Testing for time-invariant unobserved heterogeneity in nonlinear panel-data models

Francesco Bartolucci[1], University of Perugia (IT)
Federico Belotti, Tor Vergata University (IT)
Franco Peracchi, Tor Vergata University and EIEF (IT)

Seminar at Université de Genève
October 5, 2012

[1] bart@stat.unipg.it

Outline of the presentation

Motivation

Proposed approach

Simulations

Empirical example

Conclusions

Main references

Why time-invariant unobserved heterogeneity?

- An important element when modeling panel data is the treatment of unobserved heterogeneity; this form of heterogeneity is typically interpreted as the effects of unobservable factors on the outcome of interest

- A popular approach is to include in the model time-invariant unit-specific effects, which may be treated as fixed or random

- The assumption of time-invariant effects is difficult to justify in certain applications (e.g., health and labor economics), especially with long panels (Hyslop, 1999; Heiss, 2008; Stowasser et al., 2011)

- Biased parameter estimates may result if the unobservable effects are in fact not constant over time

Time-varying effects in non-linear panel-data models

- Hyslop (1999) proposes a dynamic random-effects probit model that allows for serial correlation in the error term

- Heiss (2008) proposes a class of limited dependent variable models where unit-specific parameters follow an AR(1) process

- Bartolucci & Farcomeni (2009) propose a multivariate extension of the dynamic logit model by including a vector of unit-specific parameters which follow a first-order homogeneous Markov chain

- These approaches require assumptions about the distribution of the unit-specific effects, thus raising problems of computational complexity and robustness

Our contribution

- We propose a simple procedure to test the null hypothesis of time-invariant effects in non-linear panel-data models against the alternative of time-varying effects

- The proposed test is attractive because:
  - It can be used with dependent variables of different types (we focus on binary and ordinal outcomes)
  - It does not require distinguishing between fixed and random effects
  - It does not require assumptions on the distribution of unit- and time-specific effects
  - It can be viewed as a specification test against time-varying omitted variables that are possibly correlated with the included covariates

The statistical model

- Basic notation:
  - n: sample size
  - T: number of occasions of observation
  - y_{it}: binary or ordinal outcome (with J categories)
  - $\mathbf{x}_{it}$: vector of k covariates
- We interpret the observable outcome y_{it} as related to a latent continuous outcome y^*_{it}
- y^*_{it} is assumed to follow the linear model

$$y^*_{it} = \alpha_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T$$

  - $\boldsymbol{\beta}$: parameter vector for the covariates
  - α_{it}: unobservable unit-specific time-varying effects
  - ε_{it}: random errors which are iid with 0 mean

- y_{it} is tied to y_{it}^* through the observation rule

$$y_{it} = \sum_{j=0}^{J-1} j \cdot 1\{\tau_j \leq y_{it}^* < \tau_{j+1}\}$$

  - $1\{A\}$: indicator function equal to 1 if the event A is true
  - $\tau_0 < \tau_1 < \cdots < \tau_{J-1} < \tau_J$: thresholds ($\tau_0 = -\infty$, $\tau_J = \infty$)

- When $J = 2$, y_{it} is a binary 0-1 indicator

Distributional assumptions

- We assume that, conditionally on α_{it} and $\mathbf{x}_{it}$, the errors ε_{it} are iid with a standard logistic distribution
- When $J = 2$ (binary case) a logistic regression with unobservable effects results:

$$\log \frac{p(y_{it} = 1|\alpha_{it}, \mathbf{x}_{it})}{p(y_{it} = 0|\alpha_{it}, \mathbf{x}_{it})} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}$$

- With $J > 2$ (ordinal case) a proportional odds regression model (McCullagh, 1980) based on global-logits results:

$$\log \frac{p(y_{it} \geq j|\alpha_{it}, \mathbf{x}_{it})}{p(y_{it} < j|\alpha_{it}, \mathbf{x}_{it})} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta} + \gamma_j, \quad j = 1, \dots, J-1$$

  - γ_j: intercepts related to the thresholds τ_j

Hypotheses of interest

- Null hypothesis (H_0): Unit-specific unobserved heterogeneity is constant over time

$$\alpha_{i1} = \alpha_{i2} = \cdots = \alpha_{iT} = \alpha_i, \qquad i = 1, \ldots, n$$

- Alternative hypothesis (H_1): Unit-specific unobserved heterogeneity is time-varying, with no *a priori* assumptions on how it evolves over time

- The test is based on the comparison of standard and pairwise Conditional Maximum Likelihood (CML) estimators of β

- As the the sample size (n) increases, both estimators are consistent under H_0, but they converge to different points under H_1

Standard CML estimator

- When the observed outcomes y_{it} are binary, the estimator is based on the maximization of the conditional log-likelihood for the fixed-effects logit model (Andersen, 1970; Chamberlain, 1980):

$$\ell_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log p(\mathbf{y}_i | \mathbf{X}_i, y_{i+})$$

  - $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$: matrix of all individual covariates
  - $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})$: response configuration of sample unit i
  - $y_{i+} = \sum_t y_{it}$: sufficient statistic for the unit-specific effect α_i
  - $p(\mathbf{y}_i | \mathbf{X}_i, y_{i+})$: conditional probability of $\mathbf{y}_i$ given $\mathbf{X}_i$ and y_{i+}

- By conditioning on y_{i+} we remove the dependence of $\mathbf{y}_i$ on α_i:

$$p(\mathbf{y}_i | \alpha_i, \mathbf{X}_i, y_{i+}) = \frac{p(\mathbf{y}_i | \alpha_i, \mathbf{X}_i)}{\sum_{z_+ = y_{i+}} p(\mathbf{y}_i = \mathbf{z} | \alpha_i, \mathbf{X}_i)} = p(\mathbf{y}_i | \mathbf{X}_i, y_{i+})$$

- When the observed outcome is ordinal, CML is applied after that the outcomes are dichotomized in all the possible ways (Brant, 1990; Baetschmann et al., 2011):

$$y_{it}^{(j)} = 1\{y_{it} \geq j\}, \quad j = 1, \ldots, J-1$$

- The conditional log-likelihood that is maximized is:

$$\ell_1(\boldsymbol{\beta}) = \sum_{j=1}^{J-1} \sum_{i=1}^{n} \log p(\mathbf{y}_i^{(j)} | \mathbf{X}_i, y_{i+}^{(j)})$$

  - $\mathbf{y}_i^{(j)} = (y_{i1}^{(j)}, \ldots, y_{iT}^{(j)})$: vector of dichotomized outcomes at level j

- The standard CML estimator is denoted by $\hat{\boldsymbol{\beta}}_1$:
  - under H_0, $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}$
  - under H_1, $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_{1*} \neq \boldsymbol{\beta}$

Pairwise CML estimator

- When y_{it} is binary, the pairwise version of the $\ell_1(\boldsymbol{\beta})$ is

$$\ell_2(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{t=2}^{T} \log p(y_{i,t-1}, y_{it} | \mathbf{x}_{i,t-1}, \mathbf{x}_{it}, y_{i,t-1} + y_{it})$$

- When y_{it} is ordinal, the pairwise version of the $\ell_1(\boldsymbol{\beta})$ is

$$\ell_2(\boldsymbol{\beta}) = \sum_{j=1}^{J-1}\sum_{i=1}^{n}\sum_{t=2}^{T} \log p(y_{i,t-1}^{(j)}, y_{it}^{(j)} | \mathbf{x}_{i,t-1}, \mathbf{x}_{it}, y_{i,t-1}^{(j)} + y_{it}^{(j)})$$

- The pairwise CML estimator is denoted by $\hat{\boldsymbol{\beta}}_2$:
  - under H_0, $\hat{\boldsymbol{\beta}}_2 \xrightarrow{p} \boldsymbol{\beta}$
  - under H_1, $\hat{\boldsymbol{\beta}}_2 \xrightarrow{p} \boldsymbol{\beta}_{2*} \neq \boldsymbol{\beta}$, with $\boldsymbol{\beta}_{2*} \neq \boldsymbol{\beta}_{1*}$ (in general)

Reshaping the data: Consecutive pairwise observations

a) Original data

i	t	y_{it}	x_{it}
1	1	1	-.564
1	2	0	.234
1	3	0	-.764
1	4	1	.921
1	5	1	.218

$\rightarrow$

b) Paired data

i	t	y_{it}	x_{it}
1	1	1	-.564
1	2	0	.234
2	1	0	.234
2	2	0	-.764
3	1	0	-.764
3	2	1	.921
4	1	1	.921
4	2	1	.218

- Panel (a) shows the personal identifier ($i = 1$), the observed outcome (y_{it}) and a covariate (x_{it}) for a specific subject observed for $T = 5$ periods
- Panel (b) shows the four consecutive pairs obtained from the original data

Test implementation

- We first consider the difference: $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$
- Under H_0, both estimators are consistent: $\hat{\boldsymbol{\delta}} =\xrightarrow{p} \mathbf{0}$ and $\sqrt{n}\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V})$
- Under H_1, the two estimators diverge: $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_* \neq \mathbf{0}$ and $\sqrt{n}\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\delta}_*, \mathbf{V}_*)$
- The second conclusion follows from the fact that, since $\ell_2(\boldsymbol{\beta})$ is maximized using only consecutive pairs of observations, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ will have different probability limits in general (a formal proof is under preparation)
- The test statistic (Hausman, 1978) is then

$$U = n\hat{\boldsymbol{\delta}}'\hat{\mathbf{V}}^{-}\hat{\boldsymbol{\delta}}$$

- $\hat{\mathbf{V}}^{-}$: generalized inverse of the estimate of $\mathbf{V}$

Test statistic

- The null asymptotic variance-covariance matrix of $\sqrt{n}\hat{\boldsymbol{\delta}}$ is estimated by a sandwich formula:

$$\hat{\mathbf{V}} = \mathbf{D}'\hat{\mathbf{W}}\mathbf{D}$$

$$\hat{\mathbf{W}} = n\mathbf{H}^{-1}\mathbf{S}\mathbf{H}^{-1} = n \begin{bmatrix} \mathbf{H}_1^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{H}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_1'\mathbf{S}_1 & \mathbf{S}_1'\mathbf{S}_2 \\ \mathbf{S}_2'\mathbf{S}_1 & \mathbf{S}_2'\mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{H}_1^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{H}_2^{-1} \end{bmatrix}$$

$$\mathbf{S}_m = \frac{\partial \boldsymbol{\ell}_m(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}, \qquad \mathbf{H}_m = \frac{\partial^2 \ell_m(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'}, \qquad m = 1, 2$$

  - $\mathbf{D} = [\mathbf{I}_k, -\mathbf{I}_k]$: matrix such that $\hat{\boldsymbol{\delta}} = \mathbf{D}(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$
  - $\boldsymbol{\ell}_m(\boldsymbol{\beta})$: vector of individual likelihood components (e.g., $\log p(\mathbf{y}_i | \mathbf{X}_i, y_{i+})$, $i = 1, \ldots, n$)

- Under H_0, $U \xrightarrow{d} \chi^2_{k^*}$
  - $k^* = \text{rank}(\mathbf{V}) \leq k$

Power of the test

- Being a pure specification test, the test may lack power in some cases (Holly, 1982)
- One such case is when unobserved heterogeneity effects are serially correlated but the first-order autocorrelation is zero (e.g., some form of seasonality)
- To handle this case, we may generalize the test by considering all possible pairs of observations for the same unit (in progress)
- We analyze the size and power of the proposed test using a set of Monte Carlo experiments

Set-up of the Monte Carlo

- Two data generating processes (DGPs) for the latent variable with only one covariate:

  1. DGP$_1$: $y^*_{it} = \alpha_{it} + x_{it}\beta + \varepsilon_{it}$
  2. DGP$_2$: $y^*_{it} = \alpha_{it} + x_{it}\beta + \sum_{k=1}^{T} d_{tk}\phi_k + \varepsilon_{it}$

  and covariates chosen as:

  - $x_{it} \sim N(0,1)$
  - $\varepsilon_{it} = \log[u_{it}/(1-u_{it})]$, $u_{it} \sim U(0,1)$
  - α_{it} follows an AR(1) process

$$\alpha_{it} = \begin{cases} v_{it}, & t = 1 \\ \alpha_{i,t-1}\rho + v_{it}\sqrt{1-\rho^2}, & t = 2, \ldots, T \end{cases}$$

  - $v_{it} \sim N(0,1)$
  - $d_{tk} = 1\{t = k\}$: dummy for the inclusion of time effects

Parameters of the Monte Carlo

- Under both DGPs, we considered binary ($J = 2$) and ordinal ($J = 5$) observable outcomes, with thresholds chosen as:
  - $\tau = 0 \rightarrow$ for binary outcomes
  - $\tau = (-2, -0.75, 0.75, 2) \rightarrow$ for ordinal outcomes
- In each experiment we keep fixed the values of the continuous covariate parameter $\beta = 1$ and of the time dummy parameters ϕ_k (equally spaced sequence between -2 and 2)
- We investigate the effect of:
  - different values of sample size (n=1000,2000,4000) and panel length (T=5,7,10)
  - different values of the AR(1) parameter ($\rho = 1, 0.75, 0.5, 0.25$)
  - If $\rho = 1$, then $\alpha_{i1} = \alpha_{i2} = \cdots = \alpha_{iT}$ and the unit-specific unobserved heterogeneity becomes time-invariant (H_0)

Table: Binary outcome (DGP_1), one covariate

(under H_0: $E(U) = 1$, $V(U) = 2$, $\sqrt{V(U)} = 1.414$)

	$T = 5$			$T = 7$			$T = 10$		
n	mean	sd	power	mean	sd	power	mean	sd	power
				$\rho = 1$					
1000	1.045	1.480	.054	1.038	1.488	.056	1.062	1.602	.053
2000	1.055	1.521	.050	1.065	1.596	.058	1.040	1.442	.059
4000	1.061	1.478	.054	.984	1.397	.047	.987	1.452	.055
				$\rho = .75$					
1000	1.638	2.005	.120	2.932	3.103	.280	5.528	4.236	.577
2000	2.198	2.421	.202	4.812	3.895	.519	10.081	5.702	.877
4000	3.654	3.501	.370	8.714	5.523	.812	19.607	7.935	.993
				$\rho = .5$					
1000	1.780	2.140	.146	2.914	3.059	.302	4.830	3.917	.512
2000	2.494	2.611	.233	4.853	3.967	.527	8.781	5.337	.820
4000	4.168	3.843	.423	8.837	5.611	.803	16.825	7.473	.985
				$\rho = .25$					
1000	1.328	1.732	.087	1.688	2.146	.124	2.245	2.492	.202
2000	1.620	2.102	.132	2.455	2.599	.225	3.566	3.305	.370
4000	2.314	2.688	.226	3.910	3.617	.404	6.212	4.529	.644

Table: Binary outcome (DGP_2), one covariate + time dummies

(under H_0: $E(U) = 5, 7, 10,\ V(U) = 10, 14, 20,\ \sqrt{V(U)} = 3.162, 3.742, 4.472$)

	$T = 5$			$T = 7$			$T = 10$		
n	mean	sd	power	mean	sd	power	mean	sd	power
				$\rho = 1$					
1000	5.117	3.334	.049	7.219	3.829	.059	10.182	4.599	.056
2000	4.877	3.219	.050	6.997	3.794	.054	9.974	4.509	.052
4000	5.005	3.125	.052	6.977	3.797	.051	9.857	4.341	.040
				$\rho = .75$					
1000	5.474	3.092	.056	8.353	4.252	.096	13.298	5.291	.174
2000	6.396	3.502	.106	10.647	5.123	.230	17.936	6.417	.432
4000	8.218	4.296	.242	14.832	6.250	.507	26.648	8.537	.845
				$\rho = .5$					
1000	5.675	3.302	.073	8.643	4.206	.102	12.956	5.108	.143
2000	6.913	3.779	.151	10.969	5.176	.249	16.946	6.515	.371
4000	9.298	4.837	.318	15.792	6.606	.551	25.027	8.463	.790
				$\rho = .25$					
1000	5.243	3.196	.058	7.544	3.721	.049	10.719	4.544	.062
2000	5.771	3.325	.082	8.443	4.211	.106	12.108	5.240	.122
4000	6.963	3.984	.152	10.321	5.031	.211	14.991	6.172	.274

Table: Ordinal outcome (DGP_1), one covariate

(under H_0: $E(U) = 1$, $V(U) = 2$, $\sqrt{V(U)} = 1.414$)

	$T = 5$			$T = 7$			$T = 10$		
n	mean	sd	power	mean	sd	power	mean	sd	power
				$\rho = 1$					
1000	1.136	1.497	.060	.959	1.399	.048	1.001	1.383	.054
2000	1.126	1.649	.061	.980	1.376	.054	.986	1.379	.049
4000	1.061	1.499	.054	.981	1.392	.047	.951	1.484	.049
				$\rho = .75$					
1000	1.815	2.208	.144	3.364	3.158	.349	7.294	4.864	.733
2000	2.638	2.822	.261	6.061	4.295	.635	13.499	6.875	.947
4000	4.361	3.883	.454	11.541	6.251	.905	25.929	9.582	.997
				$\rho = .5$					
1000	1.995	2.330	.175	3.505	3.252	.362	6.333	4.492	.653
2000	3.017	3.030	.308	6.291	4.504	.645	11.733	6.457	.914
4000	5.169	4.191	.538	11.735	6.413	.907	22.279	8.789	.997
				$\rho = .25$					
1000	1.391	1.816	.097	1.931	2.217	.163	2.822	2.850	.282
2000	1.821	2.241	.147	3.015	2.968	.317	4.671	3.965	.503
4000	2.759	2.974	.274	5.182	4.235	.533	8.275	5.319	.779

Table: Ordinal outcome (DGP$_2$), one covariate + time dummies

(under H_0: $E(U) = 5, 7, 10$, $V(U) = 10, 14, 20$, $\sqrt{V(U)} = 3.162, 3.742, 4.472$)

	$T = 5$			$T = 7$			$T = 10$		
n	mean	sd	power	mean	sd	power	mean	sd	power
				$\rho = 1$					
1000	5.192	3.326	.064	7.134	3.801	.057	10.116	4.661	.058
2000	5.095	3.314	.055	6.977	3.752	.050	10.137	4.480	.055
4000	4.969	3.179	.048	6.925	3.632	.051	10.048	4.660	.063
				$\rho = .75$					
1000	6.023	3.552	.094	9.841	4.969	.183	16.089	6.030	.315
2000	7.559	4.135	.202	13.504	5.965	.433	23.357	7.891	.718
4000	10.806	5.338	.422	20.710	7.624	.795	37.636	10.401	.976
				$\rho = .5$					
1000	6.522	3.824	.124	10.044	4.915	.192	15.470	6.027	.299
2000	8.402	4.565	.257	13.889	5.787	.460	21.799	7.610	.651
4000	12.419	5.797	.540	21.488	7.667	.833	34.598	9.978	.953
				$\rho = .25$					
1000	5.492	3.318	.060	8.088	4.133	.083	11.610	4.929	.098
2000	6.343	3.636	.105	9.513	4.497	.161	13.943	5.697	.206
4000	8.201	4.461	.231	12.620	5.474	.368	18.499	6.931	.476

Summary of the Monte Carlo

The available results suggest that the test generally performs well and it shows:

- small size bias
- good power properties, especially for ordinal outcomes and as n and T increase

Empirical illustration

- Data from the Health and Retirement Study (HRS), a longitudinal panel study that surveys a representative sample of more than 26,000 Americans over the age of 50 every two years
- We estimate the same model as Heiss (2008) but using a longer panel
- We employ the RAND HRS Data File (Version L), a user-friendly version of the data produced by the RAND Center for the Study of Aging, which contains all 10 waves from 1992 to 2010
- The selected sample is a balanced panel of 4,060 respondents
- The outcome variable is the self-rated health (SRH) with $J = 5$ categories
- Two model specifications:
  - M_1: age splines, BMI, number of GP visits
  - M_2: M_1 + wave dummies

Table: Summary statistics ($n = 4,060$; $T = 10$)

Variable	mean	sd	min	max
SRH	3.4	1.0	1	5
Age	65.2	7.0	50	93
Female	.52	.50	0	1
High school	.34	.47	0	1
Some college	.20	.40	0	1
College degree+	.22	.41	0	1
Non white	.15	.36	0	1
BMI	27.4	4.8	12.8	52.3
GP visits	7.2	8.9	0	100

Table: Logit (under H_0)

	M_1	M_2
Standard		
Age splines: 50+	-.107 ***	-.102
Age splines: 60+	.048 ***	.040 **
Age splines: 70+	-.044 ***	-.049 ***
Age splines: 80+	.035	.036
Age splines: 90+	-1.008 *	-1.041 *
BMI	-.026 ***	-.027 ***
GP visits	-.046 ***	-.046 ***
Pairwise		
Age splines: 50+	-.115 ***	-.160
Age splines: 60+	.041 **	.020
Age splines: 70+	-.027	-.055 **
Age splines: 80+	.056	.032
Age splines: 90+	-.027	-.068
BMI	-.024 *	-.026 **
GP visits	-.034 ***	-.034 ***
Wave dummies	No	Yes
U	56.9	178.8
p-value	.000	.000

Table: AR(1) logit (under H_1)

	M_1	M_2
Age splines: 50+	-.114 ***	-.052 ***
Age splines: 60+	.053 ***	.044 **
Age splines: 70+	-.043 ***	-.062 ***
Age splines: 80+	.031	.020
Age splines: 90+	-1.137 **	-1.172 **
BMI	-.066 ***	-.064 ***
GP visits	-.052 ***	-.052 ***
Female	.142	.194 *
High school	1.444 ***	1.472 ***
Some college	1.951 ***	1.997 ***
College degree+	2.779 ***	2.819 ***
Non white	-1.106 ***	-1.102 ***
Wave dummies	No	Yes
σ	2.736 ***	2.752 ***
ρ	.951 ***	.951 ***
Log-lik	-43181.8	-43063.3

Summary of the empirical illustration

- The null hypothesis of time-invariant unobserved heterogeneity is strongly rejected, thus confirming Heiss (2008)'s results (even if based on a shorter panel)
- A more plausible model for the data is one where SRH depends on true unobservable health, and this latent variable follows a time-series process with decaying autocorrelation

Conclusions

- Pros of the proposed approach:
  - It can be used with dependent variables of different types (we focus on binary or ordinal outcomes); indeed it may be used with any model based on a canonical link function (in the sense of GLM)
  - The approach does not require assumptions on the distribution of unit- and time-specific effects

- Cons of the proposed approach:
  - We assume a parametric link (logit)
  - The use of conditional likelihood methods implies that:
    - time-variations in the outcome for a given unit are needed
    - no time-invariant regressors can be included in the model
  - Unit-specific effects enter additively

Further steps and extensions

- Theoretical study of the asymptotic distribution of the CML estimators (standard and pairwise) under H_1

- A more comprehensive set of Monte Carlo experiments to investigate the effect of heteroskedasticity and different time-series models for the individual effects (e.g., AR(2), first-order Markov chain, etc.)

- Extension of the approach to handle more complex structures (e.g., all pairwise combinations to handle certain forms of seasonality)

Main references

- Andersen, E. (1970). Asymptotic properties of conditional maximum likelihood estimators. *JRSS-B*, **32**, pp. 283-301.
- Bartolucci, F. (2006). Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. *JRSS-B*, **68**, pp. 155-178.
- Bartolucci, F. and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *JASA*, **104**, pp. 816-831.
- Bartolucci, F., Farcomeni, A. and Pennoni, F. (2012). *Latent Markov Models for longitudinal data*. Chapman & Hall/CRC Press.
- Baetschmann, G., Staub, K. E. and Winkelmann, R. (2011). Consistent estimation of the fixed effects ordered logit model. *IZA Discussion paper*, 5443.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordered logistic regression. *Biometrics*, **46**, pp. 1171-1178.
- Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, **41**, pp. 164-171.

- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, **47**, pp. 225-238.
- Hausman, J. A. (1978). Specification tests in Econometrics. *Econometrica*, **46**, pp. 1251-1271.
- Heiss, F. (2008). Sequential numerical integration in nonlinear state space models for microeconometric panel data. *Journal of Applied Econometrics*, **23**, 373-389.
- Holly, A. (1982). A remark on Hausmans specification test. *Econometrica*, **50**, pp. 749-759.
- Hsiao, C. (2005). *Analysis of Panel Data*. New York: Cambridge University Press.
- Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, **67**, pp. 1255-1294.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion). *JRSS-B*, **42**, pp. 109-142.
- Stowasser, T., Heiss, F., McFadden, D. and Winter, J. (2011). Healthy, Wealthy and Wise? Revisited: An Analysis of the Causal Pathways from Socio-economic Status to Health. NBER Working Papers 17273 National Bureau of Economic Research, Inc.