell in each time period $t = \underline{t}, ..., \bar{t}$, $\mathbf{P} = \{P^t\}_{t=\underline{t}}^{\bar{t}}$, and their job-finding probabilities at each duration $d = 0, ..., \bar{d}$, $\mathbf{T}^t = \{T_d^t\}_{d=0}^{\bar{d}}$, where $\underline{t}$ to $\bar{t}$ is the sample period and $\bar{d}$ is an upper bound on the duration of unemployment that we impose. The job-finding probabilities T_d^t for a given type may vary both across spells and during the unemployment spell. The probability for an individual to start an unemployment spell at time t and still be unemployed at duration d thus is $P_d^t = P^t \prod_{\delta=0}^{d-1}(1 - T_\delta^t)$ given the respective probabilities for her type. At this stage, we do not impose any structure or distributional assumptions on the joint distribution of P's and T's.[4] Estimating the full matrix of correlations of these probabilities in the population would be challenging.

Our focus is on characterizing and identifying key statistics that capture relevant heterogeneity in job finding T_d^t in the context of this model. More specifically, we are interested in the following three moments: First, as a baseline, we would like to identify the overall *heterogeneity in job-finding*

[4] Technically, we impose that $T_d^t < 1$ and $P_d^t > 0$ (i.e., a strict inequality), which ensures that every type has a positive measure of unemployed in each cohort and at each duration.

probabilities at a given moment in time, as characterized by the variance in job finding probabilities for a given cohort t at a given duration d, $var_d^t(T_d^t)$. Note that superscript t and subscript d on the variance refer to the corresponding sample of unemployed individuals. Second, we would like to identify the heterogeneity in job finding probabilities that is *persistent* – over the unemployment spell and/or across cohorts – as characterized by the covariance of job finding probabilities between durations d and d' and/or cohorts t and t', $cov_d^t(T_d^t, T_{d'}^{t'})$. As shown below, this moment is crucial for identifying the extent of dynamic selection over the unemployment spell as well as for identifying the extent of unobserved heterogeneity across unemployment spells. Finally, we are interested in identifying *heterogeneity in the dynamics* of job finding probabilities – over the unemployment spell as well as across cohorts – as characterized by the variance in the ratios of job-finding probabilities between durations d and d' and/or cohorts t and t', $var_d^t\left(\frac{T_{d'}^{t'}}{T_d^t}\right)$. This allows us to test the common assumption of proportionality, implying that the variance of the ratios is zero, and to characterize deviations from proportionality both over the unemployment spell and across cohorts. In what follows, we think about these objects in more detail and outline their implications.

Duration dependence in job finding. The heterogeneity in job finding within a given cohort t is key for the dynamic selection of job seekers into longer unemployment durations. Following Mueller et al. [2021], we can decompose the observed changes in job-finding probabilities over the unemployment spell into true duration dependence and dynamic selection as follows:

$$\underbrace{E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right)}_{\text{`observed' duration dependence}} = \underbrace{E_d^t\left[T_d^t - T_{d+1}^t\right]}_{\text{`true' duration dependence}} + \underbrace{E_d^t\left(T_{d+1}^t\right) - E_{d+1}^t\left(T_{d+1}^t\right)}_{\text{dynamic selection}}, \tag{1}$$

which can be rewritten as:

$$E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) = E_d^t\left[T_d^t - T_{d+1}^t\right] + \frac{cov_d^t\left(T_d^t, T_{d+1}^t\right)}{1 - E_d^t\left(T_d^t\right)}. \tag{2}$$

Equation (2) highlights that what matters for dynamic selection is not the overall variance in job-finding probabilities at a given duration, but instead the covariance of job-finding probabilities across durations. Along with the average job-finding probability in the first period, this moment identifies the extent to which dynamic selection contributes to the observed duration dependence. We label the covariance term $cov_d^t(T_d^t, T_{d'}^t)$ for $d' > d$ as *persistent heterogeneity over the spell of unemployment*. Intuitively, it identifies the extent to which a person with a high job finding probability at duration d still has high-job finding probability at a longer duration d'. In the limit, when the covariance goes to zero, then heterogeneous job finding probabilities do not contribute to dynamic selection as the selection into longer unemployment based on the job finding probability at duration d is uncorrelated with the job finding probability at duration d'.

The overall variance in job-finding probabilities can be decomposed into a persistent and transitory term:

$$\underbrace{var_d^t(T_d^t)}_{\text{heterogeneity in job finding}} = \underbrace{cov_d^t(T_d^t, T_{d'}^t)}_{\text{persistent heterogeneity}} + \underbrace{cov_d^t(T_d^t, T_d^t - T_{d'}^t)}_{\text{transitory heterogeneity}}. \tag{3}$$

For the purpose of studying dynamic selection, it is not sufficient to simply characterize the heterogeneity in job finding. Furthermore, the transitory heterogeneity is interesting by itself as it relates to the extent of heterogeneity in the dynamics of job finding, $var_d^t\left(\frac{T_{d'}^t}{T_d^t}\right)$.

Dynamics of job finding over time. Just like for the heterogeneity within a cohort, we are interested in identifying the statistics that characterize the heterogeneity across cohorts and in particular how it evolves over the business cycle. We can again decompose the observed changes in job-finding probabilities across cohorts into 'true' changes and dynamic selection as follows:

$$\underbrace{E_d^t\left(T_d^t\right) - E_d^{t'}\left(T_d^{t'}\right)}_{\text{'observed' dynamics in job finding}} = \underbrace{E_d^t\left[T_d^t - T_d^{t'}\right]}_{\text{'true' dynamics}} + \underbrace{E_d^t\left(T_d^{t'}\right) - E_d^{t'}\left(T_d^{t'}\right)}_{\text{dynamic selection}}. \tag{4}$$

Analogue to the 'true' duration dependence, the true dynamics capture how job finding changes for a given type across cohorts. When comparing cohorts of unemployed job seekers, the selection term is no longer fully determined by the persistent heterogeneity in job finding probabilities, as the selection depends on the probability of becoming unemployed, P^t, and thus how that correlates with job finding probabilities. The overall heterogeneity in job finding depends again on both persistent and transitory heterogeneity across cohorts:

$$\underbrace{var_d^t(T_d^t)}_{\text{heterogeneity in job finding}} = \underbrace{cov_d^t(T_d^t, T_d^{t'})}_{\text{persistent heterogeneity}} + \underbrace{cov_d^t(T_d^t, T_d^{t'} - T_d^t)}_{\text{transitory heterogeneity}}, \tag{5}$$

where we define $cov_d^t(T_d^t, T_d^{t'})$ as the *persistent heterogeneity across cohorts.* To be clear, our definition of a cohort refers to individuals who become unemployed at the same time t and thus individuals can be part of multiple cohorts. For this reason, in the context of the multiple spell analysis below, we refer to this also as the persistent heterogeneity across spells. The transitory heterogeneity is again intrinsically linked to heterogeneity in the dynamics across cohorts, like for instance the cyclicality in job finding. As will become clear below, the distinction is particularly relevant in the presence of multiple spell data, as it identifies the heterogeneity in job finding that is fixed between two spells of unemployment.

There are two main issues with identifying the moments outlined in the prior paragraphs: First, we do not observe the probabilities but only the random outcome thereof, which makes it difficult to identify the distributional statistics outlined above. Second, we observe outcomes only in the selected sample of job seekers who become unemployed at time t and are still unemployed at duration d. It is in principle not possible to look at the covariance of job finding across durations d and d' for all those unemployed at duration d, nor to look at the covariance and dynamics across cohorts t and t' for all those unemployed at time t. In the next two sub-sections, we describe how we can overcome these identification issues, by using a rich set of observable characteristics, by using multiple spell data and ultimately combining both of these approaches for identification.

2.2 Identification with Observables

We model the job-finding probability for an individual who becomes unemployed at time t and is still unemployed at duration d as

$$T_d^t = T_d^t(X_d^t) + \varepsilon_d^t,$$

where $T_d^t(X_d^t) = E_d^t\left(T_d^t|X_d^t\right)$ is the individual's job finding probability based on observable characteristics X_d^t at time t and duration d. We use the sub-index on the moment indicator as short-hand notation to refer to the considered sample of individuals, i.e., people unemployed at time t and duration d. The average job finding probability for a sample of unemployed job seekers corresponds to the sample's discrete-time hazard rate. ε_d^t captures the unobservable heterogeneity, orthogonal to the observable characteristics, $E_d^t\left(\varepsilon_d^t|X_d^t\right) = 0$.

While we cannot observe an individual's job-finding probability, we can observe whether or not an individual has found a job. For an individual of type T_d^t, the realization of the probability is

$$F_d^t = \begin{cases} 1 & \text{with prob } T_d^t \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We posit a prediction model $F_d^t(.)$ such that

$$F_d^t = F_d^t(X_d^t) + e_d^t, \tag{7}$$

where e_d^t is a prediction error. Note that the source of prediction error is both sampling error as well as randomness in the outcome variable, which is the random realization of the underlying probability. If the prediction model is unbiased, we have $E_d^t(\hat{F}_d^t|X_d^t) = T_d^t(X_d^t)$.[5]

2.2.1 Heterogeneity

We are interested in evaluating the ex-ante heterogeneity in job finding types, $var_d^t\left(T_d^t\right)$. The challenge is to separate the ex-ante heterogeneity from the heterogeneity in ex-post outcomes, $var_d^t\left(F_d^t\right) = E_d^t\left(T_d^t\right)\left(1 - E_d^t\left(T_d^t\right)\right)$. We can bound the ex-ante heterogeneity by the variation in ex-post outcomes that is predictable ex-ante using observables:

Proposition 1. ***Lower bound for the variance in types.*** *The hold-out sample R-squared between the observed realization of the probability and its prediction provides a lower-bound estimate of the variance in (observable) types relative to the variance in realizations:*

$$R^2(F_d^t, \hat{F}_d^t) \leq \frac{var_d^t(T_d^t(X_d^t))}{var_d^t(F_d^t)} \leq \frac{var_d^t\left(T_d^t\right)}{var_d^t(F_d^t)}. \tag{8}$$

Proof. See Appendix B.

The R-squared equals the covariance between the realizations and predictions scaled by their respective variances, $R^2(F_d^t, \hat{F}_d^t) = \frac{cov_d^t(F_d^t, \hat{F}_d^t)^2}{var_d^t(F_d^t)var_d^t(\hat{F}_d^t)}$. By evaluating the covariance in a hold-out sample rather than in the sample used for estimating the prediction model, we avoid any confounding

[5] We refer to a predictor to be unbiased only for the variables X_d^t that are also included in the prediction model.

correlation between the sample error underlying individual predictions and their specific outcomes. The main argument in the proof relies on the Cauchy-Schwarz inequality, which implies the lower bound. This lower bound becomes tight when the predictor is unbiased. The hold-out sample covariance of the observed realization and the prediction then equals the variance in observable types, $cov_d^t(F_d^t, \hat{F}_d^t) = var_d^t(T_d^t(X_d^t))$ (see Proposition B1 in Appendix B).

2.2.2 Persistent heterogeneity

We are interested in separating out the heterogeneity that is persistent and the heterogeneity that is transitory over the spell of unemployment or across cohorts. As discussed, persistent heterogeneity over the spell crucially determines the potential for selection out of unemployment and how it changes the average job finding dynamics among the unemployed. Transitory heterogeneity implies that individuals are subject to different within-individual changes and thus causes individual job finding dynamics to be heterogeneous. Just like we can infer the overall heterogeneity from the covariance between the contemporaneous predictions and job finding realizations, we can infer the *persistent* heterogeneity from the covariance between predictions and realizations, but using lags or leads of the predictions instead:

Proposition 2. ***Persistent heterogeneity.*** *If the predictor is unbiased, i.e.* $E_d^t(\hat{F}_d^t|X_d^t) = T_d^t(X_d^t)$*, then the hold-out sample covariance of the observed realization for cohort* t *at duration* d *and the prediction model for cohort* t' *at duration* d' *evaluated in the sample of all unemployed in cohort* t *at duration* d *is an estimate of the covariance in observable types between duration* d *and* d' *and/or across cohorts* t *and* t'*:*

$$cov_d^t(F_d^t, \hat{F}_{d'}^{t'}) = cov_d^t(T_d^t(X_d^t), T_{d'}^{t'}(X_{d'}^{t'})). \tag{9}$$

Proof. See Appendix B.

Note that we do not need to observe the same individuals in both (d, t) and (d', t'), but we only require common support in terms of the respective sample. This proposition relies on the unobserved heterogeneity at state (d, t) to be orthogonal to the predictions at state (d', t'). This is trivial when the predictor is unbiased and uses only observables that are fixed across states (d, t). In our empirical application of the dynamics of job finding over the unemployment spell, we only use predictors from the year prior to the start of the unemployment spell and thus do not change over the unemployment spell. We can thus use the covariance in equation (9) to identify an upper bound for the individual-level decline in job finding over the unemployment spell. Following Proposition 2, we can prove:

Corollary 1. ***Upper bound for true duration dependence over the unemployment spell.*** *If the predictor is unbiased, i.e.* $E_d^t(\hat{F}_d^t|X^t) = T_d^t(X^t)$ *and depends on observables that are determined prior to the spell* X^t*, and the unobserved heterogeneity is weakly persistent, i.e.* $cov_d^t(\varepsilon_d^t, \varepsilon_{d+1}^t) \geq 0$*, then the hold-out sample moments of the observed duration dependence in job finding and the covariance of the observed realization at duration* d *and the prediction model at duration* $d+1$ *provide an upper bound for the individual-level decline in job finding over the unemployment spell, as follows:*

$$E_d^t\left(T_d^t - T_{d+1}^t\right) \leq E_d^t\left(F_d^t\right) - E_{d+1}^t\left(F_{d+1}^t\right) - \frac{cov_d^t\left(F_d^t, \hat{F}_{d+1}^t\right)}{1 - E_d^t\left(F_d^t\right)}. \tag{10}$$

Proof. See Appendix B.

We would like to note that we do not make any assumptions on the shape of the true duration dependence over the spell of unemployment and our approach is fully flexible in this respect. At the same time, the upper bound nature of our estimates of true duration dependence between durations d and $d+1$ prevents us from pinning down its exact shape over the spell of unemployment.

The empirical analysis of duration dependence in Section 4 estimates the persistence in heterogeneity over the spell of unemployment and uses it to construct the upper bound following equation (10). In the same spirit, we study in Section 5 how persistent job finding rates are across cohorts at different stages of the business cycle, but we cannot use that directly to study how much sorting into unemployment contributes to cyclicality in the average job finding rate, as discussed further above. Instead, to study selection effects over the business cycle, we use equation (4) and look at selection through the lens of the prediction model by looking at changes in the predicted job-finding probability that come from changes in observable characteristics in the pool of unemployed, i.e. $E_d^t\left(\hat{F}_d^{t'}\right) - E_d^{t'}\left(\hat{F}_d^{t'}\right)$.[6] Of course, this measure of selection does not take into account selection based on unobservable types, i.e. $E_d^t\left(\varepsilon_d^{t'}\right) - E_d^{t'}\left(\varepsilon_d^{t'}\right)$, and we cannot use the same bounding argument as for dynamic selection over the spell of unemployment. Still, we believe that this exercise has merit because of the richness of our data and the high predictive power of our model, which leaves much less room for selection on unobservables than in previous studies who used only a limited set of observable characteristics.

2.2.3 Heterogeneity in job-finding dynamics

The estimated selection on observables allows us to bound the dynamics in the *average* job finding rate, as we have just discussed. We can also leverage the observables to directly study heterogeneity in the dynamics in job finding rates across individuals. That is, we can gauge the potential heterogeneity in the dynamics of job finding by relating individuals' predictions across dynamic states, $\hat{F}_d^t$ and $\hat{F}_{d'}^{t'}$, given their observables X_d^t. This then also allows us to determine which specific observables are predictive of stronger dynamics and whether these dynamics are proportional to the baseline job finding rates, i.e., $T_{d'}^{t'}(X_d^t) = \beta_{d,d'}^{t,t'} T_d^t(X_d^t)$ for any X_d^t, where $\beta_{d,d'}^{t,t'}$ is a proportionality factor that is independent of X_d^t. Appendix B.4 discusses in detail how this relates to the continuous-time proportional hazard model Cox [1972].

It is important to qualify the analysis here in two dimensions: First, the analysis ignores dynamic selection on unobservable types and would attribute this to the dynamics in observable types. Clearly, the better the prediction model and the less unobserved heterogeneity is left, the less of a caveat this is. Second, we hold observables fixed. This is exactly what allows us to use predictions for all individuals unemployed in state (t,d) *and* (t',d') and thus address the issue of selection based on observables.[7]

[6]Note that $E_d^t\left(\hat{F}_d^{t'}\right) - E_d^{t'}\left(\hat{F}_d^{t'}\right) = E_d^t\left(\hat{F}_d^{t'}(X_d^t)\right) - E_d^{t'}\left(\hat{F}_d^{t'}(X_d^{t'}))\right) = \left[E_d^t\left(T_d^{t'}(X_d^t)\right) - E_d^t\left(T_d^{t'}(X_d^{t'})\right)\right] + \left[E_d^t\left(T_d^{t'}(X_d^{t'})\right) - E_d^{t'}\left(T_d^{t'}(X_d^{t'})\right)\right]$. The first term in square brackets implies that our measure of selection over time includes changes in selection that come from changes in observables at the individual level. Thus, our empirical measure of selection refers to selection on observables rather than selection at the individual level alone (the second term).

[7]While this is a major advantage of our approach, it captures only the dynamics of job finding probability holding observable characteristics constant and thus not the dynamics that come from changes in observables at the individual level. In our empirical setting, this only applies to the analysis of job finding over the business cycle, since in the context of dynamics over the spell we only use observables that are pre-determined relative to the unemployment spell in question.

To test for the proportionality assumption, we can simply compute – for a given pair of states (t, d) *and* (t', d') as well as for a given set of observables in the baseline state, X_d^t – the variance of $\frac{\hat{F}_{d'}^{t'}(X_d^t)}{\hat{F}_d^t(X_d^t)}$ and test whether it is positive. Given the large number of possible pairwise comparisons in our data, we would like to go beyond the pairwise comparisons and leverage the full extent of our data. We do so by using the individual predictions $\hat{F}_{d'}^{t'}(X_d^t)$ across different states (d, t) and relate them to the duration of the unemployment spell, the time of the unemployment spell or the unemployment rate at the time of the unemployment spell, as follows:

$$\log(\hat{F}_{d'}^{t'}(X_d^t)) = \beta_0(X_d^t) + \beta_\Delta(X_d^t) \times \delta + \eta_{d'}^{t'} \tag{11}$$

where δ (Δ) is either the duration d (D), the time t (Tr), the unemployment rate u_t (U) or a vector that combines some of these variables. We run this linear regression for each individual in the sample with characteristics X_d^t at duration d in baseline period t and thus can test whether $var_d^t(\hat{\beta}_\Delta(X_d^t)) = 0$, while correcting for the estimation error underlying $\hat{\beta}_\Delta(X_d^t)$.[8]

We note again that when rejecting proportionality, we can in principle not distinguish between individual dynamics or selection on unobservables driving the non-proportionality. We discuss these issues in more detail in Sections 4 and 5.

2.3 Identification with Multiple Spell Data

The identification approach using observables does not capture all of the individual-level heterogeneity, but only the heterogeneity that is predictable given the available observables. We can, however, capture both observed and unobserved heterogeneity with data on multiple unemployment spells per person and thus observing the outcome of job finding, F_d^t, at the same duration d on two or more separate occasions. This is the idea underlying the use of multiple unemployment spells as proposed by Honoré [1993] and more recently Güell and Lafuente [2022] and Alvarez et al. [forthcoming]. In the context of our model, we can show that

$$var_d^{t_1,t_2}(T_d^{t_1}) = \underbrace{cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2})}_{\text{persistent heterogeneity across the two spells}} + \underbrace{cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_1} - T_d^{t_2})}_{\text{transitory heterogeneity in first spell}}, \tag{12}$$

where t_1 refers to the time of the first unemployment spell and t_2 refers to the time of the second unemployment spell. We can show that the persistent heterogeneity across two spells is identified by the covariance of job finding outcomes in the two spells:

Proposition 3. ***Persistent heterogeneity across cohorts in multiple spell data.*** *For two randomly chosen spells for each individual, the covariance of job finding outcomes in spell 1 and 2 identifies the heterogeneity that is persistent between the two spells:*

$$cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) = cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2}). \tag{13}$$

Proof. See Appendix B.

[8]See Proposition B4 in the Appendix for more details on this test.

Corollary 2. ***Lower bound with multiple spell data.*** *For two randomly chosen spells for each individual, if some of the heterogeneity in job finding is transitory and specific to a spell, i.e.* $cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_1} - T_d^{t_2}) > 0$*, then the covariance of job finding outcomes in spell 1 and 2 identifies a lower bound for the heterogeneity in job finding*

$$cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) < var_d^{t_1,t_2}(T_d^{t_1}). \tag{14}$$

Proof. See Appendix B.

The multiple spell approach thus only identifies heterogeneity that is persistent between the time of the first and second unemployment spell. Moreover, it relies on a potentially selected sample of individuals that are observed to be unemployed in those two cohorts, which is captured by the superscript t_1, t_2 on the variance and covariance terms. We can also apply the approach with multiple spell data for outcomes at different durations, say d and $d+1$, but this is for an even more selected sample as not all individuals who are unemployed at duration d in the first spell are unemployed at duration $d+1$ in the second spell. That is also why this approach cannot use multiple observations for individuals at different durations in the same unemployment spell, as individuals who find a job at duration d are no longer unemployed at duration $d+1$ in the same spell. Finally, we would like to note that an assumption that is implicit in the multiple spell approach is the absence of lagged duration dependence, i.e. the duration and frequency of past spells does not affect the job-finding probability in the current spell. For the prediction model using observables, the fact that we include information on past unemployment spells in the prediction model is not in contradiction with this assumption. Instead, the approach using observables for identification relies on the frequency and duration of past unemployment spells providing useful signal of the underlying type of the worker.

2.4 Identification with Observables and Multiple Spell Data

The method using observables misses out on the unobservable heterogeneity, but allows capturing heterogeneity that is transitory across spells. This transitory heterogeneity potentially is non-negligeable as individuals may change occupations, accumulate skills through training or school, or experience changes in their financial wealth, their family life or their health status between two unemployment spells. It is important to note that this transitory heterogeneity across cohorts can still be persistent over the spell of unemployment and thus be relevant for understanding the dynamic selection over the spell. While the multiple spell approach misses out on this transitory heterogeneity, it allows capturing unobservable heterogeneity that is persistent across spells and that is missed by the approach using observables. We can thus leverage the complementarities between the two approaches to better understand the different sources of heterogeneity and impose tighter bounds on the overall heterogeneity.

First, we can assess empirically how important transitory heterogeneity is across two spells relative to the heterogeneity that is persistent. The total observable heterogeneity can be decomposed into heterogeneity that is persistent and transitory across two spells:

$$\underbrace{var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}))}_{\text{Observable Heterogeneity (OH)}} = \underbrace{cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2}))}_{\text{Persistent Observable Het. (POH)}} + \underbrace{cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_1}(X_d^{t_1}) - T_d^{t_2}(X_d^{t_1}))}_{\text{Transitory Observable Het. (TOH)}}. \tag{15}$$

As shown in Proposition B1, the approach using observables can identify the observable heterogeneity within a cohort with the covariance of the predicted job finding probability with the outcome within the same cohort, but using a random hold-out sample, i.e. $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) = var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}))$. We can also apply this method across different spells to find the observable heterogeneity that remains persistent. If the unobservable heterogeneity is orthogonal to the observable heterogeneity, we can identify the persistent observable heterogeneity through $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2})$. The combination of the two approaches thus allows to gauge the importance of transitory heterogeneity.

Second, mutatis mutandum, we can assess empirically how important unobservable heterogeneity is relative to observable heterogeneity, at least for the heterogeneity that is persistent across spells. Indeed, we can decompose the persistent heterogeneity into observable and unobservable persistent heterogeneity:

$$\underbrace{cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2})}_{\text{Persistent Heterogeneity (PH)}} = \underbrace{cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2}))}_{\text{Persistent Observable Het. (POH)}} + \underbrace{cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2})}_{\text{Persistent Unobservable Het. (PUH)}}. \tag{16}$$

As shown in Proposition 3, the approach to identification using multiple spell data can identify the persistent heterogeneity, both observed and unobserved, with the covariance in job finding outcomes across two spells, i.e. $cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) = cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2})$. This can be compared to the persistent observable heterogeneity identified using the predictions based on observables through $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2})$. The combination of the two methods thus allows to shed light on the importance of unobservable heterogeneity.[9]

The two equations (15) and (16) illustrate the comparative advantage of the two approaches to identification and taken together allow to provide a better understanding of the overall heterogeneity. Again, equation (16) shows the advantage of using multiple spell data, which is to capture also the persistent heterogeneity across spells, including the heterogeneity that is unobserved (PUH). Equation (15) shows the advantage of the approach using observables, which is to identify the observable heterogeneity, including the heterogeneity that is transitory across spells (TOH). The moment conditions using observables are in principle not restricted to the sample of individuals with multiple spells, but can be applied to this specific sample. The persistent observable heterogeneity (POH) in job finding is the common denominator of the two approaches, whereas the Transitory Unobservable Heterogeneity (TUH), $cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_1} - \varepsilon_d^{t_2})$, is missing from both approaches. Still, by combining the two approaches we can identify a *tighter* lower bound on the total heterogeneity in job finding:

Proposition 4. ***Lower bound with observables and multiple spell data.*** *For two randomly chosen spells for each individual, if the predictor is unbiased, i.e.* $E_d^{t_1,t_2}(\hat{F}_d^{t_i} | X_d^{t_i}) = T_d^{t_i}(X_d^{t_i})$ *for* $i = 1, 2$, *and if the unobserved heterogeneity is orthogonal to observable characteristics, i.e.* $E_d^{t_1,t_2}(\varepsilon_d^{t_i} | X_d^{t_j}) = 0$

[9] Note that the decomposition above again relies on the observable heterogeneity being independent of the unobservable heterogeneity across spells, but this is not crucial for our purposes here as shown below.

for all combinations of $i = 1, 2$ and $j = 1, 2$, the following lower bound for the true variance in types holds:

$$L = cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) + cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) - cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) \leq var_d^{t_1,t_2}(T_d^{t_1}).$$

Proof. See Appendix B.

The proposition requires that the two spells are randomly selected, which implies symmetry in the covariances (e.g., $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) = cov_d^{t_1,t_2}(F_d^{t_2}, \hat{F}_d^{t_1})$), and that that the unobserved type in each spell is independent of the observable characteristics.[10] Intuitively, the lower bound adds up equations (15) and (16) by adding $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1})$ and $cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2})$ and then subtracts the common component of persistent observable heterogeneity, identified by $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2})$, in order to avoid double counting of this term.

Further Opportunities. Proposition 4 showcases the complementarity between the two approaches for understanding different sources of heterogeneity, but further opportunities arise from combining the two approaches.

First, both approaches miss out on the unobserved heterogeneity in job finding that is transitory and specific to a spell. However, we can make assumptions comparable to proportionality assumptions in the literature studying the role of selection on observables vs. unobservables (e.g., Altonji et al. [2005], Oster [2019], Finkelstein et al. [2021]). The advantage in our setting is that we can directly assess the relative importance of unobservable heterogeneity to observable heterogeneity, at least for the heterogeneity that is persistent across spells. We can then gauge the magnitude of the transitory unobservable heterogeneity, assuming that the relative importance of the unobservable vs. observable heterogeneity is the same for the persistent and transitory component of heterogeneity. That is, $TUH = TOH \times \frac{PUH}{POH}$. Such an assumption remains speculative in nature, but we can evaluate its plausibility further by studying whether the relative importance of unobserved vs. observed heterogeneity for the persistent component remains the same in different sub-samples and at different durations.

In a similar spirit, we can evaluate how much heterogeneity estimates depend on sample compositions, which can constrain the multiple spell approach, by leveraging the approach using observables. More specifically, Proposition 4 limits the analysis of heterogeneity to the sample of individuals with multiple unemployment spells, due to the nature of the multiple spell approach. To correct for the potential differences in heterogeneity due to sample selection, we can compare the observable heterogeneity in the sample of individuals with only one spell of unemployment to the observable heterogeneity in the sample of individuals with at least two spells of unemployment, i.e., $cov_d^{t_1}(F_d^{t_1}, \hat{F}_d^{t_1})$ vs. $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1})$.

Finally, in the approach with multiple spells we exploit the heterogeneity that is persistent across cohorts t for individuals with multiple spells. As discussed in detail above, to study the selection across

[10] We note that when the observable characteristics used in the prediction can change across spells, we do not require orthogonality of observable characteristics and unobserved heterogeneity across spells when correcting the lower bound. Proposition B2 in the Appendix B establishes this modified lower bound as $\Lambda = L + cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) - cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) \leq var_d^{t_1,t_2}(T_d^{t_1})$. We note that the orthogonality across spells rules out history dependence where prior job finding affects future job finding. While this is implicitly assumed in the literature using multiple spells for identification, this can be partially relaxed to the extent that the history dependence is captured through the observables.

durations, we need to identify the heterogeneity that is persistent over the spell of unemployment, $cov_d^t(T_d^t, T_{d+1}^t)$. We can in principle apply the same approach in multiple spell data and identify the heterogeneity that is persistent across both spells and durations and thus identify the unobservable component of job finding that is persistent over the spell of unemployment and thus contributes to selection, $cov_{d,d+1}^{t_1,t_2}(F_d^{t_1}, F_{d+1}^{t_2})$.[11]

While our main focus in the empirical analysis is on using observables for identification, we will harness the complementary value of the multiple spell approach to further improve our understanding of the differences in job finding across job seekers, over the unemployment spell and over the business cycle.

3 The Predictability of Long-Term Unemployment

This section presents the prediction model, evaluates its predictive value and shows the complementary value of using multiple spells in estimating 'unobserved' heterogeneity. We start by describing the data and institutional context.

3.1 Data and Context

We merge several data sources from Sweden for the universe of prime-age job seekers starting an unemployment spell between 1992 and 2016.

First of all, we use data on unemployment spells from the Public Employment Service (PES), merged with data on UI benefit payments from the UI funds in Sweden. Unemployment benefits replace 80% of pre-unemployment earnings for workers who have worked for at least 6 months prior to being displaced and contributed to the UI system for at least 12 months. The unemployment benefit level is subject to a maximum and a minimum. Before 2001, the benefits were constant during the unemployment spell. Downward steps have been introduced in subsequent reforms for both the replacement rate and the maximum level. UI benefits are typically received for 450 business days, after which the unemployed must accept to participate in counselling activities and, potentially, active labor market programs (ALMP).[12] The PES organizes various ALMPs for unemployed workers with training programs being the cornerstone of Swedish labor market policy for many years. The ALMPs are targeted to the long-term unemployed or those who are 'typically at risk' of long-term unemployment [Richardson and van den Berg, 2013]. Our data contains information on the date the unemployed registered with the PES (which is a pre-requisite to start receiving UI benefits), unemployment benefits received and participation in the ALMPs. We define an unemployment spell as a spell of non-employment during which an individual receives unemployment benefits. To define the start date of an unemployment spell, we use the registration date at the PES. The end of a spell is defined as finding any employment or leaving the PES for other reasons.

[11] In parallel to equation (16), one can decompose the persistent heterogeneity across two spells at durations d and $d+1$ as $cov_{d,d+1}^{t_1,t_2}(T_d^{t_1}, T_{d+1}^{t_2}) = cov_{d,d+1}^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_{d+1}^{t_2}(X_{d+1}^{t_2})) + cov_{d,d+1}^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_{d+1}^{t_2})$ + cross terms. Note that the sample selection is more severe in this case compared to the case illustrated in equation (16), because not only we need to observe two spells for the same individual but also it requires that the unemployed job seeker did not find a job between durations d and $d+1$ in the second spell.

[12] See Kolsrud et al. [2018] for more details on the UI schedule in Sweden.

Second, the unemployment data are linked with the longitudinal dataset LISA which merges several administrative and tax registers for the universe of Swedish individuals aged 16 and above. In addition to socio-demographic information (such as age, family situation, education, county of residence, immigration, etc.), LISA contains exhaustive information on earnings, taxes and transfer and capital income on an annual basis. It also includes data on the occupation and industry of workers. We use the matched employer-employee register (RAMS) to obtain further information on workers' employers and their tenure prior to becoming unemployed.

Third, for selected years and sub-samples, we have access to additional data sources. This includes granular data on asset portfolios, including liquid bank accounts, outstanding debt and other financial and real asset holdings for the universe of households, but only up to 2007 when the wealth tax in Sweden was abolished. We use data on union membership and contributions to the UI system [Landais and Spinnewijn, 2021]. Sweden is with Iceland, Denmark and Finland, one of the only four countries in the world to have a voluntary UI scheme, in which an important role is played by the unions who administer the unemployment benefits. Workers who have not contributed to obtain the comprehensive UI coverage receive the minimum benefit level instead. The premium for comprehensive UI coverage was heavily subsidized, but this subsidy has been reduced in a reform in 2007. Around 80-90 percent of workers have been covered by comprehensive UI. Finally, we have IQ data for men from military enlistments and we have also merged survey data from SILC (Survey of Income and Living Conditions) to our administrative data with questions related to health and mental well-being.

Table 1 lists the range of variables used in the baseline model, which are generally available for all years in the sample period and thus allows estimating the baseline model for every year in a consistent manner.[13] We only use pre-determined variables, i.e. variables that predate the unemployment spell in question. The set of variables thus includes basic socio-demographics, which are usually available in labor force survey data; migration history, including citizenship and years since immigration; yearly income, both wage and non-wage, as well as income from other household members, for each year over the 5 years pre-dating the year of the unemployment spell; employment history over the five years, pre-dating the year of the unemployment spell, including the employment status in November of each year, the number of unemployment spells, days on UI, days on DI, and the number of employers. In our baseline model, we include these variables for both over the last 2 and the last 5 years to capture the timing of the employment histories. We also include job tenure at the pre-unemployment firm and its characteristics (size and employment growth and layoff rate); finally, our baseline model includes 3-digit industry and municipality dummies. Additional variables – which are not available in all years or only for a subset of individuals – are used in extensions of the baseline model for the year 2006.

Table 2 shows descriptive statistics for the sample of unemployed job seekers in our sample. Panel A compares the sample of unemployed to the overall population and shows that the sample of unemployed is selected towards the young, foreign and low-education and low-income population. Overall, our data over the years 1992 to 2016 features over 7 million unemployment spells. An important observation is that many of these spells include the participation in ALMPs including training and job search

[13] A limitation to this is that, since the LISA panel only starts in 1990 and even later for days spent on UI and DI, income and employment histories are partly censored for the earliest years in our sample. We impute these censored observations using the individual's history when partially observed, falling back on the population mean in 1995 when the full history is missing.

Table 1: Variables Included in Baseline Prediction Model

Socio - demographics	Migration history	Income (last 5 years)	Employment (last 5 years)	Industry and municipality
Educational attainment	Years since migration	Labour Income	Number of unemployment spells	3-digit industry code
Age	Citizenship	Income from other sources	Days on UI	Municipality
Gender		Income of other HH members	Days on DI	
Marital status			Employment status	
Number of kids			Number of employers	
Age of kids			Pre-unemployment job tenure	
Foreign citizenship			Pre-unemployment firm (size, growth rate, layoff rate)	

Notes: For income histories, we include variables capturing annual income during the year before the unemployment spell and total income during the preceding 4 years. For employment histories, we compute the variables over the last 2 years and last 5 years.

assistance, though a majority of them start at 12 months of unemployment or later. We include these spells in our baseline prediction exercise, but in a series of robustness checks we perform the prediction exercise without spells that enter ALMPs in the first 6 months of the unemployment spell. In Panel B, we show additional statistics for employment histories of the unemployed in our sample, including days on UI and DI, number of employers and employment spells, the pre-unemployment tenure, prior employer size and its layoff rate in the prior year. Overall, the table shows that there are substantial differences in these variables across the unemployed in our sample.

3.2 Prediction Model

Our baseline model predicts the long-term unemployment risk for spells starting in 2006. To be precise, we predict the probability that someone does not leave unemployment in the first 6 months of the unemployment spell. We will mostly report the complement of this probability, which we will refer to as the 6-month *job-finding probability* at the start of the unemployment spell. We will probe the robustness of our results to more narrow definitions of job finding and find that the predictive power of our model remains very similar. We then redo the prediction exercise for every year in the data and also estimate the 6-month job-finding probabilities for workers at 6 and 12 months of the unemployment spell in 2006. As in any prediction exercise, there is a trade-off between improving predictive power and over-fitting when including more variables. To address this issue, we use machine learning techniques to optimally choose variables to include in the prediction model and generate predictions. More specifically, we use an Ensemble Model (see, e.g., Einav et al. [2018]), which combines three different Machine Learning (ML) algorithms: LASSO, Gradient-Boosted Decision Trees, and Random Forest. These models take different approaches for the selection of variables, but also allow differently for non-linearities and interactions between these variables. The Ensemble Model trains each ML algorithm separately and, in a final step, assigns each of these three algorithms a linear weight. All results presented here use the predictions of this Ensemble Model, but evaluated in the hold-out sample.

Table 2: Descriptive Statistics

A. Unemployed Sample vs. Population	**Mean**			
	Unemployed	Population		
Age	36.0	40.0		
Female	48%	49%		
Foreign	16%	8%		
Primary Edu.	21%	16%		
Secondary Edu.	57%	54%		
Tertiary Edu.	21%	29%		
Labour Income (2006 SEK)	100,500	199,700		
Other Income (2006 SEK)	12,100	14,900		
Household Income (2006 SEK)	106,200	156,200		
Number of unemployment spells	7,259,797			
Spells interrupted by training	1,628,080			

B. Employment History	**Mean**	**Percentile**		
		25th	50th	75th
Days on UI (2y)	113	0	48	190
Days on UI (5y)	251	0	156	409
Days on DI (2y)	34	0	0	0
Days on DI (5y)	66	0	0	19
Unemp. Spells (5y)	1.4	0	1	2
Employers (5y)	1.9	1	2	3
Tenure (years)	1.8	1	1	3
Firm Size	4,460	15	147	3,063
Firm Layoff Rate	35%	17%	25%	48%

Notes: Descriptive statistics for the baseline sample for the years 1992-2016 and ages 25-54.

We provide more detail on the Ensemble Model and the tuning and estimation of the underlying prediction algorithms in Appendix C.

To evaluate the accuracy of our prediction model, we compare predictions and outcomes in the hold-out sample for the year 2006 in Panel A of Figure 1. The figure shows a binned scatter plot of the averages of the 6-month job-finding rate for 20 equally sized bins of the predicted 6-month job-finding probability. Typically, attenuation of outcomes from the 45-degree lines suggest issues with sampling error due to limited sample size, which is not a significant issue here as the prediction exercise yields outcomes that are close to the 45-degree line. The individual-level linear regression, shown as a red line, of a dummy for finding a job within 6 months on the predicted 6-month job-finding probability has a slope of 1.08 (0.01).

Our prediction algorithm thus seems to work very well, producing predictions that are subject to minimal bias. In Appendix C, we gauge this further by performing this evaluation for different groups separately. For the subgroup analysis, we split the sample by income, citizenship, gender, education, days on UI and days on DI, but we also consider a split into 36 groups, based on income decile, gender and citizenship, and into 144 groups based on income decile, gender, citizenship, days on DI and days on UI. The slope remains close to one in each of these separate analyses, showing that our prediction exercise also does well for different prediction models and sub-groups.

Finally, we also evaluate the accuracy of the three predictions algorithms separately and compare it to a linear model estimated with the same variables as in the baseline ML model. As the least square estimates of the linear model are unbiased, this is assessing the importance of sampling error in attenuating the relation between predictions and outcomes. As shown in Appendix C, we again find that the slopes are very close to 1, just like for the Ensemble Model. This also suggests that our approach does not suffer from any biases that may arise in the non-linear ML algorithms.

Overall, we conclude that our prediction exercise performs very well and we do not detect any biases in the prediction, neither for job seekers with specific predicted job-finding probabilities, nor for job seekers belonging to specific observable groups.

3.3 Predictive Value

We use our prediction model to assess the heterogeneity in long-term unemployment risk. Panel B of Figure 1 shows the distribution of the predicted 6-month job-finding probability in the hold-out sample for the year 2006. The figure shows substantial dispersion in the job-finding probability, with our predictions covering almost the entire range from 0 to 100 percent. The average job-finding probability equals 71 percent. For the 80th percentile of the sample, the predicted job-finding probability is as high as 85 percent. The distribution has a long left tail. The 20th percentile corresponds to a predicted job-finding probability of 58 percent, the 5th percentile to a probability of 36 percent. The predicted job-finding probability is subject to error in the prediction model and thus Figure 1 can in principle exaggerate the dispersion in the 6-month job-finding risk. As shown above, the covariance of the predicted job-finding rate with actual job finding in the hold out sample gives an accurate estimate of the variance in job-finding risk. Our estimate of the covariance is 0.029, which is substantial and slightly above the variance of 0.027. This suggests that the prediction error is rather small, confirming the results from Panel A of Figure 1. We also find that the R-squared of the predicted job-finding rate with actual job finding is 0.150. To address the issue of sampling error, we also bootstrapped the 95% confidence interval for these moments and find its fairly tight with [0.029, 0.030] for the covariance and [0.146, 0.154] for the R-squared.[14] The R-squared may appear fairly low, but the dependent variable is a binary, random realization of a probability. The low R-squared thus reflects the noise associated with this binary realization.[15] A complementary way to evaluate the predictive power of the model is to consider the area under the ROC. This curve contrasts the false-positive rate to the true-positive rate in the hold-out sample, depending on the threshold used to map the predicted job-finding probabilities into a binary outcome. The area under the ROC equals 0.73, compared to a lower bound of 0.5 for a random prediction model and an upper bound of 1 for a model with perfect foresight.

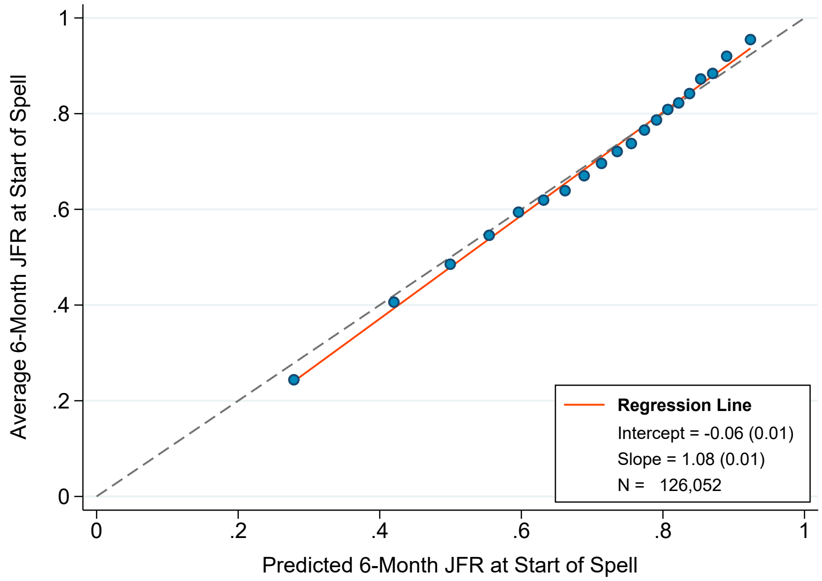
Our baseline model studies job-finding probabilities at the start of unemployment spells that

[14] We draw 500 bootstrap samples from the hold-out sample in 2006, following Mullainathan and Spiess [2017].

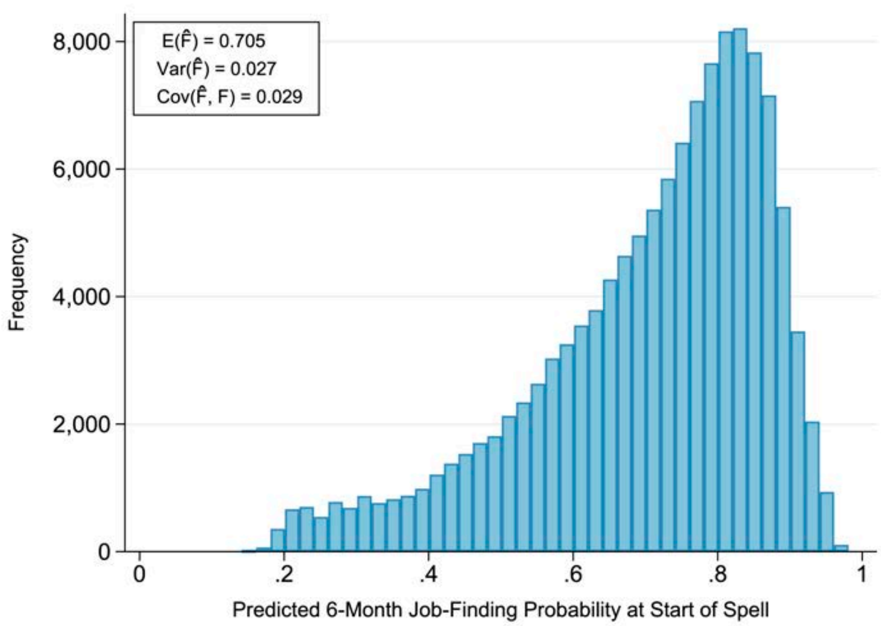
[15] For example, if T is uniformly distributed on interval [0, 1], then $R^2(F, T) = 1/3$.

Figure 1: Baseline Prediction Model in 2006

A. Comparing Predictions to Outcomes



B. Distribution of Predicted Job-Finding Probability



Notes: Panel A presents a binned scatter plot of observed and predicted job finding. 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. Panel B shows the distribution of the predicted job-finding probability in the hold-out sample for the year 2006.

initiated in 2006, which is prior to the Great Recession. Panels A and B of Figure 2 illustrate how the distribution of predicted job-finding rates changes when considering instead job seekers who are long-term unemployed or became unemployed during the Great Recession. In particular, Panel A compares the distributions for individuals six months into the spell vs. at the start of the spell, while Panel B compares the distributions for individuals at the start of the spell in 2009 vs. 2006. Not surprisingly, both for the long-term unemployed and the recession year, the job-finding rates are substantially lower on average. In both cases, few individuals remain who are almost certain to find a job in the next six months, while the predicted job-finding probabilities become more compressed in the bottom range of the distribution. As discussed in the conceptual framework, the observed differences in job finding can result from compositional changes in the pool of unemployed job seekers, but also from dynamics in job-finding chances, which can be heterogeneous too. We aim to separate the different forces in the next two sections.

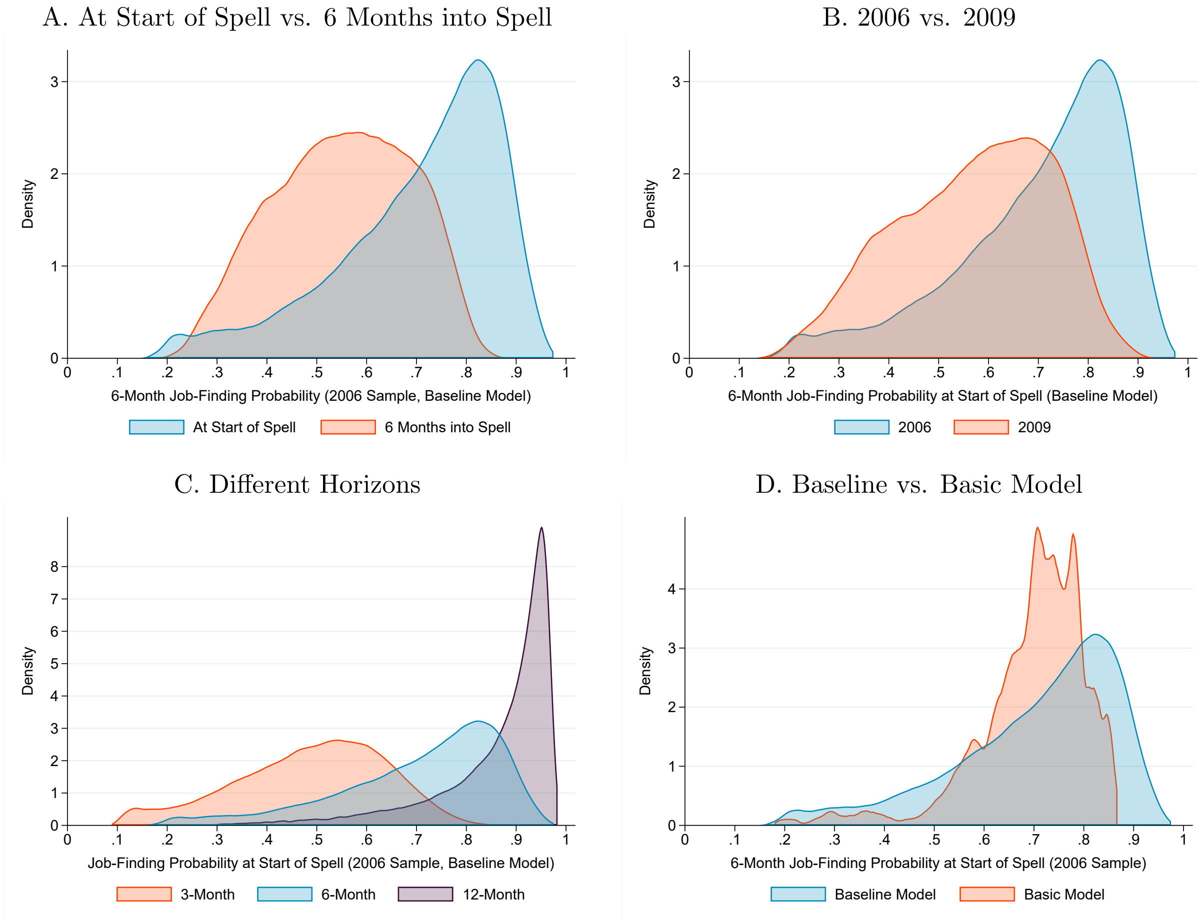
Further Robustness. We perform a series of robustness checks where we evaluate the predictive value and accuracy over different horizons, for different samples and for different models.

First, while our focus is on job finding over a 6 month-horizon, as it corresponds to the standard measurement of long-term unemployment risk, we can evaluate the job-finding rates over different horizons too. Panel C of Figure 2 shows substantial dispersion in predicted job finding when considering instead job finding over the first 3 or 12 months of the unemployment spell. The prediction exercise yields a similar predictive power when the job-finding rate is measured at 3, 6 and 12 months of the unemployment spell (see Appendix Table A1). Moreover, when we compute the R-squared of the prediction for the 3- and 12-month job-finding rate with actual job finding over the first 6 months, we find that the R-squared is very similar compared to using the 6-month model. This is re-assuring and suggests that our results do not rely on a particular horizon chosen for the job-finding rate.

Second, we gauge how our prediction is affected by our definition of job finding and the transition through active labor market policies. We first consider active labor market policies (ALMP) started in the first 6 months of the unemployment spell. Figure A1 in the Appendix shows that these spells are not strongly correlated with predicted long-term unemployment risk. Appendix Table A2 also shows that the R-squared and covariance change little when we exclude these spells from the hold-out sample. We also re-estimate the model after excluding all spells with ALMP starting in the first 6 months and we find that the R-squared remains similar both in the hold-out sample of spells with the ALMP spells included ($R^2 = 0.147$) and excluded ($R^2 = 0.146$) (see also Appendix Figure A1).

Third, we evaluate the predictive power of our baseline model when we use additional information on the end of the unemployment spell. As shown in Appendix Table A2, when we re-define job finding in the hold-out sample as those spells which ended but did not take up education other than training, the predictive power of the baseline model remains nearly unaffected. When we re-define job finding as those spells which ended but did not take up education and did not end for unknown reasons, the predictive power is somewhat lower, but still high with an R-squared of 0.128. Furthermore, when we exclude the about 10% of recalls to the previous employer from the sample, the R-squared remains nearly unaffected. This also remains true when we retrain the model excluding all observations

Figure 2: Distribution of Predicted Job-Finding Probability



Notes: This figure reports the distribution of various predicted job-finding probabilities. In all four panels, the baseline (in blue) is the predicted 6-month job-finding probability at the start of the spell for the 2006 holdout sample. Panel A compares the baseline with the predicted distribution 6 months into the spell. Panel B compares with the predicted distribution in 2009. Panel C shows the predicted job-finding probabilities over 3-month and 12-months horizons. Finally, Panel D contrasts the baseline model, which uses all the variables shown in Table 1, with the basic model, which only uses the socio-demographic variables shown in the table.

that ended in a recall.[16] The fact that the recall/new job distinction does not affect our results on observable heterogeneity is re-assuring.

Finally, we evaluate the predictive performance of the different sub-models underlying the Ensemble Model and also compare them to the linear model. Appendix Table A2 shows that all sub-models perform well, with the R-squared being the highest for the gradient-boosted model (0.150) and the lowest for the LASSO (0.134). Appendix Figure C4 confirms this ranking plotting the area under the receiver operating characteristic curve (AUC), which is a standard metric to compare predictive power of ML models. The different models also provide predictions that are highly correlated, with the correlation coefficient between 0.92 and 0.96 (see Appendix Figure C5). The Ensemble Model uses a linear combination of the three sub-models, but the gain in predictive performance relative to the

[16]We find that the recall rates in Sweden are relatively low and only slowly declining in comparison with Katz [1986]. See also Nekoei and Weber [2015].

separate prediction models is minimal.[17] Interestingly, we find that also the linear model performs very well with an R-squared of 0.137. It also provides predictions that are again highly correlated with the submodels, with a correlation coefficient of 0.89 and higher. Overall, this suggests that over-fitting is not a first-order issue when the sample of unemployment spells is large. Moreover, the potential non-linearities and interactions between variables leveraged by ML methods are not particularly important either for the high predictive power of our baseline model. Instead, what matters is the rich data going into the model.

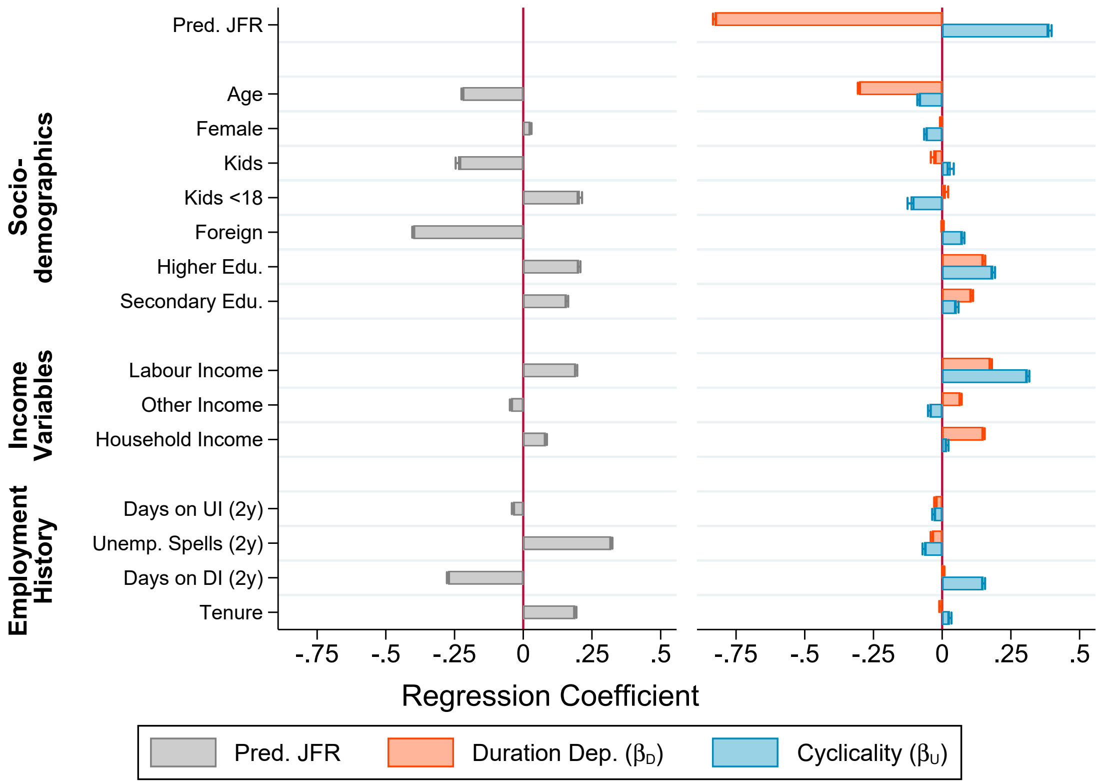
3.4 Predictive Variables

While our prediction model does not allow us to evaluate the causal impact of job seekers' characteristics on their unemployment risk, we are interested in how much different sets of variables contribute to the predictive value of the model. Prior work using labor force surveys to predict long-term unemployment has been mostly limited to socio-demographic variables, including age, gender, family composition and education. A simple linear regression already reveals that various other characteristics relate significantly to long-term unemployment risk, conditional on these socio-demographic variables. To make the estimates comparable and help with the interpretation, we first standardize both the explanatory variables and the predicted job-finding probabilities. The left panel of Figure 3 shows that on average job seekers face a higher risk of long-term unemployment when they are older and less educated. This risk is also further increased for job seekers who had lower income and tenure in their prior employment and spent more days on UI and DI prior to the unemployment spell. The strength of our analysis comes from the data-rich environment, which allows us to observe more characteristics of job seekers, but also prior employment and unemployment outcomes, which may further proxy for otherwise unobservable heterogeneity across job seekers.

We can evaluate more formally how the predictive value of our baseline model compares to a basic model, which only uses socio-demographic variables (see Column 1 of Table 1). Panel D of Figure 2 shows that the predicted dispersion using more limited information is substantially smaller. To quantify how much smaller, Table 3 reports the R-squared in the hold out sample for various sub-models. The basic model with only socio-demographic variables has a predictive power that is less than half of our full model (R-squared of 0.071). We then sequentially add variables that are increasingly unlikely to be available in surveys and have been rarely used in earlier work. As shown in the top panel of the table, the subsequent inclusion of income variables and the employment history substantially improves the predictive power of the model and these variables thus seem key for the predictability of long-term unemployment risk. Moreover, once we have added these variables, further adding income history, migration history, and location and industry effects do not add much predictive power. Clearly, the ordering of variables used for various sub-models in Table 3 matters as different predictors are correlated. To compare the marginal contributions of different sets of variables, we add each of them separately to the socio-demographics in the basic model. The bottom panel of the

[17] Note that the linear weight given to the LASSO predictions is negative for the baseline model estimated on the 2006 sample of job seekers at the start of the spell, reflecting the collinearity among the predictions. The respective weights are 0.710 for the gradient-boosted model, 0.338 for the random forest and −0.048 for the LASSO. These weights vary across years and unemployment durations, as shown in Appendix Figure C3. Restricting the weights to be positive does not meaningfully affect the predictive value of the Ensemble Model.

Figure 3: Heterogeneity in Job Finding, Duration Dependence and Cyclicality



Notes: This figure presents results from linear regressions of the predictions on a subset of observables. Both right-hand-side and left-hand-side variables have been standardized by subtracting the sample mean and dividing by the sample standard deviation, so the coefficients can be interpreted as the standard-deviation change in the outcome associated with a one-standard-deviation change in the covariate. The left panel shows the OLS coefficients of a regression of the predicted 6-month job-finding probability at the start of the spell, from the baseline model in 2006, on the variables listed on the y-axis. The right panel shows the coefficients from regressions of the duration dependence parameter β_D (see Section 4) and the cyclicality parameter β_U (see Section 5) on the same covariates and the predicted job-finding rate. In both panels, standard 95% confidence intervals are shown around the point estimates.

table reveals that adding either the income variables, the migration history, or industry fixed effects increases the R-squared by about 30 percent and thus realizes a third of the gain in predictive power when expanding from the basic to the baseline model.[18]

The most predictive set of variables regards the individuals' employment history prior to the unemployment spell. This includes number of days spent unemployed, the number of unemployment spells, but also DI receipt and number of job switches (see column 4 of Table 1). A further decomposition of the prior history shows that information in the year prior to unemployment is sufficient to realize

[18] As non-linearities do not seem very relevant, we can evaluate the contribution of different sets of variables by sequentially adding them to a linear regression of the predictions and observing how the R^2 changes. For this, we calculate a Shapley-Owen decomposition of the R^2, as described by Grömping [2007] and Huettner and Sunder [2012]. Intuitively, when the regressors are correlated, the change in the R^2 that we attribute to each variable depends on the order in which it is introduced into the regression. The Shapley-Owen decomposition overcomes this limitation by computing the average change across all possible orderings of the variables. The conclusions are very similar, as shown in Appendix Table A9.

most of the gain in predictive power, with an increase in R-squared of 67% relative to an increase of 6% or less for any additional year (see Appendix Table A4). The marginal contribution is highest when using the most recent year, but still sizeable at 13% when only adding information from five years earlier.[19]

We explore to what extent the predictive power of the baseline model can be improved further by adding additional variables available in 2006. As these variables are not available for all years between 1992 and 2016 and some are only available for a subsample, we did not include them in the baseline model. Panel C of Table 3 shows that using further information on the prior occupation (at the 3-digit level) or union membership adds limited predictive power to the baseline model. Strikingly, also adding detailed information on individuals' financial and real assets adds basically no predictive power. The same is true for IQ, which we observe for men from military enlistments. Adding information on UI benefits (only observed for 54.6% percent of the spells) does increases the R-squared by 4.4%. This is a more significant increase, but still arguably small in light of the attention given to the design of UI benefits and the corresponding moral hazard. Of course, the UI benefit level received depends on the pre-unemployment earnings and employment history, which are included in the baseline model. However, even relative to the basic model including only basic socio-demographics, the UI benefits and thus job seekers' potential search responses explain relatively little of the variation in employment outcomes (see Appendix Table A6). Adding workers' choices to get comprehensive UI or not increases the R-squared by only 0.4%. However, we note that most workers in our sample do get the comprehensive UI (70.3%) and prior work has shown relatively limited risk-based selection into comprehensive UI with most of it explained by observables included in our prediction model [Landais et al., 2021]. The final column in Panel C of Table 3 shows that the additional information from the various administrative registers jointly increases the R-squared for our prediction to 0.156 (relative to 0.150 for our baseline model). Other than information on DI receipt, our prediction model uses no information on job seekers' health. In Appendix Table A5 we briefly explore the potential predictive power of health status using survey data. The results indicate that information on mental health in particular adds additional explanatory power above and beyond our model predictions, although we cannot perfectly address concerns of over-fitting in the small survey samples.[20]

Overall, this confirms that our prediction model seems saturated along some key dimensions of heterogeneity. However, it also reminds us that the predicted heterogeneity remains a lower bound on the overall heterogeneity across job seekers.

3.5 Identifying Unobserved Heterogeneity in Multiple Spell Data

So far we have focused on unemployment spells that initiated in 2006. However, we observe the universe of unemployment spells for the period 1992-2016 and thus observe multiple unemployment spells for a sizeable share (69%) of the individuals in our data. This allows us to estimate heterogeneity following the multiple spell approach, which we can combine with our approach to obtain a tighter estimate of the overall heterogeneity as outlined in the conceptual framework in Section 2.4. We follow

[19] Panel C in Appendix Table A4 also shows that none of the specific features of the pre-unemployment history (e.g., UI receipt, number of unemployment spells, employment switches) by themselves are driving the predictive power.

[20] After constructing a mental health index using principal component analysis, we find that adding this index to our model predictions in a linear regression model increases the R-squared by 14 percent.

Table 3: R^2 FOR VARIOUS MODELS IN THE YEAR 2006

A. Sub-models of Baseline - Sequential

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_0, F_0)$	0.071	0.099	0.106	0.140	0.142	0.147	0.149	0.150
Change (j) vs ($j-1$)	-	+38.5%	+7.3%	+32.5%	+1.1%	+4.1%	+1.2%	+0.6%
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

B. Sub-models of Baseline - Marginal

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_0, F_0)$	0.071	0.099	0.081	0.137	0.100	0.089	0.098	0.078
Change (j) vs (1)	-	+38.5%	+14.5%	+92.2%	+40.3%	+24.7%	+37.9%	+9.1%
Socio-demographics	X	X	X	X	X	X	X	X
Labour Income		X						
Other Income			X					
Employment History				X				
Income History					X			
Migration History						X		
Industry							X	
Municipality								X

C. Extensions of Baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_0, F_0)$	0.150	0.151	0.151	0.150	0.150	0.151	0.157	0.156
Change (j) vs (1)	-	+0.7%	+0.3%	-0.3%	+0.1%	+0.4%	+4.4%	+3.9%
Baseline Variables	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. Panel A starts from the basic model in (1) and adds variable groups sequentially until all of the groups included in the baseline model are incorporated in (8). Panel B adds the same variable groups one at a time. Finally, Panel C starts from the baseline model in (1) and adds additional information from other administrative data sets, first one at a time and then all at once in column (8).

the same terminology and classify heterogeneity by being persistent or transitory across spells for the respective observable and unobservable components.

Table 4 puts together our estimates of the covariances between observed and/or predicted job finding rates, either across or within spells and shows how to quantify the different sources of heterogeneity. We focus on job finding rates at the start of the spell, but extend this to job finding rates later in the spell in Section 4. The analysis is also restricted to the hold-out sample of individuals with at least two spells, choosing two spells at random if the individual has more than two. To gauge sample selection, we also compare the estimates of the observable heterogeneity for the multiple spell sample and the total sample used in our prediction analysis.

Table 4 provides a number of insights. The covariance in observed job finding between two spells in the lower left corner (PH) of Table 4 identifies the heterogeneity in types that is persistent between the two spells for this particular sample. It is estimated to be 0.030. This is substantial and slightly higher than the observable heterogeneity we find in the reference sample, which includes job seekers who only experience one spell. As discussed before, under the assumption of unbiasedness in prediction, the covariance between the predicted and actual job finding for the same cohort, but evaluated in a hold-out sample, identifies the observable heterogeneity ($\tilde{OH}$), which is estimated to be 0.027. The tilde refers to the reference sample for which it is evaluated. Of course, these two estimates are difficult to compare by themselves. However, we can quantify the importance of the different subcomponents of heterogeneity by combining the two approaches on the same sample.

First, the covariance in the predicted job finding in the first spell with the actual job finding in the second spell in the upper left corner (POH) identifies the persistent observable component of heterogeneity for the multiple spell sample. It is estimated to be 0.017 or about 56% of the total persistent heterogeneity, PH. This thus shows how unobservable heterogeneity remains important in our data-rich environment and highlights the advantage of using multiple spells for identification.

Second, we can estimate the total observable heterogeneity for this multiple spell sample, using again the covariance between the predicted and actual job finding for the same cohort, which is estimated to be 0.023 (see OH in Table 4). This is 37% higher than the persistent component POH for the same multiple spell sample. This is due to transitory heterogeneity, either coming from changes in observable characteristics or changes in how observable characteristics relate to job finding across spells. While less important than the unobservable component, the transitory component is also sizeable and highlights a key advantage of using observables. Moreover, building on the approach using observables, we can compare the estimate of the observable heterogeneity for the multiple spell sample and the reference sample. Including individuals with only one spell increases the estimated covariance between predicted and actual job finding for the same cohort from 0.023 to increases to 0.027 or by an additional 17%. This suggests that there is additional heterogeneity in the total sample, which is missed when restricting the sample to individuals with multiple spells.[21]

Third, following Proposition 4, we can combine the estimated covariances using both approaches

[21] To be precise, for every pair of spells in the two-spell sample, we draw a random spell from the same year as the second spell, but including individuals with one spell only. This keeps the distribution of number of spells by year the same in the two-spell and reference sample. This explains why this estimate deviates slightly from the estimate of 0.029 reported earlier for the year 2006.

Table 4: Observable, Unobservable, Persistent and Transitory Components of Heterogeneity in Job Finding

	Multiple Spells Sample			Reference Sample
	Persistent	**Transitory**	**Total**	**Total**
Observable	$POH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_1}, F_0^{t_2})$ **0.017**	$OH - POH$ **0.006**	$OH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.023**	$\tilde{OH} = Cov_0^{t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.027**
Unobservable	$PH - POH$ **0.013**			
Total	$PH = Cov_{0,0}^{t_1,t_2}(F_0^{t_1}, F_0^{t_2})$ **0.030**		*By Proportionality* *0.042*	*By Proportionality* *0.048*

Notes: This table reports key statistics for the sample of individuals with multiple unemployment spells between 1992 and 2016. The sample size consists of 735,797 individuals. For the first three columns, we construct the two-spell sample by restricting the analysis to individuals with at least two spells in our hold-out sample. We take two of the spells, choosing at random if there are three or more, and randomly label them as "first" and "second". For the last column, we construct a reference sample by matching each second spell with a spell chosen at random from the hold-out sample of the same calendar year, without excluding unique spells. F_0^{t1} ($\hat{F}_0^{t1}$) resp. F_0^{t2} ($\hat{F}_0^{t2}$) refers to actual (predicted) job finding over the first 6 months in the first resp. second spell.

to establish a tighter lower bound for the multiple spell sample:

$$L = OH + PH - POH = 0.036. \tag{17}$$

Or alternatively, having found that the observable heterogeneity is more pronounced for the reference sample than for the multiple spell sample, we can make the same assumption for the unobservable heterogeneity and find an even tighter bound for the reference sample:

$$\tilde{L} = \tilde{OH} + PH - POH = 0.040. \tag{18}$$

Expressing this lower bound relative to the variance in job finding outcomes in the 2006 sample, this implies that at least 19% of the variation in job finding realizations can be determined 'ex ante', which can be directly compared to the R-squared of 15% for our prediction model. This highlights again the strong degree of complementarity between the two approaches. We find comparable lower bounds on the overall heterogeneity when either using the identification approach relying on multiple spells or the identification approach relying on observables. Combining the two approaches, however, we can establish a lower bound which is substantially higher.

Finally, the one component of heterogeneity that we cannot identify directly is the unobserved heterogeneity that is transitory across spells. As discussed in Section 2.4, one can gauge the potential importance of this term, by assuming that the ratio of the observable vs. unobservable heterogeneity is the same for the persistent and transitory components. Given this – admittedly somewhat speculative – assumption, we find a variance of the transitory unobservable component of 0.006, which implies a total variance of 0.042 for the multiple spell sample.[22] Assuming that the same proportionality

[22] We can assess the reasonability of the proportionality assumption by studying how the ratio of the observable vs. unobservable heterogeneity changes when studying the persistent component for different samples. Table A7 shows that the share of unobservable heterogeneity increases from 0.44 to 0.47 (0.58) when restricting the two-spell sample further to spells two (five) years apart and thus considering persistence over a longer horizon. Table A8 shows that the proportion remains unchanged when restricting the two-spell sample to those who have been LT unemployed in (at

assumption holds between the sample of multiple spells and the sample of all spells (including single spell individuals), we get an estimate of the total variance in job finding rates of 0.048 for the reference sample, increasing the explained share of variation in job finding outcomes to 23%. This suggests that our lower-bound estimate is tight relative to the overall degree of heterogeneity.

To sum up, while both approaches highlight a large role for heterogeneity in job-finding risk and of similar magnitude, they identify different elements of heterogeneity. The two approaches are thus highly complementary and jointly imply a larger role for heterogeneity in job-finding risk than identified by the methods separately.

4 Job Finding over the Unemployment Spell

The substantial heterogeneity in job-finding rates at the start of the spell implies that the long-term unemployed will differ from the short-term unemployed. This section studies how the job-finding prospects change over the unemployment spell and revisits the question how much changes in the composition of unemployed job seekers contribute to this. We also study heterogeneity in the dynamics of job finding that individuals experience over the unemployment spell.

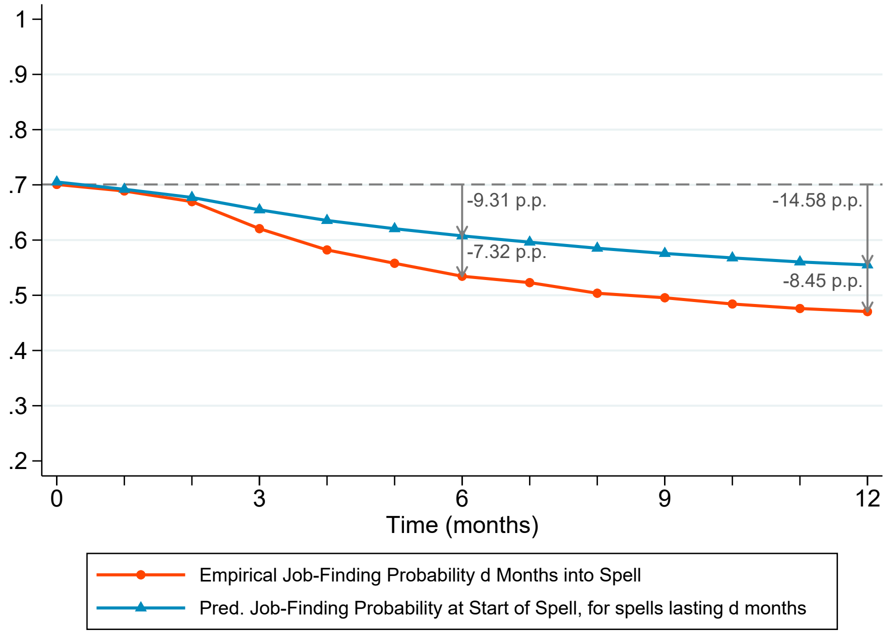
4.1 Dynamic Selection into Long-Term Unemployment

We first study how much compositional changes contribute to the average decrease in job-finding rates over the unemployment spell. Figure 4 provides a graphical illustration of the dynamic selection, using the prediction model estimated in the prior section for job seekers at the start of the unemployment spell. The figure shows how the average of these predictions changes for the pool of job seekers still unemployed at different unemployment durations. The decline is substantial. For the job seekers still unemployed six months into the spell the average predicted job finding is 9.3 percentage points lower compared to all the unemployed starting a spell. This is 56 percent of the observed drop in job-finding rates of 16.6 percentage points. For those still unemployed after one year, the average predicted job finding is 14.6 percentage points lower, corresponding now to 63 percent of the observed drop. Of course, such analysis is constrained by the observables used in our prediction model. Using our basic model with socio-demographic variables only, we explain a drop of 8 instead of 14.6 percentage points and would thus assess the dynamic selection as potentially only half as important.

The actual contribution of dynamic selection to the decline in job finding rates, however, depends on the heterogeneity in job finding rates that remains *persistent* over the spell, as discussed in Section 2. That is, in order to interpret the residual drop in Figure 4 as the dynamic effect of unemployment, i.e., as *true* duration dependence, the job-finding rates would need to be persistent over the spell. If not, the mean reversion in job-finding rates among the surviving unemployed would imply that the residual drop underestimates the true-duration dependence. Following Proposition 2 and Corollary 1, we can separate out the role of dynamic selection and provide a lower-bound by estimating the persistent observable heterogeneity, now not across unemployment spells like in Table 4, but over

least) one of the spells (0.44) and thus considering a component that remains persistent not just across spells, but also over the spell of unemployment. Overall, the share of the unobserved heterogeneity is thus fairly stable across these different samples.

Figure 4: Dynamic Selection over the Unemployment Spell



Notes: The figure compares the evolution of the empirical 6-month job-finding rate d months into the spell with the predicted 6-month job-finding rate at the beginning of the spell for individuals who reach the d-th month of unemployment, in the 2006 hold-out sample.

the unemployment spell. For this we need to predict job finding rates for job seekers at longer unemployment durations. We can then evaluate how persistent differences in job finding rates are over the unemployment spell by comparing the relative predictive value of the prediction model estimated on the unemployed later in the spell vs. on the unemployed earlier in the spell for the outcomes for the latter. This is shown in Table 5, focusing again on the sample of job seekers in 2006.

We first consider job seekers at the start of the spell and the predictive value of predictions estimated six months into the spell. We find $R_0^2(F_0, \hat{F}_6) = 0.124$, which compares to $R_0^2(F_0, \hat{F}_0) = 0.150$ when using the contemporaneous predictions. The R-squared is evaluated in the hold-out sample of newly unemployed and not just those who remain unemployed for at least 6 months. Our R-squared estimates suggest that more than 80 percent of the observable heterogeneity is persistent over the first six months of the unemployment spell. Only a small fraction of the heterogeneity estimated at the start of the spell is transitory. Following Corollary 1, we can then calculate $cov_0(F_0, \hat{F}_6)/(1 - E_0(F_0))$ to provide a lower-bound of 7.3 percentage points on the average decline in job finding due to selection. This is a lower bound as it only considers persistent observable heterogeneity, but it relies on the unobservable heterogeneity being persistent too. This conservative lower-bound implies that true duration dependence can explain at most 51% of the observed decline in the 6-month job-finding rate between the start of the spell and at 6 months of the spell.

We perform the same calculation for the decline in the job-finding rate between 6 and 12 months of the spell and find a lower-bound decline due to selection of 0.025. This is about 38% of the total decline in the 6-month job-finding rate from 0.55 to 0.49 and thus less than during the first six months of the

Table 5: Predictable Heterogeneity during the Unemployment Spell

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	126,052	0.701	0.705	0.027	0.029	0.150
	6M Model	126,052	0.701	0.604	0.018	0.022	0.124
	12M Model	126,052	0.701	0.536	0.009	0.013	0.093
6M into Spell	6M Model	42,823	0.552	0.551	0.019	0.019	0.076
	12M Model	42,823	0.552	0.504	0.009	0.011	0.056
12M into Spell	12M Model	20,831	0.485	0.489	0.009	0.009	0.037

Notes: The table reports summary statistics about models trained on different unemployment durations in 2006. 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.

unemployment spell. The declining role of selection over the unemployment spell can be attributed to the fact that dynamic selection affects the sample composition of long-term unemployed. The reason is that, if heterogeneity is persistent, the variance in predicted job-finding rates among survivors should decline over the unemployment spell. Indeed, we find that the hold-out sample variance in predicted 6-month job-finding rates is 0.027 at 0 months of unemployment, but declines to 0.019 at 6 months and 0.009 at 12 months of unemployment.

As mentioned, our analysis relies on the richness of the observables used in our prediction model. Using a prediction model with only basic socio-demographic variables, we find some persistence over the unemployment spell, but we would attribute much less of the decline in job finding to dynamic selection. For the first six months of the sample for example, we would explain only 22% of the decline compared to about half of the decline using our baseline prediction model (see Appendix Table A9). By the same token, we can use our estimates of the unobserved heterogeneity using multiple spells to turn the lower-bound estimate into an estimate of the overall contribution of dynamic selection to the decline in job finding rates. Indeed, we showed in Table 4 above that the extent of observed heterogeneity at the start of the spell in the multiple spell sample is about 56% of the extent of total heterogeneity. If one assumes that the selection into long-term unemployment on observables and unobservables is proportional, we can scale the covariance in Column 4 of Table 5 up by a factor of $\frac{1}{0.56} = 1.76$. Using the same calculation as above, this would imply that we can even account for 88% of the decline in job finding over the spell of unemployment.

Robustness. We perform three robustness checks on these results. First, we address issues of sampling error due to the smaller sample sizes of unemployed at 6 and 12 months of duration.[23]

[23]Note that these predictions models are indeed estimated on much smaller samples and thus potentially subject to more sampling error. However, as can be seen from Panels B and C of Appendix Figure C6, there is basically no attenuation when comparing outcomes to predictions. Moreover, splitting the sample again into 36 groups based on

To do this, we pool the years 2006 and 2007 and redo our prediction exercise. Appendix Table A10 reports the corresponding results, which are very similar to the results in Table 5. In fact, we find that $R^2(F_0^{t_0}, \hat{F}_6^{t_0}) = 0.136$ in the hold-out sample of 2006 and 2007, compared to $R^2(F_0^{t_0}, \hat{F}_0^{t_0}) = 0.162$, and following Corollary 1, we find a very similar number for $cov_0^{t_0}(F_0^{t_0}, \hat{F}_6^{t_0})/(1 - E_0^{t_0}(F_0^{t_0}))$, accounting for 45% of the observed decline in the 6-month job-finding rate between the start of the spell and at 6 months of the spell. This suggests that sampling error is not biasing our conclusions.

Second, we redo the same analysis for 2009-2010 when both unemployment and LTU risk were higher and report the results in Table A11. We find that $R^2(F_0^{t_0}, \hat{F}_6^{t_0}) = 0.097$ in the hold-out sample of 2009-2010, compared to $R^2(F_0^{t_0}, \hat{F}_0^{t_0}) = 0.126$. While the persistent share is estimated to be equally important, the predictable heterogeneity is thus somewhat smaller in these recession years. The observed duration dependence in job-finding rates is, however, smaller as well during the recession years: the relative decline in the 6-month job-finding rate 6 months into the spell equals 18% in 2009-2010 compared to 24% in 2006-2007. This is consistent with prior evidence for the US in Krueger et al. [2014]. Separating out the role of dynamic selection following Corollary 1, we find that it accounts for at least 40% of this observed decline in the 6-month job-finding rate between the start of the spell and at 6 months of the spell. This is only slightly lower than the lower bound during the pre-recession years 2006-2007.

Third, we can use the multiple spell approach more directly to account for unobservable heterogeneity that is persistent over the spell. For this, we consider again individuals who experience at least two spells, but are also long-term unemployed in at least one of the spells. We can then compare their job finding at the start and later into the spell across these two spells, as shown in Appendix Table A8. We find a covariance $Cov_{6,0}^{t_1,t_2}(F_6^{t_1}, F_0^{t_2}) = 0.016$. We first note that this is smaller than the corresponding covariance at the start of the spell, $Cov_{0,0}^{t_1,t_2}(F_0^{t_1}, F_0^{t_2}) = 0.030$ (see Table 4), simply confirming that not all heterogeneity that is persistent across spells is also persistent over the spell. At the same time, the multiple spell covariance $Cov_{6,0}^{t_1,t_2}(F_6^{t_1}, F_0^{t_2}) = 0.016$ misses heterogeneity that is persistent over the spell, but transitory across spells. We can evaluate this using our observable approach for this sample, $Cov_{6,0}^{t_1,t_2}(\hat{F}_6^{t_2}, F_0^{t_2}) = 0.014$, and compare this to the common component underlying both covariances, $Cov_{6,0}^{t_1,t_2}(\hat{F}_6^{t_1}, F_0^{t_2}) = 0.009$. Empirically, we thus find that the observable heterogeneity that is transitory across spells ($0.014 - 0.009 = 0.005$) is almost as important as the unobservable heterogeneity that is persistent across spells ($0.016 - 0.009 = 0.007$). Both sources of heterogeneity thus contribute to dynamic selection over the spell and the decline in observed job-finding rates. This also directly supports our assumption in Corollary 1 that unobservable heterogeneity is (weakly) persistent over the unemployment spell and thus the lower bound nature for the observable heterogeneity that is persistent over the unemployment spell.

We can combine the two approaches to provide a tighter lower bound. But, of course, when using the multiple spell approach to speak directly to persistent heterogeneity over the spell, the sample becomes even more selected, as we illustrate in Appendix Table A8. Interestingly, also for this selected sample the ratio of observable heterogeneity to total heterogeneity is 0.56, which turns out to be exactly the same as what we found at the start of the spell (see Table 4). This lends credibility to our exercise above where we invoke the proportionality assumption instead. Indeed, the ratio of

income decile, gender and citizenship and computing the average observed and predicted job-finding rate for each group in the hold-out sample, we find again that the slope is very close to one, suggesting that the predictions remain unbiased.

observable and unobservable heterogeneity appears to be stable across different types of heterogeneity (persistent heterogeneity over spell vs. overall heterogeneity).

Discussion. Overall, our estimates show that heterogeneity is important in explaining the observed dynamics of job finding over the spell of unemployment. This corroborates the findings in recent work. Mueller et al. [2021] document substantial predictability based on elicited beliefs about the job-finding probability in U.S. survey data. Using a model of beliefs, they then show that selection accounts for 85% of the observed decline of job finding over the first 12 months of the unemployment spell. Of course, we should not necessarily expect the role of dynamic selection to be the same in the US and Sweden, but their estimates are very close to our estimates when we account for unobserved heterogeneity too. Also Alvarez et al. [forthcoming] and Güell and Lafuente [2022] find a large role for heterogeneity using repeated unemployment spells. As discussed, their method identifies both observed and unobserved heterogeneity in job-finding risk, but only to the extent that it is persistent across spells.[24]

Our analysis has highlighted the importance of using a rich set of observables to assess the role of dynamic selection for the observed decline in job finding. For example, Kroft et al. [2016] find little selection into long-term unemployment based on educational attainment, age, race and gender in U.S. data, comparable to what we find when limiting the set of observables. Our findings also point to a lesser role for true duration dependence than often suggested in earlier work. For example, using a resume-audit study in the US with job seekers ages 40 and younger, Kroft et al. [2013] find that call-back rates to interviews decline by about 40% over the first 12 months of the unemployment spell. We find that, in the Swedish context, dynamic selection can account for at least a 10 percentage point decline of the observed decline in job finding from 70% to 49% over the first 12 months of the unemployment spell. This leaves at most a 11 percentage point decline of job finding, or 16% decline relative to the initial job-finding rate, for true duration dependence in job finding. Assuming contexts are comparable, this implies that true duration dependence in job finding is substantially less important than duration dependence in call back rates. This is in line with Jarosch and Pilossoph [2018] who show that a decline in call-back rate may not result in a meaningful decline in the job-finding rate.[25]

[24] Alvarez et al. [forthcoming] consider unemployment spells in Austria and find that dynamic selection is particularly important over the first 6 months of the unemployment spell, which we find as well in our data. Güell and Lafuente [2022] consider unemployment spells in Spain and explicitly quantify that 18% of the 'ex-post' variation in unemployment spell durations is due to 'ex-ante' heterogeneity. Comparing our results with Güell and Lafuente [2022], the R-squared quantifies the share of variation in realized job finding that is predictable using observables. This is 15% in our baseline sample. Using the multiple spell approach instead, we would estimate the heterogeneity to explain about 14% of the variation in realized job finding (scaling PH in Table 4 by the variance of job-finding outcomes in 2006). This share is somewhat smaller than their estimate. Note of course that there is no one-to-one mapping between the variance of job finding and the variance in unemployment durations, due to non-linearities in the relationship between the two concepts, but we verified in simulations that the two compositions give similar though not equivalent answers.

[25] Note that Kroft et al. [2013] find that the decline in call-back rates is less pronounced when the unemployment rate is high. Comparing our estimates prior and during the Great Recession, we find that the observed duration dependence indeed decreased, but our lower-bound on the role of dynamic selection, if anything, decreases.

4.2 Heterogeneity in Duration Dependence

We can add more structure to the dynamics in job finding and use the data to estimate the observable heterogeneity in these dynamics. In particular, we assume the following model for predicted job-finding rates by duration of unemployment d for a given cohort t (we drop the latter superscript here for ease of notation):

$$log(\hat{F}_d(X)) = \beta_0(X) + \beta_D(X)d + \eta_d, \tag{19}$$

where $\hat{F}_d(X)$ is the predicted job-finding rate at unemployment duration d for an individual with observables X prior to unemployment.[26] An important feature of this exercise is that we can estimate this model for any individual simply relying on the predictions given his or her observables X prior to becoming unemployed, regardless of when he or she found a job. That is, we compute $\hat{F}_d(X)$ at durations $d = 0, 6$ and 12 months for all individuals unemployed at the start of the spell based on their X. We then use these predicted job findings to estimate an observable-specific intercept $\beta_0(X)$ and slope $\beta_D(X)$ for each X in the baseline year 2006. Of course, we should again interpret the estimated slopes as an upper-bound on *true* duration dependence, as some of the decrease in job finding for individuals with observables X who remain unemployed could be driven by selection on unobservables driving low job-finding. The estimated coefficients are measured with error and thus the dispersion in the estimated coefficients reflects both actual dispersion as well as sampling error. To address this issue, we shrink each observable-specific prediction using the standard errors of the same regressions, as follows:

$$\tilde{\hat{\beta}}_j(X) = E(\hat{\beta}_j(X)) + \sqrt{\frac{var(\hat{\beta}_j(X)) - (\hat{\sigma}_j(X))^2}{var(\hat{\beta}_j(X))}}[\hat{\beta}_j(X) - E(\hat{\beta}_j(X))], \tag{20}$$

for $j \in \{0, D\}$. $E(\hat{\beta}_j(X))$ and $var(\hat{\beta}_j(X))$ are equal to the sample mean and variance of the predictions and $\hat{\sigma}_j(X)$ is the standard error corresponding to prediction $\hat{\beta}_j(X)$.

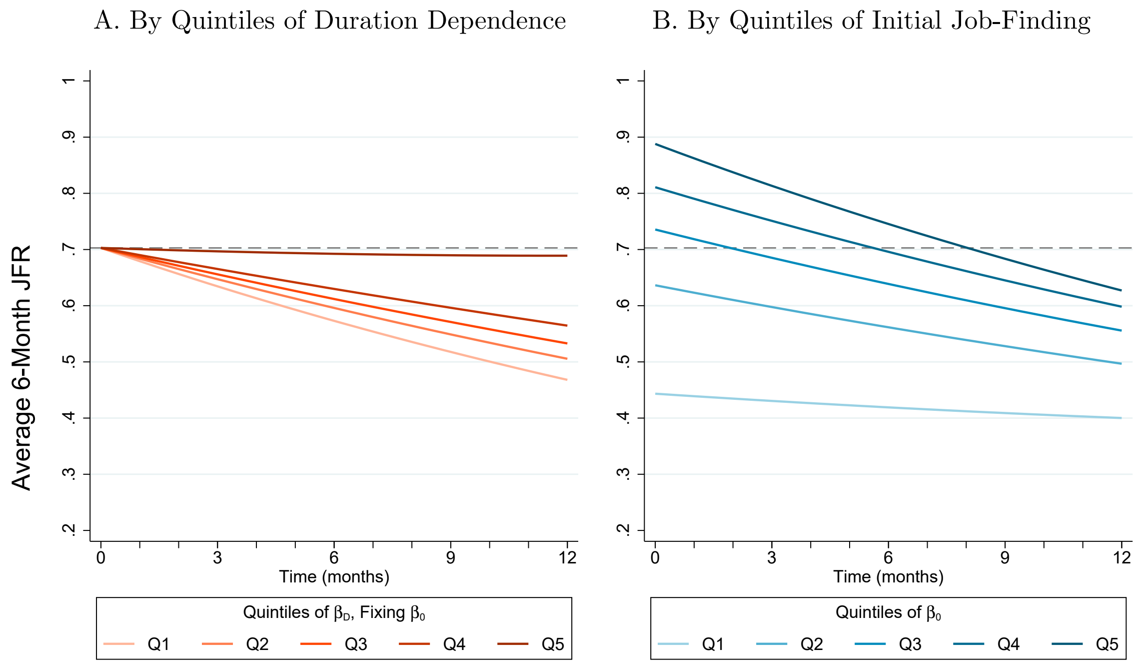
Our estimates reveal substantial heterogeneity in individual job finding dynamics. The distribution of the (shrunken) estimates of $\hat{\beta}_D(X)$, shown in Appendix Figure A5, displays significant dispersion. We find an estimated standard deviation of 0.0131 with a bootstrapped confidence interval of [0.0130, 0.0132].[27] Because we use our predictions in logs in equation (19) above, this by itself implies that job-finding rates do not vary proportionally throughout the unemployment spell. We also carry out a second test for the heterogeneity in individual dynamics that does not impose the linearity in duration d in the above equation, by computing the log of the ratio of the predicted job finding rate for a subset of observables in our data, i.e. $\log\left(\frac{\hat{F}_6(X)}{\hat{F}_0(X)}\right)$. As reported in Appendix Figures B1 and B2, we find substantial heterogeneity in this ratio across different X's, which again is in contradiction with the assumption of proportionality in proportional hazard models.[28] Importantly, this rejects a common assumption in models of job search. Indeed, a large number of papers have estimated

[26] Since we implement this test only for the baseline year 2006, we dropped the t in the notation. See Appendix B.4 for further details.

[27] Again following Mullainathan and Spiess [2017], we draw 500 bootstrap samples from the 2006 hold-out sample.

[28] See also Appendix B.4 for the theoretical underpinnings derived from the proportional hazard model Cox [1972].

Figure 5: Heterogeneity in Duration Dependence



Notes: Panel A shows the predicted individual change in the job-finding rate for the five quintiles of the distribution of β_D, assuming a job-finding rate of 0.703 at the start of the unemployment spell. Panel B shows the predicted change at the individual level in the job-finding rate for the five quintiles of the distribution of β_0.

mixed proportional hazard models, assuming full proportionality in the job-finding rates over the unemployment spell along observable dimensions. Our rejection corroborates the findings in Alvarez et al. [forthcoming], who find evidence of non-proportionality in data on repeated unemployment spells in Austria.[29]

Panel A of Figure 5 illustrates the magnitude of the estimated heterogeneity in individual dynamics by showing how the job-finding rates evolve over the unemployment spell for the five quintiles of $\hat{\beta}_D(X)$, all starting from the average job-finding rate of 70% for comparability. This clearly evinces the significant dispersion. In fact, for the top quintile, our exercise predicts that the job-finding rate stays basically the same, whereas for the bottom quintile it declines by around 33% over a 12-month spell. Panel B of Figure 5 shows how the predicted individual level dynamics relate to the initial job-finding rate. More specifically, it shows the predicted changes by quintile of the distribution of the intercept $\hat{\beta}_0(X)$. Clearly, the decline in the job-finding rate is strongly correlated with the initial job-finding rate. In fact, for the bottom quintile of the initial job-finding rate, we predict a small decline of 4 percentage points. For the top quintile, however, we predict a decline of 26 percentage points. Hence, the predicted job-finding rates would converge as the spell continues for *all* job seekers

[29] Alvarez et al. [forthcoming] test for proportionality by looking at the job-finding hazard at a given duration in the second spell conditional on finding a job early vs. late in the first spell. The proportional hazard assumption implies that the hazard in the second spell for those who find a job early in the first spell should dominates the hazard for those who find a job late in the first spell at all durations in the second spell. They strongly reject the proportionality assumption. At the same time, their test requires that there is no transitory heterogeneity across spells. For this reason, we view our test complementary to theirs as it holds even with transitory heterogeneity across spells.

in the spell. As a result, both the heterogeneous dynamics and the dynamic selection contribute to the compression of the job finding distribution among job seekers who remain unemployed for longer. We note again that we cannot rule out that the heterogeneity in the predicted individual dynamics is driven by differential dynamic selection based on unobservables. However, it seems unlikely that the dynamic selection would be so important for job seekers with the highest predicted job-finding rates and fully absent for job seekers with the lowest predicted job-finding rates at the start of the spell.

Finally, we briefly study which type of job seekers are more at risk of declining job-finding rates. Following the same approach as for the heterogeneity in job finding at the start of the spell, we correlate observable characteristics with the dynamic component $\hat{\beta}_D(X)$ in the right panel of Figure 3. Like in the left panel, we again standardize both the outcome and explanatory variables. We confirm the finding that the most pronounced gradient is in the individuals' job finding at the start of the spell. However, the dynamics in job-finding rates do not only differ across initial job-finding rates, but also conditional on initial job-finding rates. This necessarily leads to rank reversals in predicted job finding over the unemployment spell. Indeed, conditional on initial job finding, we for example find that job seekers experience stronger declines in job finding during unemployment when they are older, have lower income and are less educated.[30] While several observables correlate significantly with the dynamic component $\hat{\beta}_D(X)$, the estimates are all relatively small.[31]

To sum up, we find significant heterogeneity in duration dependence across individuals with different observed characteristics. Our findings reject the common assumption of proportionality used in standard models of job finding.

5 Job Finding over the Business Cycle

This section turns from the dynamics in job finding over the unemployment spell to the dynamics in job finding over the business cycle. Like for the dynamics over the spell, an important, outstanding question is to what extent the dynamics over the business cycle are driven by compositional changes in the pool of unemployed or by changes in the job-finding rates themselves. In the latter case, the natural follow-up question is whether the cyclicality in job finding varies across job seekers.

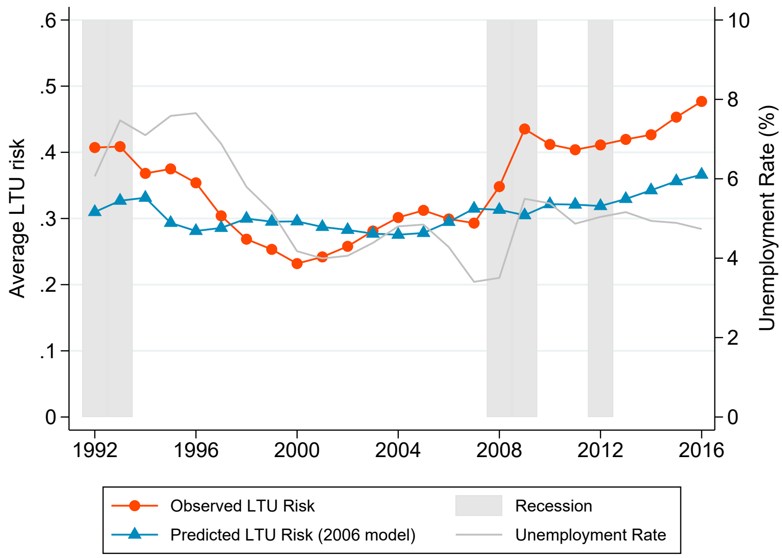
5.1 Cyclical Selection into Unemployment

We first study the role that selection into the pool of unemployed plays for the cyclicality of the average long-term unemployment risk. Figure 6 shows the average long-term unemployment risk for each year in our sample period as well as the aggregate unemployment rate. There is a strong positive correlation between the two, with a correlation coefficient of 0.3. The increase in long-term unemployment risk is particularly notable during the Great Recession period, as well as the decrease

[30] Our findings are somewhat different from Eriksson and Rooth [2014] who find using randomized CVs that the call back rate is significantly lower only for job seekers after 9 months into unemployment, though they do find heterogeneity in the duration profile as the call backs only decline for job seekers with low-medium skills jobs. Again, they study only call-backs, which may not translate into job finding.

[31] Appendix Figure A8 shows the corresponding bivariate correlations for comparison. We also report the Shapley-Owen decomposition of the R^2 to assess the explanatory power of the different groups of variables in Appendix Figure A9, mirroring the analysis done for the baseline model. The results again confirm that the job-finding rates at the start of the spell jump out in explaining the variation in duration dependence.

Figure 6: Compositional Effects over the Business Cycle



Notes: The figure shows the averages of 1 minus the 6-month job-finding rate in the hold-out sample for the years 1992-2016, the averages of the predicted long-term unemployment risk using the 2006 model, and the yearly averages of the unemployment rate (see Appendix Figure A4 for a comparison with OECD data). The grey shaded areas correspond to periods with two consecutive quarters of negative growth in Gross Domestic Product.

after the large recession in Sweden in the beginning of the 1990s. The substantial variation in long-term unemployment risk over the business cycle is also a feature of other major developed economies, including the U.S. (see, e.g., Elsby et al. [2009], Shimer [2012], among many others).

We revisit the heterogeneity hypothesis, i.e., whether compositional changes in the pool of unemployed do translate into higher LTU risk in recessions, by using the 2006 model to predict the long-term unemployment risk of newly unemployed job seekers in each year in our sample. Figure 6 shows how the average long-term unemployment risk predicted by the 2006 model but in the sample of newly unemployed in each year changes over time.[32] By keeping the prediction model fixed, changes in this counter-factual long-term unemployment risk only reflect compositional changes in the pool of unemployed. As is clear from the figure, there appears to be no distinct relationship of the predicted long-term unemployment risk with the aggregate unemployment rate, much in contrast to the observable long-term unemployment risk. As shown in Appendix Table A12 the raw correlation of the predicted and the aggregate unemployment rate is very close to 0. The bi-variate regression coefficient with the log unemployment rate as dependent variable grows slightly larger (0.082) when we control for a linear time trend but is only at 12% of the one of the regression with observable LTU risk (0.700).[33] The lack of correlation is even more apparent during the Great Recession period, when

[32] Since the LISA panel only starts in 1990 and even later for days spent on UI and DI, we impute the pre-unemployment variables for spells in the earliest years of our sampling variable. In particular, we use the individual's history when partially observed, but use the population mean in 1995 when the individual's history is entirely missing. Given our finding that employment histories from the prior year are the most predictive, the imputation of earlier years does not seem restrictive.

[33] As reported in the same table, we find that the contribution of selection is 22% if we extend the sample to the first three years where we do not have complete income and employment histories available.

LTU risk increased sharply by 13.6 p.p. from 2006 to 2009, but the predicted risk barely moved and increased by just 1.0 p.p. over the same time period, accounting for only 7% of the movement in the observed LTU risk over the Great Recession.

Overall, this shows that – even when using a rich set of observables that is highly predictive of job finding – there is little support for the heterogeneity hypothesis. That is, most of the increase in long-term unemployment risk in recessions cannot be attributed to observable changes in the composition of the pool unemployed workers.[34] Of course, our analysis using multiple spells has shown the importance of unobserved heterogeneity, but the considered set of observables captures more than half of the overall heterogeneity. Hence, we cannot exclude compositional changes along these unobservable factors, but for compositional changes to contribute meaningfully to the cyclicality in the long-term unemployment risk, we would need the selection on unobservable heterogeneity underlying individuals' job finding to be substantially different than the selection on observable heterogeneity.[35]

In sum, our analysis suggests that compositional shifts in the pool of unemployed account only to a small extent for the cyclical movements in the aggregate job finding rate, at least to the extent predicted by the observables included in our model. Our findings thus speak against structural factors being important for movements of the unemployment rate as the proportion of those who are likely to become long-term unemployed stays relatively constant over the business cycle.

5.2 Heterogeneity in Cyclicality

When compositional changes cannot account for the large increase in long-term unemployment risk in recessions, this implies that the same individuals face varying risks depending on when they become unemployed. An important question taken up in the literature is whether individuals differ in their cyclicality of job finding and long-term unemployment risk. Heterogeneity in the cyclicality of job finding could be driven by differential labor supply responses to recessions or labor demand changing differentially for workers of different skills. Prior work, for example, has found that those who tend to work fewer hours have a lower cyclicality of job finding out of unemployment [Bils et al., 2012], but that there are little differences in the cyclicality based on education and prior wages [Mueller, 2017]. We revisit this question using richer data and we leverage the extendability of our approach repeating the prediction exercise for each year from 1995 to 2016.[36]

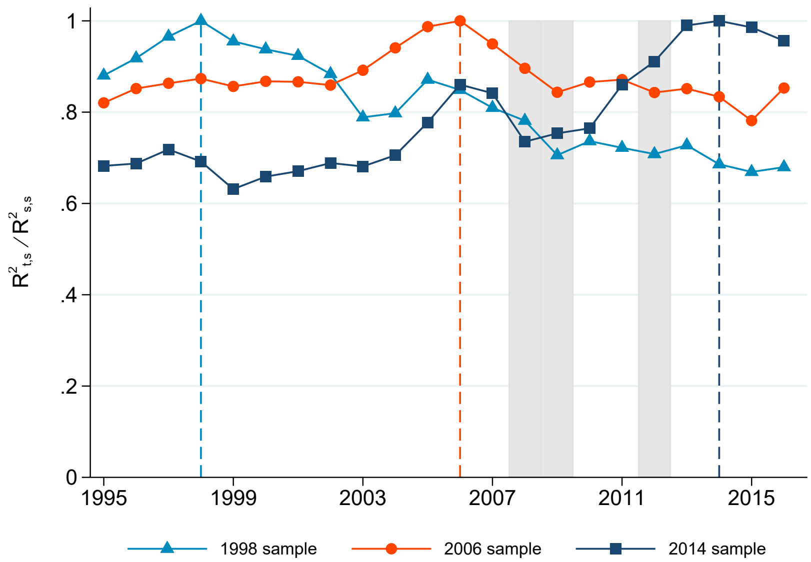
Persistence. To preface the analysis of heterogeneity in cyclicality, we first consider again how persistent the heterogeneity in job finding is, but now across cohorts of unemployed rather than over the unemployment spell. As shown in Section 2.4, the combination of the multiple spell approach with the use of observables, allowed us to separate the observable heterogeneity into a component that is persistent across spells and one that is transitory. Using all multiple spells, we find that about

[34] Note that selection into long-term unemployment is driven both by selection into unemployment, i.e. at the start of the spell, and selection that occurs over the spell. We find, however, similar patterns when we look at the predicted risk of remaining unemployed from 6 to 12 months or from 12 to 18 months, see Appendix Figure A3.

[35] In Appendix B.5, we perform calculations under the assumption that the distributions of observable and unobservable heterogeneity, scaled by their standard deviations, move identically over the business cycle. We show that the compositional changes based on observables should be inflated by a scaling factor of 1.87 to account for shifts in unobservables in the pool of unemployed. We find that compositional shifts then can explain 22% over the period 1995-2016 and 14% over the period of the Great Recession of the observed changes in LTU risk.

[36] We start only in 1995 to reduce the censoring of pre-unemployment histories.

Figure 7: Persistence in the Predictive Value over Time



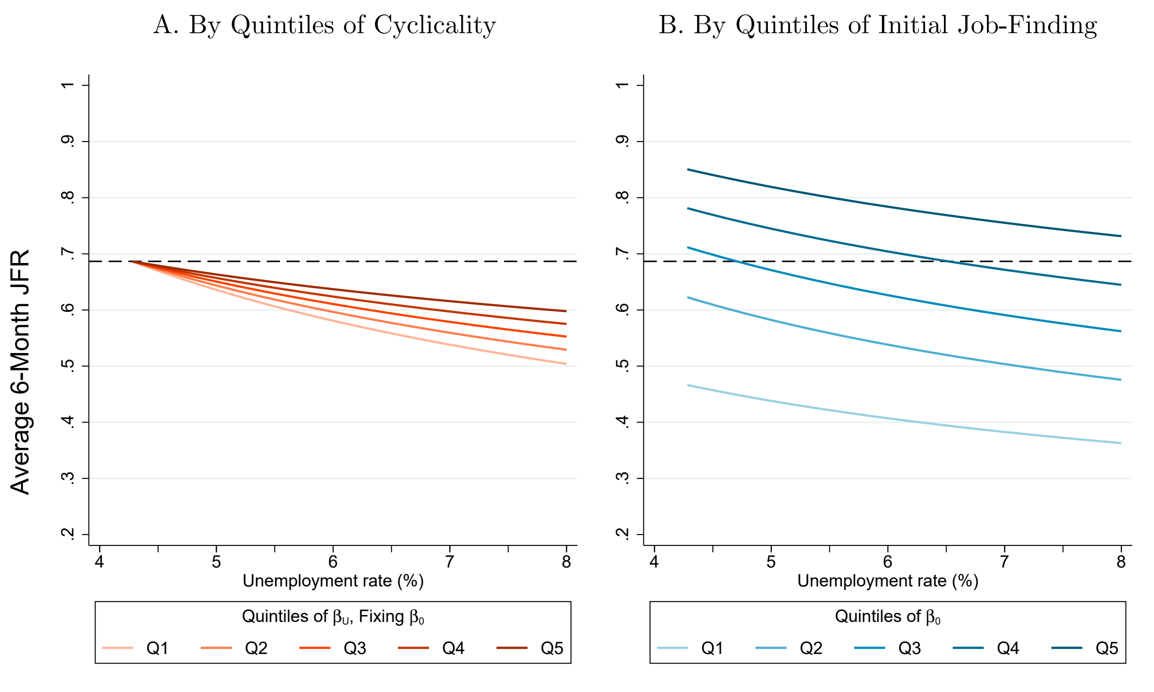
Notes: The figure shows the R^2 in the hold out sample of 1998, 2006 and 2014 using the prediction model from each year *relative to* R^2 in the hold out sample from the same year as the prediction model.

three quarters of the observable heterogeneity is persistent across spells as measured by the ratio of the covariances POH/OH, reported in Table 4. However, as we consider spells that are further apart, the share of observable heterogeneity that remains persistent goes down and it becomes less than 50 percent when considering spells that are at least 5 years apart.[37] To study this further, we can consider the R-squared instead, scaling the covariance by the respective variances in job findings in either year, and compare the relative predictive value of models estimated in a different year t than the year of the hold-out sample. Figure 7 shows the relative predictive value of models estimated in a different year t for three different hold-out samples (resp. 1998, 2006 and 2014). For example, the orange dotted line shows the R-squared in the hold-out sample of 2006 for models from different years relative to the R-squared for the model of 2006. Not surprisingly, the model from 2006 does best but the fall off in predictive power for other years is relatively modest. Moreover, even when using models estimated in more distant years, the decrease in the predictive power is limited. E.g., for the 2014 hold-out sample, the predictive power of the model estimated 20 years earlier is still around 70 percent of the predictive power of the model estimated in 2014. We also see slight reductions in the predictive values, for all three hold-out samples, during the Great Recession. Overall, the evidence, suggests that the features that predict long-term unemployment are relatively stable and that while job-finding rates are cyclical, there are only limited rank reversals in predicted job-finding over time.

Cyclicality. We can now use the predictions based on observables to estimate how heterogeneity in individual cyclicality in parallel to our analysis of heterogeneity in individual duration dependence

[37]Appendix Table A7 repeats Table 4, but restricting the multiple spell sample further to spells two or five years apart. We find that the further apart the spells, the lower are the estimated covariances across spells in Columns 1 and 2, whereas the within-spell covariances in Columns 3 and 4 are hardly affected.

Figure 8: Heterogeneity in Cyclicality



Notes: Panel A shows the predicted individual change in the job-finding rate for the five quintiles of the distribution of β_U, normalizing the job-finding rate to the one in 2006. Panel B shows the predicted change at the individual level in the job-finding rate for the five quintiles of the distribution of the intercept β_0.

in the previous section. An advantage of our prediction exercise is that we can predict a hypothetical job-finding rate for each individual and each year in our data, conditional on their observable characteristics. We can then estimate for each individual in the sample the cyclicality in the predicted 6-month job-finding rates based on her observables X^{t_0} prior to unemployment in baseline year t_0 by relating it to the log of the aggregate unemployment rate, u_t, as follows:

$$\log(\hat{F}_0^t(X^{t_0}) = \beta_0(X^{t_0}) + \beta_U(X^{t_0})(u_t - u_{t_0}) + \beta_{Tr}(X^{t_0})(t - t_0) + \eta_0^t, \tag{21}$$

where we use 2006 as the reference year t_0. Note that we include only individuals who were actually unemployed in this reference year and still focus on job finding rates at the start of the unemployment spell. This exercise also holds characteristics of individual job seekers constant over time. As we have shown that there is little movement in composition of types, this suggests that the conclusions from our exercise are not much affected by these restrictions. We again shrink the distributions of $\hat{\beta}_j(X^{t_0})$'s using the standard errors of the same regressions, as for the duration-dependence estimates (see equation 20 above).

Panel A in Figure 8 illustrates the heterogeneity in the individual cyclicality of job finding by splitting the sample into the five quintiles based on the cyclicality estimate $\hat{\beta}_U(X^{t_0})$, normalizing the job-finding rate to the one for the 2006 unemployment rate. It shows that there is only moderate dispersion in how the predicted job-finding rates vary over the range of the unemployment rates observed in Sweden in the sample period, especially in comparison to the dispersion in duration dependence discussed earlier (see Figure 5).[38] Panel B of Figure 8 illustrates how the predicted individual level dynamics relate to the job-finding rate in the reference year. More specifically, it shows the predicted changes by quintile of the distribution of the intercept $\hat{\beta}_0(X^{t_0})$. The declines in the job-finding rate are quite similar across the five quintiles, very much in contrast with the dynamics over the unemployment spell, where we found large differences in the slopes.[39]

In the right panel of Figure 3 we show again how observable characteristics correlate with the cyclical component $\hat{\beta}_U(X^{t_0})$. We confirm that in contrast to the duration dependence analysis job seekers with higher job-finding rates do not experience larger, but slightly smaller (relative) declines as the unemployment rate increases. We again find that, conditional on initial job finding, job seekers are more shielded against declines in the job-finding rate when they are more educated and had higher labor income prior to unemployment.[40]

To sum up, we find a moderate dispersion in the individual cyclicality of job finding despite the important cyclicality overall and the large predictable differences in individual job-finding rates at a given moment in time. Our findings suggest that movements in the job finding rate over time are fairly uniform across workers, at least to the extent predicted by the observables in our model. The predictable heterogeneity in job finding is thus persistent over time, which is also reflected in the persistence of the predictive power of the prediction models across different years in our sample.

6 Conclusion

This paper uses rich administrative data from Sweden to study the predictability of long-term unemployment. We find substantial predictable heterogeneity of LTU risk that is driven by the use of comprehensive data on income, employment and benefit histories and show how the predictability in LTU risk relates to issues of selection over the unemployment spell and the business cycle.

We show that, over the spell of unemployment, our results of substantial predictability have important implications for dynamic selection and the observed duration dependence in job finding. In particular, we show that the persistence in the predictability of job finding is a key statistic that pins down the extent of dynamic selection. We show empirically that the predictability of job finding is indeed very persistent over the spell of unemployment and that, as a consequence, at least 49% of the observed decline in job finding is driven by dynamic selection. This finding complements recent research that has found an important role for selection but with different and complimentary ap-

[38] Appendix Figure A6 shows the distributions of the estimated permanent component ($\hat{\beta}_0(X^{t_0})$)and cyclical component ($\hat{\beta}_U(X^{t_0})$) of the job-finding rate. The distribution of the permanent component in Panel A closely resembles the distribution of predicted job-finding rates for the year 2006 in Figure 2, as we choose the year 2006 as our reference year. Panel B shows the distribution of the cyclical component. The average cyclicality equals -0.353, implying that if the unemployment rate doubles, the job-finding rate decreases by 35.3 percent. For the cyclical component, the shrunken standard deviation is 0.1065, which is moderate, and the bootstrapped 95% confidence interval is tight at [0.1061, 0.1069]. As in previous exercises, we draw 500 bootstrap samples from the 2006 hold-out sample for this last step.

[39] In parallel to the duration-dependence analysis, we also carry out a test that relies on a pairwise comparison between year 2009 and 2006 based on $\log\left(\frac{\hat{F}_0^{2009}(X^{2006})}{\hat{F}_0^{2006}(X^{2006})}\right)$. As shown in Appendix Figures B1 and B2, we find significant heterogeneity in this log ratio, but this can also be driven by differential time trends in addition to differential cyclicality across individuals with different observables.

[40] The estimated Shapley values in Appendix Figure A9 confirm the relative importance of characteristics other than the 2006 predicted job-finding in explaining the individual cyclicality.

proaches [Güell and Lafuente, 2022; Alvarez et al., forthcoming; Mueller et al., 2021]. In fact, using the sample of job seekers with repeated unemployment spells, we find a substantial role for unobserved heterogeneity. Taken together, dynamic selection can account for as much as 88% of the observed decline in job finding over the spell of unemployment.

We also examine to what extent the cyclicality in job finding and LTU risk is driven by cyclical changes in the pool of unemployed. Prior work has found little role for composition, but was limited to socio-demographic variables in survey data. Despite the richer data and the high predictive power for job finding in the cross-section of job seekers, we find little evidence that the rise in LTU risk is driven by composition and thus confirm the previous evidence over the business cycle (e.g., Baker [1992] and Kroft et al. [2016]). Our approach of estimating prediction models for job finding at different stages of the unemployment spell or the business cycle also allows us to predict the duration dependence and the cyclicality of job finding for each individual in our data. We find substantial heterogeneity in the dynamics over the unemployment spell that is inconsistent with the common assumption of proportionality in models of job search. Over the business cycle, instead, we only find limited heterogeneity in the cyclicality of job finding.

The predictability of LTU risk has important policy implications too. Many countries use long-term unemployment as a criterion for the assignment of active labor market policies. The high predictability of LTU risk suggests that we can target ALMPs better and assign them earlier in the spell, as some countries try to do (e.g., Ernst et al. [2024]). A valuable avenue for future research would be to identify the causal effects of unemployment policies by predicted LTU risk and whether these change over the unemployment spell or over the business cycle.

References

Ahn, Hie Joo and James D. Hamilton, "Heterogeneity and Unemployment Dynamics," *Journal of Business & Economic Statistics*, 2020, *38* (3), 554–569.

_ , Bart Hobijn, and Ayşegül Şahin, "The Dual US Labor Market Uncovered," Working Paper 31241, National Bureau of Economic Research May 2023.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Alvarez, Fernando, Katarína Borovičková, and Robert Shimer, "A Nonparametric Variance Decomposition Using Panel Data," Technical Report October 2014.

_ , _ , and _ , "Decomposing Duration Dependence in a Stopping Time Model," *The Review of Economic Studies*, forthcoming.

Arni, Patrick, Marco Caliendo, Steffen Künn, and Robert Mahlstedt, "Predicting the Risk of Long-Term Unemployment: What can we learn from Personality Traits, Beliefs and other Behavioral Variables," Technical Report, Working Paper 2014.

Baker, Michael, "Unemployment Duration: Compositional Effects and Cyclical Variability," *The American Economic Review*, 1992, *82* (1), 313–321.

Berger, Mark C., Dan Black, and Jeffrey A. Smith, "Evaluating profiling as a means of allocating government services," in Michael Lechner and Friedhelm Pfeiffer, eds., *Econometric Evaluation of Labour Market Policies*, Physica-Verlag HD Heidelberg 2001, pp. 59–84.

Bertheau, Antoine, Edoardo Maria Acabbi, Cristina Barcelo, Andreas Gulyas, Stefano Lombardi, and Raffaele Saggio, "The Unequal Cost of Job Loss across Countries," Working Paper 29727, National Bureau of Economic Research February 2022.

Bils, Mark, Yongsung Chang, and Sun-Bin Kim, "Comparative advantage and unemployment," *Journal of Monetary Economics*, 2012, *59* (2), 150–165.

Black, Dan A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel, "Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System," *American Economic Review*, September 2003, *93* (4), 1313–1327.

Card, David, Jochen Kluve, and Andrea Weber, "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 10 2017, *16* (3), 894–931.

Cockx, Bart and Muriel Dejemeppe, "Duration dependence in the exit rate out of unemployment in Belgium. Is it true or spurious?," *Journal of Applied Econometrics*, 2005, *20* (1), 1–23.

Cox, D. R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1972, *34* (2), 187–220.

Einav, Liran, Amy Finkelstein, Sendhil Mullainathan, and Ziad Obermeyer, "Predictive modeling of U.S. health care spending in late life," *Science*, 2018, *360* (6396), 1462–1465.

Elsby, Michael W. L., Ryan Michaels, and Gary Solon, "The Ins and Outs of Cyclical Unemployment," *American Economic Journal: Macroeconomics*, January 2009, *1* (1), 84–110.

Elsby, Michael W.L., Bart Hobijn, and Ayşegül Şahin, "On the importance of the participation margin for labor market fluctuations," *Journal of Monetary Economics*, 2015, *72*, 64–82.

Eriksson, Stefan and Dan-Olof Rooth, "Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment," *American Economic Review*, March 2014, *104* (3), 1014–39.

Ernst, Sebastian, Andreas I. Mueller, and Johannes Spinnewijn, "Risk Scores for Long-Term Unemployment and the Assignment to Job Search Counseling," *AEA Papers and Proceedings*, May 2024, *114*, 572–76.

Farber, Henry S., Dan Silverman, and Till von Wachter, "Determinants of Callbacks to Job Applications: An Audit Study," *American Economic Review*, May 2016, *106* (5), 314–18.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Place-Based Drivers of Mortality: Evidence from Migration," *American Economic Review*, August 2021, *111* (8), 2697–2735.

Gregory, Victoria, Guido Menzio, and David G Wiczer, "The Alpha Beta Gamma of the Labor Market," Working Paper 28663, National Bureau of Economic Research April 2021.

Grömping, Ulrike, "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician*, 2007, *61* (2), 139–147.

Güell, Maia and Cristina Lafuente, "Revisiting the determinants of unemployment duration: Variance decomposition à la ABS in Spain," *Labour Economics*, 2022, *78* (C).

Hall, Robert E and Marianna Kudlyak, "Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics," Working Paper 25625, National Bureau of Economic Research February 2019.

Hall, Robert E. and Sam Schulhofer-Wohl, "Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers," *American Economic Journal: Macroeconomics*, January 2018, *10* (1), 1–32.

Heckman, J. and B. Singer, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 1984, *52* (2), 271–320.

Honoré, Bo E., "Identification Results for Duration Models with Multiple Spells," *The Review of Economic Studies*, 01 1993, *60* (1), 241–246.

Huettner, Frank and Marco Sunder, "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values," *Electronic Journal of Statistics*, 2012, *6* (none), 1239 – 1250.

Jarosch, Gregor and Laura Pilossoph, "Statistical Discrimination and Duration Dependence in the Job Finding Rate," *The Review of Economic Studies*, 09 2018, *86* (4), 1631–1665.

Katz, Lawrence, "Layoffs, Recall and the Duration of Unemployment," NBER Working Papers 1825, National Bureau of Economic Research, Inc 1986.

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

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment*," *The Quarterly Journal of Economics*, 06 2013, *128* (3), 1123–1167.

_ , _ , _ , and Lawrence F. Katz, "Long-Term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Nonparticipation," *Journal of Labor Economics*, 2016, *34* (S1), S7–S54.

Krueger, Alan B, Judd Cramer, and David Cho, "Are the Long-Term Unemployed on the Margins of the Labor Market?," *Brookings Papers on Economic Activity*, 2014, *2014* (1), 229–299.

Kuhn and Max, "Building Predictive Models in R Using the caret Package," *Journal of Statistical Software*, 2008, *28* (5), 1–26.

Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica*, 1979, *47* (4), 939–956.

Landais, Camille and Johannes Spinnewijn, "The Value of Unemployment Insurance," *The Review of Economic Studies*, 09 2021, *88* (6), 3041–3085.

_ , Arash Nekoei, Peter Nilsson, David Seim, and Johannes Spinnewijn, "Risk-Based Selection in Unemployment Insurance: Evidence and Implications," *American Economic Review*, April 2021, *111* (4), 1315–55.

Machin, Stephen and Alan Manning, "Chapter 47 The causes and consequences of longterm unemployment in Europe," in "Handbook of Labor Economics," Vol. 3, Elsevier, 1999, pp. 3085–3139.

Morchio, Iacopo, "Work Histories and Lifetime Unemployment," *International Economic Review*, 2020, *61* (1), 321–350.

Mueller, Andreas I., "Separations, Sorting, and Cyclical Unemployment," *American Economic Review*, July 2017, *107* (7), 2081–2107.

_ , Johannes Spinnewijn, and Giorgio Topa, "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, January 2021, *111* (1), 324–63.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, May 2017, *31* (2), 87–106.

Nekoei, Arash and Andrea Weber, "Recall Expectations and Duration Dependence," *American Economic Review*, May 2015, *105* (5), 142–46.

OECD, "Statistical profiling in public employment services," Technical Report 224, OECD 2019.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, April 2019, *37* (2), 187–204.

Pavoni, Nicola and G. L. Violante, "Optimal Welfare-to-Work Programs," *The Review of Economic Studies*, 01 2007, *74* (1), 283–318.

Richardson, Katarina and Gerard J. van den Berg, "Duration Dependence versus Unobserved Heterogeneity in Treatment Effects: Swedish Labor Market Training and the Transition Rate To Employment," *Journal of Applied Econometrics*, 2013, *28* (2), 325–351.

Schmieder, Johannes F, Till M von Wachter, and Jörg Heining, "The Costs of Job Displacement over the Business Cycle and Its Sources: Evidence from Germany," Working Paper 30162, National Bureau of Economic Research June 2022.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender, "The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years," *The Quarterly Journal of Economics*, 04 2012, *127* (2), 701–752.

_ , _ , and _ , "The Effect of Unemployment Benefits and Nonemployment Durations on Wages," *American Economic Review*, March 2016, *106* (3), 739–77.

Shimer, Robert, "Reassessing the ins and outs of unemployment," *Review of Economic Dynamics*, 2012, *15* (2), 127–148.

_ and Ivan Werning, "Liquidity and Insurance for the Unemployed," *American Economic Review*, December 2008, *98* (5), 1922–42.

Spinnewijn, Johannes, "Training and search during unemployment," *Journal of Public Economics*, 2013, *99*, 49–65.

Appendix

A Additional Figures and Tables

A.1 Predictive Power: Further Results and Robustness

Table A1: Robustness: Job Finding over Different Horizons

Job Finding Horizon	N	$E(\cdot)$		$Var(\cdot)$		$Cov(\cdot)$		$R^2(\cdot)$	
		F	$\hat{F}$	F	$\hat{F}$	$\hat{F}, F$	$\hat{F}, F_{6m}$	$\hat{F}, F$	$\hat{F}, F_{6m}$
3 Months	126,052	0.483	0.487	0.250	0.023	0.025	0.026	0.110	0.142
6 Months	126,052	0.701	0.705	0.210	0.027	0.029	0.029	0.150	0.150
12 Months	126,052	0.861	0.863	0.120	0.015	0.016	0.020	0.138	0.134

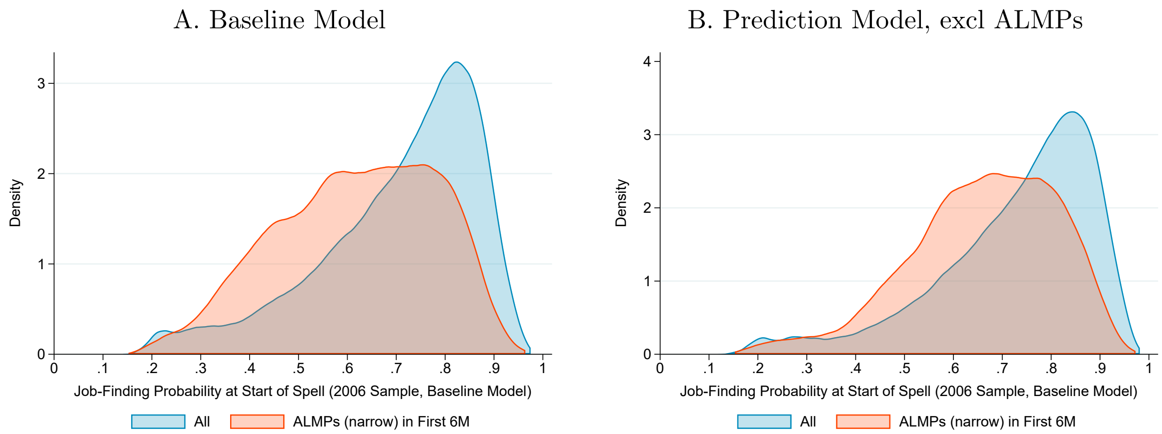
Notes: The table reports summary statistics about observed and predicted job-finding probabilities at the start of the spell over different horizons. We consider job finding over three horizons: three months, six months (the baseline) and twelve months since the beginning of the spell.

Table A2: Robustness: ALMPs, Job Finding, Prediction Algorithm

Model	Sample	N	$E(F_0)$	$E(\hat{F}_0)$	$Var(F_0)$	$Var(\hat{F}_0)$	$Cov(\hat{F}_0, F_0)$	$R^2(\hat{F}_0, F_0)$
A. Baseline								
Baseline	All	126,052	0.701	0.705	0.210	0.027	0.029	0.150
B. Robustness to ALMPs								
Baseline	No ALMPs	116,608	0.730	0.712	0.197	0.026	0.027	0.144
No ALMPs	All	126,052	0.701	0.728	0.210	0.026	0.028	0.147
No ALMPs	No ALMPs	116,608	0.730	0.733	0.197	0.026	0.027	0.146
C. Robustness to job finding definition								
Baseline	No AvOrs 7-8	108,541	0.683	0.704	0.217	0.027	0.029	0.146
Baseline	No AvOrs 5-8	67,861	0.740	0.745	0.192	0.021	0.023	0.128
Baseline	No recalls	113,242	0.701	0.704	0.210	0.027	0.029	0.152
No Recalls	All	126,052	0.701	0.708	0.210	0.028	0.030	0.150
No Recalls	No recalls	113,242	0.701	0.706	0.210	0.029	0.030	0.152
D. Robustness to functional form								
Linear	All	126,052	0.701	0.701	0.210	0.030	0.029	0.137
R. Forest	All	126,052	0.701	0.710	0.210	0.027	0.028	0.143
B. Gradient	All	126,052	0.701	0.703	0.210	0.027	0.029	0.150
LASSO	All	126,052	0.701	0.702	0.210	0.024	0.026	0.134
E. Robustness to sample split								
Baseline: 10-30-10								
↑ Weights: 10-20-20	All	126,052	0.701	0.703	0.210	0.027	0.029	0.148
↑ Tuning: 20-20-10	All	126,052	0.701	0.705	0.210	0.027	0.029	0.148
↑ Training: 10-40-10	All	100,873	0.702	0.702	0.209	0.029	0.030	0.151
F. Robustness to ensemble weights								
Positive weights	All	126,052	0.701	0.702	0.210	0.027	0.029	0.150

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. Panel A presents results for the baseline 2006 model ("Baseline") on the full 2006 hold-out sample ("All"). Panel B shows statistics for the baseline model and a model trained on the subset of spells that did not include ALMPs during the first six months of unemployment ("No ALMPs"), on the full holdout sample and excluding spells that include ALMPs during the first six months ("No ALMPs"). Panel C considers the baseline model and a model trained on unemployed that were not recalled by their previous employer ("No recalls") on four different samples: full, excluding spells that ended because the job seeker entered education other than training or died ("No AvOrs 7-8"), 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"), and excluding unemployed that were recalled by their previous employer ("No recalls"). Panel D looks at the linear regression model ("Linear") and the three Machine Learning predictors that underlie the baseline or ensemble model, all on the full sample. Panel E considers three alternative ensemble predictors constructed with different sample splits: for the first two, we preserve the full holdout sample and take 10% of the sample from training to either tune parameters or estimate weights; for the third one, we move 10% from the holdout sample to the training sample and compute the statistics on this smaller ("Reduced") holdout sample. Finally, Panel F shows the performance of the ensemble model when we constrain the weights to be positive (in the baseline exercise, they are only constrained to sum to 1).

Figure A1: Distribution of Predicted JFP: Exit to ALMPs



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

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)
$R^2(\hat{F}_0, F_0)$	0.071	0.099	0.106	0.140	0.142	0.147	0.149	0.150
Change (j) vs $(j-1)$	-	+38.5%	+7.3%	+32.5%	+1.1%	+4.1%	+1.2%	+0.6%

B. Linear model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_0, F_0)$	0.063	0.085	0.092	0.123	0.125	0.129	0.134	0.137
Change (j) vs $(j-1)$	-	+35.1%	+8.2%	+33.0%	+1.6%	+3.6%	+3.6%	+2.4%
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).

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

A. Groups of variables: sequential sub-models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$R^2(\hat{F}_0, F_0)$	0.071	0.119	0.126	0.129	0.131	0.132	0.136
Change (j) vs $(j-1)$	-	+66.6%	+6.3%	+2.7%	+1.1%	+0.6%	+3.4%
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)
$R^2(\hat{F}_0, F_0)$	0.071	0.119	0.097	0.088	0.083	0.081	0.100
Change (j) vs (1)	-	+66.6%	+36.8%	+23.6%	+16.0%	+13.2%	+40.4%
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)
$R^2(\hat{F}_0, F_0)$	0.071	0.084	0.091	0.109	0.121
Change (j) vs $(j-1)$	-	+17.6%	+8.3%	+20.4%	+10.9%
Basic Socio-demographics	X	X	X	X	X
Days on UI (last 2 years)		X	X	X	X
# Unempl. Spells (last 2 years)			X	X	X
# Employers(last 2 years)				X	X
Days on DI (last 2 years)					X

D. Individual variables: marginal sub-models

	(1)	(2)	(3)	(4)	(5)
$R^2(\hat{F}_0, F_0)$	0.071	0.084	0.084	0.091	0.088
Change (j) vs (1)	-	+17.6%	+17.9%	+28.4%	+24.1%
Basic Socio-demographics	X	X	X	X	X
Days on UI (last 2 years)		X			
# Unempl. Spells (last 2 years)			X		
# Employers (last 2 years)				X	
Days on DI (last 2 years)					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 employers) 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.

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	1.063	1.005		0.989	0.912		0.989	0.930	
	(0.108)	(0.112)		(0.340)	(0.347)		(0.340)	(0.362)	
Health PC1		0.021	0.052					0.022	0.058
		(0.011)	(0.012)					(0.044)	(0.043)
Mental Health PC1					0.037	0.056			
					(0.034)	(0.035)			
R^2	0.138	0.142	0.026	0.093	0.106	0.030	0.093	0.096	0.022
Adj. R^2	0.125	0.128	0.025	0.082	0.084	0.018	0.082	0.074	0.010
N	742	742	742	84	84	84	84	84	84

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 (+3%), 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 14% and adjusted R^2 by 2%, 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 a 3% increase 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.

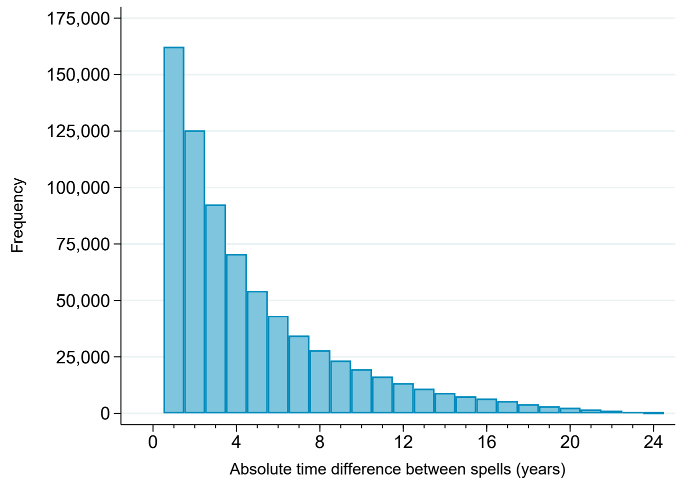
Table A6: R^2 FOR EXTENDED MODELS: STARTING FROM BASIC

	Extensions of Basic							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R^2(\hat{F}_0, F_0)$	0.071	0.090	0.077	0.080	0.075	0.078	0.079	0.101
Change (j) vs (1)	-	+27.1%	+7.7%	+12.2%	+5.5%	+9.8%	+11.0%	+41.4%
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.2 Multiple Spell Analysis: Sample and Results

Figure A2: 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 735,797 individuals.

Table A7: Observable, Unobservable, Persistent and Transitory Components of Heterogeneity in Job Finding: Robustness

A. Baseline

	Multiple Spells Sample			Control Sample
	Persistent	**Transitory**	**Total**	**Total**
Observable	$POH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_1}, F_0^{t_2})$ **0.017**	$OH - POH$ **0.006**	$OH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.023**	$\tilde{OH} = Cov_0^{t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.027**
Unobservable	$PH - POH$ **0.013**			
Total	$PH = Cov_{0,0}^{t_1,t_2}(F_0^{t_1}, F_0^{t_2})$ **0.030**		*By Proportionality* *0.042*	*By Proportionality* *0.048*

B. Spells more than 2 years apart

	Multiple Spells Sample			Control Sample
	Persistent	**Transitory**	**Total**	**Total**
Observable	$POH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_1}, F_0^{t_2})$ **0.014**	$OH - POH$ **0.011**	$OH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.024**	$\tilde{OH} = Cov_0^{t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.027**
Unobservable	$PH - POH$ **0.012**			
Total	$PH = Cov_{0,0}^{t_1,t_2}(F_0^{t_1}, F_0^{t_2})$ **0.025**		*By Proportionality* *0.046*	*By Proportionality* *0.051*

C. Spells more than 5 years apart

	Multiple Spells Sample			Control Sample
	Persistent	**Transitory**	**Total**	**Total**
Observable	$POH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_1}, F_0^{t_2})$ **0.008**	$OH - POH$ **0.015**	$OH = Cov_{0,0}^{t_1,t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.023**	$\tilde{OH} = Cov_0^{t_2}(\hat{F}_0^{t_2}, F_0^{t_2})$ **0.028**
Unobservable	$PH - POH$ **0.011**			
Total	$PH = Cov_{0,0}^{t_1,t_2}(F_0^{t_1}, F_0^{t_2})$ **0.019**		*By Proportionality* *0.055*	*By Proportionality* *0.066*

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.

Table A8: Observable, Unobservable, Persistent and Transitory Components of Heterogeneity in Job Finding: Long-Term Unemployment

	Multiple Spells, LTU Sample			Multiple Spells Sample	Control Sample
	Persistent	**Transitory**	**Total**	**Total**	**Total**
Observable	$C_1 = Cov_{6,0}^{t_1,t_2}(\hat{F}_6^{t_1}, F_0^{t_2})$ **0.009**	$C_3 - C_1$ **0.005**	$C_3 = Cov_{6,0}^{t_1,t_2}(\hat{F}_6^{t_2}, F_0^{t_2})$ **0.014**	$C_4 = Cov_{0,0}^{t_1,t_2}(\hat{F}_6^{t_2}, F_0^{t_2})$ **0.016**	$C_5 = Cov_0^{t_2}(\hat{F}_6^{t_2}, F_0^{t_2})$ **0.020**
Unobservable	$C_2 - C_1$ **0.007**				
Total	$C_2 = Cov_{6,0}^{t_1,t_2}(F_6^{t_1}, F_0^{t_2})$ **0.016**		*By Proportionality* *0.025*	*By Proportionality* *0.028*	*By Proportionality* *0.034*

Notes: The first three columns of this table restrict the multiple spells sample described in Table 4 to pairs where the first spell lasted more than 6 months, and computes C_1 and C_2 using predicted and actual job-finding probabilities from the 6th to the 12th month of unemployment during the first spell. Columns 4 and 5 presents results for the multiple spells and control samples, respectively.

A.3 Selection over the Spell: Robustness

Table A9: Models Trained on Different Variables: Basic Model

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	126,052	0.701	0.696	0.013	0.014	0.071
	6M Model	126,052	0.701	0.570	0.008	0.010	0.059
	12M Model	126,052	0.701	0.513	0.004	0.007	0.052
6M into Spell	6M Model	42,823	0.552	0.546	0.009	0.009	0.038
	12M Model	42,823	0.552	0.497	0.005	0.006	0.032
12M into Spell	12M Model	20,831	0.485	0.490	0.004	0.004	0.017

Notes: The table reports summary statistics for models trained on different unemployment durations in 2006 using only basic socio-demographic variables - a basic model. The estimates can be compared to the results for the baseline model in Table 5. 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 A10: Models Trained on 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	226,837	0.704	0.708	0.031	0.032	0.162
	6M Model	226,837	0.704	0.598	0.017	0.022	0.136
	12M Model	226,837	0.704	0.542	0.010	0.014	0.103
6M into Spell	6M Model	74,412	0.536	0.538	0.018	0.020	0.083
	12M Model	74,412	0.536	0.501	0.010	0.012	0.061
12M into Spell	12M Model	38,257	0.478	0.483	0.010	0.010	0.040

Notes: The table reports summary statistics about models trained on different unemployment durations in 2006 and 2007. The estimates can be compared to the results for the baseline model using only 2006 in Table 5. 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 A11: Models Trained on Different Samples: Pooled 2009-2010 Data

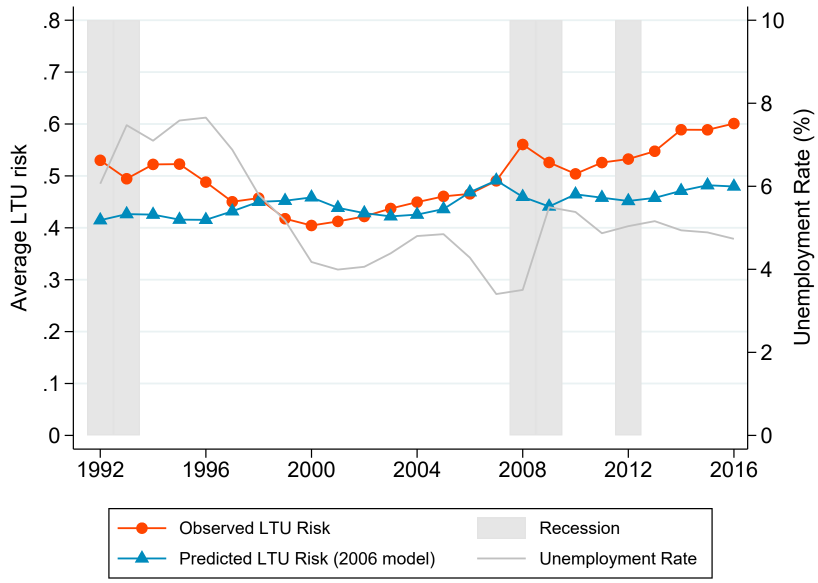
Sample	**Model**	N	$E(F)$	$E(\hat{F})$	$Var(\hat{F})$	$Cov(\hat{F}, F)$	$R^2(\hat{F}, F)$
At Start of Spell	0M Model	236,043	0.576	0.578	0.029	0.030	0.126
	6M Model	236,043	0.576	0.513	0.013	0.018	0.097
	12M Model	236,043	0.576	0.471	0.009	0.012	0.063
6M into Spell	6M Model	99,369	0.470	0.472	0.012	0.014	0.061
	12M Model	99,369	0.470	0.444	0.008	0.009	0.038
12M into Spell	12M Model	48,905	0.427	0.424	0.008	0.009	0.041

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.

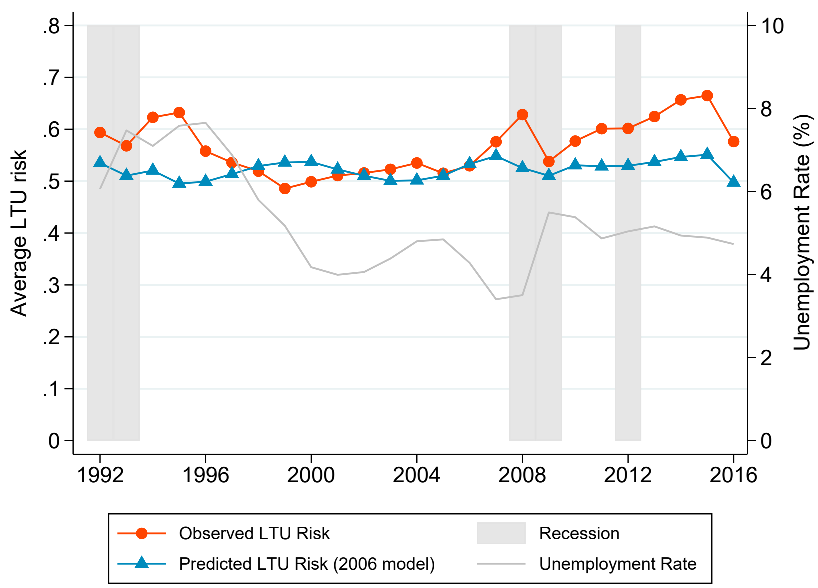
A.4 Selection over the Business Cycle: Robustness

Figure A3: Average Risk and Selection into Long-Term Unemployment

A. 6 Months into Spell

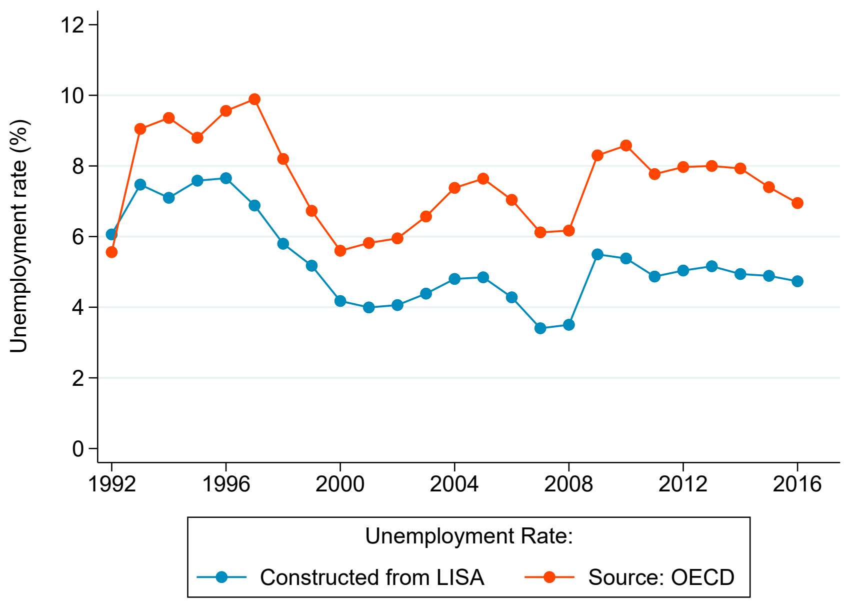


B. 12 Months into Spell



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 A4: 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 A12: Relationship between Unemployment and Long-Term Unemployment Risk

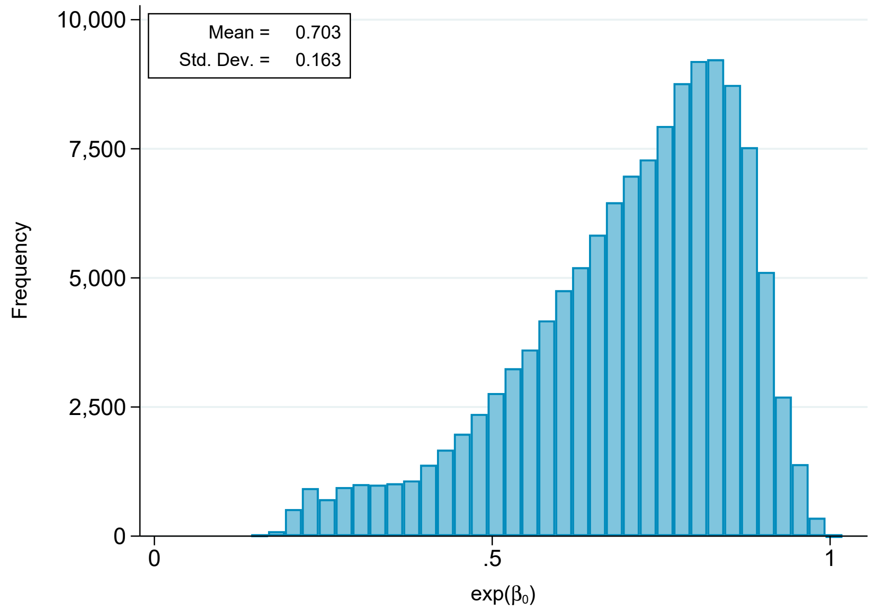
	Predicted log LTU risk (2006)		Observable log LTU risk	
	(1)	(2)	(3)	(4)
	Panel A. 1995-2016			
Log unemployment rate	-0.050	0.082	0.309	0.700
	(0.085)	(0.052)	(0.226)	(0.086)
Time trend		0.011		0.034
		(0.002)		(0.003)
R^2	0.017	0.708	0.085	0.892
Adj. R^2	-0.033	0.677	0.040	0.881
Observations	22	22	22	22
	Panel B. 1992-2016			
Log unemployment rate	0.006	0.146	0.342	0.791
	(0.072)	(0.068)	(0.185)	(0.136)
Time trend		0.008		0.026
		(0.002)		(0.004)
R^2	0.000	0.402	0.129	0.679
Adj. R^2	-0.043	0.348	0.091	0.650
Observations	25	25	25	25

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 restricts to 1995-2016 to avoid early censoring of income and employment histories in the first three years of the sample, whereas Panel B uses every year in our sample period.

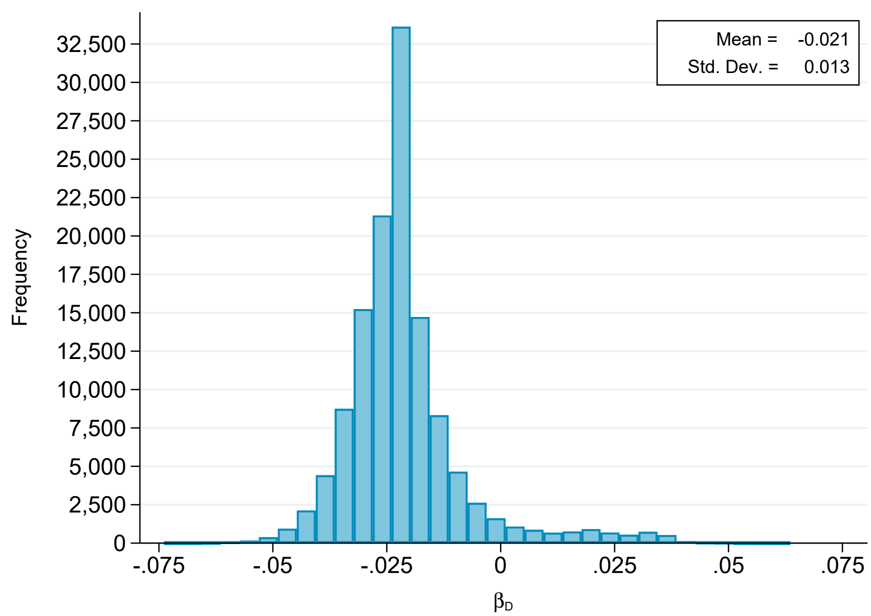
A.5 Heterogeneity in Dynamics over Spell: Additional Results

Figure A5: Distribution of Permanent and Duration-Dependent Component of Job-Finding Risk

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



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

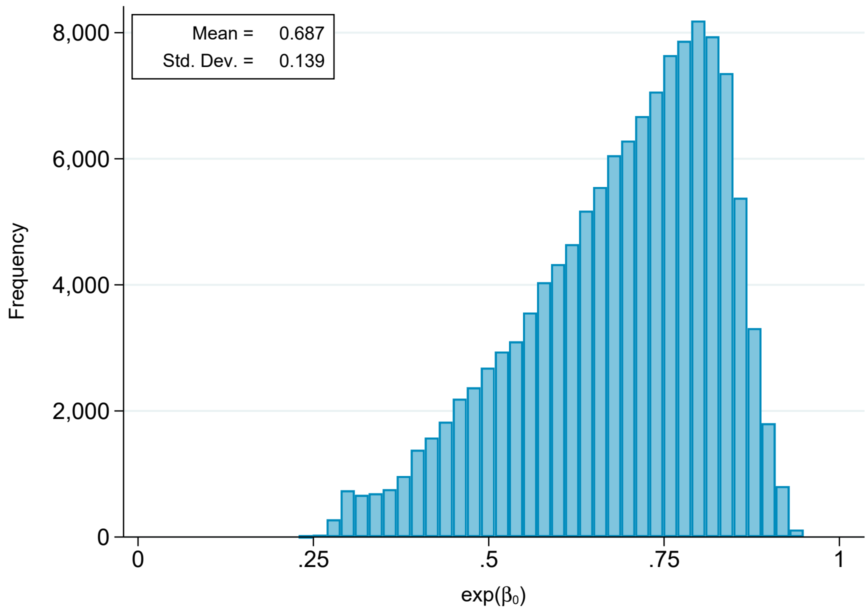


Notes: The figure shows the distribution of the coefficients from the regressions outlined in equation 19, after applying the shrinkage in equation 20. Panel A shows the histogram of the exponential of the intercept $\exp(\hat{\beta}_0)$, while Panel B shows the histogram of the duration-dependence coefficient $\hat{\beta}_D$.

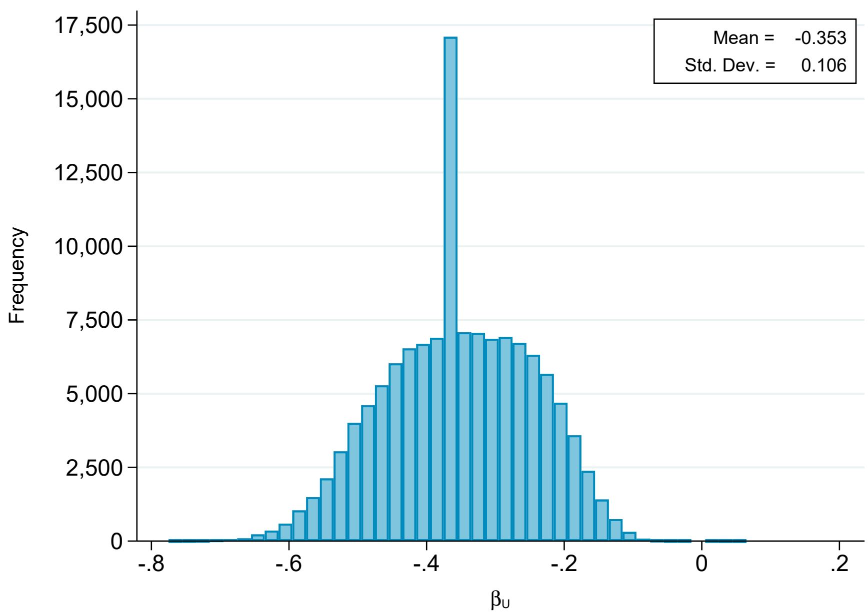
A.6 Heterogeneity in Dynamics over Business Cycle: Additional Results

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

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

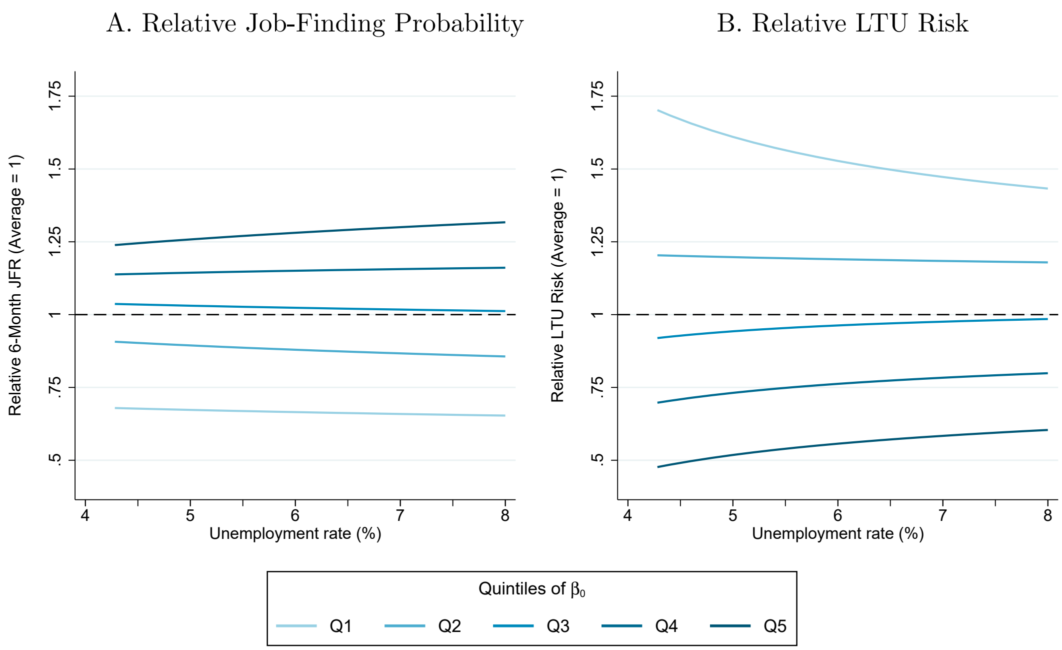


B. Histogram of $\hat{\beta}_U$



Notes: The figure shows the distribution of the coefficients from the regressions outlined in equation 21, after applying the shrinkage in equation 20. 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 $\hat{\beta}_U$.

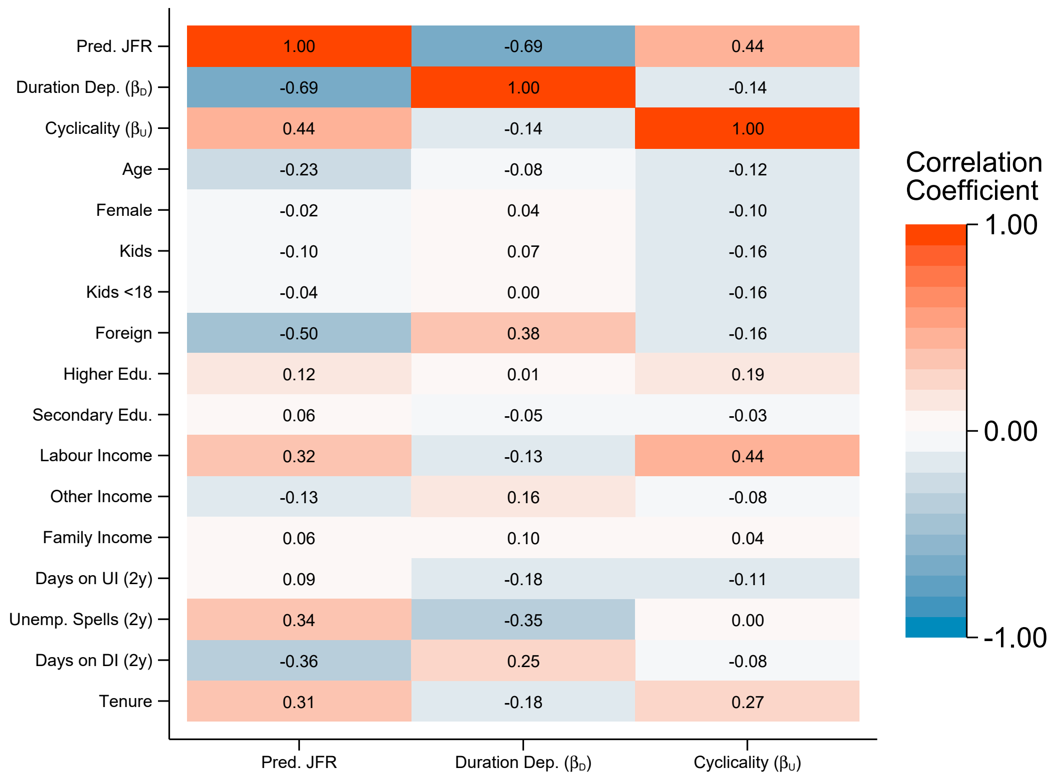
Figure A7: Heterogeneity in Individual Cyclicality: Relative To Average



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.

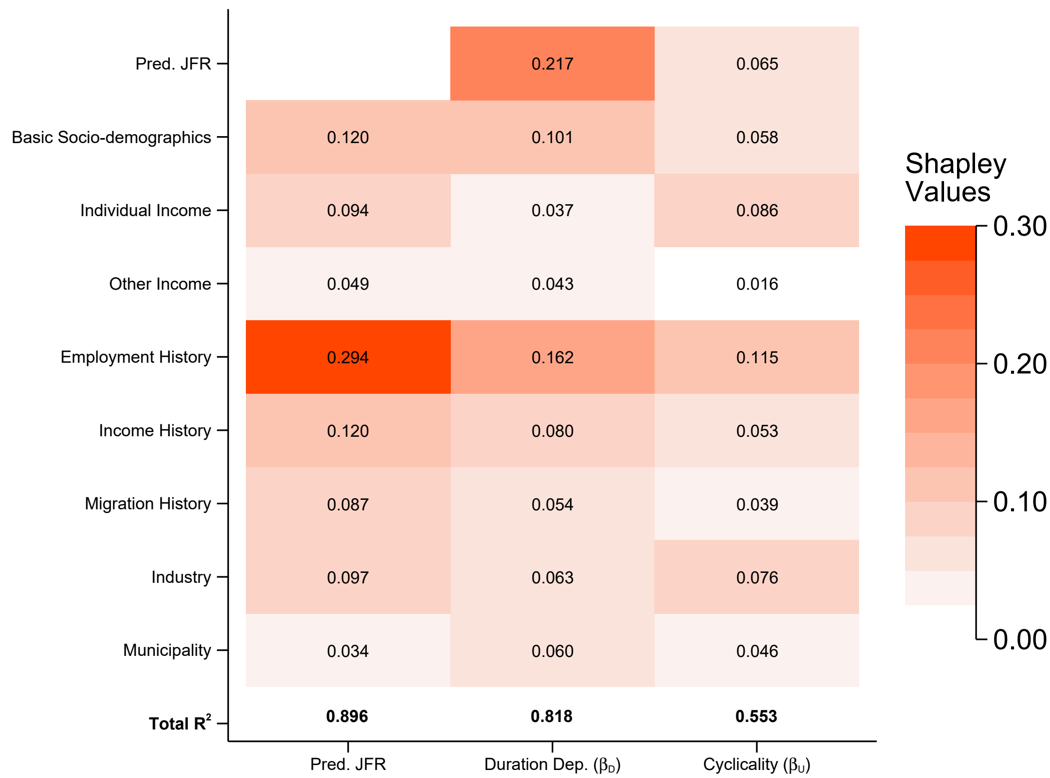
A.7 Heterogeneity and Relationship to Observables: Additional Results

Figure A8: 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 A9: 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.

B Proofs and Further Analysis

B.1 Identification with Observables

Proof of Proposition 1. We first note that

$$\begin{aligned}
cov_d^t\left(F_d^t,\hat{F}_d^t\right) &= E_d^t\left(F_d^t\hat{F}_d^t\right)-E_d^t\left(F_d^t\right)E_d^t\left(\hat{F}_d^t\right)\\
&= E_d^t\left(E_d^t\left(F_d^t\hat{F}_d^t|T_d^t\right)\right)-E_d^t\left(T_d^t\right)E_d^t\left(\hat{F}_d^t\right)\\
&= E_d^t\left(\Pr\left(F_d^t=1|T_d^t\right)E_d^t\left(\hat{F}_d^t\times 1|T_d^t\right)+\Pr\left(F_d^t=0|T_d^t\right)E_d^t\left(\hat{F}_d^t\times 0|T_d^t\right)\right)-E_d^t\left(T_d^t\right)E_d^t\left(\hat{F}_d^t\right)\\
&= E_d^t\left(E_d^t\left(T_d^t\hat{F}_d^t|T_d^t\right)\right)-E_d^t\left(T_d^t\right)E_d^t\left(\hat{F}_d^t\right)\\
&= E_d^t\left(T_d^t\hat{F}_d^t\right)-E_d^t\left(T_d^t\right)E_d^t\left(\hat{F}_d^t\right)\\
&= cov_d^t\left(T_d^t,\hat{F}_d^t\right).
\end{aligned}$$

Next, we use the assumption that $E\left(\varepsilon_d^t|X_d^t\right)=0$ to show that

$$\begin{aligned}
cov_d^t\left(T_d^t,\hat{F}_d^t\right) &= cov_d^t\left(T_d^t(X_d^t)+\varepsilon_d^t,\hat{F}_d^t\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)+cov_d^t\left(\varepsilon_d^t,\hat{F}_d^t\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)+E_d^t\left(\varepsilon_d^t\hat{F}_d^t\right)-E_d^t\left(E_d^t\left(\varepsilon_d^t|X_d^t\right)\right)E_d^t\left(\hat{F}_d^t\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)+E_d^t\left(\varepsilon_d^t\hat{F}_d^t\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)+E_d^t\left(E_d^t\left(\varepsilon_d^t\hat{F}_d^t|X_d^t\right)\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)+E_d^t\left(E_d^t\left(\varepsilon_d^t|X_d^t\right)E_d^t\left(\hat{F}_d^t|X_d^t\right)\right)\\
&= cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right).
\end{aligned}$$

Combining the fact that $cov_d^t\left(T_d^t,\hat{F}_d^t\right)=cov_d^t\left(\hat{F}_d^t,T_d^t(X_d^t)\right)$ and that $cov_d^t\left(F_d^t,\hat{F}_d^t\right)=cov_d^t\left(T_d^t,\hat{F}_d^t\right)$, we get that:

$$cov_d^t\left(F_d^t,\hat{F}_d^t\right) = cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right).$$

Now we can use the Cauchy-Schwarz inequality,

$$\begin{aligned}
var_d^t\left(T_d^t(X_d^t)\right)var_d^t\left(\hat{F}_d^t\right) &\geq cov_d^t\left(T_d^t(X_d^t),\hat{F}_d^t\right)^2\\
&= cov_d^t\left(F_d^t,\hat{F}_d^t\right)^2.
\end{aligned}$$

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

$$\frac{var_d^t\left(T_d^t(X_d^t)\right)}{var_d^t\left(F_d^t\right)}\geq\frac{cov_d^t\left(F_d^t,\hat{F}_d^t\right)^2}{var_d^t\left(F_d^t\right)var_d^t\left(\hat{F}_d^t\right)}=R^2\left(F_d^t,\hat{F}_d^t\right).$$

Given our assumption that $E_d^t(\varepsilon_d^t|X_d^t) = 0$, we also have that $var_d^t(T_d^t) \geq var_d^t(T_d^t(X_d^t))$ and thus

$$R^2\left(F_d^t, \hat{F}_d^t\right) \leq \frac{var_d^t(T_d^t(X_d^t))}{var_d^t(F_d^t)} \leq \frac{var_d^t(T_d^t)}{var_d^t(F_d^t)}.$$

QED.

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

Proposition B1. *If the predictor is unbiased, i.e.* $E_d^t(\hat{F}_d^t|X_d^t) = T_d^t(X_d^t)$*, 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:*

$$cov_d^t(F_d^t, \hat{F}_d^t) = var_d^t(T_d^t(X_d^t)). \tag{B1}$$

Proof. We take from the proof of Proposition 1 that $cov_d^t(F_d^t, \hat{F}_d^t) = cov_d^t(T_d^t(X_d^t), \hat{F}_d^t)$ and then use the fact that $E_d^t(\hat{F}_d^t|X_d^t) = T_d^t(X_d^t)$ as follows

$$\begin{aligned}
cov_d^t(F_d^t, \hat{F}_d^t) &= cov_d^t(T_d^t(X_d^t), \hat{F}_d^t,) \\
&= E_d^t\left(T_d^t(X_d^t)\hat{F}_d^t\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(\hat{F}_d^t\right) \\
&= E_d^t\left(E_d^t(T_d^t(X_d^t)\hat{F}_d^t|X_d^t)\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(E_d^t(\hat{F}_d^t|X_d^t)\right) \\
&= E_d^t\left(T_d^t(X_d^t)E_d^t(\hat{F}_d^t|X_d^t)\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(E_d^t(\hat{F}_d^t|X_d^t)\right) \\
&= E_d^t\left(T_d^t(X_d^t)^2\right) - E_d^t\left(T_d^t(X_d^t)\right)^2 \\
&= var_d^t(T_d^t(X_d^t)).
\end{aligned}$$

QED.

Proof of Proposition 2. We take from the proof of Proposition 1 that, for any t and d, $cov_d^t(F_d^t, \hat{F}_d^t) = cov_d^t(T_d^t(X_d^t), \hat{F}_d^t)$. By extension $cov_d^t(F_d^t, \hat{F}_{d'}^{t'}) = cov_d^t(T_d^t(X_d^t), \hat{F}_{d'}^{t'})$ when $E\left(\varepsilon_d^t|X_{d'}^{t'}\right) = 0$. The latter assumption holds trivially when X are variables that are fixed across states t and d. We then use the fact that $E_d^t(\hat{F}_{d'}^{t'}|X_d^t) = T_{d'}^{t'}(X_d^t)$ and proceed as follows,

$$\begin{aligned}
cov_d^t(F_d^t, \hat{F}_{d'}^{t'}) &= cov_d^t(T_d^t(X_d^t), \hat{F}_{d'}^{t'}) \\
&= E_d^t\left(T_d^t(X_d^t)\hat{F}_{d'}^{t'}\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(\hat{F}_{d'}^{t'}\right) \\
&= E_d^t\left(E_d^t(T_d^t(X_d^t)\hat{F}_{d'}^{t'}|X_d^t)\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(E_d^t(\hat{F}_{d'}^{t'}|X_d^t)\right) \\
&= E_d^t\left(T_d^t(X_d^t)E_d^t(\hat{F}_{d'}^{t'}|X_d^t)\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(E_d^t(\hat{F}_{d'}^{t'}|X_d^t)\right) \\
&= E_d^t\left(T_d^t(X_d^t)T_{d'}^{t'}(X_d^t)\right) - E_d^t\left(T_d^t(X_d^t)\right) E_d^t\left(T_{d'}^{t'}(X_d^t)\right) \\
&= cov_d^t(T_d^t(X_d^t), T_{d'}^{t'}(X_d^t)).
\end{aligned}$$

QED.

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

$$\begin{aligned}
E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) &= E_d^t\left[T_d^t - T_{d+1}^t\right] + E_d^t\left(T_{d+1}^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \int T_{d+1}^t dG_d^t(T_d^t) - \int T_{d+1}^t dG_{d+1}^t(T_{d+1}^t) \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \int T_{d+1}^t dG_d^t(T_d^t) - \frac{\int T_{d+1}^t(1-T_d^t)dG_d^t(T_d^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \frac{(1-E_d^t(T_d^t))E_d^t(T_{d+1}^t)}{1-E_d^t(T_d^t)} - \frac{\int T_{d+1}^t(1-T_d^t)dG_d^t(T_d^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \frac{(1-E_d^t(T_d^t))E_d^t(T_{d+1}^t)}{1-E_d^t(T_d^t)} - \frac{E_d^t(T_{d+1}^t) - E_d^t(T_{d+1}^t T_d^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \frac{E_d^t(T_{d+1}^t T_d^t) - E_d^t(T_{d+1}^t)E_d^t(T_d^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left[T_d^t - T_{d+1}^t\right] + \frac{cov_d^t(T_d^t, T_{d+1}^t)}{1-E_d^t(T_d^t)},
\end{aligned}$$

where we used the fact that $dG_{d+1}^t(T_{d+1}^t) = \frac{(1-T_d^t)dG_d^t(T_d^t)}{\int(1-T_d^t)dG_d^t(T_d^t)}$. The equation above can be re-arranged to

$$\begin{aligned}
E_d^t\left(T_d^t - T_{d+1}^t\right) &= E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t, T_{d+1}^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t(X^t)+\varepsilon_d^t, T_{d+1}^t(X^t)+\varepsilon_{d+1}^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t(X^t)+\varepsilon_d^t, T_{d+1}^t(X^t))}{1-E_d^t(T_d^t)} - \frac{cov_d^t(T_d^t(X_d^t)+\varepsilon_d^t, \varepsilon_{d+1}^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t(X^t), T_{d+1}^t(X^t))}{1-E_d^t(T_d^t)} - \frac{cov_d^t(T_d^t(X_d^t)+\varepsilon_d^t, \varepsilon_{d+1}^t)}{1-E_d^t(T_d^t)} \\
&= E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t(X^t), T_{d+1}^t(X^t))}{1-E_d^t(T_d^t)} - \frac{cov_d^t(\varepsilon_d^t, \varepsilon_{d+1}^t)}{1-E_d^t(T_d^t)} - \frac{cov_d^t(T_d^t(X^t), \varepsilon_{d+1}^t)}{1-E_d^t(T_d^t)}
\end{aligned}$$

The last covariance term $cov_d^t(T_d^t(X^t), \varepsilon_{d+1}^t)$ equals 0 by the assumption that any unobserved heterogeneity is orthogonal to the observables, $E(\varepsilon_d^t|X^t) = 0$ for any d. This indeed implies that unobserved heterogeneity and observable heterogeneity across adjacent periods are orthogonal for a given set of individuals.

If $cov_d^t(\varepsilon_d^t, \varepsilon_{d+1}^t) \geq 0$, then the equation above implies

$$E_d^t\left(T_d^t - T_{d+1}^t\right) \leq E_d^t\left(T_d^t\right) - E_{d+1}^t\left(T_{d+1}^t\right) - \frac{cov_d^t(T_d^t(X^t), T_{d+1}^t(X^t))}{1-E_d^t(T_d^t)}.$$

Next, we use the fact that, in the hold-out sample, $E_d^t\left(F_d^t\right) = E_d^t\left(T_d^t\right)$, $E_{d+1}^t\left(F_{d+1}^t\right) = E_{d+1}^t\left(T_{d+1}^t\right)$, and $cov_d^t(F_d^t, \hat{F}_{d+1}^t) = cov_d^t(T_d^t(X^t), T_{d+1}^t(X^t))$ (from Proposition 2) and thus get

$$E_d^t\left(T_d^t - T_{d+1}^t\right) \leq E_d^t\left(F_d^t\right) - E_{d+1}^t\left(F_{d+1}^t\right) - \frac{cov_d^t\left(F_d^t, \hat{F}_{d+1}^t\right)}{1 - E_d^t\left(F_d^t\right)}.$$

QED.

B.2 Identification with Multiple Spell Data

Proof of Proposition 3.

$$\begin{aligned}
cov_d^{t_1,t_2}\left(F_d^{t_1},F_d^{t_2}\right) &= E_d^{t_1,t_2}\left[F_d^{t_1}F_d^{t_2}\right]-E_d^{t_1,t_2}\left(F_d^{t_1}\right)E_d^{t_1,t_2}\left(F_d^{t_2}\right)\\
&= E_d^{t_1,t_2}\left[E_d^{t_1,t_2}\left(F_d^{t_1}F_d^{t_2}|T_d^{t_1},T_d^{t_2}\right)\right]-E_d^{t_1,t_2}\left(T_d^{t_1}\right)E_d^{t_1,t_2}\left(T_d^{t_2}\right)\\
&= E_d^{t_1,t_2}\Big[E_d^{t_1,t_2}\left(1\times 1|T_d^{t_1},T_d^{t_2}\right)\Pr\left(F_d^{t_1}=1\ \&\ F_d^{t_2}=1|T_d^{t_1},T_d^{t_2}\right)\\
&\quad +E\left(1\times 0+0\times 1+0\times 0|T_d^{t_1},T_d^{t_2}\right)\left(1-\Pr\left(F_d^{t_1}=1\ \&\ F_d^{t_2}=1|T_d^{t_1},T_d^{t_2}\right)\right)\Big]\\
&\quad -E_d^{t_1,t_2}\left(T_d^{t_1}\right)E_d^{t_1,t_2}\left(T_d^{t_2}\right)\\
&= E_d^{t_1,t_2}\left[\Pr\left(F_d^{t_1}=1\ \&\ F_d^{t_2}=1|T_d^{t_1},T_d^{t_2}\right)\right]-E_d^{t_1,t_2}\left(T_d^{t_1}\right)E_d^{t_1,t_2}\left(T_d^{t_2}\right)\\
&= E_d^{t_1,t_2}\left[T_d^{t_1}T_d^{t_2}\right]-E_d^{t_1,t_2}\left(T_d^{t_1}\right)E_d^{t_1,t_2}\left(T_d^{t_2}\right)\\
&= cov_d^{t_1,t_2}\left(T_d^{t_1},T_d^{t_2}\right). \qquad \text{(B2)}
\end{aligned}$$

QED.

Proof of Corollary 2. One can decompose the variance in a persistent and transitory part as follows:

$$var_d^{t_1,t_2}\left(T_d^{t_1}\right) = cov_d^{t_1,t_2}\left(T_d^{t_1},T_d^{t_2}\right)+cov_d^{t_1,t_2}\left(T_d^{t_1},T_d^{t_1}-T_d^{t_2}\right).$$

Using Proposition 3 from above, and if $cov_d^{t_1,t_2}\left(T_d^{t_1},T_d^{t_1}-T_d^{t_2}\right)<0$, then this implies that $cov_d^{t_1,t_2}\left(F_d^{t_1},F_d^{t_2}\right) < var_d^{t_1,t_2}\left(T_d^{t_1}\right)$.
QED.

B.3 Identification with Observables and Multiple Spell Data

Proof of Proposition 4. If the predictor is unbiased, i.e. $E_d^{t_1,t_2}(\hat{F}_d^{t_i}|X_d^{t_i}) = T_d^{t_i}(X_d^{t_i})$ for $i = 1, 2$, then following the proof of Proposition 2, we get

$$\begin{aligned}
cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) &= E_d^{t_1,t_2}(F_d^{t_1}\hat{F}_d^{t_2}) - E_d^{t_1,t_2}(F_d^{t_1})E_d^{t_1,t_2}(\hat{F}_d^{t_2}) \\
&= E_d^{t_1,t_2}(F_d^{t_1}T_d^{t_2}(X_d^{t_2})) - E_d^{t_1,t_2}(F_d^{t_1})E_d^{t_1,t_2}(T_d^{t_2}(X_d^{t_2})) \\
&= E_d^{t_1,t_2}(T_d^{t_1}T_d^{t_2}(X_d^{t_2})) - E_d^{t_1,t_2}(T_d^{t_1})E_d^{t_1,t_2}(T_d^{t_2}(X_d^{t_2})) \\
&= cov_d^{t_1,t_2}(T_d^{t}, T_d^{t_2}(X_d^{t_2})), \qquad \text{(B3)}
\end{aligned}$$

and

$$\begin{aligned}
cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) &= E_d^{t_1,t_2}(\hat{F}_d^{t_1}\hat{F}_d^{t_2}) - E_d^{t_1,t_2}(\hat{F}_d^{t_1})E_d^{t_1,t_2}(\hat{F}_d^{t_2}) \\
&= E_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})T_d^{t_2}(X_d^{t_2})) - E_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}))E_d^{t_1,t_2}(T_d^{t_2}(X_d^{t_2})) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})). \qquad \text{(B4)}
\end{aligned}$$

and from Proposition 1, we have

$$cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) = var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})). \qquad \text{(B5)}$$

.

Using equations B2, B3 and B5 and the independence assumption that $E_d^{t_1,t_2}(\varepsilon_d^{t_i}|X_d^{t_j}) = 0$ for all combinations of $i = 1, 2$ and $j = 1, 2$, we get

$$\begin{aligned}
cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) \qquad \text{(B6)} \\
cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2}) + \varepsilon_d^{t_2}) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t}), T_d^{t_2}(X_d^{t_2})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}) \qquad \text{(B7)} \\
cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_1}(X_d^{t_1})) \\
&= var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})), \qquad \text{(B8)}
\end{aligned}$$

which implies

$$\begin{aligned}
L &= cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) + cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) - cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) \\
&= var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}). \qquad \text{(B9)}
\end{aligned}$$

Using the Cauchy-Schwarz inequality, we get

$$var_d^{t_1,t_2}(\varepsilon_d^{t_1})var_d^{t_1,t_2}(\varepsilon_d^{t_2}) \geq cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2})^2.$$

Under random selection of first and second spell, we get

$$var_d^{t_1,t_2}(\varepsilon_d^{t_1}) = var_d^{t_1,t_2}(\varepsilon_d^{t_2}) \geq cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}),$$

and thus

$$\begin{aligned} L &= var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}) \\ &\leq var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + var_d^{t_1,t_2}(\varepsilon_d^{t_1}) = var_d^{t_1,t_2}(T_d^{t_1}) = var_d^{t_1,t_2}(T_d^{t_2}), \end{aligned} \tag{B10}$$

where the last equality follows again from random selection of the first and second spell.
QED.

It is possible to relax the assumption of independence *across* spells in the presence of time-varying observable characteristics. We can prove the following proposition:[1]

Proposition B2. ***Alternative lower bound.*** *For two randomly chosen spells for each individual, if the predictor is unbiased, i.e.* $E_d^{t_1,t_2}(\hat{F}_d^{t_i}|X_d^{t_i}) = T_d^{t_i}(X_d^{t_i})$ *for* $i = 1, 2$*, the following lower bound for the true variance in types holds:*

$$\Lambda = L + cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) - cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) \leq var_d^{t_1,t_2}(T_d^{t_1}),$$

where L *is the lower bound from Proposition 4.*

Proof. Using equations B2, B3, B4 and B5, we get

$$\begin{aligned} cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) \\ cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\ &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_2}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\ cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, F_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}) \\ &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2}) + \varepsilon_d^{t_2}) \\ cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2}) + \varepsilon_d^{t_2}) \\ cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) &= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_1}(X_d^{t_1})) \\ &= var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) = var_d^{t_1,t_2}(T_d^{t_2}(X_d^{t_2})). \end{aligned}$$

Moreover, under random selection of spells, we have perfect symmetry, i.e. $cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) = cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, F_d^{t_2})$. Define $C \equiv cov_d^{t_1,t_2}(F_d^{t_1}, F_d^{t_2}) + cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) - 2cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2})$, then under

[1] Of course, we still impose the independence assumption for characteristics from the same spell, i.e. $E_d^{t_1,t_2}(\varepsilon_d^{t_i}|X_d^{t_i}) = 0$ for $i = 1, 2$.

symmetry we get:

$$
\begin{aligned}
C &= cov_d^{t_1,t_2}(F_d^t, F_d^{t_2}) + cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, \hat{F}_d^{t_2}) - cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_2}) - cov_d^{t_1,t_2}(\hat{F}_d^{t_1}, F_d^{t_2}) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2}) + \varepsilon_d^{t_2}) \\
&+ cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) \\
&- cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}) + \varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\
&- cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2}) + \varepsilon_d^{t_2}) \\
&= cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2})) + cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), \varepsilon_d^{t_2}) \\
&+ cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) \\
&- cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) - cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, T_d^{t_2}(X_d^{t_2})) \\
&- cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), T_d^{t_2}(X_d^{t_2})) - cov_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1}), \varepsilon_d^{t_2}) \\
&= cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}). \qquad \text{(B11)}
\end{aligned}
$$

This implies:

$$cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) + C = var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}). \qquad \text{(B12)}$$

Using the Cauchy-Schwarz inequality, we get

$$var_d^{t_1,t_2}(\varepsilon_d^{t_1}) var_d^{t_1,t_2}(\varepsilon_d^{t_2}) \geq cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2})^2.$$

Under random selection of the first and second spell, we get

$$var_d^{t_1,t_2}(\varepsilon_d^{t_1}) = var_d^{t_1,t_2}(\varepsilon_d^{t_2}) \geq cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}),$$

and thus

$$
\begin{aligned}
\Lambda = cov_d^{t_1,t_2}(F_d^{t_1}, \hat{F}_d^{t_1}) + C &= var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + cov_d^{t_1,t_2}(\varepsilon_d^{t_1}, \varepsilon_d^{t_2}) \\
&\leq var_d^{t_1,t_2}(T_d^{t_1}(X_d^{t_1})) + var_d^{t_1,t_2}(\varepsilon_d^{t_1}) \\
&= var_d^{t_1,t_2}(T_d^{t_1}) = var_d^{t_1,t_2}(T_d^{t_2}), \qquad \text{(B13)}
\end{aligned}
$$

where the last equality follows again from random selection of the first and second spell.
QED.

B.4 Testing for Proportionality in Job-Finding Hazards

Our paper tests for proportionality of discrete-time job-finding probabilities to observables characteristics. In this Appendix, we develop the theoretical underpinnings of these tests and how they relate to the proportional hazard model as first introduced by Cox [1972].

We define the continuous-time job finding hazard for cohort t and duration d for an individual with characteristics X^t as $\lambda_d^t(X^t)$. Note that we make here the assumption that the characteristics X do not change over the spell of unemployment as in our empirical application. We also abstract here from unobserved heterogeneity, as discussed in the main text. The survival function related to this hazard is

$$S_d^t(X^t) = \exp\bigg\{-\int_0^d \lambda_y^t(X^t)dy\bigg\}. \tag{B14}$$

The discrete-time job-finding hazard at a duration d over a horizon h for individuals who became unemployed at calendar time t then is

$$\begin{aligned} T_d^t(X^t) &= \frac{S_d^t(X^t) - S_{d+h}^t(X^t)}{S_d^t(X^t)} = 1 - \frac{S_{d+h}^t(X^t)}{S_d^t(X^t)} \\ &= 1 - \frac{\exp\Big\{-\int_0^{d+h}\lambda_y^t(X^t)dy\Big\}}{\exp\Big\{-\int_0^{d}\lambda_y^t(X^t)dy\Big\}} \\ &= 1 - \exp\bigg\{-\int_d^{d+h}\lambda_y^t(X^t)dy\bigg\} \\ &= 1 - \exp\bigg\{-\Lambda_d^t(X^t)\bigg\}. \end{aligned} \tag{B15}$$

where $\Lambda_d^t(X^t) = \int_d^{d+h}\lambda_y^t(X^t)dy$ is the cumulated hazard between periods d and $d+h$.

Doing a first-order approximation of $T_d^t(X^t)$ around $h = 0$, we get

$$T_d^t(X) \approx \lambda_d^t(X^t)h. \tag{B16}$$

and thus

$$\frac{T_{d'}^{t'}(X^t)}{T_d^t(X^t)} \approx \frac{\lambda_{d'}^{t'}(X^t)}{\lambda_d^t(X^t)}.$$

for any X^t, t', t, d', and d.

Proportionality Test I - Two States. The Cox-proportional hazard model assumes that $\lambda_d^t(X^t) = \tilde{\lambda}_d^t e^{\alpha X^t}$, where $\tilde{\lambda}_d^t$ is the baseline hazard. Given the proportionality of the baseline hazard to the function with observables, we get

$$\frac{T_{d'}^{t'}(X^t)}{T_d^t(X^t)} \approx \frac{\tilde{\lambda}_{d'}^{t'}}{\tilde{\lambda}_d^t}, \tag{B17}$$

which only depends on the baseline hazard, but not on observables characteristics. Note that we evaluate the function $T^t_d(.)$ for the same set of characteristics in both t and t'.

Proposition B3. *For any X^t, t', t, d', and d, if the job finding hazard is proportional to observables, i.e. $\lambda^t_d(X^t) = \tilde{\lambda}^t_d e^{\alpha X^t}$, and the predictor of the discrete job-finding hazard is unbiased, i.e. $E^t_d(\hat{F}^t_d|X^t) = \hat{F}^t_d(X^t) = T^t_d(X^t)$, then*

$$T^R(X^t, t', t, d', d) \equiv \frac{\hat{F}^{t'}_{d'}(X^t)}{\hat{F}^t_d(X^t)} \approx \frac{\tilde{\lambda}^{t'}_{d'}}{\tilde{\lambda}^t_d} \tag{B18}$$

and thus independent of X^t.

Proof. Under proportionality we have that $\frac{T^{t'}_{d'}(X^t)}{T^t_d(X^t)} \approx \frac{\lambda^{t'}_{d'}(X^t)}{\lambda^t_d(X^t)}$. If the predictor is unbiased then, we can replace the T's and get $\frac{E^t_d(\hat{F}^{t'}_{d'}|X^t)}{E^t_d(\hat{F}^t_d|X^t)} \approx \frac{\tilde{\lambda}^{t'}_{d'}}{\tilde{\lambda}^t_d}$.QED.

We implement the test by regressing $\log(T^R(X^t, t', t, d', d))$ on a subset of observables, $\tilde{X}^t \subset X^t$, for either durations $d = 0$ and $d' = 6$ or years $t = 2006$ and $t' = 2009$. We test whether $var^t_d(\log(T^R(\tilde{X}^t, t', t, d', d))) = 0$ in the sample unemployed in baseline year t and duration d. Note that the log transformation is not necessary but we do it for consistency with our other test further below.

Note that the test is not exact as it relies on an approximation in the neighborhood of $h = 0$, but one can show that the following equation holds exactly:

$$\frac{\log(1 - T^{t'}_{d'}(X^t))}{\log(1 - T^t_d(X^t))} = \frac{\Lambda^{t'}_{d'}(X^t)}{\Lambda^t_d(X^t)}.$$

and thus an alternative (exact) test would be to show that

$$\frac{\log(1 - E^t_d(\hat{F}^{t'}_{d'}|X^t))}{\log(1 - E^t_d(\hat{F}^t_d|X^t))} = \frac{\tilde{\Lambda}^{t'}_{d'}}{\tilde{\Lambda}^t_d}$$

is independent of X^t, where we defined $\Lambda^{t'}_{d'}(X^t) = \tilde{\Lambda}^{t'}_{d'} e^{\alpha X^t}$ and $\tilde{\Lambda}^{t'}_{d'} = \int_d^{d+h} \tilde{\lambda}^t_y dy$ is the cumulated baseline hazard between period d and $d + h$.

Proportionality Test II - Multiple States. One disadvantage of the two-state proportionality test above is that one can only implement it between pairs (t, d) and (t', d'). This is particularly an issue for the time dimension t as we would like to distinguish between trend and cylce, which is not possible in the comparison of two periods only. For this reason, we develop an additional test, which imposes structural assumptions on the baseline hazard of the following form:

$$\lambda^t_d = e^{\beta_C + \beta_D d + \beta_U(u_t - \bar{u}_{t_0}) + \beta(t - t_0)}, \tag{B19}$$

where u_t is the unemployment rate and t a linear time trend. Given the first-order approximation in equation B16, this implies that

$$\log(T_d^t(X^{t_0})) \approx \beta_0 + \alpha X^{t_0} + \beta_D d + \beta_U(u_t - \bar{u}_{t_0}) + \beta_{Tr}(t - t_0), \tag{B20}$$

where $\beta_0 = \beta_C + log(h)$.

Proposition B4. *For any X^{t_0}, t and d, if the job finding hazard is proportional to observables, i.e. $\lambda_d^t(X^{t_0}) = \tilde{\lambda}_d^t e^{\alpha X^{t_0}}$, and the predictor of the discrete job-finding hazard is unbiased, i.e. $E_d^t(\hat{F}_d^t|X^{t_0}) = \hat{F}_d^t(X^{t_0}) = T_d^t(X^{t_0})$, then*

$$\log(\hat{F}_d^t(X^{t_0})) \approx \beta_0(X^{t_0}) + \beta_D d + \beta_U(u_t - \bar{u}_{t_0}) + \beta_{Tr}(t - t_0) \tag{B21}$$

and thus b_D, b_U and b_{Tr} are independent of X^{t_0}.

Proof. Replacing $\hat{F}_d^t(X^{t_0}) = T_d^t(X^{t_0})$ in equation B20 above gives equation B21. QED.

We implement the test by computing

$$\log(\hat{F}_d^{t_0}(X^{t_0})) = \beta_0(X^{t_0}) + \beta_D(X^{t_0})d + \eta_d^{t_0} \tag{B22}$$

and

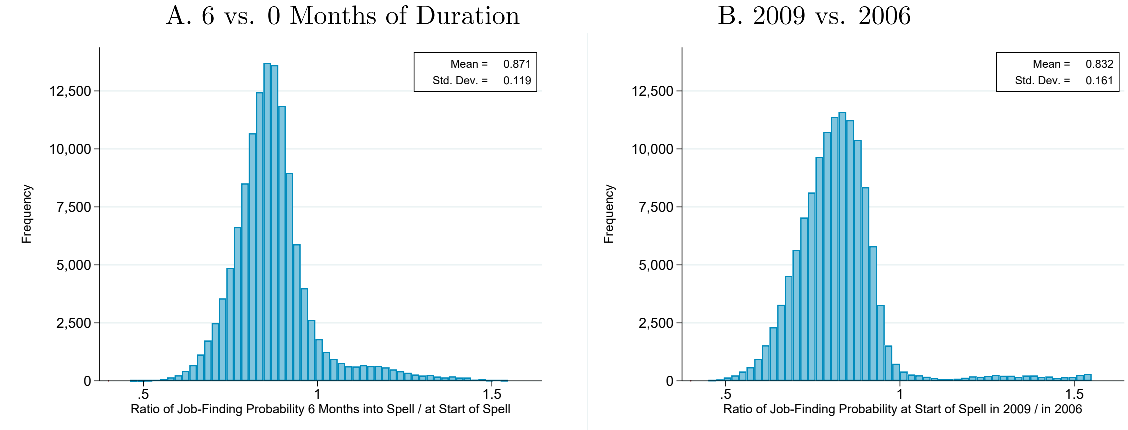
$$\log(\hat{F}_0^t(X^{t_0})) = \beta_0(X^{t_0}) + \beta_U(X^{t_0})(u_t - \bar{u}_{t_0}) + \beta_{Tr}(X^{t_0})(t - t_0) + \eta_0^t \tag{B23}$$

through a linear regression for each X^{t_0} in the sample of unemployed in baseline year t_0 and duration $d = 0$ but using predictions for $d > 0$ and $t \neq t_0$.[2] The error term in the equations captures the approximation error.[3] We test whether $var_0^{t_0}(\hat{\beta}_D(X^{t_0})) = 0$ and $var_0^{t_0}(\hat{\beta}_U(X^{t_0}) = 0$ in the sample of unemployed in baseline year $t_0 = 2006$ and duration $d = 0$, where the hat denotes the fact that the coefficients are estimated. As explained in the main text, we shrink our estimates of $\hat{\beta}_D(X^{t_0})$ and $\hat{\beta}_U(X^{t_0})$ to adjust for sampling error.

[2]We could implement the test in one regression using the prediction models for all combinations of d and t (instead of two regressions) but this is computationally more demanding as we need to compute the prediction model for each duration in each cohort.

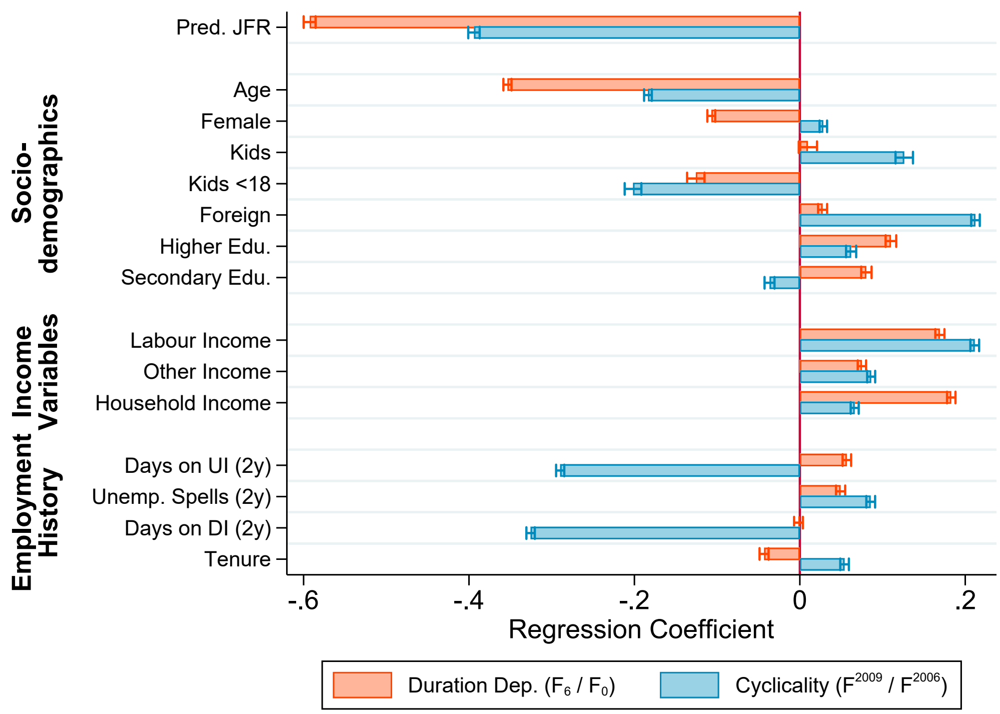
[3]Of course, it may also reflect specification error or deviations from unbiasedness, but we carry out the test under the assumption that the specification is correct and that the predictions are unbiased.

Figure B1: Proportional Hazards Test: Distributions



Notes: Panel A shows the distribution of ratios of the predicted job finding at 6 months of job finding at 0 months of duration (in 2006). Panel B shows the distribution of the predicted job finding in 2009 and the predicted job finding in 2006 (for duration $d = 0$).

Figure B2: Proportional Hazards Test: Regressions



Notes: The figure shows the regression coefficients of a regressions of $\log \frac{\hat{F}_6^{2006}}{\hat{F}_0^{2006}}$ (duration dependence) and $\log \frac{\hat{F}_0^{2009}}{\hat{F}_0^{2006}}$ (cyclicality) on a subset of variables contained in the prediction model and for the sample of individuals in the hold-out sample of 2006 at duration 0 months.

B.5 Selection on Unobservables over the Business Cycle

As discussed in the main text of the paper, our test for the heterogeneity hypothesis is based on observables characteristics and thus omits the potential importance of selection based on unobservables. Our analysis of the sample of multiple unemployment spells in Section 3.5 shows, however, that observables capture about 56 percent of the total heterogeneity in job finding (see Table 4). In this section, we show that based on the assumption that – scaled by their respective variances – the distribution of the unobservable heterogeneity is identical and moves identically to the demeaned distribution of observable heterogeneity, we can estimate the contribution of selection on unobservables over the business cycle.

We start with our model $T_d^t = T_d^t(X_d^t) + \varepsilon_d^t$, where $G_d^t(\tilde{T}_d^t(X_d^t))$ is the CDF of the demeaned observable job-finding rate, $\tilde{T}_d^t(X_d^t)$, and $Q_d^t(\varepsilon_d^t)$ the CDF of unobservable heterogeneity. We impose here the assumption that $G_d^t\left(\frac{\sigma_x}{\sigma_\varepsilon}\varepsilon_d^t\right) = Q_d^t(\varepsilon_d^t)$, i.e. that the demeaned observable and unobservable distributions are identical, up to a scaling factor of their relative standard deviations. Given this assumption and an estimate of the relative standard deviations, we can compute the change in the job-finding rate that comes from observable and unobservable factors between period t_2 and a base period t_1:

$$\Delta(T)^{obs}_{comp} = \int T dG_d^{t_2}(T) - \int T dG_d^{t_1}(T) \tag{B24}$$

$$\begin{aligned}
\Delta(T)^{unobs}_{comp} &= \int \varepsilon dG_d^{t_2}\left(\frac{\sigma_x}{\sigma_\varepsilon}\varepsilon\right) - \int \varepsilon dG_d^{t_1}\left(\frac{\sigma_x}{\sigma_\varepsilon}\varepsilon\right) \\
&= \int \frac{\sigma_\varepsilon}{\sigma_x} T dG_d^{t_2}(T) - \int \frac{\sigma_\varepsilon}{\sigma_x} T dG_d^{t_1}(T) \\
&= \frac{\sigma_\varepsilon}{\sigma_x}\left[\int T dG_d^{t_2}(T) - \int T dG_d^{t_1}(T)\right] \\
&= \frac{\sigma_\varepsilon}{\sigma_x}\Delta(T)^{obs}_{comp}.
\end{aligned} \tag{B25}$$

This implies that given an estimate of the compositional change due to observables, $\Delta(T)^{obs}_{comp}$, and an estimate of the relative variances of observable and unobservable composition, one can compute the shift the job-finding rate due to unobservable factors. Furthermore, the total compositional shift then is

$$\begin{aligned}
\Delta(T)^{tot}_{comp} &= \Delta(T)^{obs}_{comp} + \Delta(T)^{unobs}_{comp} \\
&= \frac{\sigma_x + \sigma_\varepsilon}{\sigma_x}\Delta(T)^{obs}_{comp}.
\end{aligned} \tag{B26}$$

Given the estimates in Table 4 for the observable and unobservable heterogeneity, we get $\sigma_x = \sqrt{0.17} = 0.1304$ and $\sigma_\varepsilon = \sqrt{0.13} = 0.1140$ and thus a total contribution of composition to changes in job finding of

$$\Delta(T)^{tot}_{comp} = 1.87\Delta(T)^{obs}_{comp}.$$

For our regressions in Table A12, this implies a coefficient of $1.87 * 0.082 = 0.153$ for the change in LTU risk due to both unobservable and unobservable shifts in the pool of unemployed, compared to an actual change of 0.700. Under the assumption of identical shifts in the distribution of observable and unobservable heterogeneity, we thus estimate that 22% can be accounted for by compositional changes.[4] For the Great Recession period, the change in the observed LTU risk of 13.6 p.p. between 2006 and 2009 contrasts with the LTU risk due to compositional shifts of only 1.87 p.p. (1.0 p.p. * 1.87) and thus accounting for 14% of the change in the observed LTU risk. We conclude that under the assumption of identical distributional shifts the heterogeneity hypothesis can account for between 14 and 22 percent of the observed change of the job finding and LTU risk over the business cycle. We believe this is a useful exercise and goes beyond prior work because (1) it relies on a prediction model with substantial predictive power and (2) our analysis with multiple spells allows to quantify the importance of unobservable heterogeneity relative to observable heterogeneity. In sum, our analysis suggests that not compositional shifts in the pool of unemployed, but rather aggregate factors in the labor market are responsible for changes of in the aggregate job finding rate/LTU risk over the business cycle.

[4]If we extend the sample to the period 1992-2016, where in the first three years we do not have the full income and employment histories available, the contribution increases to 35% ($=1.87 * 0.146/0.791$).

C Prediction Model: Details and Robustness

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 decision trees and LASSO model. To avoid overfitting, we train and calibrate the prediction algorithm on a training sample, for which we use 50% 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 three 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. Finally, we combine the three obtained predictions into one using a linear combination that we calibrate in the data. We describe each of the three 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 in a separate tuning set, consisting of 10% of the overall sample. We use the package `caret` [Kuhn and Max, 2008] in R, which allows for standardized tuning of a large class of models. In particular, we use the following models (internal `caret` names in brackets):

1. Random forest (`ranger`): random forest is an ensemble model constructed by aggregating a large number of decision trees using a technique known as bootstrap aggregation or "bagging". In a nutshell, the model is constructed as an average of decision trees trained on separate bootstrap samples. To further reduce over-fitting, to which individual trees are prone, the algorithm also selects a random subset of covariates to split on at each node of the tree. We tune two parameters: the minimal node size (`min.node.size`), from the set $\{1, 3, 5, 7, 9, 12, 14\}$; and the number of variables used at each node (`mtry`), from $\{10, 20, 30, 40, 50\}$. This gives us a total of 35 combinations to choose from.

2. Gradient-boosted decision trees (`xgbTree`): gradient boosting is an alternative technique to aggregate simple estimators ("weak learners", in the ML jargon), decision trees in this case, into a stronger ensemble prediction model. The algorithm involves iteratively fitting a decision tree to the residuals of the previous round of boosting. We tune the learning rate (`eta` in the package), a parameter that governs the weight given to the new tree after every boosting round. Smaller values of `eta` shrink the contribution of each round to the ensemble, thus making the model more conservative. We choose `eta` from a grid of 50 log-spaced values between 0.0001 and 0.9.

3. LASSO (`glmnet`): the well-known LASSO is a penalized least squares estimator with an L^1 (i.e., absolute distance) penalty, scaled by a multiplicative parameter (usually denoted by λ, `lambda` in the package). Larger values of λ lead to more shrinkage and a smaller subset of covariates with non-zero coefficients in the model. We choose this penalty or shrinkage parameter from a grid of 50 log-spaced values between 0.0001 and 0.1.

For each of these models, we optimize among the listed alternatives using 3-fold cross validation, where the objective is the area under the receiver operating characteristic curve (AUC).[5] Thus, for each of the parameter values we want to test, and for each fold, predictions are generated using a model trained in the remaining two folds. We then evaluate the AUC in the full tuning sample. The results of this grid search procedure are shown in Figure C1 for the baseline model in 2006. The objective functions for the gradient boosting and LASSO models seem well-behaved, with highest performance in the middle of the grids (`eta` $\in [0.01, 0.1]$ and `lambda` $\in [0.001, 0.01]$) and sharp deterioration as we approach the upper bounds. The random forest, in contrast, exhibits high AUCs throughout the grid, so the objective function looks comparatively flat. The optimal values for this tuning set are: `min.node.size` $= 3$, `mtry` $= 40$, `eta` $= 0.0219$ and `lambda` $= 0.0034$.

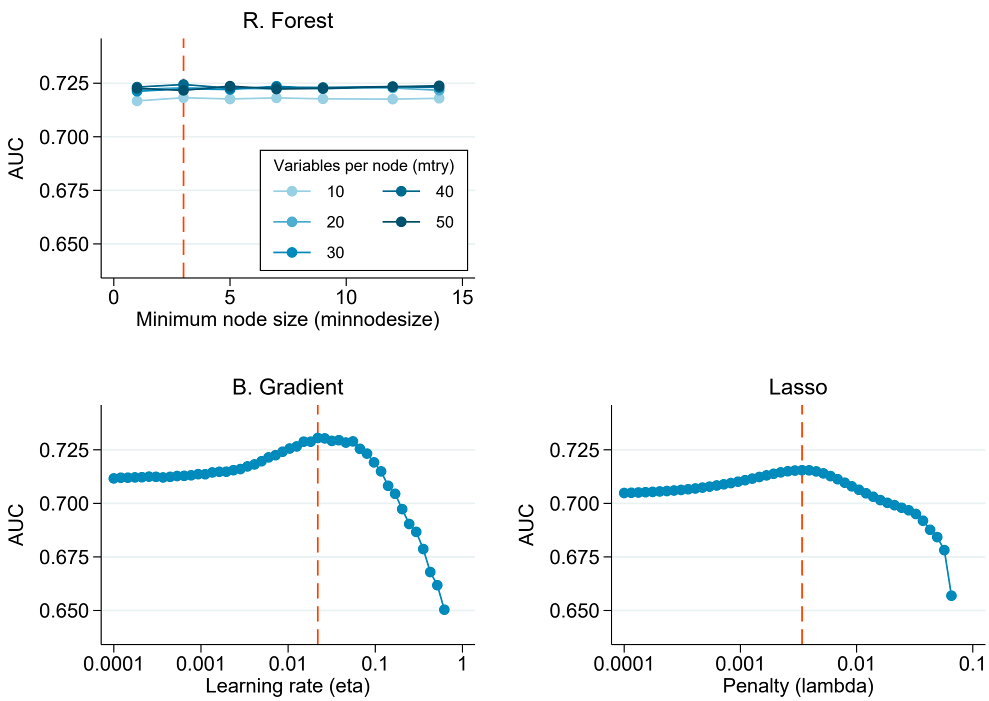
We repeat this procedure for each year and set of variables separately, using 10% of the sample at the start of the spell. That is to say, the baseline model including all of the variables in Table 1 is tuned once for every year between 1992 and 2016, while each of the models in Table 3 is tuned separately in 2006. Then, the same optimal parameter values are used at other unemployment durations (namely, 6 months and 12 months into the spell) for the same year and set of variables. This is done to reduce variance: since sample size declines sharply with unemployment duration, tuning the parameters in these smaller samples would introduce unnecessary noise into the prediction models. Figure C2 shows how the optimal values evolve over time for the baseline model. In line with the shape of the cross-validated AUC in Figure C1, the optimal values of `eta` and `lambda` are quite stable over time, while the two parameters in the random forest model exhibit larger jumps around the grid.

Estimating the Models Using these tuned parameter values, all models are estimated using 30% of the sample (for models that predict job-finding probabilities at the start of the spell) or 40% of the sample (for models that predict at other unemployment durations). These training sets do not include any observations used to tune the parameters in the previous step. Once trained, the models are used to generate predictions for the remaining 60% of the sample.

Obtaining Ensemble Predictor Finally, we combine the predictions from the random forest, gradient-boosted decision 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. 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. We do not constraint the weights to be positive, as this seems to have small effects on the performance

[5] This is a common metric used in the machine learning literature to measure the performance of binary prediction models.

Figure C1: Cross-validation Results for the Baseline Model: 2006

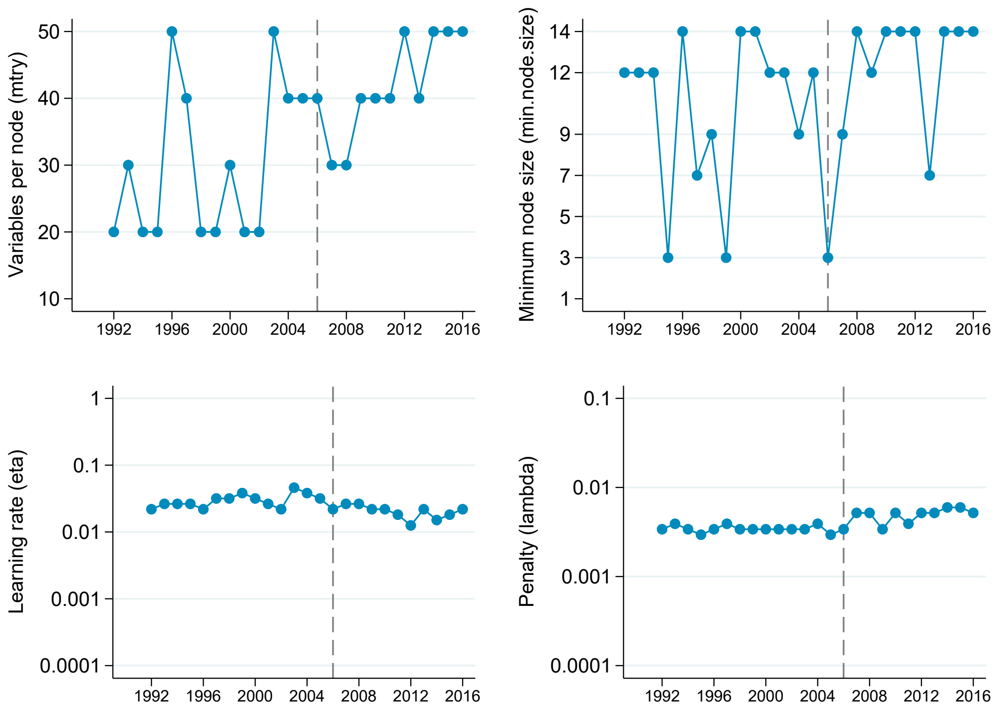


Notes: This figure shows the cross-validated AUC for each of the three Machine Learning models that comprise the baseline ensemble model, for each of the parameter values in our grids. This AUC is evaluated in the tuning sample for the year 2006.

of the model (see panel F in Appendix Table A2). For this step, we use a further 10% of the sample, held out from the previous steps. Again, this is done separately for each year, set of variables and unemployment duration.

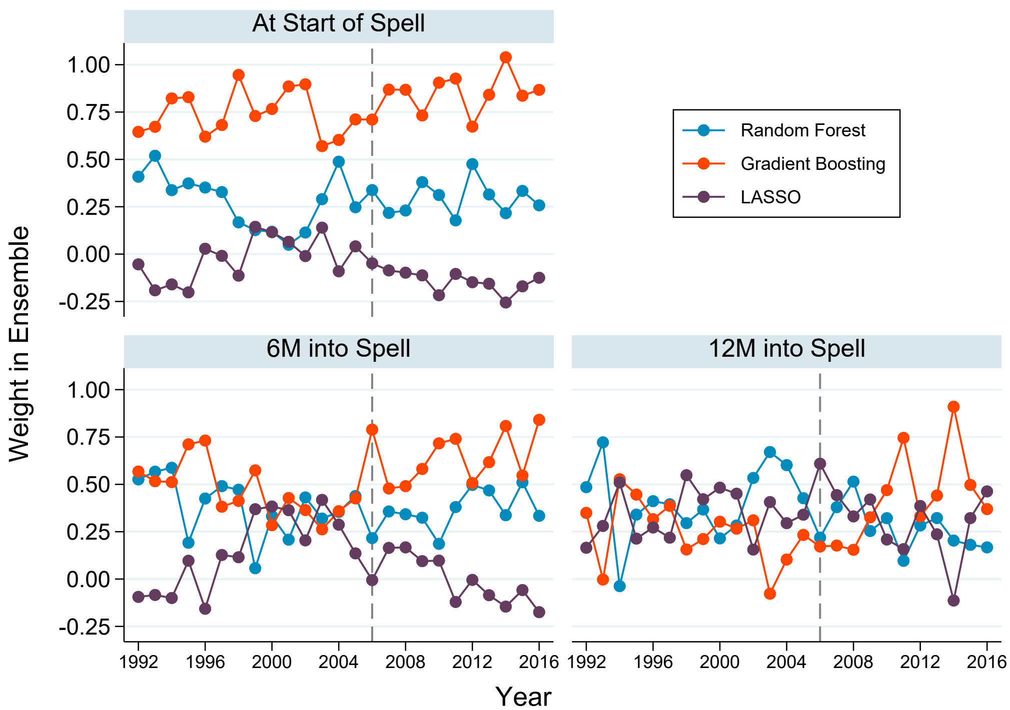
Figure C3 plots the weights in the baseline model across time and unemployment durations. For predictions at the start of the spell, the weights are broadly stable over time and mirrors the ranking of the three models in terms of predictive value (see discussion below). At other unemployment durations, the series are noticeable noisier, with frequent rank reversals from year to year. This suggests higher variance in the weight calibration step, which is performed in much smaller samples 6 and 12 months into the unemployment spell. For 2006 at the start of the spell, the weights are $\hat{\beta}_{rf} = 0.338$, $\hat{\beta}_{gb} = 0.710$ and $\hat{\beta}_{lasso} = -0.048$.

Figure C2: Optimal Tuning Parameters for the Baseline Model: 1992-2016



Notes: This figure shows the optimal value of the tuned parameters in the baseline model, for every year in our sample period (1992-2016).

Figure C3: Ensemble Weights for the Baseline Model: 1992-2016

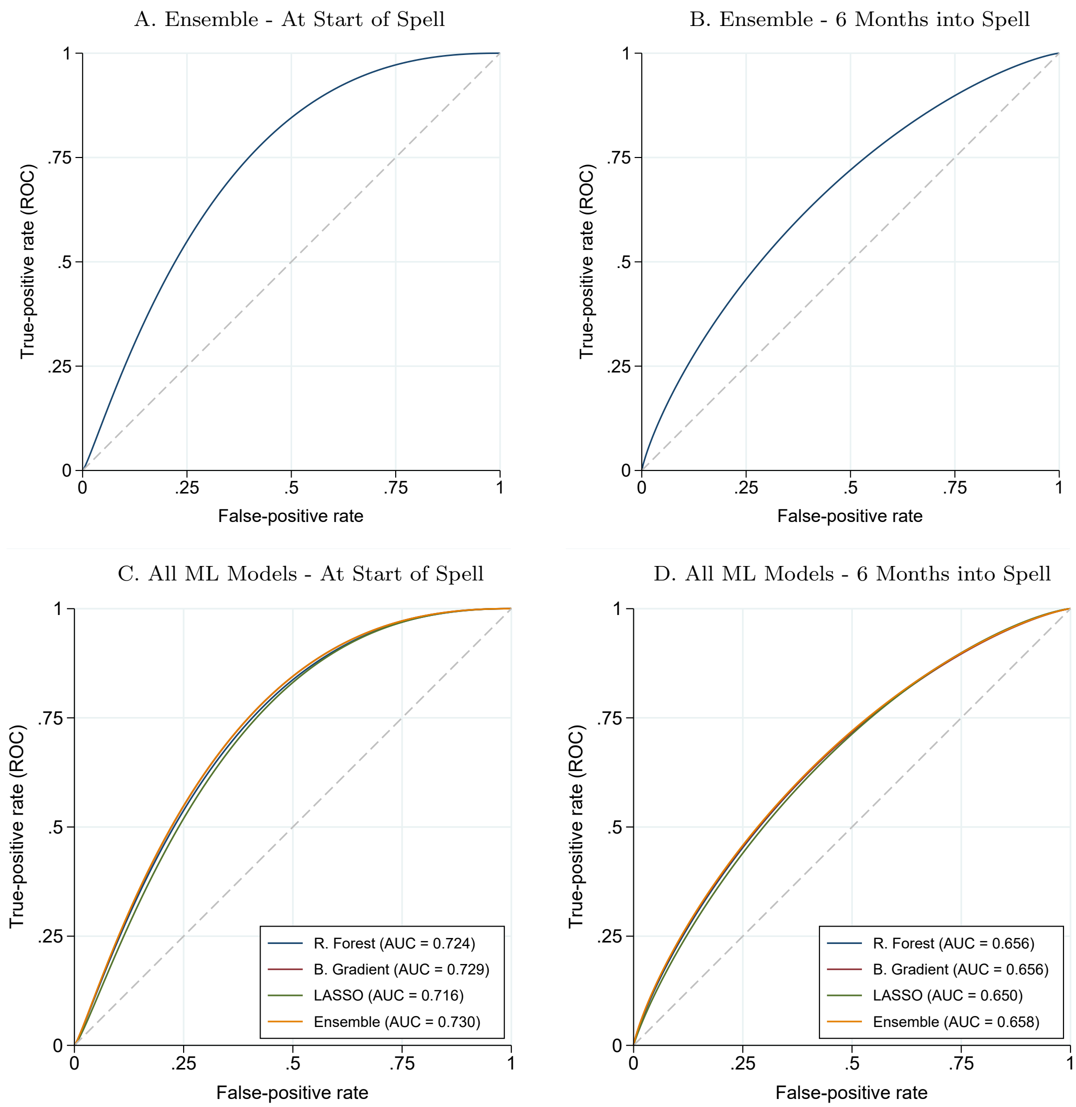


Notes: This figure shows the chosen weights for the Machine Learning models in the ensemble. This is done for every year in the sample period and every unemployment duration (0, 6 and 12 months into the spell).

C.2 Further Results on Predictive Value

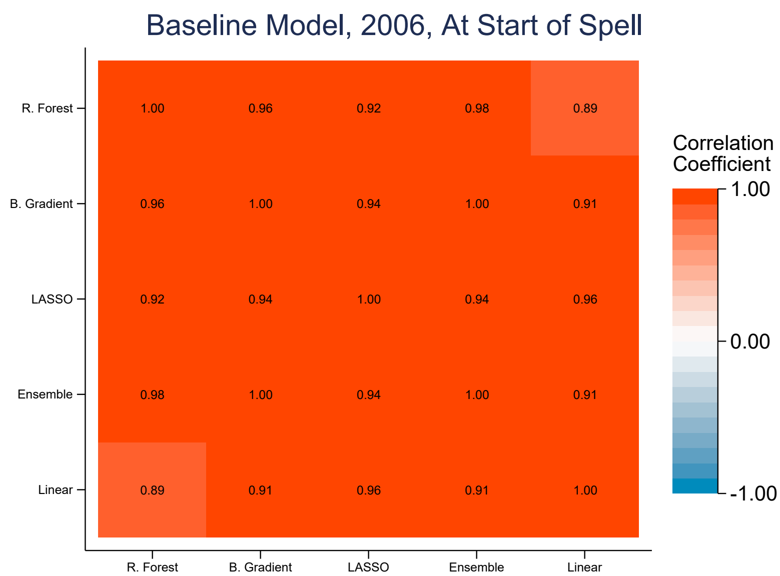
Figure C4 shows receiver operating characteristic curves (ROC) for the baseline ensemble and the three underlying models in the year 2006, evaluated in the respective hold-out sample. The curves show the ensemble slightly improves on the three individual models. Also, the ranking of these in terms of AUC agrees with the main R^2 criterion used in the paper, as shown in Table A2: gradient boosting performs best, followed by random forest and, finally, LASSO. For more discussion, see 3.3. Figure C5 also shows that the predictions from all three models are highly correlated, with coefficients above 0.92.

Figure C4: 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.

Figure C5: Correlation Coefficients between Predictors



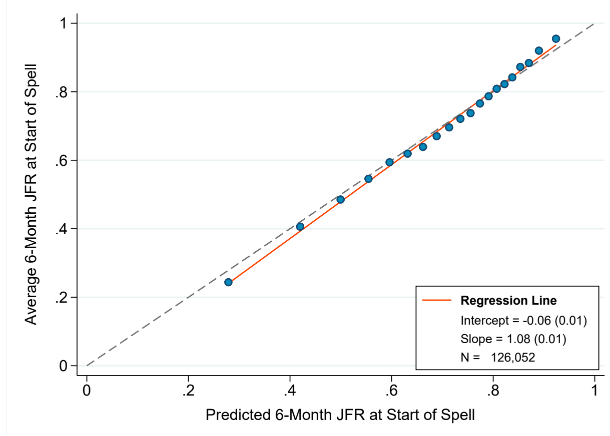
Notes: This figure shows the correlation coefficients between pairs of predictions for the baseline model in 2006.

C.3 Further Results on Accuracy

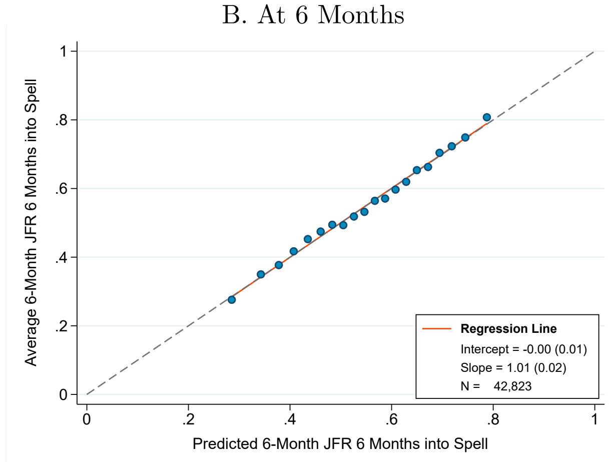
While Panel A of Figure 1 compares predictions and outcomes to assess the accuracy of the Ensemble Model for the entire sample, Figures C6 to C13 show the same plots for different prediction models, by different durations and for different subgroups. Even when we construct 144 groups by income decile, gender, citizenship, days on DI and days on UI, we see from Panel A, B and C of Figure C15 that average predicted probabilities within the groups remain well calibrated. Overall, this analysis confirms the accuracy of the underlying prediction algorithms and provides re-assurance 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 C6: Comparing Predictions to Outcomes: by Duration

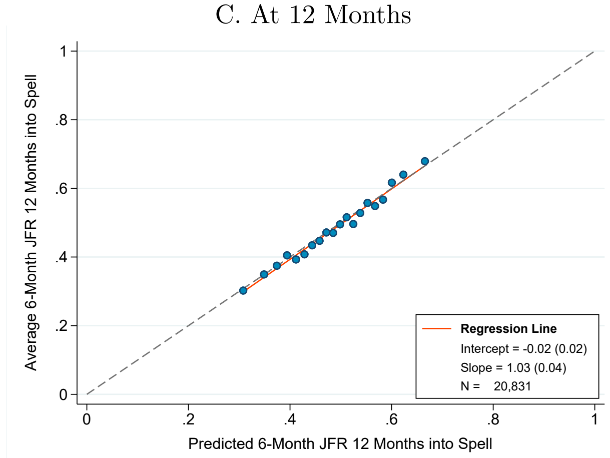
A. At 0 Months



B. At 6 Months

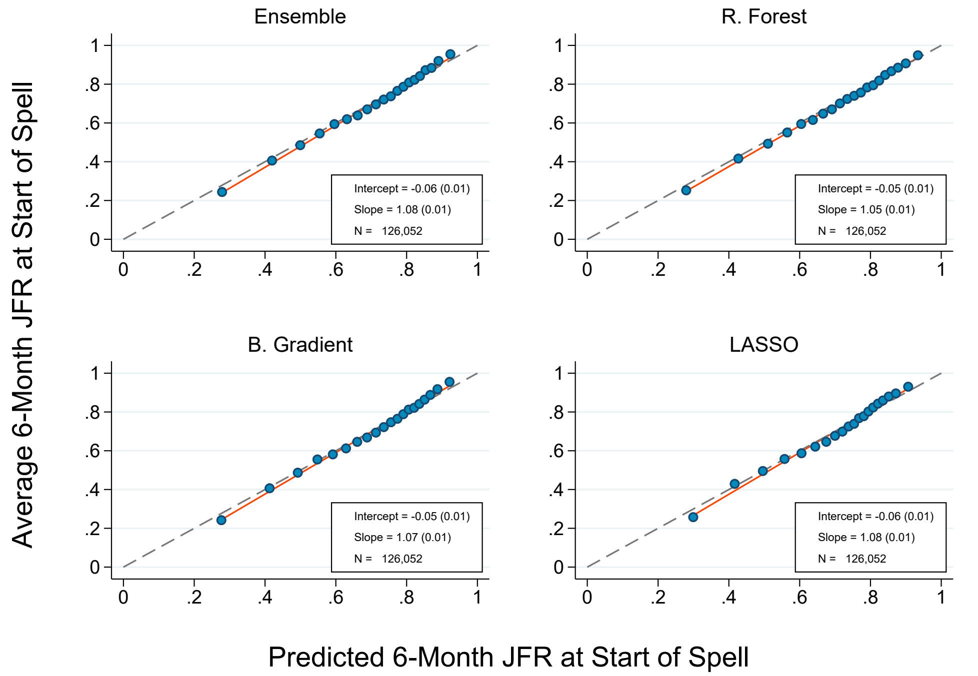


C. At 12 Months



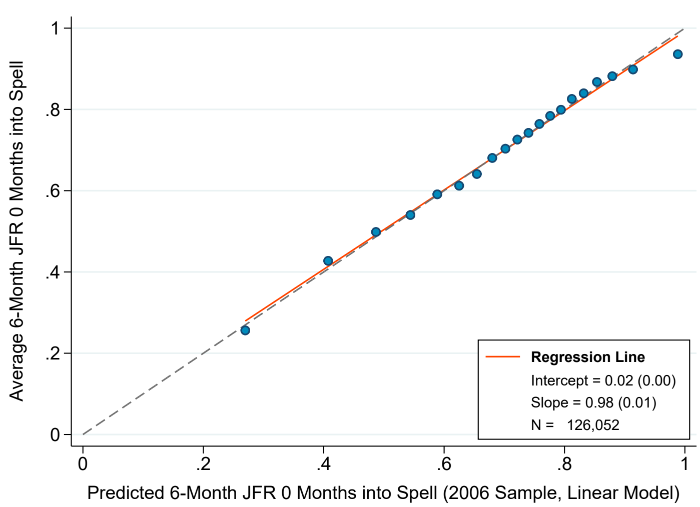
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates, as in Figure 1, for various unemployment durations. Panel A simply reproduces Panel A of 1, while Panels B and C show the predictions 6 and 12 months into the unemployment spell, respectively. All results correspond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

Figure C7: Comparing Predictions to Outcomes: by Prediction Model



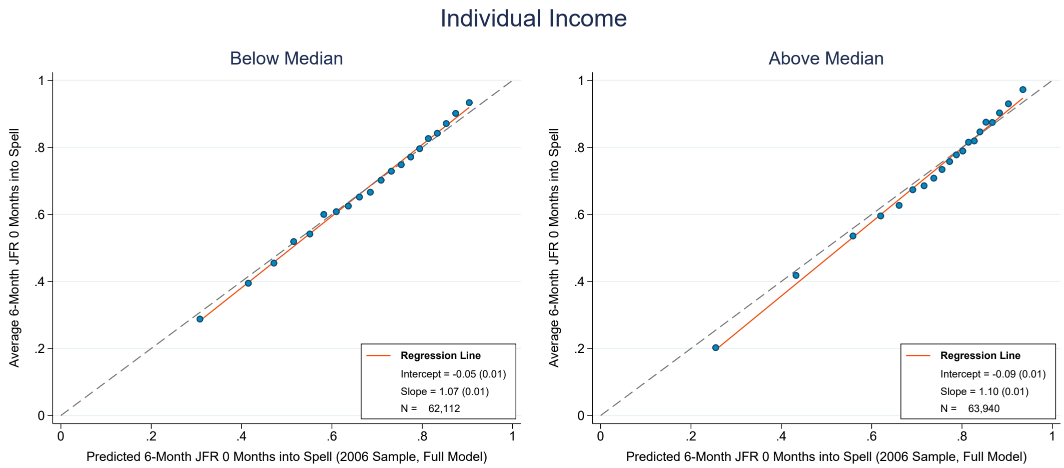
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but for the different underlying prediction models separately.

Figure C8: Comparing Predictions to Outcomes: Linear Model



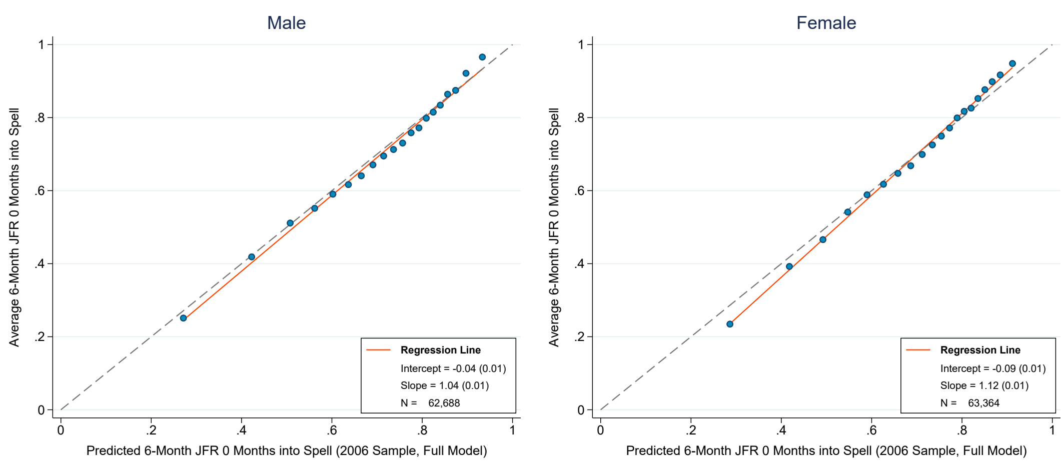
Notes: The figure shows a binned scatter plot of observed job finding and the predictions of the linear model.

Figure C9: Comparing Predictions to Outcomes: by Income



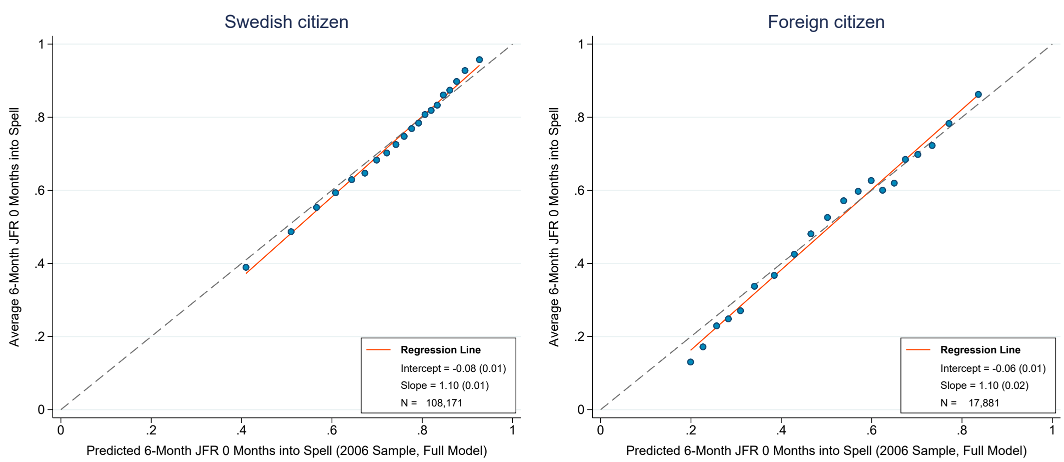
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by individual labour income.

Figure C10: Comparing Predictions to Outcomes: by Gender



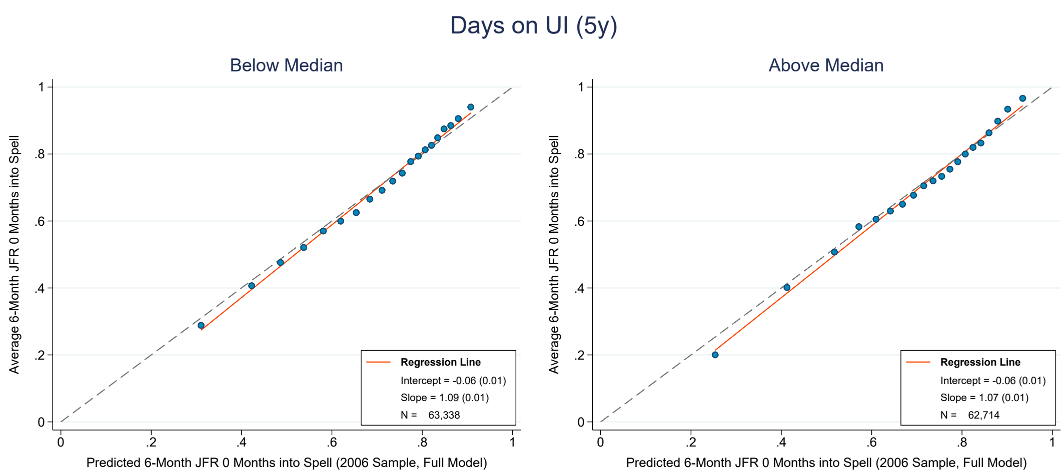
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by gender.

Figure C11: Comparing Predictions to Outcomes: by Citizenship



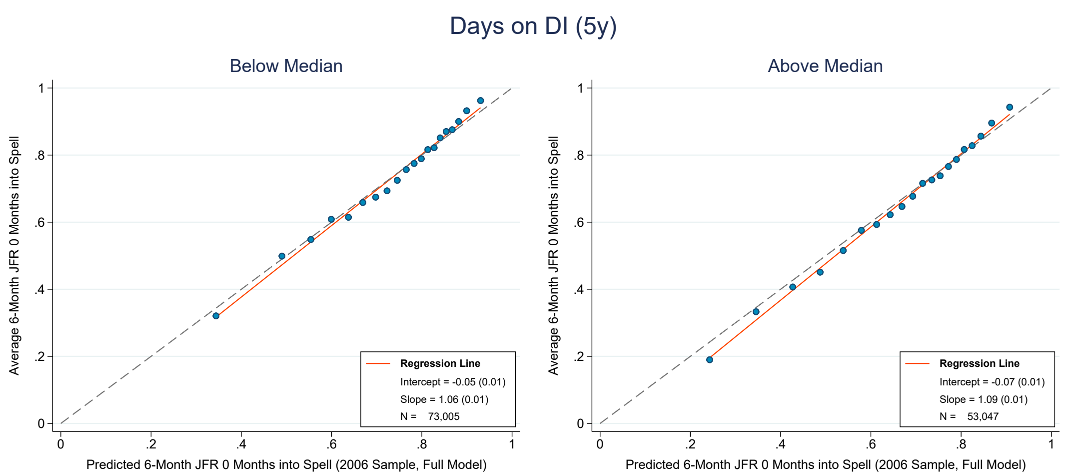
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by citizenship.

Figure C12: 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.

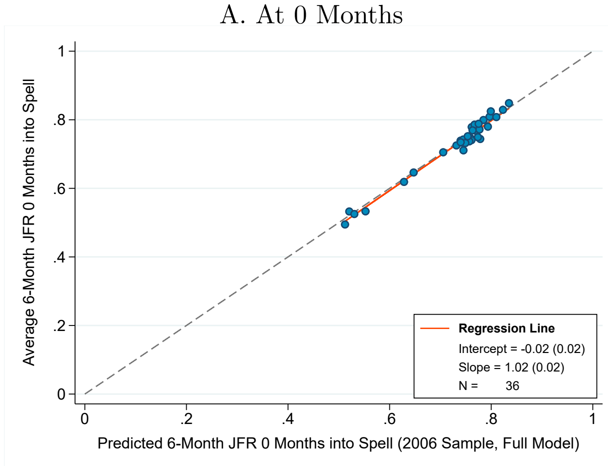
Figure C13: Comparing Predictions to Outcomes: by Days on DI



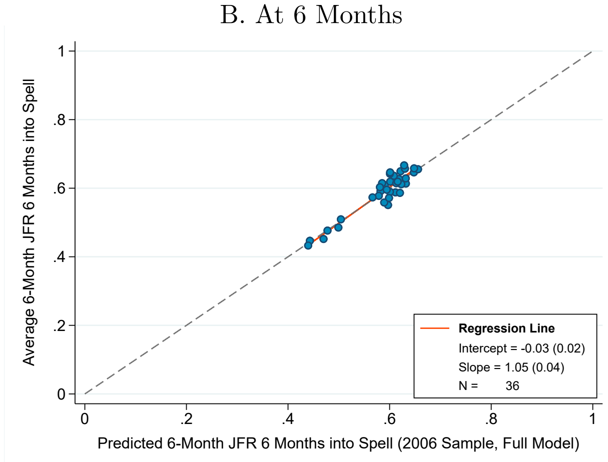
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates at the start of the spell, as in Figure 1, but splitting the 2006 hold-out sample into two bins by days on DI during the 5 years preceding the unemployment spell.

Figure C14: Comparing Predictions to Outcomes: by 36 groups

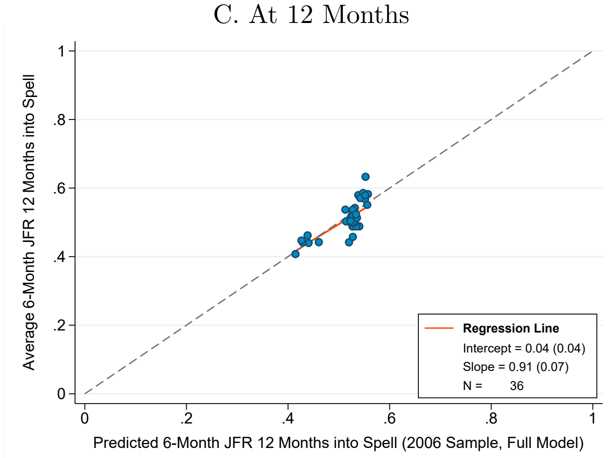
A. At 0 Months



B. At 6 Months



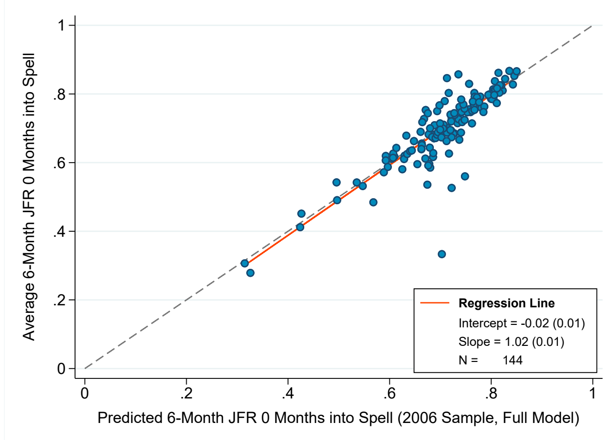
C. At 12 Months



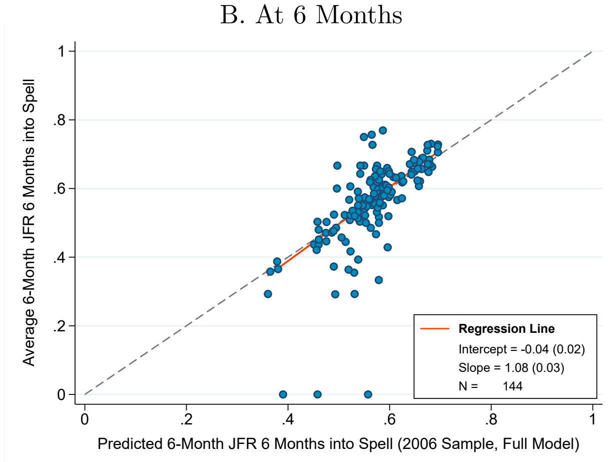
Notes: The figure presents binned scatter plots of observed and predicted job-finding rates for various unemployment durations. Here we construct 36 bins by deciles of labour income, gender and citizenship, and we report average observed and predicted job-finding rates for each bin. The regression output 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.

Figure C15: Comparing Predictions to Outcomes: by 144 groups

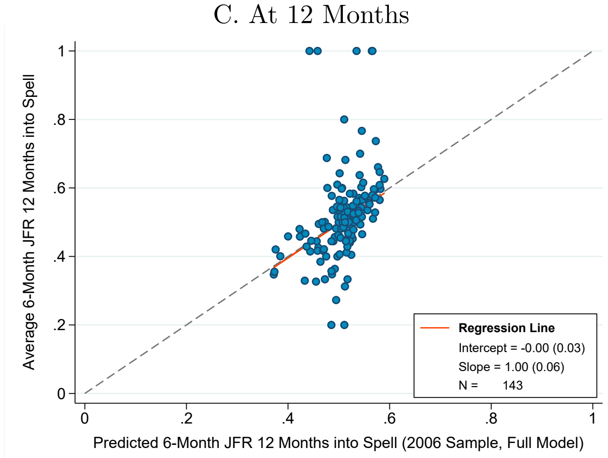
A. At 0 Months



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.