

Supplementary Information for the paper:
“Parsimonious parametrizations of transition
matrices of Markov chain and hidden Markov
models”

Francesco Bartolucci^{1*}†, Silvia Pandolfi^{1†} and Fulvia Pennoni^{2†}

¹Department of Economics, University of Perugia, Via A. Pascoli, 20,
Perugia, 06123, Italy.

²Department of Statistics and Quantitative Methods, University of
Milano-Bicocca, Via Bicocca degli Arcimboldi 8, Milan, 20126, Italy.

*Corresponding author(s). E-mail(s): francesco.bartolucci@unipg.it;
Contributing authors: silvia.pandolfi@unipg.it; fulvia.pennoni@unimib.it;

†These authors contributed equally to this work.

In the following, we illustrate some other results for both the Markov Chain (MC) and Hidden Markov (HM) models for the applicative examples presented in [Sections 6.1 and 6.2](#) of the paper.

1 Application

We show some descriptive statistics of the Health and Retirement Study (HRS)¹ longitudinal data described in the paper. Table [1](#) presents the marginal distribution of the [Self-Reported Health Status](#) (named SRHS) over the eight interviews, and Table [2](#) reports the empirical transition matrix. Table [3](#) shows some descriptive statistics of the time-fixed and time-varying individual covariates.

¹More details on the study and questionnaire can be found at the website: <http://hrsonline.isr.umich.edu/>

Table 1 Empirical distribution of the response variable over the occasions of interview in terms of percentage frequencies.

SRHS category	Occasion of interview							
	1	2	3	4	5	6	7	8
poor	0.048	0.047	0.048	0.061	0.055	0.059	0.072	0.081
fair	0.115	0.130	0.131	0.170	0.156	0.169	0.192	0.202
good	0.272	0.289	0.281	0.320	0.306	0.320	0.322	0.320
very good	0.308	0.326	0.342	0.311	0.336	0.322	0.300	0.297
excellent	0.257	0.208	0.198	0.137	0.147	0.130	0.114	0.100

Table 2 Conditional empirical distributions of the response variable at time t given the response at time $t - 1$, with $t = 2, \dots, T$.

SRHS at $t - 1$	SRHS at t				
	poor	fair	good	very good	excellent
poor	0.512	0.326	0.095	0.042	0.026
fair	0.131	0.499	0.272	0.076	0.023
good	0.029	0.165	0.527	0.235	0.044
very good	0.011	0.050	0.257	0.550	0.132
excellent	0.010	0.025	0.109	0.333	0.523

1.1 Markov chain analysis

In the following we report additional results of the MC model illustrated in [Section 6.1](#) of the paper. Table 4 shows the estimates of the parameters affecting the initial probabilities. At the first interview, black people tend to belong to the 1st, 2nd, or 3rd state, grouping individuals with worst perceived health conditions: the odds ratio to belong to the 4th state for black people versus white is equal to $\exp(-0.557) = 0.573$, thus showing that blacks have the poorest self-rated health. Tables 5 and 6 show those affecting the transition probabilities of the selected model DIST1-COV. We refer to Section 3 of the paper for details on the model.

Table 3 Distribution of the time-fixed and time-varying covariates.

Variable	Category	Proportion
<i>gender:</i>	male	0.419
	female	0.581
<i>race:</i>	white	0.829
	black	0.137
	others	0.034
<i>education:</i>	high school diploma	0.226
	high school graduate	0.051
	some college	0.332
	college and above	0.197
		0.194
	Mean	St.Dev.
<i>age in 1992:</i>	54.803	5.459
<i>age in 1994:</i>	58.669	5.453
<i>age in 1996:</i>	56.664	5.448
<i>age in 1998:</i>	60.527	5.450
<i>age in 2000:</i>	62.481	5.450
<i>age in 2002:</i>	64.627	5.456
<i>age in 2004:</i>	66.577	5.457
<i>age in 2006:</i>	68.562	5.440

Table 4 Estimated regression coefficients for the initial probabilities ($\hat{\alpha}_u, \hat{\beta}_u$) of the MC under constraint DIST1-COV (significance indicated at the [†]10%, *5%, and **1% level).

Covariate	$u=2$	$u=3$	$u=4$	$u=5$
Intercept	0.727	1.632*	1.858**	2.514**
Gender (female)	-0.033	-0.285*	-0.201	-0.243 [†]
Race (black)	0.196	-0.139	-0.557**	-1.113**
Race (others)	-0.245	-0.245	-0.865**	-0.801**
Education (diploma)	0.625*	0.628*	0.966**	1.173**
Education (high school graduate)	0.445**	1.124**	1.667**	1.912**
Education (some college)	0.622**	1.532**	2.180**	2.581**
Education (college and above)	0.635 [†]	2.082**	3.076**	3.722**
Age	-0.002	-0.008	-0.019	-0.039**

1.2 Hidden Markov analysis

In the following we report additional results of the HM model illustrated in [Section 6.2](#) of the paper. We refer to Section 3 of the paper for the names of the constraints and their features. [Table 7](#) reports the estimated standard errors of the conditional probabilities of the HM model under constraint CONST-COV with $k = 5$ hidden states obtained through a parametric bootstrap based on $B = 200$ samples. [Table 8](#) shows the estimates of the model parameters for the initial probabilities under the

Table 5 Estimated intercepts for the transition probabilities ($\hat{\gamma}_{uv}$) of the MC model under constraint DIST1-COV (significance indicated at the † 10%, *5%, and **1% level).

State	$v=1$	$v=2$	$v=3$	$v=4$	$v=5$
$u=1$	0.000	0.229	-1.806	-3.483	-5.259
$u=2$	-0.951	0.000	0.022	0.041	-1.462
$u=3$	-2.608**	-3.973**	0.000	-0.770	-0.121
$u=4$	-0.226 †	-0.947**	-1.555	0.000	-2.511
$u=5$	-1.152	-0.822**	-0.915**	-0.560**	0.000

Table 6 Estimated regression coefficients for the transition probabilities of the MC model under constraint DIST1-COV (significance indicated at the † 10%, *5%, and **1% level).

Covariate	$\hat{\delta}_{-4}$	$\hat{\delta}_{-3}$	$\hat{\delta}_{-2}$	$\hat{\delta}_{-1}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
Gender (female)	-1.051**	-0.188	-0.196**	-0.035	-0.083**	-0.342**	-0.387*	-1.150*
Race (black)	0.749	0.943**	0.666**	0.176**	-0.174**	-0.000	0.498**	-0.434
Race (others)	0.672	0.271	0.493**	0.237**	-0.088	-0.019	-0.537	-13.006
Education (diploma)	-0.849	-0.861**	-0.501**	-0.194**	0.115 †	-0.031	-0.527	0.434
Education (high school g.)	-2.545**	-1.645**	-0.935**	-0.368**	0.170**	0.171*	-0.143	1.568**
Education (some coll.)	-2.961**	-1.858**	-1.197**	-0.500**	0.220**	0.304**	-0.527 †	0.636
Education (college and a.)	-3.828**	-2.469**	-1.634**	-0.682**	0.307**	0.342**	-0.078	1.333
Age	-0.026	0.036**	0.017**	0.009**	-0.011**	-0.028**	-0.033**	-0.029

CONST-COV constraint, while Tables 9 and 10 show the estimates of those for the transition probabilities.

Table 7 Standard errors for the estimates of the conditional response probabilities obtained by the parametric bootstrap based on $B=200$ samples generated from the HM model under constraint CONST-COV with $k = 5$ hidden states.

Standard errors of the conditional response probabilities ($\hat{\phi}_{1y,u}$)					
Category (y)	$u=1$	$u=2$	$u=3$	$u=4$	$u=5$
poor	0.014	0.012	0.005	0.002	0.001
fair	0.005	0.009	0.008	0.003	0.001
good	0.001	0.004	0.006	0.006	0.001
very good	0.000	0.001	0.005	0.006	0.004
excellent	0.000	0.001	0.003	0.006	0.007

Table 8 Estimated regression coefficients for the initial probabilities ($\hat{\alpha}_u, \hat{\beta}_u$) of the HM model under constraint CONST-COV (significance indicated at the $^\dagger 10\%$, $^* 5\%$, and $^{**} 1\%$ level).

Covariate	$u=2$	$u=3$	$u=4$	$u=5$
Intercept	1.081	1.815*	1.645*	2.699**
Gender (female)	-0.168	-0.473**	-0.316*	-0.425**
Race (black)	0.126	-0.128	-0.822**	-1.467**
Race (others)	-0.270	-0.376	-1.217**	-0.817*
Education (diploma)	0.559	0.457	1.090**	1.218**
Education (high school graduate)	0.723**	1.482**	2.193**	2.390**
Education (some college)	0.672*	1.756**	2.576**	3.078**
Education (college and above)	0.962	2.595	3.825*	4.554*
Age	-0.008	-0.011	-0.019	-0.044**

Table 9 Estimated intercepts on the transition probabilities ($\hat{\gamma}_{u,v}$) of the HM model under constraint CONST-COV (significance indicated at the $^\dagger 10\%$, $^* 5\%$, and $^{**} 1\%$ level).

Latent state	$v=1$	$v=2$	$v=3$	$v=4$	$v=5$
$u=1$	0.000	-2.447**	-2.788**	-4.643**	-5.430**
$u=2$	-10.484**	0.000	-27.169**	-2.835**	-2.236**
$u=3$	-4.363**	-4.962**	0.000	-5.047**	-5.375**
$u=4$	-3.195**	-1.748**	-3.448**	0.000	-25.901**
$u=5$	-6.187**	-5.384**	-3.703**	-1.137**	0.000

Table 10 Estimated regression coefficients for the transition probabilities of the HM model under constraint CONST-COV (significance indicated at the $^\dagger 10\%$, $^* 5\%$, and $^{**} 1\%$ level).

Covariate	$\hat{\delta}$
Gender (female)	-0.232**
Race (black)	0.181**
Race (others)	0.326**
Education (diploma)	-0.331**
Education (high school graduate)	-0.683**
Education (some college)	-0.774**
Education (college and above)	-1.036**
Age	0.006 †