**Adjusting for Selection Bias in Longitudinal Analyses using Simultaneous Equations Modelling: The Relationship between Employment Transitions and Mental Health**

Fiona Steele, Robert French and Mel Bartley

**Online Supplemental Material**

# eAppendix A: Further Details of Models A-D in Table 1

Model A

Model A makes no adjustment for direct or indirect selection. The model for $H\_{ti}$, the mental health score of individual $i$ at year $t$ $(t>1)$, takes the form:

$H\_{ti}=β\_{0}+β\_{2}ΔE\_{t-1i}+β\_{3}x\_{t-1i}+u\_{i}^{(H)}+e\_{ti}$(e1)

where $β\_{0}$ is an intercept, $ΔE\_{t-1i}$ is a vector of dummy variables for the change in employment status between $t-1$ and $t$ with a row-vector of coefficients $β\_{2}$, $x\_{t-1i} $is a vector of time-varying covariates with coefficients $β\_{3}$, $u\_{i}^{(H)}$ is a normally distributed random effect representing unmeasured time-invariant individual characteristics, and $e\_{ti}$ is a normally distributed time-varying residual.

Compared to the mental health model given by eq. (1) in the text, eq. (e1) does not include the lagged response $H\_{t-1i}$ as a predictor. For this reason, the ‘initial conditions’ problem does not arise. However, for comparability with the other models considered, eq. (e1) is fitted for $t>1$ only and estimated jointly with a model for $t=1$:

$H\_{1i}=γ\_{0}+γ\_{1}x\_{1i}+λ^{(H)}u\_{i}^{(H)}+e\_{1i}$ (e2)

where $λ^{(H)}$ is a random effect loading. Note that eq. (e2) is identical to equation (3) in the text.

Model A is completely specified by eq. (e1) and (e2). No model is specified for employment transitions, which implies that $ΔE\_{t-1i}$ is assumed to be uncorrelated with the time-invariant influences on $H\_{ti}$ represented by $u\_{i}^{(H)}$.

Model B

Model B extends Model A by conditioning on mental health at $t-1$. This is the most common approach to adjusting for direct selection in previous research. Eq. (e1) for $t>1$ is therefore replaced by

$H\_{ti}=β\_{0}+β\_{1}H\_{t-1i}+β\_{2}ΔE\_{t-1i}+β\_{3}x\_{t-1i}+u\_{i}^{(H)}+e\_{ti}$(e3)

which is identical to eq. (1) in the text.

As in Model A, the model in eq. (e3) is estimated jointly with the model for $t=1$ given by eq. (e2).

Model C

Models C and D are simultaneous equation models in which Model B for mental health is estimated jointly with a model for employment transitions to adjust for indirect selection.

Model C consists of eq. (e3) and (e2) and a model for employment transitions between $t-1$ and $t$. We denote the employment status of individual $i$ at year $t$ by $E\_{ti}$ (coded as 1 = employed, 2 = economically inactive, and 3 = unemployed) with category probabilities $π\_{ti}^{(k)}=Pr⁡(E\_{ti}=k)$ for $k=1, 2, 3$. Lagged employment status is denoted by three indicator variables $E\_{l,t-1i}$ $\left(l=1, 2, 3\right)$, which are stacked into column vector $E\_{t-1i}$. The model for employment transitions is specified as a multinomial logit model which consists of two equations contrasting the probabilities of being economically inactive $(k=2)$ or unemployed $(k=3)$ with the probability of being employed $(k=1)$ at $t$. In Model C, the employment transitions model for $t>1$ takes the form

$log\left(\frac{π\_{ti}^{\left(k\right)}}{π\_{ti}^{\left(1\right)}}\right)=α\_{1}^{\left(k\right)}E\_{t-1i}+α\_{4}^{\left(k\right)}z\_{t-1i}+u\_{i}^{(Ek)}, k=2, 3$ (e4)

where $z\_{t-1i}$ is a vector of time-varying covariates, $α\_{1}^{\left(k\right)}$ and $α\_{4}^{\left(k\right)}$ are row-vectors of coefficients, and $(u\_{i}^{\left(E2\right)}, u\_{i}^{\left(E3\right)})$ are normally distributed individual-level random effects.

Eq. (e4) is estimated jointly with the initial conditions model

$log\left(\frac{π\_{1i}^{\left(k\right)}}{π\_{1i}^{\left(1\right)}}\right)=θ\_{0}^{\left(k\right)}+θ\_{1}^{\left(k\right)}z\_{1i}+λ^{(Ek)}u\_{i}^{(Ek)}, k=2, 3$ (e5)

which is identical to eq. (4) in the text.

Eq. (e4) has the same form as eq. (2) in the text, but without the interaction between $E\_{t-1i}$ and mental health at $t-1$ (fitted as a quadratic with $H\_{t-1i}$ and $H\_{t-1i}^{2}$). Thus Model C does not allow for a direct effect of mental health at $t-1$ on employment transitions between $t-1$ and $t$, a potential source of direct selection. However, Model C does allow for *indirect* selection due to shared unmeasured time-invariant influences on mental health and employment transitions. This is achieved by allowing for non-zero correlation between the individual random effects. Specifically $(u\_{i}^{\left(H\right)}, u\_{i}^{\left(E2\right)}, u\_{i}^{\left(E3\right)}) $ are assumed to follow a multivariate normal distribution with all pairwise correlations freely estimated, and where $corr(u\_{i}^{\left(H\right)},u\_{i}^{\left(E2\right)})$ and $corr(u\_{i}^{\left(H\right)},u\_{i}^{\left(E3\right)})$ allow for indirect selection.

Model D

Model D is the full simultaneous equation model represented by eq. (1)-(4) in the text. This model extends Model C by including mental health at $t-1 $as a predictor of employment transitions between $t-1$ and $t$. Specifying the effect of mental health at $t-1 $as a quadratic function, eq. (e4) is replaced by

$log\left(\frac{π\_{ti}^{\left(k\right)}}{π\_{ti}^{\left(1\right)}}\right)=α\_{1}^{\left(k\right)}E\_{t-1i}+α\_{2}^{\left(k\right)}E\_{t-1i}H\_{t-1i}+α\_{3}^{\left(k\right)}E\_{t-1i}H\_{t-1i}^{2}+α\_{4}^{\left(k\right)}z\_{t-1i}+u\_{i}^{(Ek)}$ (e6)

for $ k=2, 3$ , which is identical to eq. (3) in the text. Eq. (e6) is estimated jointly with eq. (e3), (e2) and (e5) of Model C, again allowing for non-zero correlation between the individual random effects.

For $E\_{t-1i}=\left(E\_{1,t-1i}, E\_{2,t-1i}, E\_{3,t-1i}\right)^{T}$ and letting$ α\_{l}^{\left(k\right)}=\left(α\_{1l}^{\left(k\right)},α\_{2l}^{\left(k\right)},α\_{3l}^{\left(k\right)}\right)^{T}$ for $l=1, 2, 3$, the part of eq. (e6) involving employment status and health at $t-1$ can be written:

$$α\_{11}^{\left(k\right)}E\_{1,t-1i}+α\_{21}^{\left(k\right)}E\_{1,t-1i}H\_{t-1i}+α\_{31}^{\left(k\right)}E\_{1,t-1i}H\_{t-1i}^{2}+$$

$$α\_{12}^{\left(k\right)}E\_{2,t-1i}+α\_{22}^{\left(k\right)}E\_{2,t-1i}H\_{t-1i}+α\_{32}^{\left(k\right)}E\_{2,t-1i}H\_{t-1i}^{2}+$$

$$α\_{13}^{\left(k\right)}E\_{3,t-1i}+α\_{23}^{\left(k\right)}E\_{3,t-1i}H\_{t-1i}+α\_{33}^{\left(k\right)}E\_{3,t-1i}H\_{t-1i}^{2}$$

where the coefficients $\left(α\_{m1}^{\left(k\right)},α\_{m2}^{\left(k\right)},α\_{m3}^{\left(k\right)}\right)$ together define the quadratic effect of prior health on the log-odds of being economically inactive $(k=2)$ or unemployed $(k=3)$ rather than employed at $t$ for men in employment category $m$ at $t-1$.

# eAppendix B: Estimation of Selected Models using the aML software

|  |
| --- |
| The aML software and manual may be downloaded from <http://www.applied-ml.com> The raw2aml program supplied with aML converts an ascii data file into aML format. The first variable in the input file must be the individual identifier, referred to in the model statements as **\_id**. Other variables must be order of the level at which they are defined, starting with time-varying variables (defined at the wave level) and then individual-level variables. Note that in aML, levels are labelled from top to bottom. Here level 1 is the individual and level 2 is wave. The following syntax files (with extension .aml) read in the aML data file bhps and specify the model to be estimated (Model B or D). Note that starting values must be provided for each parameter, so it is important to build models gradually using estimates from simpler models as starting values for extended models. The update program can be used to update the starting values in a syntax file using estimates from a specified output file.  |

**Model B**

**/\*Request standard errors based on numerical evaluation of the Hessian matrix upon convergence \*/**

option numerical standard errors;

**/\*Read in the input data file, converted to aML format using the raw2aml program \*/**

dsn = bhps.dat;

**/\*Define loading for individual random effect in initial condition equation (3) \*/**

define parameter lambdaH;

**/\*Specify variables to be included in the GHQ model for t>1, eq (1). ‘1’ is a constant (with coefficient the intercept). The variable l\_ghq is lagged GHQ, and et\_i\_e to et\_u\_i are dummies for employment transitions. i denotes economic inactivity (out of the labour market), u denotes unemployment, and e employment. So et\_e\_u indicates a transition from employment to unemployment. X1 and x2 are covariates. \*/**

define regressor set BetaX;

var = 1 l\_ghq et\_i\_e et\_u\_e et\_e\_i et\_i\_i et\_e\_u et\_i\_u et\_u\_u et\_u\_i x1 x2;

**/\*Specify variables to be included in the GHQ model for t=1, eq (3), with covariates x11 and x12. \*/**

define regressor set GammaX;

var = 1 x11 x12;

**/\*Individual random effects (uh) and time-varying residuals (e) are assumed to follow normal distributions. The residuals are referred to by these names in the model statements that follow. \*/**

define normal distribution; dim=1; name=uH;

define normal distribution; dim=1; name=e;

**/\*Specify model for GHQ for t>1. Residual terms are specified at the individual level (identified by \_id, the name given by aML to the first variable in the input file) and wave level (\_iid). \*/**

continuous model;

 keep if t>1;

 outcome = ghq;

 model = regset BetaX + res(draw=\_id,ref=uH) + res(draw=\_iid,ref=e);

**/\*Specify model for GHQ for t=1 \*/**

continuous model;

 keep if t==1;

 outcome = ghq;

 model = regset GammaX +

 par lambdaH \* res(draw=\_id,ref=uH) + res(draw=\_iid,ref=e);

**/\*Give starting values for all parameters. The ‘T’ indicates that the parameter is to be freely estimated. (An ‘F’ can be used to constrain a parameter.) The ‘\*’ needs to be replaced by a numeric value. \*/**

starting values;

lambdaH T \*

**/\*Model for GHQ for t>1 \*/**

constant T \*

l\_ghq T \*

et\_i\_e T \*

et\_u\_e T \*

et\_e\_i T \*

et\_i\_i T \*

et\_e\_u T \*

et\_i\_u T \*

et\_u\_u T \*

et\_u\_i T \*

x1 T \*

x2 T \*

**/\*Model for GHQ for t=1 \*/**

constant T \*

x11 T \*

x12 T \*

/\***Standard deviations of the residuals uH and e** \*/

siguH T \*

sige T \*

;

**Model D**

|  |
| --- |
| Additions to Model B syntax are highlighted in blue. |

**/\*Request standard errors based on numerical evaluation of the Hessian matrix upon convergence \*/**

option numerical standard errors;

**/\*Read in the input data file, converted to aML format using the raw2aml program \*/**

dsn = bhps.dat ;

**/\*Define loading for individual random effect in initial condition equation (3) for health\*/**

define parameter lambdaH;

**/\*Define loadings for individual random effects in initial condition equations (4) for employment \*/**

define parameter lambdaE2;

define parameter lambdaE3;

**/\*Specify variables to be included in the GHQ model for t>1, eq (1) \*/**

define regressor set BetaX;

 var = 1 l\_ghq et\_i\_e et\_u\_e et\_e\_i et\_i\_i et\_e\_u et\_i\_u et\_u\_u et\_u\_i x1 x2;

**/\*Specify variables to be included in the GHQ model for t=1, eq (3) \*/**

define regressor set GammaX; var = 1 x11 x12;

**/\*Specify variables to be included in the employment model for t>1, eq (2). l\_e1, l\_e1ghq and l\_e1ghq2 are the dummy for lagged employment status category 1, and its interaction with GHQ and GHQ2. Similar variables are included for lagged employment status categories 2 and 3. Note that a separate set of variables may be specified for each contrast with category 1.**

**\*/**

define regressor set Alpha2Z;

 var = l\_e1 l\_e1ghq l\_e1ghq2 l\_e2 l\_e2ghq l\_e2ghq2

 l\_e3 l\_e3ghq l\_e3ghq2

 z1 z2;

define regressor set Alpha3Z;

 var = l\_e1 l\_e1ghq l\_e1ghq2 l\_e2 l\_e2ghq l\_e2ghq2

 l\_e3 l\_e3ghq l\_e3ghq2

 z1 z2;

**/\*Specify variables to be included in the employment model for t=1, eq (4).**

**\*/**

define regressor set Theta2Z;

 var = 1 z11 z12;

define regressor set Theta3Z;

 var = 1 z11 z12;

**/\*Individual random effects for health (uH) and employment (uE2 and uE3) are assumed to follow a trivariate normal distribution. Random effects in the multinomial logit model for employment must be ‘integrated out’. The residuals are referred to by these names in the model statements that follow. \*/**

define normal distribution; dim=3;

 name=uh;

 name=uE2;

 name=uE3;

 number of integration points = 12;

**/\*Normally distributed time-varying residual for health (e) \*/**

define normal distribution; dim=1;

 name=e;

**/\*Specify model for GHQ for t>1 \*/**

continuous model;

 keep if t>1;

 outcome = ghq;

 model = regset BetaX + res(draw=\_id,ref=uH) + res(draw=\_iid,ref=e);

**/\*Specify model for GHQ for t=1 \*/**

continuous model;

 keep if t==1;

 outcome = ghq;

 model = regset GammaX +

 par lambdaH \* res(draw=\_id,ref=uH) + res(draw=\_iid,ref=e);

**/\*Specify multinomial logit model for employment status for t>1 \*/**

multinomial logit model;

 keep if t>1;

 outcome = empstat;

 model 2 = regset Alpha2Z + intres(draw=\_id,ref=uE2);

 model 3 = regset Alpha3Z + intres(draw=\_id,ref=uE3);

**/\*Specify multinomial logit model for employment status for t=1 \*/**

multinomial logit model;

 keep if t=1;

 outcome = empstat;

 model 2 = regset Theta2Z + par lambdaE2 \* intres(draw=\_id,ref=uE2);

 model 3 = regset Theta3Z + par lambdaE3 \* intres(draw=\_id,ref=uE3);

**/\*Give starting values for all parameters. The ‘\*’ needs to be replaced by a numeric value. siguH and sige are labels given to the standard deviations of the residuals uH and e \*/**

starting values;

lambdaH T \*

lambdaE2 T \*

lambdaE3 T \*

**/\*Model for GHQ for t>1 \*/**

constant T \*

l\_ghq T \*

et\_i\_e T \*

et\_u\_e T \*

et\_e\_i T \*

et\_i\_i T \*

et\_e\_u T \*

et\_i\_u T \*

et\_u\_u T \*

et\_u\_i T \*

x1 T \*

x2 T \*

**/\*Model for GHQ for t=1 \*/**

constant T \*

x11 T \*

x12 T \*

**/\*Model for employment status category 2 vs 1 for t>1 \*/**

l\_e1 T \*

l\_e1ghq T \*

l\_e1ghq2 T \*

l\_e2 T \*

l\_e2ghq T \*

l\_e2ghq2 T \*

l\_e3 T \*

l\_e3ghq T \*

l\_e3ghq2 T \*

z1 T \*

z2 T \*

**/\*Model for employment status category 3 vs 1 for t>1 \*/**

l\_e1 T \*

l\_e1ghq T \*

l\_e1ghq2 T \*

l\_e2 T \*

l\_e2ghq T \*

l\_e2ghq2 T \*

l\_e3 T \*

l\_e3ghq T \*

l\_e3ghq2 T \*

z1 T \*

z2 T \*

**/\*Model for employment status category 2 vs 1 for t=1 \*/**

Constant T \*

z11 T \*

z12 T \*

**/\*Model for employment status category 3 vs 1 for t=1 \*/**

Constant T \*

z11 T \*

z12 T \*

/\***Standard deviations of the individual random effects uH, uE2 and uE3** \*/

siguH T \*

siguE2 T \*

siguE3 T \*

**/\*Correlations between individual random effects \*/**

rhoHE2 T \*

rhoHE3 T \*

rhoE2E3 T \*

/\***Standard deviations of the time-varying residual for GHQ** \*/

sige T \*

;

# eAppendix C: Additional results

*eTable 1*. Descriptive statistics for outcome variables and covariates

|  |  |  |
| --- | --- | --- |
| Variable |  Mean |  SD |
| *Mental health and employment status* |  |  |
| GHQ | 10.489 | 5.075 |
| Employment status: |  |  |
|  Employed | 0.817 |  |
|  Out of LM | 0.121 |  |
|  Unemployed | 0.062 |  |
| Employment transitions between $t-1 $ and $t$: |  |  |
|  Employed in both  | 0.787 |  |
|  Employed → economically inactive | 0.020 |  |
|  Employed → unemployed  | 0.019 |  |
|  Economically inactive → employed  | 0.013 |  |
|  Economically inactive in both  | 0.094 |  |
|  Economically inactive → unemployed  | 0.007 |  |
|  Unemployed → employed  | 0.022 |  |
|  Unemployed → economically inactive  | 0.011 |  |
|  Unemployed in both  | 0.028 |  |
|  |  |  |
| *Covariates* |  |  |
| Age | 40.145 | 12.773 |
| Has a partner | 0.720 |  |
| Dependent co-resident children: |  |  |
|  No children | 0.684 |  |
|  Youngest child aged 0 to 4 | 0.149 |  |
|  Youngest child aged 5 to 15 | 0.167 |  |
| Non-manual household | 0.787 |  |
| Local unemployment rate | 0.025 | 0.015 |
| Note: In the models age and GHQ are mean centred, local unemployment rate is standardised. |

*eTable 2*. Point estimates and 95% confidence intervals from random effects model of GHQ score at $t$, before and after adjusting for indirect selection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable |  | Model B |  | Model D |
| **Model for** $t>1$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Employment transitions between $t-1 $ and $t$: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  Employed in both (reference) |  | 0 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |
|  |  Employed → economically inactive |  | 0.921 |  | ( | 0.681 | to  | 1.161 | ) |  | 0.591 |  | ( | 0.347 | to  | 0.835 | ) |
|  |  Employed → unemployed  |  | 2.452 |  | ( | 2.209 | to  | 2.694 | ) |  | 2.218 |  | ( | 1.971 | to  | 2.464 | ) |
|  |  Economically inactive → employed  |  | -0.825 |  | ( | -1.121 | to  | -0.529 | ) |  | -1.085 |  | ( | -1.383 | to  | -0.788 | ) |
|  |  Economically inactive in both  |  | 1.033 |  | ( | 0.878 | to  | 1.188 | ) |  | 0.408 |  | ( | 0.233 | to  | 0.583 | ) |
|  |  Economically inactive → unemployed  |  | 1.054 |  | ( | 0.650 | to  | 1.458 | ) |  | 0.455 |  | ( | 0.044 | to  | 0.866 | ) |
|  |  Unemployed → employed  |  | -1.166 |  | ( | -1.394 | to  | -0.937 | ) |  | -1.361 |  | ( | -1.592 | to  | -1.129 | ) |
|  |  Unemployed → economically inactive  |  | 1.467 |  | ( | 1.134 | to  | 1.799 | ) |  | 0.870 |  | ( | 0.529 | to  | 1.210 | ) |
|  |  Unemployed in both  |  | 0.888 |  | ( | 0.657 | to  | 1.120 | ) |  | 0.401 |  | ( | 0.154 | to  | 0.648 | ) |
|  | GHQ at $t-1$ |  | 0.224 |  | ( | 0.214 | to  | 0.233 | ) |  | 0.225 |  | ( | 0.215 | to  | 0.234 | ) |
| Age at $t$ |  | 0.010 |  | ( | 0.005 | to  | 0.015 | ) |  | 0.014 |  | ( | 0.009 | to  | 0.018 | ) |
| Has a partner at $t-1 $ |  | 0.420 |  | ( | 0.303 | to  | 0.537 | ) |  | 0.449 |  | ( | 0.331 | to  | 0.566 | ) |
| Dependent children at $t-1$: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  No children (reference) |  | 0 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |
|  Youngest child aged 0 to 4  |  | 0.199 |  | ( | 0.080 | to  | 0.319 | ) |  | 0.202 |  | ( | 0.083 | to  | 0.321 | ) |
|  Youngest child aged 5 to 15  |  | 0.340 |  | ( | 0.235 | to  | 0.445 | ) |  | 0.340 |  | ( | 0.235 | to  | 0.445 | ) |
| Constant |   | -0.905 |  | ( | -1.011 | to  | -0.799 | ) |  | -0.783 |  | ( | -0.890 | to  | -0.676 | ) |
| **Model for** $t=1$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Age at $t$ |  | 0.049 |  | ( | 0.041 | to  | 0.058 | ) |  | 0.045 |  | ( | 0.037 | to  | 0.053 | ) |
| Has a partner at $t$ |  | -0.377 |  | ( | -0.617 | to  | -0.137 | ) |  | -0.192 |  | ( | -0.433 | to  | 0.049 | ) |
| Dependant children in household at $t$: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  No children (reference) |  | 0 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |
|  Youngest child aged 0 to 4  |  | 0.314 |  | ( | 0.029 | to  | 0.599 | ) |  | 0.387 |  | ( | 0.103 | to  | 0.672 | ) |
|  Youngest child aged 5 to 15  |  | 0.636 |  | ( | 0.339 | to  | 0.932 | ) |  | 0.637 |  | ( | 0.341 | to  | 0.933 | ) |
| Constant |   | -0.467 |  | ( | -0.656 | to  | -0.277 | ) |  | -0.615 |  | ( | -0.806 | to  | -0.425 | ) |
|  | Random effect loading, $λ^{(H)}$  |  | 1.144 |  | ( | 1.095 | to  | 1.192 | ) |  | 1.140 |  | ( | 1.092 | to  | 1.188 | ) |

Note: Model B adjusts for direct selection through inclusion of GHQ at $t-1$. Model D is the simultaneous equations model for GHQ and employment transitions, with residual correlations to adjust for indirect selection and a direct effect of GHQ on employment transitions.

*eTable 3*. Point estimates and 95% confidence intervals from random effects multinomial logit model of employment status at $t$ (estimated jointly with Model D for GHQ at $t$)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable |  | Economically inactive vs Employed ($k=2$) |  | Unemployed vs Employed ($k=3$) |
| **Model for** $t>1$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Employment status at $t-1$: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  Employed |  | -3.224 |  | ( | -3.367 | to | -3.080 | ) |  | -3.334 |  | ( | -3.503 | to | -3.164 | ) |
|  Economically inactive  |  | 1.213 |  | ( | 0.998 | to  | 1.427 | ) |  | -1.155 |  | ( | -1.386 | to  | -0.923 | ) |
|  Unemployed  |  | -0.864 |  | ( | -1.562 | to  | 1.427 | ) |  | -1.329 |  | ( | -1.562 | to  | -1.097 | ) |
|  GHQ × Employed  |  | -0.018 |  | ( | -0.034 | to  | -0.001 | ) |  | -0.001 |  | ( | -0.017 | to  | 0.015 | ) |
|  GHQ2 × Employed  |  | 0.004 |  | ( | 0.003 | to  | 0.005 | ) |  | 0.002 |  | ( | 0.000 | to  | 0.003 | ) |
|  GHQ × Economically inactive  |  | 0.018 |  | ( | -0.003 | to  | 0.039 | ) |  | 0.019 |  | ( | -0.012 | to  | 0.049 | ) |
|  GHQ2 × Economically inactive  |  | -0.00003 |  | ( | -0.002 | to  | 0.002 | ) |  | -0.0001 |  | ( | -0.003 | to  | 0.002 | ) |
|  GHQ × Unemployed  |  | -0.037 |  | ( | -0.063 | to  | -0.011 | ) |  | -0.049 |  | ( | -0.070 | to  | -0.027 | ) |
|  GHQ2 × Unemployed  |  | 0.002 |  | ( | 0.000 | to  | 0.004 | ) |  | 0.001 |  | ( | -0.001 | to  | 0.003 | ) |
| Age at $t$ |  | 0.073 |  | ( | 0.067 | to  | 0.079 | ) |  | 0.008 |  | ( | 0.002 | to  | 0.014 | ) |
| Has a partner at $t-1$ |  | -0.623 |  | ( | -0.764 | to  | -0.482 | ) |  | -0.698 |  | ( | -0.846 | to  | -0.551 | ) |
| Dependent children at $t-1$: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  No children (reference) |  | 0 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |
|  Youngest child aged 0 to 4  |  | -0.278 |  | ( | -0.456 | to  | -0.101 | ) |  | 0.219 |  | ( | 0.052 | to  | 0.386 | ) |
|  Youngest child aged 5 to 15  |  | -0.328 |  | ( | -0.467 | to  | -0.188 | ) |  | -0.011 |  | ( | -0.155 | to  | 0.134 | ) |
| Non-manual household at $t-1$ |  | -0.539 |  | ( | -0.659 | to  | -0.420 | ) |  | -0.736 |  | ( | -0.859 | to  | -0.614 | ) |
| Local unemployment rate at $t-1$ |  | 0.149 |  | ( | 0.101 | to  | 0.198 | ) |  | 0.375 |  | ( | 0.323 | to  | 0.426 | ) |
| **Model for** $t=1$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Local unemployment rate at $t-1$ |  | 0.264 |  | ( | 0.164 | to  | 0.365 | ) |  | 0.369 |  | ( | 0.282 | to  | 0.456 | ) |
| Age at $t$ |  | 0.061 |  | ( | 0.054 | to  | 0.068 | ) |  | -0.032 |  | ( | -0.039 | to  | -0.024 | ) |
| Constant |   | -2.910 |  | ( | -3.113 | to  | -2.708 | ) |  | -3.055 |  | ( | -3.246 | to  | -2.863 | ) |
| Random effect loading, $λ^{(Ek)}$  |  | 1.205 |  | ( | 1.042 | to  | 1.368 | ) |  | 0.991 |  | ( | 0.877 | to  | 1.104 | ) |

**Predicted probabilities of employment transitions between** $t-1$ **and** $t$ **by GHQ at** $t-1$

The following graphs show predicted probabilities of being employed, economically inactive or unemployed at $t$ among men who were employed (eFigure 1) or unemployed (eFigure 2) at $t-1$, according to their GHQ score at $t-1$. Predicted transition probabilities from economically inactive are not shown because there was little evidence to suggest that these were influenced by GHQ at $t-1$. The predictions are based on the parameter estimates from the random effects multinomial logit model of employment status at $t$ ($t>1$), which is estimated jointly with Model D for GHQ at $t$ (shown in eTable 3). The values of all predictors other than employment status and GHQ at $t-1$ are held constant at their sample means and the individual random effects were fixed at zero. The figures also show 95% confidence intervals for the predicted probabilities, which were computed using the delta method.[[1]](#footnote-1)

|  |  |
| --- | --- |
| ModelD_E_E.wmf(a) Pr(E at $t$ |E at $t-1$) | ModelD_E_EI.wmf(b) Pr(EI at $t$ |E at $t-1$) |
| ModelD_E_UE.wmf(c) Pr(UE at $t$ |E at $t-1$) |

*eFigure 1*. Predicted probabilities of (a) employment, (b) economic inactivity, and (c) unemployment at $t$ by GHQ at $t-1$, given employed at $t-1$

|  |  |
| --- | --- |
| ModelD_UE_E.wmf(a) Pr(E at $t$ |UE at $t-1$) | ModelD_UE_EI.wmf(b) Pr(EI at $t$ |UE at $t-1$) |
| ModelD_UE_UE.wmfPr(UE at $t$ |UE at $t-1$) |

*eFigure 2*. Predicted probabilities of (a) employment, (b) economic inactivity, and (c) unemployment at $t$ by GHQ at $t-1$, given unemployed at $t-1$

1. For full details of the computation of standard errors for predicted probabilities from a multinomial logit model using the delta method, see: Fox J, Andersen R. Effect displays for multinomial and proportional-odds logit models. *Sociological Methodology* 2006;**36**:225-255. [↑](#footnote-ref-1)