Imputation of Counterfactual Outcomes in the Presence of Predictable Errors

Sílvia Gonçalves, McGill University
Serena Ng, Columbia University and NBER

2024 NBER Summer Institute

Outline

1. Introduction and Main Results
2. Linear Prediction: from BLUP to PLUP
3. Imputing Counterfactual Outcomes with PLUP and PUP
4. Unconditional and Conditional Inference

Setup

Pseudo conditional mean function m_{it}

$$\begin{aligned} \text{not treated} \quad Y_{it}(0) &= m_{it} + e_{it}, \quad E[e_{it}] = 0, \quad t = 1, \dots, T. \\ \text{treated} \quad Y_{it}(1) &= m_{it} + e_{it} + \underbrace{\delta_{it}}_{\substack{\text{individual} \\ \text{treatment effect}}} \quad t > T_0. \end{aligned}$$

a. Estimation of m_{it}: regressions, sc, factors/matrix completion
b. Imputation of $Y_{i,T_0+h}(0)$: e_{it} often assume unpredictable.
c. Testing of δ_{iT_0+h}: asymptotic, conformal, subsampling, etc.

In (c), we carefully account for sampling uncertainty from (a), but total imputation error is dominated by out-of-sample error in (b).

Example: $Y_{it}(0) = x_{it}'\beta + e_{it} \Rightarrow \hat{Y}_{iT_0+1}(0) = x_{iT_0+1}'\hat{\beta}_{OLS}$.

$$\text{var}(\hat{e}_{i,T_0+1}) = \sigma_e^2 + x_{T_0+1}'\text{var}(\hat{\beta}_{OLS})x_{T_0+1} \to \sigma_e^2.$$

e_{it} can be correlated across i and t

1. Model misspecification

   i. $Y_{it}(0)$ is AR(2) but we assume an AR(1) model.

   ii. $Y_{it}(0) = \lambda_i F_t + v_{it}$, but we assume $Y_{it}(0) = \lambda_i + F_t + v_{it}$.

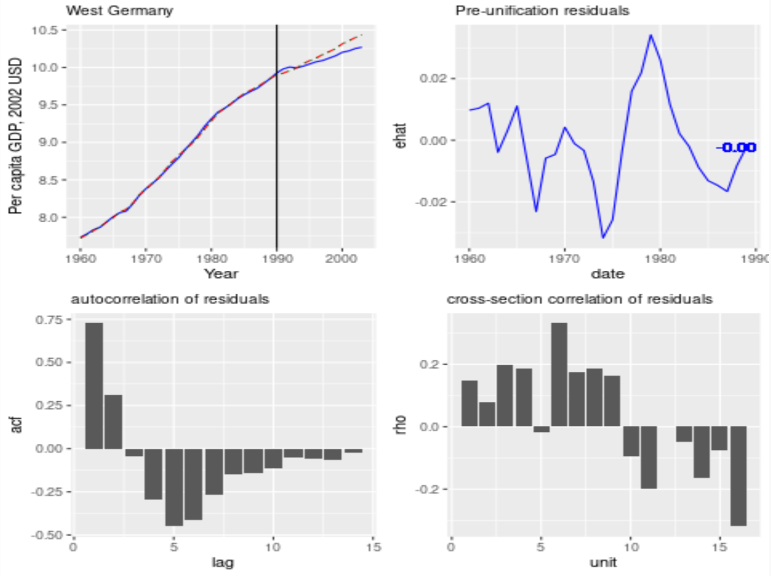
   iii. $Y_{it}(0) = \phi_1 Y_{i,t-1}(0) + v_{it}$ but we let $m_{1t} = \sum_{j=2}^{N+1} \omega_j Y_{jt}$, $e_{1t} = \phi_1 Y_{1,t-1}(0) + v_{1t} - \sum_{j=2}^{N+1} \omega_j Y_{jt}(0)$ is persistent.

2. Genuine correlation

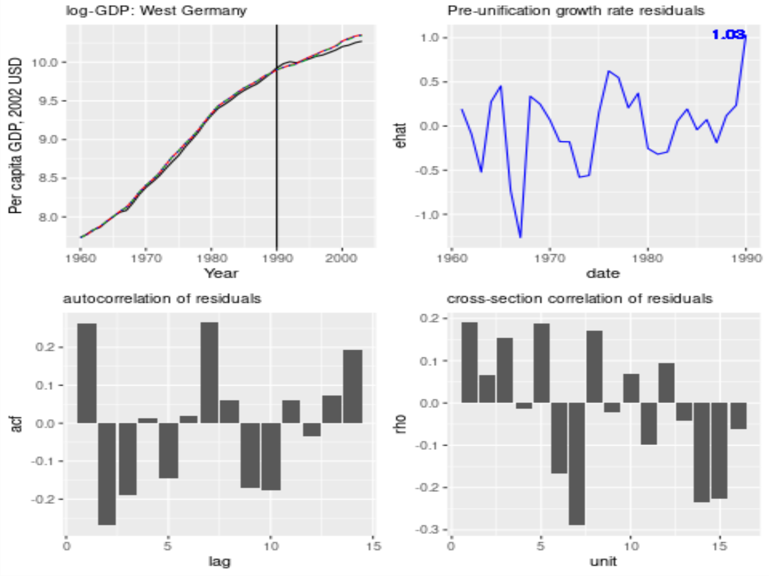
   - $Y_{it}(0) = \Lambda_i' F_t + e_{it}$ and e_{it} is serially or spatially correlated.

3. Temporal aggregation of data.

Germany log GDP imputed from level model



Germany log GDP imputed from growth rate model



This paper: exploit predictability in e_{it}

- Imputation of $Y_{i,T_0+h}(0)$ is a special (out-of-sample) prediction problem where the true values are not sampled.

- Goldberger proposed infeasible BLUP for linear prediction. We derive PLUP and PUP for potential outcomes.
  - i PUP does not require linearity and can be used with any consistent estimator of m_{it}.
  - ii Asymptotic mse expansions for stationary mixing processes show PLUP/PUP have smaller mse than standard predictions.

- Implications of ignoring predictability for inference.
  - i Standard prediction is inefficient but unconditionally unbiased.
  - ii Conditional bias may distort inference.

Linear Model: $y = X\beta + e$, $e \sim (0, \Omega)$ is autocorrelated.

- Goldberger's Best Linear Unbiased Predictor

  $y_{m|T_0} = A'y$ and $E[A(X - x_m)\beta] = 0$ gives

  $$y^*_{m|T_0} = x'_m\beta_{GLS} + \omega'\Omega^{-1}e_{GLS}, \quad \omega = E[e_m e_{GLS}].$$

- first order serial dependence $e_t = \phi_1 e_{t-1} + v_t$:

  $$y^*_{T_0+h|T_0} = x'_{T_0+h}\beta_{GLS} + \phi_1^h e_{GLS,T_0}.$$

  Cochrane-Orcutt (1949) used $\phi_1^h(y_{T_0} - x'_{T_0}\hat{\beta})$

- cross-section correlation: random effects $\Rightarrow$ Empirical Bayes.

Prediction errors: AR(1) case $e_t = \phi_1 e_{t-1} + v_t$

- Standard (OLS) prediction $\hat{y}_{T_0+1|T_0}$ has error

$$\begin{aligned} \hat{e}_{T_0+1|T_0} &= e_{T_0+1} - x'_{T_0+1}(\hat{\beta}_{OLS} - \beta) \\ &= e_{T_0+1} + o_p(1) \end{aligned}$$

- Feasible BLUP $\hat{y}^*_{T_0+1|T_0} = x'_{T_0+1}\hat{\beta}_{GLS} + \hat{\phi}_1 \hat{e}_{GLS,T_0}$ has error

$$\begin{aligned} \hat{e}^*_{T_0+1|T_0} &= v_{T_0+1} - (x_{T_0+1} - \phi_1 x_T)'(\hat{\beta}_{GLS} - \beta) - \left(\hat{\phi}_{1,GLS} - \phi_1\right) \hat{e}_{GLS,T_0} \\ &= v_{T_0+1} + o_p(1) \end{aligned}$$

Feasible BLUP is asymptotically more efficient because

$$\sigma_v^2 \leq \sigma_e^2.$$

- Observation: gains are due to $\phi_1 e^*_{T_0}$, not $\hat{\beta}_{OLS}$ vs $\hat{\beta}_{GLS}$.

From blup to plup (Practical Linear Unbiased Predictor)

- Source of efficiency gains motivates

$$\underbrace{\hat{y}^{+}_{T_0+1|T_0}}_{\text{PLUP}} = \underbrace{x'_{T_0+1}\hat{\beta}_{OLS}}_{\text{OLS prediction}} + \underbrace{\hat{\rho}_1 \hat{e}_{T_0}}_{\text{OLS-based correction}}$$

$\hat{e}_{T_0}$ and $\hat{\rho}_1$ are based on in-sample OLS residuals, $\hat{\rho}_1$ is first order sample autocorrelation coefficient.

- PLUP error $\hat{e}^{+}_{T_0+1|T_0} = \hat{e}_{T_0+1|T_0} - \hat{\rho}_1 \hat{e}_{T_0}$

$$\begin{aligned} \hat{e}^{+}_{T_0+1|T_0} &= e_{T_0+1} - \rho_1 e_{T_0} + o_p(1) \\ &= (\phi_1 - \rho_1) e_{T_0} + v_{T_0+1} + o_p(1). \end{aligned}$$

$e^{+}_{T_0+1|T_0}$ is asymptotically v_{T_0+1} when $\phi_1 = \rho_1$.

Lemma 1: (h=1)

Assumption A1: (a) $E\left|e_t\right|^r < \infty$ for some $r > 2$ and for all t;
(b) $\{e_t\}$ is a strictly stationary strong mixing process.

- Under Assumption A1 such that $\hat{\beta} \xrightarrow{p} \beta$ and $\hat{\rho}_1 \xrightarrow{p} \rho_1 \equiv \frac{\gamma_1}{\gamma_0}$,

$$
\begin{array}{rcl}
\text{(ols)} \quad & \hat{e}_{T_0+1|T_0} & \sim \quad (0, \gamma_0) \\
(\textsc{plup}) \quad & \hat{e}^{+}_{T_0+1|T_0} & \sim \quad (0, \gamma_0(1-\rho_1^2)).
\end{array}
$$

Result also holds if e_t is mixing (not necessarily AR(1)).

Lemma 2 ($h > 1$)

Three predictors:

- standard: $\hat{y}_{T_0+h|T_0} = X'_{T_0+h}\hat{\beta}$
- iterated PLUP : $\hat{y}^{+}_{T_0+h|T_0} = X'_{T_0+h}\hat{\beta} + \hat{\rho}_1^h \hat{e}_{T_0}, \qquad \hat{\rho}_1 = \hat{\gamma}_1/\hat{\gamma}_0.$
- direct PLUP: $\hat{y}_{T_0+h|T_0} = X'_{T_0+h}\hat{\beta} + \hat{\rho}_h \hat{e}_{T_0}, \qquad \hat{\rho}_h = \hat{\gamma}_h/\hat{\gamma}_0$

Remarks: Absent sampling uncertainty:

- all predictions are asymptotically unbiased provided $E[e_t] = 0$.
- asymptotic mse of direct PLUP is never worst.
- additional lags of e_{T_0} may be better, but even one lag helps.
- gains from PLUP $\downarrow$ with h because of mixing (ergodicity).

Table 1: DGP: $y_t = X_t'\beta + e_t, \qquad e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + v_t$.

(T,N)=(50,20)

Unconditional Errors

h	best	noadj	ols	pulpl	plupd	best	noadj	ols	plupl	plupd
	bias					mse				
	(ϕ_1, ϕ_2)=(0.80,0.00)									
1	-0.00	-0.01	-0.00	-0.00	-0.00	0.05	0.14	0.14	0.05	0.05
2	0.00	-0.00	0.00	0.00	0.00	0.05	0.14	0.14	0.08	0.08
5	0.00	0.00	0.00	0.00	0.00	0.05	0.13	0.14	0.12	0.12
10	0.00	-0.00	-0.00	-0.00	0.00	0.05	0.14	0.14	0.14	0.14
avg	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.07	0.08	0.06	0.06
	(ϕ_1, ϕ_2)=(1.30,-0.40)									
1	-0.00	-0.01	-0.00	-0.00	-0.00	0.05	0.43	0.43	0.06	0.06
2	0.00	-0.00	-0.00	-0.00	-0.00	0.05	0.43	0.43	0.16	0.16
5	0.00	-0.00	0.00	-0.00	0.00	0.05	0.42	0.43	0.37	0.35
10	0.00	-0.01	-0.00	-0.00	-0.00	0.05	0.43	0.44	0.48	0.44
avg	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.28	0.29	0.21	0.20

(T,N)=(50,20)

Conditional on $(e_{T_0-1}, e_{T_0}) = (0.5, 1.0)$

h	best	noadj	ols	pulpl	plupd	best	noadj	ols	plupl	plupd
	bias					mse				
	(ϕ_1, ϕ_2)=(0.80,0.00)									
1	-0.00	0.80	0.79	0.01	0.01	0.05	0.68	0.68	0.05	0.05
2	0.00	0.64	0.64	0.01	0.02	0.05	0.49	0.49	0.09	0.09
5	0.00	0.33	0.33	0.01	0.02	0.05	0.23	0.23	0.13	0.13
10	0.00	0.11	0.10	0.00	0.02	0.05	0.15	0.15	0.14	0.15
avg	-0.00	0.36	0.35	0.01	0.02	0.00	0.18	0.18	0.06	0.06
	(ϕ_1, ϕ_2)=(1.30,-0.40)									
1	-0.00	1.10	1.09	0.18	0.18	0.05	1.25	1.26	0.09	0.09
2	0.00	1.03	1.03	0.18	0.24	0.05	1.19	1.20	0.17	0.20
5	0.00	0.62	0.62	-0.04	0.20	0.05	0.72	0.73	0.34	0.38
10	0.00	0.21	0.21	-0.23	0.08	0.05	0.47	0.48	0.49	0.45
avg	-0.00	0.61	0.61	-0.04	0.17	0.00	0.56	0.56	0.19	0.22

Outline

From Prediction to Imputation

- Goldberger's BLUP is for linear prediction of data not yet observed but eventually available.
- Chow-Lin: interpolate (impute) missing values of mixed frequency data that may or may not be observed.
- For $i \in [1, N_1]$, we want to impute missing Y_{i,T_0+h} from
    - T_0 outcomes for N_1 treated units
    - T outcomes for $N - N_1$ control units.

Setup for Potential Outcomes

$$\underbrace{\mathcal{Y}}_{n\times 1}(0) = \begin{pmatrix} Y_{1,1:T_0}(0) \\ \vdots \\ Y_{N,1:T_0}(0) \\ Y_{N_1+1,T_0+1:T}(0) \\ \vdots \\ Y_{N,T_0+1:T}(0) \end{pmatrix} = \begin{pmatrix} \underbrace{Y^{pre}(0)}_{(NT_0\times 1)} \\ \\ \underbrace{Y^{post}(0)}_{(N-N_1)T_1\times 1} \end{pmatrix} = \mathcal{M} + \mathcal{E}$$

$$\mathcal{E} = \begin{pmatrix} \mathcal{E}^{pre} \\ \mathcal{E}^{post} \end{pmatrix} \sim (0, \Gamma) \qquad \Gamma = E[\mathcal{E}\mathcal{E}'] = \begin{pmatrix} \Gamma_{pre,pre} & \Gamma_{pre,post} \\ \Gamma_{post,pre} & \Gamma_{post,post} \end{pmatrix}$$

Let $n = NT - N_1 T_1$. The $n \times n$ covariance matrix Γ depends on

a. $\Sigma = E[e_t e_t']$, the $N \times N$ covariance matrix of $e_t = (e_{1t}', e_{0t}')'$
b. $\Omega_i = E[e_i e_i']$, the $T_{0_i} \times T_{0_i}$ time series covariance of unit i.

Assumptions

A1 strict stationarity, strong mixing.

A2 $\hat{m}_{it} = \hat{Y}_{it}(0)$ is any consistent estimator such that
$\hat{m}_{i,T_0+h} - m_{i,T_0+h} = o_p(1);\ T_0^{-1} \sum_{t=1}^{T_0} (\hat{m}_{i,t} - m_{i,t})^2 = o_p(1)$.

Under A1 and A2, the standard imputation has error

$$\hat{\delta}_{i,T_0+h} - \delta_{i,T_0+h} = \underbrace{(m_{i,T_0+h} - \hat{m}_{i,T_0+h})}_{(i)} + \underbrace{e_{i,T_0+h}}_{(ii)}.$$

(i) is estimator dependent, may need White/HAC, but is $o_p(1)$.

(ii) is not captured by model, dominates (i) asymptotically.

blup for Counterfactual Outcome

Re-doing the Goldberger problem with $\mathcal{M} = \mathcal{X}\beta$ gives

BLUP: adjusted counterfactual outcome

$$m^{+}_{i,T_0+h} = X'_{T_0+h}\beta_{GLS} + \omega'_{ih}\Gamma^{-1}\mathcal{E}_{GLS}.$$

BLUP is infeasible. But under Assumption A, we have residuals

- $\hat{e}_{jt} = Y_{jt}(0) - \hat{m}_{jt}$ for any j not treated, $t = 1, \ldots, T$.
- $\hat{e}_{it} = Y_{it}(0) - \hat{m}_{it}$, any i treated, $t = 1, \ldots, T_0$.

We can construct PLUP that is asymptotically equivalent to BLUP.

Practical Unbiased Predictor, pup: $\hat{Y}^+_{it}(0) = \hat{m}^+_{it}$

* Correction can be done even for non-linear imputation models.

Let $\hat{m}_{i,T_0+h}$ be any standard imputation satisfying A1 and A2:

- time series PUP: $\hat{m}^+_{i,T_0+h} = \hat{m}_{i,T_0+h} + \sum_{k=0}^{p} \hat{\rho}_i^k \hat{e}_{i,T_0-k}$.
- cross-section PUP $\hat{m}^+_{i,T_0+h} = \hat{m}_{i,T_0+h} + \sum_{j>N_1} \hat{\theta}_{i,j} \hat{e}_{j,T_0+h}$.
- time series and cross-section PUP:

$$\hat{m}^+_{i,T_0+h} = \hat{m}_{i,T_0+h} + \sum_{s=0}^{p_i} \rho_{is} \hat{e}_{i,T_0-s} + \sum_{j \neq i}^{N} \sum_{s=-p_j}^{p_j} \theta_{ij,s} \hat{e}_{j,T_0+h-s}.$$

Related Work

- PUP can use any consistent $\hat{m}_{it}$ already developed. No need to abandon static procedures.

- PUP type corrections have been suggested.
  - Chernozhukov-Wuthrith-Zhu (2021): suggest to add $\rho(L)\epsilon_{t-1}$, no mse analysis, limited proof of concept.
  - Fan-Masini-Medeiros (2022): mutually correlated idiosyncratic errors $Y_{it}(0) = \gamma_i' W_{it} + \lambda_i' F_t + \theta_1' \hat{\epsilon}_{-i,t} + v_{it}$ by LASSO.
  - Ferman (2023): DiD with omitted spatial correlation.

We provide analytical motivation and study their properties.

Simulations, $Y_{it}(0) = F_t'\Lambda_i + e_{it}$, serial correlation

h	best	noadj	pca	pupl	pupd	best	noadj	pca	pupl	pupd
	$e_{1t} = 0.6e_{1t-1} + \epsilon_{1t}$									
	Unconditional bias of $\hat{\delta}_{i,T_0+h}$					mse				
1	0.00	-0.00	0.00	0.01	0.01	0.25	0.37	0.43	0.33	0.33
2	-0.01	-0.01	-0.00	-0.00	-0.00	0.33	0.37	0.45	0.42	0.42
5	-0.00	-0.00	0.01	0.01	0.01	0.37	0.39	0.48	0.47	0.48
10	-0.01	-0.01	-0.00	-0.00	-0.00	0.39	0.40	0.50	0.50	0.51
avg	-0.00	-0.00	0.00	0.00	0.00	0.12	0.13	0.18	0.17	0.17
	$e_{1t} = 0.6e_{1t-1} + \epsilon_{1t},\ e_{T_0}$=1.00									
	Conditional bias of $\hat{\delta}_{i,T_0+h}$					mse				
1	0.00	0.60	0.56	0.15	0.15	0.25	0.61	0.66	0.36	0.36
2	-0.01	0.35	0.32	0.12	0.14	0.33	0.45	0.53	0.45	0.47
5	-0.00	0.21	0.19	0.09	0.13	0.37	0.42	0.51	0.48	0.52
10	-0.01	0.12	0.09	0.04	0.09	0.39	0.40	0.51	0.50	0.54
avg	-0.00	0.15	0.12	0.04	0.09	0.12	0.14	0.19	0.17	0.19

Simulations, $Y_{it}(0) = F_t' \Lambda_i + e_{it}$, cross-section correlation

h	best	noadj	pca	pupl	pupd	best	noadj	pca	pupl	pupd
	$e_{1t} = 0.5 e_{2t} + \epsilon_{1t}$									
	Unconditonal bias of $\hat{\hat{\delta}}_{i,T_0+h}$					mse				
1	0.00	-0.00	0.00	0.00	0.00	0.15	0.20	0.33	0.28	0.28
2	-0.01	-0.01	-0.00	-0.01	-0.01	0.15	0.21	0.35	0.29	0.29
5	0.00	-0.00	0.01	0.01	0.01	0.15	0.21	0.35	0.30	0.30
10	-0.01	-0.01	-0.01	0.00	0.00	0.15	0.21	0.34	0.29	0.29
avg	-0.00	-0.00	0.00	0.01	0.01	0.01	0.02	0.05	0.04	0.04
	$e_{1t} = 0.5 e_{2t} + \epsilon_{1t},\ e_{2T_0+1} = 0.27539$									
	Conditional bias of $\hat{\hat{\delta}}_{i,T_0+h}$					mse				
1	0.00	0.14	0.14	0.04	0.04	0.15	0.17	0.29	0.27	0.27
2	-0.01	0.49	0.49	0.12	0.12	0.15	0.38	0.52	0.35	0.35
5	0.00	0.47	0.48	0.13	0.13	0.15	0.37	0.52	0.35	0.35
10	-0.01	-0.45	-0.44	-0.14	-0.14	0.15	0.35	0.48	0.34	0.34
avg	-0.00	-0.12	-0.12	-0.06	-0.06	0.01	0.03	0.06	0.05	0.05

Inference

$$\widehat{\delta}_{it} - \delta_{it} = o_p(1) \text{ estimation error } + e_{it}.$$

Suppose e_{it} is normal. Given point estimate $\hat{\delta}_{it}$ with variance σ^2_{δ}, we can construction a prediction interval of the form

$$\text{point estimate } \pm 2\text{s.e.}$$

We can choose from

- Two predictors (standard vs PUP);
- Two types of inference (unconditional and conditional).

Unconditional Inference

- Unconditional inference takes average over draws of e_{T_0}. Useful for evaluating procedures.

- Proposition:

  - Let $\{e_t\}$ be a stationary mixing process. Under normality, the standard and PUP predictions all have correct unconditional coverage, ie, as $T_0 \to \infty$,

$$P\left(-z_{1-\alpha/2} \leq \frac{\hat{\delta}_{i,T_0+1} - \delta_{i,T_0+1}}{\hat{\sigma}_{e,i}} \leq z_{1-\alpha/2}\right) = \Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2}).$$

  - but the PUP intervals are shorter.

Conditional inference based on standard imputation

Suppose $e_{it} = \phi_i e_{it-1} + v_{it}$ where $v_{it} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_{v,i}\right)$.

Problem at $h = 1$:

$$\frac{e_{i,T_0+1}}{\sigma_{e,i}} \sim \mathcal{N}(\phi_i \frac{e_{i,T_0}}{\sigma_{e,i}}, \frac{\sigma^2_{v,i}}{\sigma^2_{e,i}}) \neq \mathcal{N}(0,1).$$

$$\begin{aligned} & P\left(-z_{1-\alpha/2} \leq \frac{\hat{\delta}_{i,T_0+1} - \delta_{i,T_0+1}}{\hat{\sigma}_{e,i}} \leq z_{1-\alpha/2} \middle| e_{i,T_0}\right) \\ =& \Phi\left(\frac{-\phi_i}{\sigma_{v,i}} e_{i,T_0} + z_{1-\alpha/2}\frac{\sigma_{e,i}}{\sigma_{v,i}}\right) - \Phi\left(\frac{-\phi_i}{\sigma_{v,i}} e_{i,T_0} - z_{1-\alpha/2}\frac{\sigma_{e,i}}{\sigma_{v,i}}\right). \end{aligned}$$

Suppose $e_{1,t} = \theta_1' e_{2:N,t} + v_{1t}$ is cross-sectionally correlated.

$$e_{1,T_0+1} | e_{2:N,T_0+1} \sim \mathcal{N}(\theta_1' e_{2:N,T_0+1}, \sigma_{v,1}^2).$$

$$\begin{aligned} & P\left(-z_{1-\alpha/2} \leq \frac{\hat{\hat{\delta}}_{1,T_0+1} - \delta_{1,T_0+1}}{\hat{\sigma}_{e,1}} \leq z_{1-\alpha/2} \Big| e_{2:N,T_0+1} \right) \\ =& \Phi\left(\frac{-\theta_1'}{\sigma_{v,1}} e_{2:N,T_0+1} + z_{1-\alpha/2} \frac{\sigma_{e,1}}{\sigma_{v,1}} \right) - \Phi\left(\frac{-\theta_1'}{\sigma_{v,1}} e_{2:N,T_0+1} - z_{1-\alpha/2} \frac{\sigma_{e,1}}{\sigma_{v,1}} \right). \end{aligned}$$

Conditionally biased prediction distorts inference

$$\Phi(\text{bias}_1 + \text{bias}_2\ z_{1-\alpha/2}) \neq \Phi(z_{1-\alpha/2}).$$

Conditional bias of standard prediction I

For serially correlated errors, (analytical) bias$_1$= $\phi_i^h/\omega_{h,i}$

$$(e_{T_0}, \sigma_v) = (-2.0, 0.5)$$

h	coverage	bias$_1$	coverage	bias$_1$
	AR(1) $\phi_1 = 0.8$		MA(1): $\theta_1 = 0.8$	
1	0.84	-2.26	0.58	-1.77
2	0.87	-1.41	0.95	-0.00
3	0.90	-1.01	0.95	-0.00
4	0.92	-0.76	0.95	-0.00
5	0.93	-0.59	0.95	-0.00

Conditional bias of standard prediction II

For cross-correlated errors, (analytical) $\text{bias}_1 = \theta_1' e_{-1,T_0+1}/\sigma_{v,1}$.

Example: $e_{1,T_0+h} = \theta_{12} e_{2,T_0+h} + v_{1,T_0+h}$, $(\Sigma_{11}, \Sigma_{00}) = (0.5, 0.841)$.

		$\Sigma_{01} = 0.613$		$\Sigma_{01} = -0.613$	
h	e_{2,T_0+h}	coverage	bias_1	coverage	bias_1
1	0.68	0.63	-1.64	0.89	-0.70
2	-0.83	0.81	1.07	0.86	0.85
3	-0.92	0.93	-0.38	0.84	0.95
4	0.09	0.95	0.17	0.95	-0.09
5	0.86	0.94	0.32	0.86	-0.89

PUP Coverage

For $h = 1$, conditional PLUP coverage is

$$P\left(-z_{1-\alpha/2} \leq \frac{(\hat{\delta}^{+}_{i,T_0+1} - \delta_{i,T_0+1})}{\hat{\sigma}^{+}_{e,i,1}} \leq z_{1-\alpha/2} \middle| e_{i,T_0}\right) \approx 1 - \alpha + o_p(1)$$

PUP conditional coverage is asymptotically accurate: $\forall h \geq 1$.

Coverage, DGP 1: $y_t = X_t'\beta + e_t$ with $e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + v_t$

$$(T, N) = (50, 20)$$

h	best	noadj	ols	pup	best	noadj	ols	pup
	Unconditional				Conditional on e_{T_0}, e_{T_0-1}			
	(ϕ_1, ϕ_2)=(0.80,0.00)							
1	0.96	0.95	0.95	0.95	0.96	0.40	0.41	0.95
2	0.95	0.95	0.95	0.95	0.95	0.64	0.63	0.95
5	0.96	0.95	0.95	0.95	0.96	0.88	0.88	0.95
10	0.95	0.95	0.95	0.94	0.95	0.94	0.94	0.94
avg	0.95	0.99	0.99	0.94	0.96	0.94	0.93	0.94
	(ϕ_1, ϕ_2)=(1.30,-0.40)							
1	0.96	0.95	0.95	0.95	0.96	0.79	0.77	0.92
2	0.95	0.95	0.95	0.95	0.95	0.76	0.75	0.94
5	0.96	0.95	0.95	0.95	0.96	0.87	0.87	0.96
10	0.95	0.95	0.94	0.95	0.95	0.94	0.94	0.94
avg	0.95	0.98	0.98	0.94	0.96	0.94	0.93	0.96

Coverage, DGP 2a: $X_{it} = \Lambda_i' F_t + e_{it}$ with $e_{1t} = \rho_1 e_{1t-1} + v_{1t}$

h	best	noadj	pca	pup	best	noadj	pca	pup
	Unconditional				Conditional on e_{1,T_0}			
	(T,N)=(50,20)							
1	0.95	0.94	0.92	0.93	0.96	0.87	0.86	0.92
2	0.91	0.94	0.91	0.91	0.92	0.92	0.90	0.91
5	0.88	0.93	0.90	0.88	0.90	0.94	0.91	0.89
10	0.89	0.94	0.90	0.84	0.91	0.94	0.91	0.85
avg	0.98	0.85	0.75	0.77	0.98	0.83	0.73	0.76
	(T,N)=(200,50)							
1	0.95	0.95	0.94	0.95	0.96	0.90	0.89	0.95
2	0.91	0.95	0.94	0.94	0.91	0.92	0.92	0.94
5	0.89	0.95	0.94	0.93	0.89	0.95	0.94	0.94
10	0.89	0.95	0.94	0.94	0.90	0.95	0.94	0.94
avg	0.99	0.93	0.92	0.92	0.99	0.92	0.90	0.92

- Gains are larger at $h = 1$ and when T_0 is large;
- Coverage of $\hat{\delta}_{it}$ not informative about $\bar{\delta}_i$.

Coverage, DGP 2b: $X_{it} = \Lambda_i' F_t + e_{it}$ with $e_{1t} = \theta_{12} e_{2,t} + v_{1t}$

h	best	noadj	pca	pup	best	noadj	pca	pup
	Unconditional				Conditional on $e_{2,T_0+1:T}$			
	(T,N)=(50,20)							
1	0.95	0.95	0.94	0.93	0.95	0.97	0.95	0.93
2	0.95	0.95	0.93	0.93	0.95	0.86	0.87	0.90
5	0.95	0.94	0.93	0.91	0.95	0.91	0.91	0.90
10	0.95	0.95	0.93	0.88	0.95	0.98	0.96	0.89
avg	0.88	0.88	0.79	0.78	0.88	0.81	0.75	0.77
	(T,N)=(200,50)							
1	0.95	0.96	0.95	0.95	0.95	0.94	0.94	0.95
2	0.95	0.95	0.94	0.94	0.95	0.94	0.93	0.94
5	0.95	0.95	0.95	0.95	0.95	0.75	0.78	0.93
10	0.95	0.95	0.95	0.94	0.95	0.86	0.87	0.93
avg	0.94	0.94	0.92	0.92	0.94	0.92	0.91	0.92

Comments on PUP in practice

1. PLUP gains specific to estimator and hypothesis. Details
   - coverage of $\hat{\delta}_{it}$ not good guide for $\bar{\delta}_i$ or $\bar{\delta}_t$.
   - At least for FBI, Gaussian distribution for different reasons.

2. A necessary condition for PUP improvements is predictability in the errors. We can check using LM test

$$\hat{e}_{it} = X_{it}'\delta_0 + \sum_{s=0}^{p_i} \delta_{i,s}\hat{e}_{i,t-s} + \sum_{j=1}^{N} \sum_{s=-p_j}^{p_j} \delta_{j,s}\hat{e}_{j,t-s}.$$

   - Rejection of H_0 is not sufficient for improvements.
   - Correction involves $\hat{\rho}_i^h \hat{e}_{i,T_0}$ or $\theta_i' \hat{e}_{-i,T_0+h}$.
   - The relevant $\hat{e}$'s must be non-zero, but we can check this!

Comments on Inference

1. Many procedures for unconditional inference when e_{it} is unpredictable, but developed for different estimators and different object of interest. How to compare methods?

2. Limited work on conditional inference,

    - Theory allows e_{it} to be dependent, consider this possibility.
    - modeling e_{it} improves point prediction, not just efficiency.

Summary

- Robust s.e for $\hat{m}_{it}$ useful, but prediction error variance is asymptotically dominated by $\text{var}(e_{i,T_0+h})$.
- Ignoring time and cross correlations associated with e_{i,T_0+h} can give inefficient and conditionally biased predictions, as well as misleading inference.
- PUP provides a simple way to account for correlation without knowing the true error structure. It can be used in conjunction with many existing estimators.

Summary

- Robust s.e for $\hat{m}_{it}$ useful, but prediction error variance is asymptotically dominated by $\text{var}(e_{i,T_0+h})$.

- Ignoring time and cross correlations associated with e_{i,T_0+h} can give inefficient and conditionally biased predictions, as well as misleading inference.

- PUP provides a simple way to account for correlation without knowing the true error structure. It can be used in conjunction with many existing estimators.

THANK YOU

Prediction vs Confidence Intervals

$$H_0 : \quad \frac{1}{T_1} \sum_{t=T_0+1}^{T} \delta_{it} = \Delta^0_{i,T_1}.$$

- If Δ_{i,T_1} is random for fixed T_1: construct prediction interval

$$\hat{\Delta}_i - \Delta_{i,T_1} = \left(\frac{1}{T_1} \sum_{t=T_0+1}^{T} m_{it} - \hat{m}_{it} \right) + \frac{1}{T_1} \sum_{t=T_0+1}^{T} e_{it}$$

- $\Delta_{i,\infty}$ is non-random: construct a confidence interval

$$\hat{\Delta}_i - \Delta_{i,\infty} = \left(\frac{1}{T_1} \sum_{t=T_0+1}^{T} m_{it} - \hat{m}_{it} \right) + \left(\frac{1}{T_1} \sum_{t=T_0+1}^{T} e_{it} + \delta_{it} - E[\delta_{it}] \right).$$

FBI Imputation Error, Bai and Ng (2021)

$$\widehat{\delta}_{it} - \delta_{it} = \underbrace{x_{it}'(\beta - \hat{\beta})}_{(i.a)} \underbrace{- F_t' \left(\frac{F'F}{T_0} \right)^{-1} \mathbf{B}_F \frac{1}{T_0} \left(\sum_{s=1}^{T_0} F_s \epsilon_{is} \right)}_{(i.b)}$$

$$\underbrace{- \Lambda_i' \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \mathbf{B}_\Lambda \frac{1}{N_0} \sum_{k=1}^{N_0} \Lambda_k \epsilon_{kt}}_{(i.c)} + \epsilon_{it} + O_p\left(\frac{1}{\min(N_0, T_0)}\right).$$

- error (i.a) is $O_p(\frac{1}{\sqrt{T_0}})$
- error (i.b) is error from estimating F: $O_p(\frac{1}{\sqrt{T_0}})$
- error (i.c) is error from estimating Λ: $O_p(\frac{1}{\sqrt{N_0}})$.

ϵ_{it} dominates error in estimate of individual treatment effect

Inference of Average Treatment Effect over Time

Recall: $\hat{\delta}_{i,T_0+h} - \delta_{i,T_0+h} = (m_{i,T_0+h} - \hat{m}_{i,T_0+h}) + e_{i,T_0+h}$.

$$\hat{\delta}_{i,T_1} = \frac{1}{T_1}\sum_{h=1}^{T_1}\hat{\delta}_{i,T_0+h}$$

$$= \frac{1}{T_1}\sum_{h=1}^{T_1} F'_{T_0+h}\left(\frac{F'F}{T_0}\right)^{-1}\mathbf{B}_F\frac{1}{T_0}\left(\sum_{s=1}^{T_0}F_s e_{is}\right) + \frac{1}{T_1}\sum_{h=1}^{T_1} e_{i,T_0+h} + c$$

Let $\mathbb{V}_{\hat{\Delta}_{i,T_1}} = \frac{\min(T_0,T_1)}{T_0}\bar{F}'\left(\frac{F'F}{T_0}\right)^{-1}\Phi_j\left(\frac{F'F}{T_0}\right)^{-1}\bar{F} + \frac{\min(T_0,T_1)}{T_1}\sigma_{ei}^2$.
Then

$$\min(\sqrt{T_0},\sqrt{T_1})\left(\frac{\hat{\Delta}_{i,T_1} - \Delta_{i,T_1}}{\sqrt{\mathbb{V}_{\hat{\Delta}_i}}}\right)\xrightarrow{d} N(0,1)$$

Back