Least Squares Estimation of Dynamic Games with Time Varying Unobserved Heterogeneity

A. Gupta (Essex/Queens) L. Puppi (USP)
F. Sanches (EESP-FGV) S. Srisuma (NUS)

DSE, London, 2025

Overview

- We propose a new estimator for dynamic discrete decision problems and games with time varying unobserved heterogeneity
- Our estimator is easy to compute and does not rely on instruments or finite dependence assumptions
- Main idea is to add another step to an existing 2-step approach by way of k-means clustering on CCPs
- This talk focuses on a single agent case for simplicity, simulation studies done on a game

Related Literature

- Model builds on Rust (1987), Aguirregabiria and Mira (2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)
- Many options for two-step estimators without UH: Hotz and Miller (1993), Hotz, Miller, Sanders and Smith (1994), Aguirregabiria and Mira (2002,2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)
- UH is often introduced to be time invariant (Kasahara and Shimotsu (2009), Aguirregabiria et al.) but not always, see Arcidiacono and Miller (2011), Hu and Shum (2012), Kaloupstidi, Scott, and Souza-Rodrigues (2021)

What We Do

- In a large N and T framework, consider models with unobserved time varying states
- Estimate CCP for each time period and do k-means clustering (cf. Pollard (1981,1982)), which identifies unobserved states (Bonhomme and Manresa (2015))
- A 2-step estimator then follows, use Sanches, Silva, and Srisuma (2016) for estimation as it is generally computationally efficient relative to Pesendorfer and Schmidt-Dengler (2008)

Notation and K-means Clustering

Variables:

- Action $a_{nt} \in A = \{0, 1, \ldots, K\}$
- Observed state $x_{nt} \in X = \{1, \ldots, J\}$
- Unobserved state $d_t \in D = \{1, \ldots, L\}$

Given $\{(a_{nt}, x_{nt})\}_{n=1,t=1}^{N,T}$, let $\widehat{p}_{Nt}$ be a vector in $[0,1]^{KJ}$ that estimates p_t, vector of CCPs, $\{\Pr[a_{nt} = a | x_{nt} = x, d_t]\}$ for $a > 0$ and all x. We perform k-means clustering by choosing

$\mathbf{p} \in [0,1]^{KJL}$ to minimize:

$$
\begin{aligned}
W_{NT}(\mathbf{p}) &= \frac{1}{T}\sum_{t=1}^{T} \phi\left(\widehat{p}_{Nt}, \mathbf{p}\right), \text{ where} \\
\phi(p, \mathbf{p}) &\equiv \min_{1 \leq \ell \leq L} \|p - p_\ell\|^2 .
\end{aligned}
$$

Regularity Conditions

Assumption A. $\{(a_{nt}, x_{nt}, d_t)\}_{n=1,t=1}^{N,T}$ satisfies:

(i) for all n, $\{(a_{nt}, x_{nt}, d_t)\}_{t=1}^{T}$ is a strictly stationary α-mixing process such that

$$\sum_{t=1}^{\infty} t^{\lambda-2} \alpha(t)^{\frac{\gamma}{\lambda+\gamma}} < \infty,$$

for some $\gamma > 0$ and even $\lambda \geq 2$, and for λ and γ, the bracketing numbers of F satisfy

$$\int_0^1 x^{-\frac{\gamma}{2+\gamma}} \left(N(x, \mathcal{F})\right)^{1/\lambda} dx < \infty,$$

where $\mathcal{F}$ is some class of functions (Pollard (1982), Andrews and Pollard (1994));

(ii) for all t, $\{(a_{nt}, x_{nt})\}_{n=1}^{N}$ is an i.i.d. sequence conditional on d_t;

(iii) $p_t = \sum_{\ell} p_\ell^0 \mathbf{1}\left[d_t = \ell\right]$ such that $p_\ell^0 \in (0,1)^{KJ}$ and $\left\| p_\ell^0 - p_{\ell'}^0 \right\| \neq 0$ for all ℓ, ℓ'.

Properties of CCP Estimator

Assumption B. For all t, N, $\widehat{p}_{Nt} = \sum_{\ell} \left(p_{\ell}^{0} + \frac{c_{\ell Nt}}{\sqrt{N}} + r_{\ell Nt} \right) \mathbf{1} \left[d_t = \ell \right]$ such that:

(i) For all ℓ, T, there exists some finite K where $\Pr \left[\| c_{\ell Nt} \| > \epsilon \sqrt{N} \mid d_t = \ell \right] \leq \exp \left(-NK\epsilon^2 \right)$ for all $N, t \leq T$ and $\epsilon > 0$;

(ii) For all ℓ, T, $\sup_{t \leq T} \left| \Pr \left[c_{\ell Nt} \in U \mid d_t = \ell \right] - \Pr \left[z_{\ell} \in U \right] \right| = O \left(\frac{1}{\sqrt{N}} \right)$ for all convex U where z_{ℓ} is some vector of jointly normal variables with zero mean;

(iii) For all ℓ, T, $\sup_{t \leq T} E \| r_{\ell Nt} \| = O(N)$.

Proposition 1. Under Assumption A, the frequency estimator of the CCP satisfies Assumption B.

Properties of K-means Estimator

Proposition 2. Under Assumptions A and B, suppose $\log T = o(N)$,

$$\sqrt{T}\left(\widehat{\mathbf{p}}_{NT} - \mathbf{p}_N^0\right) \xrightarrow{d} \mathrm{N}\left(0, \Gamma_N^{-1} V_N \Gamma_N^{-1}\right) \text{ as } T \to \infty,$$

w.p.a. 1 as $N \to \infty$, for some PD matrices Γ_N and V_N, and $\left\|\mathbf{p}_N^0 - \mathbf{p}^0\right\| = o(1)$.

Intuition is, for $\mathbf{p}_{NT}$ near $\mathbf{p}_N^0$:

$$\begin{aligned} W_{NT}(\mathbf{p}_{NT}) &= W_{NT}\left(\mathbf{p}_N^0\right) - \frac{1}{\sqrt{T}} Z_{NT}^{\intercal}\left(\mathbf{p}_{NT} - \mathbf{p}_N^0\right) \\ &\quad + \frac{1}{2}\left(\mathbf{p}_{NT} - \mathbf{p}_N^0\right)^{\intercal} \Gamma_N \left(\mathbf{p}_{NT} - \mathbf{p}_N^0\right) \\ &\quad + o_p\left(\frac{\left\|\mathbf{p}_{NT} - \mathbf{p}_N^0\right\|}{\sqrt{T}} + \left\|\mathbf{p}_{NT} - \mathbf{p}_N^0\right\|^2\right), \\ \widehat{\mathbf{p}}_{NT} &\approx \mathbf{p}_N^0 + \frac{1}{\sqrt{T}} \Gamma_N^{-1} Z_{NT} \text{ w.p.a. } 1 \text{ as } N \to \infty. \end{aligned}$$

Properties of Estimated Unobserved States

K-means clustering gives $\left\{\widehat{d}_t\right\}_{t=1}^{T}$. Let $\lambda_\ell^0 := \Pr\left[d_t = \ell\right]$, $\lambda_{\ell\ell'}^0 := \Pr\left[d_t = \ell, d_{t+1} = \ell'\right]$, and $\lambda_{\ell'|\ell}^0 := \Pr\left[d_{t+1} = \ell' | d_t = \ell\right]$.

Lemma 3. Under Assumptions A and B, as $N, T \to \infty$, if $\log T = o\left(N\right)$, then $\widehat{\lambda}_\ell = \lambda_\ell^0 + o_p\left(1\right)$, $\widehat{\lambda}_{\ell\ell'} = \lambda_{\ell\ell'}^0 + o_p\left(1\right)$, and $\widehat{\lambda}_{\ell'|\ell}^0 = \lambda_{\ell'|\ell}^0 + o_p\left(1\right)$ for all ℓ, ℓ'.

Let $\left(\widehat{\lambda}_M, \widehat{\lambda}_J, \widehat{\lambda}_C\right)$ and $\left(\widetilde{\lambda}_M, \widetilde{\lambda}_J, \widetilde{\lambda}_C\right)$ be respective vectors of frequency esimators estimating these probabilities with $\widehat{d}_t$ and d_t.

Lemma 4. Under Assumptions A and B, as $N, T \to \infty$, if $\sqrt{T} = o\left(N\right)$, then $\left\|\widehat{\lambda}_M - \widetilde{\lambda}_M\right\| = o_p\left(\frac{1}{\sqrt{T}}\right)$, $\left\|\widehat{\lambda}_J - \widetilde{\lambda}_J\right\| = o_p\left(\frac{1}{\sqrt{T}}\right)$, and $\left\|\widehat{\lambda}_C - \widetilde{\lambda}_C\right\| = o_p\left(\frac{1}{\sqrt{T}}\right)$.

Markov Decision Process

Assumption M.
(i) For all a, x, d, ε and some $\theta \in \Theta \subseteq R^p$,

$$\begin{aligned} u(a, x, d, \varepsilon) &= \pi_\theta(a, x, d) + \varepsilon(a), \\ u_\theta(a, x, d, \varepsilon) &= \theta^\top \pi(a, x, d), \end{aligned}$$

where Θ is compact and π known;

(ii) State transition satisfies:

$$F\left(x', d', \varepsilon' | x, d, \varepsilon, a\right) = Q\left(\varepsilon'\right) H\left(d' | d\right) G\left(x' | x, a\right),$$

where Q, H, and G are respectively the CDFs of ε_{nt}, $d_{t+1} | d_t$, and $x_{nt+1} | x_{nt}, a_{nt}$;
(iii) ε_{nt} has a known and absolutely continuous distribution with bounded density supported on R^{K+1};
(iv) The value of β is known.

Primitives of the model are (θ, β, F).

ALS - Main Idea

Consider a model of *binary actions* based on

$$a_t(\theta) = \mathbf{1}\left[v_\theta(x_t) - \varepsilon_t \geq 0\right] \quad \text{for} \quad \theta \in \Theta \subset \mathbb{R}^p,$$

where x_t and ε_t are obs and unobs *state variables* s.t. v_θ and distribution of $\varepsilon_t | x_t$ are known (say, CDF Q).

For all x,

$$P_\theta(x) := \Pr\left[a_t(\theta) = 1 | x_t = x\right] = Q\left(v_\theta(x)\right).$$

Suppose we observe a random sample $\{a_t, x_t\}_{t=1}^T$ where $a_t = a_t(\theta_0)$ for some $\theta_0 \in \Theta$. θ_0 can be estimated by minimizing the distance between $P(\cdot) := \Pr\left[a_t = 1 | x_t = \cdot\right]$ and $P_\theta(\cdot)$.

ALS - Main Idea

- When x_t is discrete, vectorizing $P(x)$ and $P_\theta(x)$ leads to

$$\widehat{\theta}_p(\mathcal{V}) = \arg\min_{\theta\in\Theta}\left(\widetilde{\mathbf{P}} - \mathbf{P}_\theta\right)^\top \mathcal{V}\left(\widetilde{\mathbf{P}} - \mathbf{P}_\theta\right),$$

  $\widetilde{\mathbf{P}}$ is a nonparametric estimator for $\mathbf{P}$.

- We prefer to minimize expected payoffs/rewards,

$$\widehat{\theta}_v(\mathcal{W}) = \arg\min_{\theta\in\Theta}\left(\widetilde{\mathbf{v}} - \mathbf{v}_\theta\right)^\top \mathcal{W}\left(\widetilde{\mathbf{v}} - \mathbf{v}_\theta\right),$$

  $\widetilde{\mathbf{v}}$ estimates $\mathbf{v}$ (vectorized $Q^{-1}(P(x))$), because when $\mathbf{v}_\theta = \mathbf{X}\theta$,

$$\widehat{\theta}_v(\mathcal{W}) = \left(\mathbf{X}^\top\mathcal{W}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathcal{W}\widehat{\mathbf{v}}$$

  N.B. dynamic models have the same structure as $v_\theta(x)$ becomes differences in expected utility flows

- SSS showed $\widehat{\theta}_v$ and $\widehat{\theta}_p$ are asymptotically equivalent (cf. Pesendorfer and Schmidt-Dengler (2008)).

ALS - Estimation

- It can be shown our model gives:

$$\mathcal{Y} = \mathcal{X}\theta_0,$$

  where $\mathcal{X}$ and $\mathcal{Y}$ are known functions of CCPs and (β, π, Q, H, G).
- Under a full rank condition $\theta_0 = (\mathcal{X}^\intercal \mathcal{X})^{-1} \mathcal{X}\mathcal{Y}$.
- Given estimators $\left(\widehat{\mathcal{Y}}, \widehat{\mathcal{X}}\right)$,

$$\begin{aligned} \widehat{\theta} &= \arg\min_{\theta\in\Theta} (\widehat{\mathcal{Y}} - \widehat{\mathcal{X}}\theta)^\top (\widehat{\mathcal{Y}} - \widehat{\mathcal{X}}\theta) \\ &= (\widehat{\mathcal{X}}^\top \widehat{\mathcal{X}})^{-1} \widehat{\mathcal{X}}^\top \widehat{\mathcal{Y}}, \end{aligned}$$

  and properties of ALSE are driven by $\widehat{H}$ (Lemma 4),

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \Sigma\right).$$

Simulations

Consider an entry game in Pesendorfer and Schmidt-Dengler (2008) with UH

- $a_{it} \in \{0,1\}$ and $x_t = (a_{1t-1}, a_{2t-1})$
- Firm $1'$s period payoffs is

$$\begin{aligned} \theta^{\intercal}\pi_1(a_t, x_t, d_t) &= a_{1t}(1-a_{2t})\mu_1 + \mu_2 a_{1t}a_{2t} \\ &\quad + a_{1t}(1-a_{1t-1})F + (1-a_{1t})a_{1t-1}W \\ &\quad + a_{1t}\sum_{\ell=1}^{L}\omega_\ell d_{\ell t}, \end{aligned}$$

  where $d_{\ell t} = 1$ iff state ℓ occurs (ow 0), and impose $\omega_L > \omega_{L-1} > \cdots > \omega_1$
- With $L = 2$, set $(\mu_1, \mu_2, F, W, \omega_1, \omega_2) = (1.2, -1.2, -0.2, 0.1, 0, 0.1)$
- Performed 1,000 simulations for different combinations of (N, T) and estimation by OLS

Equilibria

States	Equilibrium 1		Equilibrium 2		Equilibrium 3		Symmetric Eq	
	p_i	p_{-i}	p_i	p_{-i}	p_i	p_{-i}	p_i	p_{-i}
$(0,0,0)$	0.733	0.276	0.667	0.463	0.672	0.460	0.572	0.572
$(0,1,0)$	0.612	0.422	0.315	0.827	0.310	0.827	0.305	0.837
$(1,0,0)$	0.800	0.223	0.822	0.307	0.824	0.309	0.837	0.305
$(1,1,0)$	0.751	0.295	0.551	0.625	0.607	0.568	0.592	0.592
$(0,0,1)$	0.751	0.295	0.550	0.625	0.498	0.675	0.592	0.592
$(0,1,1)$	0.637	0.438	0.349	0.830	0.343	0.831	0.338	0.840
$(1,0,1)$	0.816	0.242	0.829	0.337	0.829	0.340	0.840	0.338
$(1,1,1)$	0.767	0.316	0.667	0.543	0.593	0.620	0.611	0.611
Distance	0.038	0.037	0.169	0.183	0.177	0.223	0.042	0.042

Symmetric Equilibrium

(M, T)	Estimator	F		μ_1		μ_2		ω_2		MSE	d_t bias (%)
		Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD		
(500, 10)	Infeasible	0.074	(0.091)	0.066	(0.079)	0.090	(0.110)	0.050	(0.063)	2.817	-
	Feasible	0.088	(0.109)	0.085	(0.100)	0.106	(0.126)	0.057	(0.070)	4.331	19.72
	Naive	0.068	(0.085)	0.063	(0.072)	0.095	(0.097)	-	-	2.738	-
(500, 50)	Infeasible	0.030	(0.038)	0.027	(0.033)	0.035	(0.044)	0.020	(0.025)	0.445	-
	Feasible	0.038	(0.046)	0.040	(0.048)	0.049	(0.054)	0.041	(0.049)	0.861	9.97
	Naive	0.030	(0.037)	0.050	(0.032)	0.057	(0.042)	-	-	0.931	-
(500,100)	Infeasible	0.020	(0.025)	0.017	(0.022)	0.024	(0.030)	0.013	(0.016)	0.204	-
	Feasible	0.027	(0.039)	0.033	(0.039)	0.038	(0.039)	0.035	(0.039)	0.506	7.41
	Naive	0.020	(0.025)	0.049	(0.021)	0.052	(0.029)	-	-	0.697	-
(1000, 10)	Infeasible	0.050	(0.063)	0.043	(0.054)	0.061	(0.076)	0.035	(0.043)	1.306	-
	Feasible	0.057	(0.073)	0.056	(0.070)	0.069	(0.086)	0.045	(0.053)	1.910	18.29
	Naive	0.045	(0.038)	0.053	(0.050)	0.072	(0.066)	-	-	1.520	-
(1000, 50)	Infeasible	0.019	(0.024)	0.017	(0.021)	0.023	(0.029)	0.014	(0.018)	0.187	-
	Feasible	0.023	(0.031)	0.028	(0.034)	0.031	(0.037)	0.032	(0.039)	0.365	9.49
	Naive	0.019	(0.050)	0.050	(0.029)	0.052	(0.029)	-	-	0.691	-
(1000,100)	Infeasible	0.015	(0.019)	0.013	(0.016)	0.019	(0.023)	0.010	(0.013)	0.115	-
	Feasible	0.018	(0.022)	0.021	(0.026)	0.025	(0.030)	0.024	(0.032)	0.222	6.75
	Naive	0.015	(0.019)	0.050	(0.016)	0.051	(0.022)	-	-	0.617	-

Equilibrium 1

(M, T)	Estimator	F		μ_1		μ_2		ω_2		MSE	d_t bias (%)
		Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD		
(500, 10)	Infeasible	0.043	(0.053)	0.049	(0.061)	0.061	(0.075)	0.058	(0.073)	1.257	–
	Feasible	0.049	(0.061)	0.076	(0.089)	0.068	(0.086)	0.068	(0.081)	2.200	20.43
	Naive	0.040	(0.050)	0.055	(0.049)	0.069	(0.062)	–	–	1.394	–
(500, 50)	Infeasible	0.017	(0.021)	0.019	(0.023)	0.023	(0.029)	0.023	(0.029)	0.184	–
	Feasible	0.019	(0.023)	0.039	(0.044)	0.033	(0.038)	0.045	(0.054)	0.449	10.68
	Naive	0.017	(0.021)	0.050	(0.019)	0.051	(0.026)	–	–	0.647	–
(500, 100)	Infeasible	0.012	(0.015)	0.013	(0.016)	0.017	(0.021)	0.016	(0.021)	0.094	–
	Feasible	0.014	(0.016)	0.034	(0.037)	0.026	(0.027)	0.039	(0.047)	0.290	7.25
	Naive	0.012	(0.015)	0.049	(0.014)	0.051	(0.019)	–	–	0.571	–
(1000, 10)	Infeasible	0.029	(0.037)	0.034	(0.043)	0.041	(0.052)	0.040	(0.052)	0.600	–
	Feasible	0.031	(0.040)	0.047	(0.059)	0.052	(0.066)	0.051	(0.062)	0.993	20.19
	Naive	0.027	(0.034)	0.051	(0.034)	0.056	(0.045)	–	–	0.928	–
(1000, 50)	Infeasible	0.012	(0.016)	0.014	(0.017)	0.018	(0.022)	0.016	(0.020)	0.100	–
	Feasible	0.013	(0.016)	0.026	(0.031)	0.029	(0.031)	0.037	(0.042)	0.254	10.35
	Naive	0.012	(0.016)	0.050	(0.015)	0.050	(0.020)	–	–	0.585	–
(1000, 100)	Infeasible	0.009	(0.011)	0.009	(0.012)	0.012	(0.015)	0.011	(0.014)	0.047	–
	Feasible	0.009	(0.011)	0.022	(0.027)	0.024	(0.023)	0.033	(0.037)	0.169	7.40
	Naive	0.008	(0.011)	0.050	(0.010)	0.050	(0.014)	–	–	0.544	–

Equilibrium 2

(M, T)	Estimator	F		μ_1		μ_2		ω_2		MSE	d_t bias (%)
		Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD		
(500, 10)	Infeasible	0.068	(0.085)	0.061	(0.075)	0.084	(0.104)	0.047	(0.059)	2.462	–
	Feasible	0.069	(0.085)	0.060	(0.074)	0.085	(0.104)	0.045	(0.055)	2.476	4.21
	Naive	0.067	(0.083)	0.061	(0.071)	0.103	(0.103)	–	–	2.884	–
(500, 50)	Infeasible	0.028	(0.035)	0.025	(0.031)	0.033	(0.042)	0.019	(0.026)	0.398	–
	Feasible	0.028	(0.035)	0.025	(0.031)	0.033	(0.042)	0.020	(0.025)	0.401	0.92
	Naive	0.029	(0.036)	0.046	(0.032)	0.061	(0.044)	–	–	0.937	–
(500, 100)	Infeasible	0.019	(0.023)	0.016	(0.020)	0.023	(0.028)	0.013	(0.017)	0.177	–
	Feasible	0.019	(0.023)	0.016	(0.020)	0.023	(0.028)	0.014	(0.017)	0.179	0.80
	Naive	0.020	(0.024)	0.043	(0.020)	0.058	(0.030)	–	–	0.701	–
(1000, 10)	Infeasible	0.049	(0.061)	0.043	(0.056)	0.058	(0.072)	0.034	(0.043)	1.217	–
	Feasible	0.049	(0.061)	0.043	(0.056)	0.058	(0.072)	0.034	(0.043)	1.221	0.81
	Naive	0.048	(0.060)	0.050	(0.052)	0.083	(0.077)	–	–	1.791	–
(1000, 50)	Infeasible	0.020	(0.025)	0.017	(0.022)	0.024	(0.030)	0.014	(0.017)	0.198	–
	Feasible	0.020	(0.025)	0.017	(0.022)	0.024	(0.030)	0.014	(0.017)	0.198	0.04
	Naive	0.020	(0.025)	0.044	(0.023)	0.058	(0.032)	–	–	0.733	–
(1000, 100)	Infeasible	0.013	(0.017)	0.012	(0.015)	0.016	(0.020)	0.009	(0.012)	0.092	–
	Feasible	0.013	(0.017)	0.012	(0.015)	0.016	(0.020)	0.009	(0.012)	0.092	0.06
	Naive	0.014	(0.017)	0.043	(0.015)	0.058	(0.023)	–	–	0.624	–

Equilibrium 3

(M, T)	Estimator	F		μ_1		μ_2		ω_2		MSE	d_t bias (%)
		Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD	Bias (Mean)	SD		
(500, 10)	Infeasible	0.069	(0.088)	0.062	(0.078)	0.083	(0.104)	0.047	(0.060)	2.505	–
	Feasible	0.069	(0.087)	0.061	(0.076)	0.082	(0.102)	0.046	(0.056)	2.439	3.07
	Naive	0.068	(0.086)	0.064	(0.074)	0.098	(0.101)	–	–	2.818	–
(500, 50)	Infeasible	0.027	(0.034)	0.025	(0.031)	0.032	(0.041)	0.019	(0.025)	0.376	–
	Feasible	0.027	(0.034)	0.025	(0.031)	0.032	(0.041)	0.019	(0.025)	0.378	0.81
	Naive	0.029	(0.035)	0.040	(0.031)	0.063	(0.042)	–	–	0.896	–
(500, 100)	Infeasible	0.019	(0.023)	0.017	(0.020)	0.023	(0.029)	0.013	(0.016)	0.179	–
	Feasible	0.019	(0.023)	0.017	(0.021)	0.023	(0.029)	0.014	(0.017)	0.180	0.63
	Naive	0.020	(0.024)	0.036	(0.021)	0.061	(0.030)	–	–	0.683	–
(1000, 10)	Infeasible	0.049	(0.061)	0.043	(0.054)	0.060	(0.074)	0.034	(0.043)	1.222	–
	Feasible	0.049	(0.061)	0.043	(0.054)	0.060	(0.074)	0.034	(0.043)	1.219	0.41
	Naive	0.049	(0.060)	0.051	(0.053)	0.079	(0.075)	–	–	1.688	–
(1000, 50)	Infeasible	0.020	(0.024)	0.017	(0.021)	0.024	(0.029)	0.014	(0.017)	0.192	–
	Feasible	0.020	(0.020)	0.017	(0.021)	0.024	(0.029)	0.014	(0.017)	0.192	0.03
	Naive	0.021	(0.025)	0.038	(0.023)	0.060	(0.030)	–	–	0.697	–
(1000, 100)	Infeasible	0.013	(0.017)	0.012	(0.015)	0.016	(0.020)	0.009	(0.012)	0.090	–
	Feasible	0.013	(0.017)	0.012	(0.015)	0.016	(0.020)	0.009	(0.012)	0.090	0.03
	Naive	0.014	(0.017)	0.037	(0.016)	0.059	(0.022)	–	–	0.590	–

Concluding Remarks

- Propose a new estimator with unobserved time varying heterogeneity without instruments/finite dependence.
- Estimation is very simple and direct:
  1. use k-means clustering to classify CCPs into types;
  2. do OLS/GLS as done in SSS.
- Idea readily extendable to ordered discrete games (e.g., Gowrisankaran, Lucarelli, Schmidt-Dengler and Town (2018)).
- Working with continuous observables is in principle possible, e.g., for continuous action games (Bajari, Benkard and Levin (2007), Srisuma (2013)) and/or with continuous x_t (Srisuma and Linton (2012), Buchholz, Shum and Xu (2021)).