Panel Models with Factor Structure

RES Annual Conference 2009

Iván Fernández-Val
BU

Martin Weidner
UCL

April 2019

Factor Structures or Interactive Effects

- Factor structures are convenient devices to incorporate latent variables in panel data models
- Popular in economics because they capture
  - aggregate shocks with heterogeneous effects across agents in macroeconomic models
  - multidimensional unobserved heterogeneity with time varying effect in microeconomic models
- Account for dependences in the cross section and time series dimensions in a parsimonious fashion
- Also called interactive effects in econometric panel data literature
- Overview of econometric issues and recent developments

Outline

1. Pure factor model
2. Linear model
3. Nonlinear model
4. Network data
5. Nonseparable model

1. Pure Factor Model

$$Y_{it} = \lambda_i' f_t + e_{it} = \sum_{r=1}^{R} \lambda_{ir} \, f_{tr} + e_{it}$$

- Panel data: $i = 1, \ldots, N$, $t = 1, \ldots, T$, $n = NT$
- Y_{it} is observed scalar response or outcome variable of interest
- λ_i is R-vector of unobserved factor loadings that captures individual heterogeneity
- f_t is R-vector of unobserved factors that captures aggregated shocks
- The number of factors R is know and does not vary with N and T
- e_{it} is zero-mean unobserved idiosyncratic error independent of λ_i and f_t, usually modeled as weakly dependent over i and t
- In matrix notation: $\underset{n \times T}{Y} = \underset{N \times R}{\lambda} \; \underset{(T \times R)'}{f'} + \underset{n \times T}{e}$

PCA: Principal Component Analysis (Pearson, 1901)

$$\left\{\widehat{\lambda}, \widehat{f}\right\} \in \underset{\{\lambda \in \mathbb{R}^{N\times R}, f \in \mathbb{R}^{T\times R}\}}{\operatorname{argmin}} \operatorname{Tr}(Y - \lambda f')(Y - \lambda f')' = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \lambda_i' f_t\right)^2$$

- Non-linear least squares estimator
- Resulting $\widehat{\lambda}$ and $\widehat{f}$ are principal components: eigenvectors corresponding to the largest eigenvalues of YY' and $Y'Y$

  $\Rightarrow$ easy and fast computation for case of balanced panels
- Need to impose rotation normalization, e.g.,

$$\frac{1}{T} f'f = \mathbb{I}_R \qquad\qquad \frac{1}{N} \lambda'\lambda = \text{diagonal matrix}$$

- Fixed effect approach that treats λ_i and f_t as parameters to be estimated
- Asymptotic properties are well-established in short panels ($N \to \infty$, *T.W. Anderson book*) and large panels ($N, T \to \infty$, *Bai, 2003*)

Estimation of Number of Factors

- There are different methods available to estimate R consistently
- *Bai & Ng (2002)* developed estimators based on model selection criteria (AIC, BIC)
- *Onatski (2009, 2010)* derived estimators based on random matrix theory for the distribution of eigenvalues
- *Ahn & Horenstein (2013)* proposed a related estimator based on maximizing the eigenvalue ratio (ER) of the matrix YY'

2. Linear Model

$$Y_{it} = X_{it}'\beta + \alpha_i + \gamma_t + \lambda_i' f_t + e_{it}$$

- X_{it} is K-vector of covariates which are strictly exogenous or predetermined with respect to e_{it}
- α_i and γ_t are unobserved additive individual and time effects capturing individual heterogeneity and aggregate shocks
- α_i and γ_t can be subsumed in factor structure since

  $$\alpha_i + \gamma_t + \lambda_i' f_t = \tilde{\lambda}_i' \tilde{f}_t, \quad \tilde{\lambda}_i = (\alpha_i, 1, \lambda_i')', \quad \tilde{f}_t = (1, \gamma_t, f_t')',$$

  but are often kept separately
- β is regression coefficient of interest measuring effect of X_{it} on Y_{it}
- α_i, λ_i, γ_t and f_t can be related to X_{it} causing omitted variable bias

Least Squares Estimator (Kiefer, 1980)

$$\left\{\widehat{\beta}, \widehat{\lambda}, \widehat{f}\right\} \in \underset{\{\beta \in \mathbb{R}^K,\, \lambda \in \mathbb{R}^{N \times R},\, f \in \mathbb{R}^{T \times R}\}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - X_{it}'\beta - \lambda_i' f_t\right)^2$$

- Problem can be solved iteratively
  - Given $\widehat{\beta}$, $\widehat{\lambda}$ and $\widehat{f}$ are obtained by principal components analysis
  - Given $\widehat{\lambda}$ and $\widehat{f}$, $\widehat{\beta}$ is obtained by least squares
- Additive effects can be incorporated by demeaning Y_{it} and X_{it}
- Rotation normalization for $\widehat{\lambda}$ and $\widehat{f}$ does affect $\widehat{\beta}$
- Estimators of R can be applied to

$$Y_{it} - X_{it}'\widetilde{\beta},$$

  where $\widetilde{\beta}$ is a preliminary consistent estimator of β, which requires specifying maximum number of factors $R_{\max}$, i.e. $R < R_{\max}$

Asymptotic Theory for Least Squares Estimator

- *Bai (2009)* showed that, as $n \to \infty$, $N/T \to c$ and $T/N \to c'$,

$$\widehat{\beta} \overset{a}{\sim} \mathcal{N}\left(\beta + \frac{B}{T} + \frac{D}{N}, \frac{V}{NT}\right)$$

  Moon & Weidner (2015, 2017) generalized this result to dynamic models and unknown R

- LS estimator suffer from incidental parameter bias due to the estimation of the factors and loadings

- B comes from estimation of loadings and D comes from estimation of factors

- Source of the bias is slow rate of convergence of $\widehat{\lambda}$ and $\widehat{f}$ with respect to $\widehat{\beta}$

- Debiasing is required for valid inference: *Bai (2009), Moon & Weidner (2015, 2017)*

Neyman-Scott with Factors (Chen, F-V & Weidner, 18)

$$Y_{it} \sim \mathcal{N}(\lambda_i' f_t, \sigma^2), \quad \widehat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \widehat{\lambda}_i' \widehat{f}_t \right)^2$$

where $\widehat{\lambda}_i$ and $\widehat{f}_t$ are principal components

- Asymptotic expansion of $\widehat{\lambda}_i' \widehat{f}_t$ around $\lambda_i' f_t$ as $N, T \to \infty$ yields

$$\widehat{\lambda}_i' \widehat{f}_t \approx \lambda_i' f_t + (\widehat{\lambda}_i - \lambda_i)' f_t + \lambda_i' (\widehat{f}_t - f_t)$$

- Behaves as estimator with $R(N + T)$ additive fixed effects

- Standard degrees of freedom calculation gives

$$E[\widehat{\sigma}^2] \approx \frac{(N - R)(T - R)}{NT} \sigma^2 \approx \sigma^2 - \frac{R(N + T)}{NT} \sigma^2$$

- First-order bias grows proportionally to number of factors R

Asymptotic vs Exact Bias

Bias	$N = 10$ $T = 10$	$N = 25$ $T = 10$	$T = 25$	$T = 10$	$N = 50$ $T = 25$	$T = 50$
			$R = 1$			
Asymptotic	-.19	-.14	-.08	-.12	-.06	-.04
Exact	-.20	-.14	-.08	-.12	-.06	-.04
			$R = 2$			
Asymptotic	-.36	-.26	-.15	-.23	-.12	-.08
Exact	-.39	-.27	-.16	-.24	-.12	-.08
			$R = 3$			
Asymptotic	-.51	-.38	-.23	-.34	-.17	-.12
Exact	-.55	-.40	-.23	-.35	-.18	-.12

Notes: Results obtained by 50,000 simulations

Design: $Y_{it} \sim \mathcal{N}(\lambda_i' f_t, \sigma^2)$, $\lambda_i \sim N(0, \mathbb{I}_R)$, $f_t \sim N(0, \mathbb{I}_R)$, $\sigma^2 = 1$

Inference and Bias Corrections

As $N, T \to \infty$ with $N/T \to 1$

$$\sqrt{NT}(\widehat{\sigma}^2 - \sigma^2) \overset{a}{\sim} N(-2R\sigma^2, 2\sigma^4)$$

- Undercoverage of confidence intervals constructed around $\widehat{\sigma}^2$

  $\Pr(\sigma^2 \in \widehat{\sigma}^2[1 \pm z_{1-\alpha/2}\sqrt{2/(NT)}]) \to \Pr(|N(-\sqrt{2}R, 1)| \leq z_{1-\alpha/2}) < 1-\alpha$

- Analytically debiased estimator can be formed as

$$\widetilde{\sigma}^2_{\text{ABC}} = \frac{NT}{(N-R)(T-R)}\widehat{\sigma}^2.$$

- A split-sample debiased estimator can be formed as

$$\widetilde{\sigma}^2_{\text{SBC}} = 3\widehat{\sigma}^2 - \bar{\sigma}^2_{N,T/2} - \bar{\sigma}^2_{N/2,T},$$

  where $\bar{\sigma}^2_{N,T/2}$ and $\bar{\sigma}^2_{N/2,T}$ are averages of estimators in haft-panels

- Confidence intervals around debiased estimators have right coverage

$$\sqrt{NT}(\widetilde{\sigma}^2 - \sigma^2) \overset{a}{\sim} N(0, 2\sigma^4), \quad \widetilde{\sigma}^2 \in \{\widetilde{\sigma}^2_{\text{ABC}}, \widetilde{\sigma}^2_{\text{SBC}}\}$$

Debiasing and Coverage Rates

	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover
		$N=10, T=10$				$N=25, T=10$		
$\widehat{\sigma}^2$	-0.55	0.09	0.56	0.00	-0.40	0.07	0.41	0.00
$\widetilde{\sigma}^2_{\mathrm{ABC}}$	-0.08	0.19	0.20	0.75	-0.02	0.11	0.11	0.85
$\widetilde{\sigma}^2_{\mathrm{SBC}}$	-0.09	0.20	0.22	0.71	-0.03	0.12	0.13	0.81
		$N=25, T=25$				$N=50, T=10$		
$\widehat{\sigma}^2$	-0.23	0.05	0.24	0.01	-0.35	0.05	0.35	0.00
$\widetilde{\sigma}^2_{\mathrm{ABC}}$	-0.01	0.06	0.06	0.91	-0.01	0.08	0.08	0.88
$\widetilde{\sigma}^2_{\mathrm{SBC}}$	-0.02	0.07	0.07	0.85	-0.01	0.08	0.08	0.85
		$N=50, T=25$				$N=50, T=50$		
$\widehat{\sigma}^2$	-0.18	0.04	0.18	0.00	-0.12	0.03	0.12	0.01
$\widetilde{\sigma}^2_{\mathrm{ABC}}$	-0.00	0.04	0.04	0.92	-0.00	0.03	0.03	0.93
$\widetilde{\sigma}^2_{\mathrm{SBC}}$	-0.01	0.05	0.05	0.88	-0.00	0.03	0.03	0.92

Notes: 50,000 simulations. Nominal level is 0.95

Design: $Y_{it} \sim \mathcal{N}(\lambda_i' f_t, \sigma^2)$, $\sigma^2 = 1$, $\lambda_i \sim N(0, \mathbb{I}_R)$, $f_t \sim N(0, \mathbb{I}_R)$, $R = 3$

3. Nonlinear Model

$$Y_{it} \mid X_{it} \sim f_Y(\cdot \mid z_{it}), \quad z_{it} = X_{it}'\beta + \alpha_i + \gamma_t + \lambda_i' f_t$$

- f_Y is a known density with respect to some dominating measure
- Covariates and unobserved effects enter f_Y through the single index z_{it}
- Semiparametric model: distribution of covariates and unobserved effects is unspecified
- β measures effect of X_{it} on conditional distribution of Y_{it} controlling for unobserved effects
- APEs are functions of the data, parameters and unobserved effects

$$\delta = (NT)^{-1} \sum_{i,t} \mathbb{E}[\varDelta_{it}(\beta, z_{it})]$$

where $\varDelta_{it}$ is a known function

Examples

Binary response model: $Y_{it} \in \{0,1\}$, F is a cdf

$$f_Y(y \mid z_{it}) = F(z_{it})^y [1 - F(z_{it})]^{1-y}$$

- Includes logit and probit
- Partial effect of continuous kth element of X_{it}

$$\Delta_{it}(\beta, z_{it}) = \beta_k f(z_{it}), \quad f(z) = \partial F(z)/\partial z$$

Count response model: $Y_{it} \in \{0,1,2,\ldots\}$

$$f_Y(y \mid z_{it}) = f(y; \exp z_{it})$$

- $f(y;\lambda)$ is the pmf of Poisson distribution
- Partial effect of binary kth element of X_{it}

$$\Delta_{it}(\beta, z_{it}) = \exp(z_{it} + \beta_k(1 - X_{it,k})) - \exp(z_{it} - \beta_k X_{it,k})$$

Fixed Effects Estimator (Chen, F.-V., & Weidner, 2018)

$$\left\{\widehat{\beta}, \widehat{\lambda}, \widehat{f}\right\} \in \underset{\{\beta\in\mathbb{R}^K,\, \lambda\in\mathbb{R}^{N\times R},\, f\in\mathbb{R}^{T\times R}\}}{\operatorname{argmin}} \sum_{i,t} \log f_Y(Y_{it} \mid X_{it}'\beta + \lambda_i' f_t)$$

- Conditional maximum likehood estimator
- Computation can be challenging, one solution is EM algorithm (*Chen, 2014*)
- Additive effects can be treated by including a factor of ones with unknown loading and a loading of ones with unknown factor
- Plug-in fixed effects estimator of APE

$$\widehat{\delta} = (NT)^{-1} \sum_{i,t} \Delta_{it}(\widehat{\beta}, \widehat{z}_{it}), \quad \widehat{z}_{it} = X_{it}'\widehat{\beta} + \widehat{\alpha}_i + \widehat{\gamma}_t + \widehat{\lambda}_i'\widehat{f}_t$$

- Rotation normalization for $\widehat{\lambda}$ and $\widehat{f}$ does not affect $\widehat{\beta}$ or $\widehat{\delta}$
- Other approaches: *Ando and Bai (2016)*, *Boneva and Linton (2017)*

Asymptotic Theory

- *Chen, F.V., & Weidner (2018)* showed that, as $n \to \infty$, $N/T \to c$ and $T/N \to c'$,

$$\widehat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta + \frac{B_\theta}{T} + \frac{D_\theta}{N}, \frac{V_\theta}{NT}\right), \quad \theta \in \{\beta, \delta\}$$

- $\widehat{\beta}$ and $\widehat{\delta}$ suffer from incidental parameter problem
- First order bias proportional to R in probit and logit models
- No bias in Poisson models with strictly exogenous covariates
- *Wang (2018)* characterized asymptotic distribution of $\widehat{\lambda}_i$ and $\widehat{f}_t$
- Applying ER criterion to $\widetilde{\lambda}\widetilde{f}'$, a preliminary estimator of $\lambda f'$ with a large number of factors $R_{\max} > R$, yields an estimator of R that works well in numerical simulations

4. Networks

- Network data can be modeled as panel data where i indexes sender and t (j in this literature) indices receiver, when there are no strategic interactions
- Y_{ij} is either an indicator for connection between i and j, or some measure of the intensity of the connection
- Examples:
  - Social networks: Y_{ij} is a friendship indicator
  - Banking: Y_{ij} is a bank connection indicator or a common property measure
  - International trade: Y_{ij} is trade indicator or volume of trade
- It is natural to use long-panel asymptotics because $N = T$
- Examples: *Candelaria (16), Charbonneau (14), Dzemski (17), Graham (17), Jochmans (17, 18), Shi & Chen (16), and Yan et al. (16)*

Factor Structures in Networks (Chen, F.-V. & Weidner, 18)

$$Y_{ij} \mid X_{ij} \sim f_Y(\cdot \mid z_{ij}), \quad z_{ij} = X_{ij}'\beta + \alpha_i + \gamma_j + \lambda_i' f_j$$

- Nonlinear model with factor structure can capture important network features in a parsimonious, reduced form fashion
- Homophily in observable characteristic W (geographic distance):

$$X_{ij,k} = d(W_i, W_j) = (W_i - W_j)^2$$

- Degree heterogeneity (country size): α_i and γ_j
- Homophily in unobservable characteristic ξ (industrial composition):

$$(\xi_i - \xi_j)^2 = \lambda_i' f_j, \quad \lambda_i = (\xi_i^2, 1, -2\xi_i)', \quad f_j = (1, \xi_j^2, \xi_j)'$$

- Clustering due to grouping indicator I (FTA or multinational firm):

$$\zeta_i \chi_j I_i I_j = \lambda_i f_j, \quad \lambda_i = \zeta_i I_i, \quad \lambda_j = \chi_j I_j$$

Gravity Equation: Poisson Model with Factor Structure

	$R = 0$	$R = 1$	$R = 2$	$R = 3^*$	$R = 4$	$R = 5$
Log Distance	-0.64	-0.63	-0.71	-0.69	-0.77	-0.90
	[0.07]	[0.05]	[0.06]	[0.06]	[0.08]	[0.09]
Border	0.71	0.41	0.32	0.36	0.38	0.36
	[0.16]	[0.07]	[0.06]	[0.06]	[0.06]	[0.12]
Legal	0.30	0.14	0.26	0.22	0.13	0.16
	[0.06]	[0.04]	[0.04]	[0.04]	[0.04]	[0.06]
Language	-0.17	-0.19	-0.02	0.03	-0.09	-0.03
	[0.10]	[0.07]	[0.06]	[0.06]	[0.08]	[0.12]
Colony	0.36	0.58	0.39	0.45	0.63	0.61
	[0.12]	[0.14]	[0.12]	[0.12]	[0.14]	[0.28]
Currency	0.60	0.29	1.37	1.38	0.65	0.63
	[0.30]	[0.38]	[0.41]	[0.36]	[1.16]	[1.92]
FTA	0.25	0.15	0.17	0.13	0.25	0.31
	[0.09]	[0.07]	[0.07]	[0.07]	[0.09]	[0.14]
Religion	-0.25	0.18	0.24	0.34	0.44	0.30
	[0.13]	[0.11]	[0.13]	[0.13]	[0.13]	[0.26]

Notes: all the columns include importer and exporter additive effects.
Std errors robust to reciprocity in brackets. *Estimate of R with $R_{\text{max}} = 5$
Data: *Helpman, Melitz, and Rubinstein (2008)*, 157 countries in 1986

5. Nonseparable Model

$$Y_{it} = g(X_{it}, A_i, B_t, \varepsilon_{it})$$

- g is unknown function, and A_i and B_t are vectors of unobservables
- The idiosyncratic error ε_{it} is identically distributed across i and t

$$\varepsilon_{it} \mid X_{it}, A_i, B_t \overset{d}{=} \varepsilon_{js} \mid X_{js}, A_j, B_s \qquad \text{for all } i, j \text{ and } t, s$$

- $Y_{it}(x) := g(x, A_i, B_t, \varepsilon_{it})$ is potential outcome when $X_{it} = x$
- Objects of interest are functions of the average structural functions

$$\mu_t(x) := \mathbb{E}\left[Y_{it}(x)\right], \quad \nu_{x_0,t}(x) := \mathbb{E}\left[Y_{it}(x) \mid X_{it} = x_0\right]$$

- If X_{it} is a treatment indicator, the ATE and ATT at time t are

$$\mu_t(1) - \mu_t(0), \quad \nu_{1,t}(1) - \nu_{1,t}(0)$$

- Replace $Y_{it}(x)$ by $1(Y_{it}(x) \leq y)$ to obtain distributional effects

Factor Structure Approximation

$$\mu_t(x) = \mathbb{E}\left[m(X_{it}, A_i, B_t) \mid X_{it} = x\right], \quad m(X_{it}, A_i, B_t) = \mathbb{E}\left[Y_{it} \mid X_{it}, A_i, B_t\right]$$

- m does not depend on i and t by identical distribution of ε_{it}
- By a singular value decomposition of $(a, b) \mapsto m(x, a, b)$:

$$m(x, a, b) = \sum_{r=1}^{\infty} \underbrace{\varphi_r(x, a)}_{=:\lambda_r(x)} \underbrace{\kappa_r(x, b)}_{=:f_r(x)} = \sum_{r=1}^{R} \lambda_r(x)\, f_r(x) + \epsilon(x)$$

- *Griebel and Habrecht (2011)* showed that

$$\epsilon(x) = \mathcal{O}\left(R^{\frac{1}{2} - \frac{p}{d_a \wedge d_b}}\right)$$

  where p is the degree of smoothness of m, and d_a and d_b are the dimensions of a and b
- Factor structure can be seen as a series approximation in the unobservables

Estimation

- Assume that X_{it} is binary
- For $x \in \{0, 1\}$, estimate the pure factor model

$$Y_{it} = \lambda_i(x)' f_t(x) + e_{it}$$

  using only the observations with $X_{it} = x$ and $R = R_n$

- Estimate the average structural functions at time t as

$$\widehat{\mu}_t(x) = N^{-1} \sum_{i=1}^{N} \widehat{\lambda}_i(x)' \widehat{f}_t(x), \quad \widehat{\nu}_{1,t}(x) := N_1^{-1} \sum_{i=1}^{N} X_{it} \widehat{\lambda}_i(x)' \widehat{f}_t(x),$$

  where $\widehat{\lambda}_i(x)$ and $\widehat{f}_t(x)$ are the estimators of $\lambda_i(x)$ and $f_t(x)$, and $N_0 = \sum_{i=1}^{N} X_{it}$

- Problem: we cannot use PCA due to the missing/unbalanced data

Matrix Completion / Nuclear Norm Minimization

How to estimate a pure factor model if we only observe Y_{it} for a subset $\mathcal{S}_x = \{(i,t) : X_{it} = x\}$?

- Least squares: note that $\Gamma_x = \lambda(x)\, f(x)' \;\Leftrightarrow\; \mathrm{rank}(\Gamma_x) \le R_n$

$$\widehat{\Gamma}_x = \underset{\{\Gamma \in \mathbb{R}^{N\times T} : \mathrm{rank}(\Gamma) \le R_n\}}{\operatorname{argmin}} \sum_{(i,t)\in\mathcal{S}_x} \left[Y_{it} - \Gamma_{it}\right]^2 ,$$

- LS is often computationally infeasible, the constraint is non-convex

- Replace $\mathrm{rank}(\Gamma) \le R_n$ by its convex relaxation: $\|\Gamma\|_* \le \mathrm{const}$,

$$\widehat{\Gamma}_x = \underset{\Gamma \in \mathbb{R}^{N\times T}}{\operatorname{argmin}} \left\{ \sum_{(i,t)\in\mathcal{S}_x} \left[Y_{it} - \Gamma_{it}\right]^2 + \rho_n \|\Gamma\|_* \right\} ,$$

  where $\rho_n > 0$ now replaces R_n as a "bandwidth parameter", and $\|\Gamma\|_*$ is the nuclear norm (sum of singular values) of Γ

- Based on matrix completion methods (*Recht, Fazel, & Parrilo, 2010; Hastie, Tibshirani, & Wainwright, 2015*)

Econometric Applications of Matrix Completion Methods

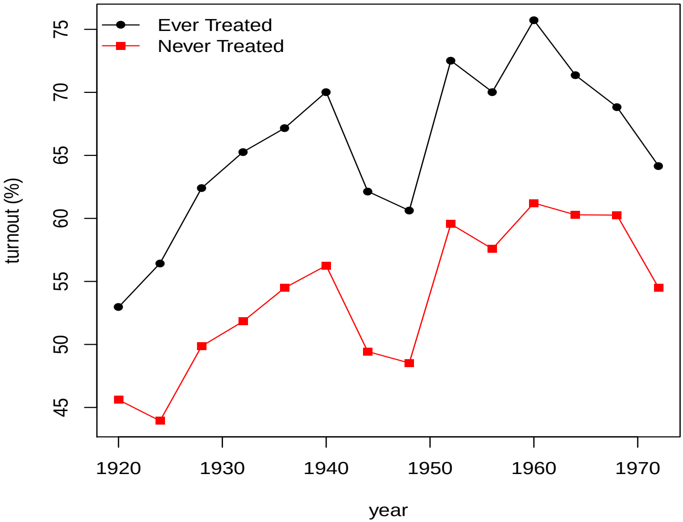
- *Athey, Bayati, Doudchenko, Imbens & Khosravi (2017)* applied matrix completion methods to estimate average treatment effects with panel data
- *Moon & Weidner (2018)* characterized the properties of nuclear norm regularized estimators of the linear model with factor structure
- *Chernozhukov, Hansen, Liao & Zhu (2018)* derived inference methods for linear panel data models with heterogenous slopes estimated using nuclear norm penalization
- *Archangelsky, Athey, Hirshberg, Imbens & Wager (2019)* derived consistency results for synthetic control estimators based on matrix completion methods

General inference after matrix completion is an active area of research

Election Day Registration (EDR) and Vote Turnout

- Effect of allowing vote registration in election day on vote turnout in the U.S. (*Xu, 2017*)
- Data: 24 presidential elections from 1920 to 2012, 47 states excluding Alaska, Hawaii and North Dakota (early adopter)
- Turnout rate is total ballots counted divided by voting-age population
- 4 waves of EDR adoption: ME, MN and WI in 1976; WY, ID and NH in 1994; MT and IA in 2008; and CT in 2012
- Focus on average treatment effect on the treated; staggered adoption (*Athey & Imbens, 2018*)
- Treated states have higher turnouts in pretreatment periods

Assessing Pretreatment Parallel Trends



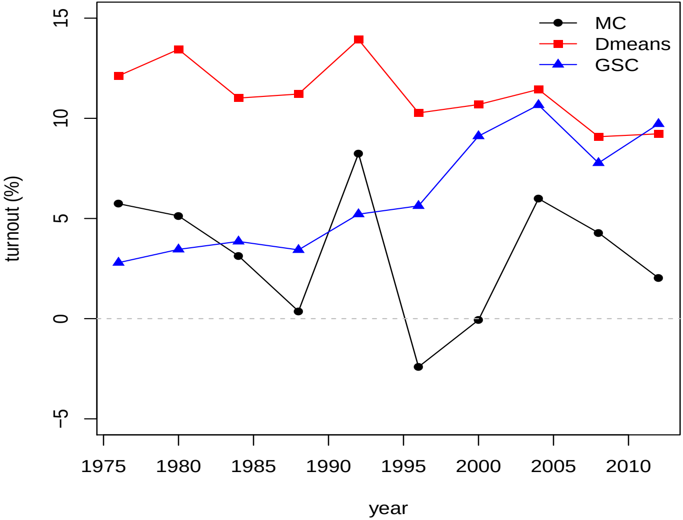
Average Treatment Effect on the Treated

Dmeans	DiD	MC	GSC
9.68	0.87	2.83	5.13
(2.18)	(2.95)	(–)	(2.27)

1. `Dmeans`: difference in means between treated and non-treated states by year
2. `DiD`: difference-in-difference or linear fixed effects method
3. `MC`: nuclear-norm matrix completion method of *Athey, Bayati, Doudchenko, Imbens & Khosravi (2017)*, implemented with the `R` package `MCPanel`
4. `GSC`: generalized synthetic control method of *Xu (2017)* based on linear models with factor structure, implemented with the `R` package `gsynth`

`MC` and `GSC` include additive state fixed effects. Clustered-by-state std. errors in parentheses

Average Treatment Effect on the Treated by Year



Concluding Remarks

- Factor structures are useful devices to incorporate multidimensional unobservables in economic models
- They count with numerous applications in empirical macro and microeconomics
- They have been recently extended to nonlinear models and network data, and combined with modern machine learning methods
- Challenges:
  - Computation and estimation of number of factors in nonlinear models
  - Inference after matrix completion (regularization) with data dependent choice of tuning parameters