

WEB APPENDIX

USING SOCIAL NETWORK ACTIVITY DATA TO IDENTIFY AND TARGET JOB SEEKERS

September 2020

This web appendix contains the following main sections:

- A. Alternative approach to the stochastic fusion PHMM
- B. Markov Chain Monte Carlo (MCMC) algorithm details for the proposed Partially Hidden Markov Model (PHMM)
- C. Posterior mean and standard deviation of the variances of the heterogeneity distribution for the 6-state deterministic fusion PHMM (112345) and the 6-state stochastic fusion PHMM (123455)
- D. LASSO and Random Forest (RF) ordered logit benchmark models
- E. Robustness analyses predictions through 5-fold cross validation
- F. PHMM estimation results for the second dataset applying the model to d-mail targeting (Section 6 in the main document)
- G. Web appendix references

Appendix A: Alternative approach for the stochastic fusion PHMM

The Stochastic PHMM in section 4.1.5 in the paper fuses the survey response by modifying the transition probabilities in the month of the survey, adding a penalization parameter to the likelihood of transition into states that do not match the “stated” job seeking status in the survey. Here we propose a second way of fusing the survey responses to the states by adding an error term to the survey response to account for possible survey response errors. Instead of constraining the transition matrix at the month of the survey, we include in the state dependent activity distribution m_{itj} another activity during the month of the survey that captures the survey response. Specifically, we take the job seeking response to the survey as an ordinal variable following an ordinal logit model during the month of the survey. More specifically,

$$\begin{aligned}
 P(Y_{it}^{survey} = 1 | S_{it} = k, \theta) &= \frac{\exp(\delta_{k,1}^{survey})}{1 + \exp(\delta_{k,1}^{survey})}, & k = 1, 2, \dots, K \\
 P(Y_{it}^{survey} = c | S_{it} = k, \theta) &= \frac{\exp(\delta_{k,c}^{survey})}{1 + \exp(\delta_{k,c}^{survey})} - \frac{\exp(\delta_{k,c-1}^{survey})}{1 + \exp(\delta_{k,c-1}^{survey})}, & k = 1, 2, \dots, K \\
 & & d = 2, \dots, C - 1 \\
 P(Y_{it}^{survey} = C | S_{it} = k, \theta) &= 1 - \frac{\exp(\delta_{k,C-1}^{survey})}{1 + \exp(\delta_{k,C-1}^{survey})}, & k = 1, 2, \dots, K
 \end{aligned} \tag{A1}$$

where Y_{it}^{survey} is the observed survey response of user i in period t , $\delta_{k,c}^{survey}$, $d = 1, 2, \dots, C - 1$, are the ordered logit thresholds, and C is the number of survey response categories. In order to fuse the survey response into the HMM, we include an additional term in the state dependent activity distribution m_{itj} in Equation (4) in the main document, by multiplying m_{itj} with the distribution in (A1) during the month(s) of the survey, and use Equation (4) in the main document without this term during the other time periods. Doing so, anchors the states to the survey responses, facilitating identification and interpretation of the latent state behavior, while allowing for response error in the survey responses. The likelihood function of the standard HMM in Equation (6) in the main document can now be used to estimate the stochastic PHMM. Thus, this way of fusing the survey responses into the latent states of behavior, requires a minor modification to a traditional HMM.

Appendix B: Markov Chain Monte Carlo (MCMC) algorithm details for the proposed Partially Hidden Markov Model (PHMM)

Given the presence of user-level heterogeneity, a Bayesian framework to estimate the PHMM is most appropriate (e.g., Netzer et al. 2017; Netzer et al. 2008; Schweidel et al. 2011; Ascarza and Hardie 2013; Zhang et al. 2014). We use a Markov Chain Monte Carlo (MCMC) algorithm to sample the posterior distribution directly through Metropolis-Hastings (MH) steps (Hastings, 1970). Given the complexity of our model due to the multivariate nature of the activity variables and the mix of distribution types, we block the parameters into separate sets and each set of parameters is updated separately using a MH step. We implement the adaptive MH algorithm as described in Atchadé and Rosenthal (2005), in short AR, which automatically adjusts the tuning parameter (the variance of the proposal density) of the MH algorithm, to improve efficiency of the MH algorithm. We next outline the main block steps of our MCMC sampler. After that we discuss the detailed implementation of each step.

(1) Update in a MH step the parameters of the logit probabilities for the discrete activity variables δ_{0pk} and δ_{1p} for $p = 1, 2, \dots, P_1, k = 1, 2, \dots, K$, using a multivariate normal proposal density with AR flexible tuning.

(2) Update in a MH step the location parameters of the type 1 Tobit models for the continuous activity variables β_{0pk} and β_{1p} for $p = 1, 2, \dots, P_2, k = 1, 2, \dots, K$, using a multivariate normal proposal density with AR flexible tuning.

(3) Update in a MH step the variance parameters of the type 1 Tobit models for the continuous activity variables σ_{pk}^2 for $p = 1, 2, \dots, P_2, k = 1, 2, \dots, K$. Here we create proposals for $\log(\sigma_{pk}^2)$, which facilitates the implementation of AR flexible tuning from a multivariate normal proposal density.

- (4) Update in a MH step the baseline logit threshold parameters for the initial state τ_j , $j = 1, 2, \dots, K - 1$ and transition probabilities ϕ_{kj} , $k = 1, 2, \dots, K$, $j = 1, 2, \dots, K - 1$ using a multivariate normal proposal density with AR flexible tuning.
- (5) Update in a MH step the user-level heterogeneity parameters of the PHMM α_i for $i = 1, 2, \dots, N$, using a multivariate normal proposal density with AR flexible tuning.
- (6) For the stochastic fusion PHMM, update in a MH step the parameters of the penalization parameters γ_c , $c = 1, 2, \dots, C$, using a multivariate normal proposal density with AR flexible tuning. We constrain the λ_c 's to be positive (through rejection sampling).
- (7) Update in a standard Gibbs step the variance-covariance matrix Σ_α of the upper-level Normal model for the user-level heterogeneity.

Metropolis-Hastings steps (1)—(6)

As steps (1)—(6) are conceptually the same, we discuss how to generate draws for these parameters in each step of the MCMC algorithm for the general case. Let ψ denote the $R \times 1$ vector of parameters to be updated (e.g., in step (1) above, ψ would contain the elements δ_{0pk} and δ_{1p} for $p = 1, 2, \dots, P_1$, $k = 1, 2, \dots, K$). Let θ be the vector containing all other model parameters excluding the parameters in ψ , and let Y be the $NP \times T$ matrix of observed activity. As the full conditional distribution $f(\psi|\theta, Y)$ does not have a closed form expression for our model, we use a MH step to generate a new value for ψ in each step of the MCMC algorithm.

We generate a proposal value for ψ , say ψ_p , from a R -variate normal proposal distribution with the current value ψ , say ψ_c , as the mean, and $\tau_{MH}\Omega_{MH}$ as the variance covariance matrix, where τ_{MH} is a (scalar) parameter and Ω_{MH} is a $R \times R$ positive definite symmetric matrix. Both τ_{MH} and Ω_{MH} are flexible tuning parameters that are updated using an algorithm proposed by Atchadé and Rosenthal (2005), which we outline below.

The proposed value ψ_p is accepted with probability:

$$\min \left\{ \frac{\exp \left(-1/2(\psi_p - \psi_0)' V_0^{-1} (\psi_p - \psi_0) \right) L(Y|\psi_p, \theta)}{\exp \left(-1/2(\psi_c - \psi_0)' V_0^{-1} (\psi_c - \psi_0) \right) L(Y|\psi_c, \theta)}, 1 \right\},$$

where $L(Y|\psi_p, \theta)$ is the value of the full-sample likelihood, given parameters (ψ_p, θ) , which is developed in the main document. Furthermore, ψ_0 and V_0 are the mean and variance-covariance matrix, respectively, of the R -variate normal prior distribution. In our study, we set $\psi_0 = 0$ and $V_0 = 100 \times I_R$ for all parameters.

The flexible tuning parameters τ_{MH} and Ω_{MH} are updated after the first 1,000 iterations. Initially, we set Ω_{MH} to the identity matrix and τ_{MH} to a value such that proposals are accepted in a broad range of 20-80%. This requires initial tuning, which we found to be easily doable. After 1,000 iterations, the parameters τ_{MH} and Ω_{MH} are automatically adjusted to target an acceptance probability of ζ , which we set to 0.28.

The main steps to automatically tune τ_{MH} and Ω_{MH} are the following (for details and proofs we refer to Atchadé and Rosenthal, 2005). Let $\epsilon_1 = 10^{-7}$, $\epsilon_2 = 10^{-6}$, $A_1 = 10^7$ and $g_{MH} = 10/l$ where l is the l -th iteration of the MCMC sampler. First, the parameter τ_{MH} is updated in the l -th iteration of the MH algorithm as:

$$\tau_{MH}^{(l+1)} = \tau_{MH}^{(l)} + g_{MH} \times (\zeta^{(l)} - \zeta) \text{ if } \epsilon_1 < \tau_{MH}^{(l+1)} < A_1,$$

$$\tau_{MH}^{(l+1)} = \epsilon_1 \text{ if } \tau_{MH}^{(l+1)} < \epsilon_1, \text{ and}$$

$$\tau_{MH}^{(l+1)} = A_1 \text{ if } \tau_{MH}^{(l+1)} > A_1.$$

Here, $\zeta^{(l)}$ is the current accept rate in the l -th iteration of the MH algorithm. In other words, if the current accept rate $\zeta^{(l)}$ is below (above) the target ζ accept rate, the updated value

$\tau_{MH}^{(l+1)}$ will be decreased (increased), which reduces (increases) the variance in the proposal distribution above. As such, the future proposed values are more (less) likely to be accepted.

The second tuning parameter Ω_{MH} is updated in the l -th iteration of the MH algorithm as:

$$\Omega_{MH}^{(l+1)} = \Gamma_{MH}^{(l+1)} + \epsilon_2 \times I_R,$$

where the parameter $\Gamma_{MH}^{(l+1)}$ is computed as:

$$\Gamma_{MH}^{(l+1)} = \Gamma_{MH}^{(l)} + g_{MH} \times \left(\left((\psi_c - \mu_{MH}^{(l)}) (\psi_c - \mu_{MH}^{(l)})' \right) - \Gamma_{MH}^{(l)} \right),$$

with

$$\mu_{MH}^{(l+1)} = \mu_{MH}^{(l)} + g_{MH} \times (\psi_c - \mu_{MH}^{(l)}).$$

Loosely speaking, $\mu_{MH}^{(l+1)}$ approximates the posterior mean and $\Gamma_{MH}^{(l+1)}$ approximates the posterior variance-covariance matrix of the parameter ψ when l becomes large. Let $d_{MH}^1 =$

$\sqrt{\sum_{i,j} (\Gamma_{MH}^{(l+1)}(i,j))^2}$, i.e., the square root of the sum of all squared elements of $\Gamma_{MH}^{(l+1)}$, then

$\Gamma_{MH}^{(l+1)} = (A_1/d_{MH}^1) \times \Gamma_{MH}^{(l)}$ if $d_{MH}^1 > A_1$. Similarly, let $d_{MH}^2 = \sqrt{\sum_i (\mu_{MH}^{(l+1)}(i))^2}$, i.e., the square

root of the sum of all squared elements of $\mu_{MH}^{(l+1)}$, then $\mu_{MH}^{(l+1)} = (A_1/d_{MH}^2) \times \mu_{MH}^{(l)}$ if $d_{MH}^2 > A_1$.

The parameters d_{MH}^1 and d_{MH}^2 prevent $\Gamma_{MH}^{(l+1)}$ and $\mu_{MH}^{(l+1)}$ from drifting away to infinity.

We note that for the variance parameters in the type 1 Tobit model we specify a log normal prior (e.g., Zellner, 1971). The advantage of such a specification in our particular case is that Atchadé and Rosenthal's (2005) algorithm for flexible tuning in the MH algorithm can be straightforwardly adapted (step 3 above). We separately update the variance parameters for the variables that we observe in each time period and the variables which we partly observe during the observation window.

Step (6) – Gibbs step to generate a draw for the variance-covariance matrix Σ_α of the upper-level Normal model for the user-level heterogeneity

The full conditional distribution of Σ_α is given by

$$p(\Sigma_\alpha | -) \sim IW(f_N, S_N^{-1}),$$

where ‘- -’ indicates all data and all other parameters, IW is the inverse Wishart probability density distribution, with degrees of freedom

$$f_N = f_0 + N,$$

where f_0 is the prior degrees of freedom and N is the number of users in the sample, and scale matrix

$$S_N = S_0 + \sum_{i=1}^N (\alpha_i - 0) (\alpha_i - 0)',$$

where α_i is the vector of user-level heterogeneity random intercepts of length (say) R and S_0 is the prior scale matrix. We set $S_0 = I_R$ and $f_0 = R + 15$ a priori.

Starting values of the MCMC algorithm

To facilitate faster convergence of the MCMC sampler, we first estimate a basic HMM using maximum-likelihood without user-level heterogeneity and no data fusion (Netzer et al. 2017). We take the maximum-likelihood estimates as starting values for the MCMC sampler for the parameters of the likelihood core of the PHMMs. We set the heterogeneity intercepts to zero and the variance-covariance matrix of the upper level model to the identity matrix.

We use the alternative stochastic fusion PHMM presented in Web Appendix A to obtain reasonable starting values for the two PHMMs presented in the main document.¹ The alternative

¹ Alternatively, one could also estimate a HMM without data fusion for the starting value.

specification is relatively easy to implement irrespective of the number of states and survey response categories, and captures the relationship between the survey response categories and the latent states. Specifically, we take the last draw of the K -state alternative stochastic fusion PHMM (Web Appendix A) as starting value for the MCMC algorithm of the K -state deterministic and stochastic fusion PHMMs in the main document. For the PHMMs with a one to many mapping between the number of states and the job seeking survey response categories, we not only need to determine the starting values of all parameters but also the mapping of the observed survey responses to the states. We manually set the mapping of the survey response categories to the states based on the estimated ordered logit probabilities of the alternative stochastic fusion PHMM (Web Appendix A). For our empirical application, we found that this procedure works reasonably well.

Convergence and label switching

We run the MCMC chain for at least one million iterations burn-in. We ran the MCMC sampler after that for another 500K iterations retaining 2,500 iterations for posterior analysis (because we ran our models on the Amazon EC2 cluster, we reduced computational/memory burden for the cloud computer and data transfer between cloud and local computer by retaining every 200th iteration). Convergence was monitored by inspection of iteration plots of the sampler outputs. We also computed Geweke's (1992) convergence diagnostic (a Z-score) for all fixed-effects parameters of the likelihood kernel (including the initial and transition probability baseline intercepts). For the four PHMMs, Table B1 summarizes the median Z-score (absolute values) across all fixed effects parameters, where values less than 2 suggest that the front and back end of the chain have similar means.

Model	Npars	Geweke
Main dataset		
Det-PHMM 12345	108	0.93
Stoch-PHMM 12345	113	1.09
Det-PHMM 112345	134	1.14
Stoch-PHMM 123455	139	1.12
Dataset used in Section 6 (targeting active job seekers)		
Det-PHMM 112345	108	1.00

Table B1. Summary of Geweke’s z-score to assess convergence of the MCMC chain. Reported values are median values of the z-scores of the ‘Npars’ fixed-effects parameters of the likelihood kernel.

Table B1 confirms the iteration plots and suggests that the MCMC chains converged. Convergence of PHMMs can become problematic as the number of states increase. While in theory we could fit seven (or higher) state PHMMs, we believe that our data is not rich enough to support a 7-state PHMM and to identify which mapping with the survey is best (e.g., 1122345, 1123345, 1123445, 1123455, 1223345, ..., 1112345, 1222345, ..., 1234555). To empirically fit such complex models, we recommend (a) a larger cross section, (b) a longer time period or a more granular time interval (e.g., bi-weekly data instead of monthly data), and (c) multiple time periods in which survey observations are observed for a subset of users.

Lastly, when the number of latent PHMM states exceeds the number of survey categories in the PHMM, one needs to control for label switching for the latent states that correspond to the same survey category in estimation, by either ordering the expected values for one of the activities (e.g., Netzer et al. 2008) or by post-processing techniques (Celeux 1998). We used the post processing approach following Celeux (1998).

Appendix C: Posterior mean and standard deviation of the variances of the heterogeneity distribution for the 6-state deterministic fusion PHMM (112345) and the 6-state stochastic fusion PHMM (123455)

	Mean	St.Dev.		Mean	St.Dev.
α_{i1}^M	2.26	(0.20)	α_{i33}^Q	1.58	(0.40)
α_{i2}^M	0.88	(0.07)	α_{i34}^Q	0.15	(0.08)
α_{i1}^π	2.13	(0.49)	α_{i35}^Q	0.76	(0.30)
α_{i2}^π	0.25	(0.18)	α_{i41}^Q	1.98	(0.59)
α_{i3}^π	0.26	(0.14)	α_{i42}^Q	0.56	(0.20)
α_{i4}^π	0.98	(0.30)	α_{i43}^Q	0.10	(0.05)
α_{i5}^π	0.38	(0.23)	α_{i44}^Q	0.14	(0.07)
α_{i11}^Q	0.13	(0.05)	α_{i45}^Q	1.71	(0.46)
α_{i12}^Q	0.40	(0.17)	α_{i51}^Q	2.93	(0.97)
α_{i13}^Q	0.10	(0.06)	α_{i52}^Q	0.12	(0.06)
α_{i14}^Q	0.11	(0.06)	α_{i53}^Q	0.58	(0.29)
α_{i15}^Q	0.47	(0.37)	α_{i54}^Q	0.24	(0.15)
α_{i21}^Q	2.02	(0.75)	α_{i55}^Q	2.16	(0.52)
α_{i22}^Q	0.35	(0.15)	α_{i61}^Q	3.79	(1.06)
α_{i23}^Q	0.12	(0.07)	α_{i62}^Q	0.22	(0.10)
α_{i24}^Q	0.50	(0.20)	α_{i63}^Q	0.44	(0.23)
α_{i25}^Q	0.30	(0.19)	α_{i64}^Q	1.84	(0.48)
α_{i31}^Q	0.74	(0.22)	α_{i65}^Q	0.10	(0.04)
α_{i32}^Q	0.58	(0.32)			

Table C1. Posterior means and standard deviations diagonal elements of Σ_α (upper level model) for the deterministic fusion PHMM (112345). α_{i1}^M is the random intercept for the variable total searches and α_{i2}^M for the variable pageviews.

	Mean	St.Dev.		Mean	St.Dev.
α_{i1}^M	1.68	(0.15)	α_{i33}^Q	1.55	(0.41)
α_{i2}^M	0.47	(0.04)	α_{i34}^Q	0.14	(0.09)
α_{i1}^π	1.74	(0.52)	α_{i35}^Q	0.83	(0.30)
α_{i2}^π	0.32	(0.16)	α_{i41}^Q	1.96	(0.50)
α_{i3}^π	0.52	(0.20)	α_{i42}^Q	0.39	(0.22)
α_{i4}^π	0.40	(0.21)	α_{i43}^Q	1.29	(0.48)
α_{i5}^π	0.52	(0.21)	α_{i44}^Q	0.67	(0.23)
α_{i11}^Q	0.76	(0.25)	α_{i45}^Q	0.13	(0.09)
α_{i12}^Q	0.81	(0.39)	α_{i51}^Q	0.11	(0.05)
α_{i13}^Q	0.61	(0.24)	α_{i52}^Q	0.15	(0.07)
α_{i14}^Q	0.23	(0.12)	α_{i53}^Q	0.22	(0.18)
α_{i15}^Q	0.13	(0.08)	α_{i54}^Q	0.11	(0.06)
α_{i21}^Q	2.17	(0.86)	α_{i55}^Q	0.71	(0.33)
α_{i22}^Q	0.19	(0.09)	α_{i61}^Q	2.09	(0.66)
α_{i23}^Q	0.75	(0.26)	α_{i62}^Q	0.21	(0.10)
α_{i24}^Q	0.14	(0.10)	α_{i63}^Q	0.35	(0.17)
α_{i25}^Q	1.43	(0.50)	α_{i64}^Q	1.58	(0.40)
α_{i31}^Q	0.55	(0.21)	α_{i65}^Q	0.90	(0.26)
α_{i32}^Q	0.26	(0.10)			

Table C2. Posterior means and standard deviations diagonal elements of Σ_α (upper level model) for the stochastic fusion PHMM (123455). α_{i1}^M is the random intercept for the variable total searches and α_{i2}^M for the variable pageviews.

Appendix D: Lasso and Random Forest (RF) ordered logit benchmark models

We used the R packages ‘glmnet’ (Archer and Williams, 2012) for the estimation of the Lasso ordered logit benchmark model and ‘ordinalForest’ (Horning, 2019) for the estimation of the RF order logit benchmark model. As reported in the main document, to calibrate the Lasso and RF models we regress the observed survey response in month 5 as an ordinal variable on the same (nine) variables that were used to calibrate the PHMMs in months 1,2, 3, and 4. To predict the job seeking status for the second survey we use the user activity in months 11,12, 13, and 14. Thus, the machine learning models include dynamics via the lagged observed activities as covariates.

For the Lasso approach, we used the backward selection methods with a maximum of 500 iterations. We selected the best model using the AIC criterion. The (non-zero) beta coefficients and ordered logit intercepts for the best Lasso models are given in Table D1.

Predicting first survey	Beta	Predicting second survey	Beta
Total searches LAG1	0.022	Total searches LAG1	0.027
Pageviews LAG1	0.044	Pageviews LAG1	0.087
Pageviews LAG3	0.006	Pageviews LAG2	-0.062
Profile updates (dum) LAG1	-0.351	Pageviews LAG3	0.025
Job searched (dum) LAG2	0.442	Profile updates (dum) LAG1	-0.586
Job searched (dum) LAG3	0.600	Profile updates (dum) LAG4	-0.021
Job searched (dum) LAG4	0.301	Job searched (dum) LAG1	0.056
Invitations received LAG2	-0.050	Job searched (dum) LAG2	0.498
Connections formed LAG2	-0.028	Job searched (dum) LAG3	0.630
Number of connections of invitee LAG1	0.040	Job searched (dum) LAG4	0.375
More invitations outside company (dum) LAG3	-0.158		
cp1	2.966	cp1	3.005
cp2	0.536	cp2	0.545
cp3	0.104	cp3	0.105
cp4	-0.209	cp4	-0.214

Table D1. Non-zero coefficients and ordered logit thresholds for the LASSO ordered logit model based on the AIC criterion for predicting the first and second survey responses.

The Lasso ordered logit results, although at large a prediction model, nevertheless provides some insights into what behaviors correlate with job search. In particular, pageviews and whether the user used the job search tool in the months prior to the survey are important predictors of the responses to the job seeking status survey. Interestingly, users that update their profile are less likely to be job seekers in the next month, probably because job seekers ‘window dressed’ their profile page a few months in advance of actively seeking a job.

For the RF ordered logit model, we mostly use the settings as recommended by Hornung (2019), including 5,000 trees. We used the proportional performance function, which is preferred if the main goal is to classify correctly as many observations as possible. We used the 100 best score sets with the highest out-of-bag prediction performance. It is possible to use only the single best score set, however, this outcome may be purely due to chance. We found for our empirical application that using the 100 best score sets led to fairly stable prediction outcomes when running the procedure several times. A lower number of score sets led to higher variability in the prediction outcomes whereas a higher number of score sets led to lower predictive performance. Table D2 lists for the RF approach the 10 most important variables for predicting next month survey responses to the job seeking question (in alphabetical order).

Predicting first survey	Predicting second survey
Connections formed LAG1	Job searched (dum) LAG1
Connections formed LAG4	Job searched (dum) LAG2
Invitations received LAG4	Job searched (dum) LAG3
Job searched (dum) LAG3	Job searched (dum) LAG4
Number of connections of invitee LAG2	Pageviews LAG1
Pageviews LAG1	Pageviews LAG2
Pageviews LAG3	Pageviews LAG4
Pageviews LAG4	Total searches LAG1
Total searches LAG1	Total searches LAG2
Total searches LAG4	Total searches LAG4

Table D2. The 10 most important predictors (in alphabetical order) for predicting the responses to the first and second survey from the RF ordinal logit approach.

Comparing the results for the RF to the Lasso approach, we see a lot of similarities. In particular, pageviews and total searches help predicting the job seeking status for users in the next month, as well as past usage of the job search tool. Similar to the Lasso approach, the RF also does not list profile updates activity as an important predictor for next month job seeking status.

Appendix E: Robustness analyses predictions through 5-fold cross validation

In this Web Appendix we perform the similar predictions to one reported in Table 6 of the main document but using a 5-fold cross validation. We create a 5-fold split of the data and fit all models 5 times. Table E1 reports the average values for the prediction metrics across the 5 sets of results. The cross-validation analysis supports the results reported in Table 6 in the main document. Hence, the prediction findings in Section 5.3 of the main document using a random split of the 491 users into a calibration sample ($N_c = 400$) and a holdout sample ($N_v = 91$) is robust to alternative splits.

		Month 5 – Survey 1				Month 14 – Survey 2				
		JI	F1	PRC	AUC	JI	F1	PRC	AUC	
Cross section	Calibration Sample 4-fold	Det. PHMM 12345	N/A				0.14	0.25	0.24	0.56
		Det. PHMM 112345					0.14	0.25	0.24	0.55
		Stoch. PHMM 12345					0.13	0.23	0.23	0.54
		Stoch. PHMM 123455					0.13	0.23	0.23	0.53
		RF					0.12	0.21	0.23	0.55
		Lasso					0.08	0.15	0.23	0.50
	Holdout sample 1-fold	Det. PHMM 12345	0.16	0.28	0.26	0.59	0.15	0.26	0.25	0.56
		Det. PHMM 112345	0.13	0.23	0.24	0.56	0.16	0.27	0.26	0.56
		Stoch. PHMM 12345	0.16	0.27	0.26	0.57	0.14	0.24	0.24	0.54
		Stoch. PHMM 123455	0.13	0.23	0.25	0.55	0.13	0.23	0.23	0.53
		RF	0.10	0.18	0.24	0.55	0.10	0.18	0.23	0.55
		Lasso	0.06	0.12	0.24	0.47	0.08	0.15	0.26	0.51

Table E1. Five-fold cross validation results holdout predictions for the four versions of the PHMM, and the RF and Lasso ordered logit benchmarks. Performance metrics (JI, F1, PRC, AUC) indicate model performance to predict the users' job seeking status in month 5 and month 14. Higher numbers indicate better performance.

Appendix F: PHMM estimation results for the second dataset applying the model to d-mail targeting (Section 6 in the main document)

We obtained a second sample of 1,621 users from the platform. For this sample we observe the users' activity for a different sample of users than the ones used in Section 5, during the time period June 2011—May 2012. These users responded to the same job search survey as discussed in Section 3 of the main document. This survey was fielded in October 2011 (5th month of the observation window). For these users we also observe whether they received a d-mail in each month and, if they received a d-mail, whether they responded to it. The activity variables we observe for this sample are similar, but not identical, to the activity variables observed in the sample used for the main analyses. We observe the following monthly activities: whether or not the user viewed a job ad (0/1 variable), whether or not the user updated his/her educational information on the profile page (0/1 variable), whether or not the user updated his/her position information on the profile page (0/1 variable), how many invitations to connect the user received, how many invitations to connect the user sent, how many page views the user made, and how many times the user's profile page was viewed. All variables are observed for the full time period.

The posterior results for the 6-state deterministic fusion PHMM (112345) are given in table F1. As in Section 5 in the main text, we report the transformed working parameters (α_i and θ). The trend parameters are reported at the working parameter level. Similar to the results in Section 5 of the main document, we masked the absolute monthly activity levels by multiplying them with the same random number, which was a single draw from a uniform distribution on the interval [0.5, 1.5].

State	1		2		3		4		5		6			
Survey Response	1		1		2		3		4		5		Trend	
Job views (dum)	0.00	(0.00)	0.02	(0.00)	0.07	(0.01)	0.27	(0.01)	0.76	(0.01)	0.21	(0.01)	0.01	(0.00)
Education modified (dum)	0.00	(0.00)	0.06	(0.01)	0.03	(0.00)	0.05	(0.01)	0.20	(0.01)	0.23	(0.01)	-0.04	(0.00)
Positions modified (dum)	0.00	(0.00)	0.15	(0.01)	0.08	(0.01)	0.13	(0.01)	0.40	(0.01)	0.43	(0.01)	-0.03	(0.00)
Invitations received	1.73	(0.06)	3.41	(0.07)	2.17	(0.04)	2.76	(0.06)	3.33	(0.10)	4.00	(0.14)	0.01	(0.00)
Invitations sent	1.73	(0.30)	7.91	(0.38)	3.08	(0.13)	2.50	(0.09)	8.71	(0.43)	39.81	(2.71)	-0.01	(0.00)
Page views	6.68	(0.72)	496.70	(11.61)	133.13	(3.09)	222.23	(3.74)	598.70	(14.83)	2,285.73	(88.94)	0.01	(0.00)
Profile views received	1.63	(0.27)	23.30	(0.70)	4.60	(0.13)	9.46	(0.22)	29.55	(0.91)	68.72	(3.27)	0.00	(0.00)
Initial state distribution	0.28	(0.01)	0.17	(0.01)	0.22	(0.01)	0.15	(0.01)	0.09	(0.01)	0.09	(0.01)		
Transition matrix	0.45	(0.02)	0.11	(0.02)	0.24	(0.02)	0.04	(0.01)	0.03	(0.01)	0.11	(0.01)		
	0.08	(0.01)	0.40	(0.02)	0.24	(0.01)	0.14	(0.01)	0.04	(0.01)	0.11	(0.01)		
	0.14	(0.01)	0.16	(0.01)	0.34	(0.01)	0.17	(0.01)	0.09	(0.01)	0.10	(0.01)		
	0.17	(0.02)	0.15	(0.02)	0.28	(0.02)	0.24	(0.01)	0.09	(0.01)	0.07	(0.01)		
	0.01	(0.01)	0.18	(0.03)	0.34	(0.03)	0.14	(0.02)	0.26	(0.01)	0.07	(0.01)		
	0.07	(0.01)	0.42	(0.02)	0.15	(0.02)	0.01	(0.00)	0.13	(0.01)	0.23	(0.01)		

Table F1. Posterior means and standard deviations (in parentheses) for the proposed 6-state deterministic fusion PHMM.

When we compare the posterior estimates of the PHMM for this dataset with the posterior estimates of the same model in the main document (Table 3A), we can see that the findings are fairly similar. The active job seekers (State 6) exhibit the highest activity. As in the main document, we find two types of non-seekers, a type that rarely uses the site, and a type that uses the site quite actively almost to the level of (casually looking) job seekers (State 5), except that they do not view jobs. Interestingly, the active job seekers (State 6) are less likely to view jobs than users that are casually looking. As before, we find that job seekers send more invitations to connect than what they receive. In addition, the posterior results for the initial state distribution and the transition probability matrix are similar to the posterior results reported in the main document for the first dataset (Table 3A). Thus, the estimation of the PHMM for this new dataset at a different time period and with somewhat different set of activities suggests that the estimates of the PHMM are fairly robust and generalizable to a new sample of users and activities.

Appendix G: Web appendix references

Archer, K. J., & Williams A.A. A., (2012). L1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. *Statistics in Medicine*, 31, 1464-1474.

Celeux, G., (1998). Bayesian inference for mixture: The label switching problem. In *Compstat* (pp. 227-232). *Physica*, Heidelberg.

Geweke, J., (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), 169-193. Oxford: Oxford University Press.

Hastings, W. K., (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.

Hornung, R., (2019). ordinalForest: Ordinal forests: Prediction and variable ranking with ordinal target variables. R package version 2.3-1. <https://CRAN.R-project.org/package=ordinalForest>

Hornung, R., (2019). Ordinal forests. *Journal of Classification*. <https://doi.org/10.1007/s00357-018-9302-x>

Zellner, A., (1971). *An introduction to Bayesian inference in econometrics*. J. Wiley and Sons, Inc.