K-means Clustering of CCPs for Estimating Dynamic Discrete Choice Models with Unobserved Heterogeneity

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

Hong Kong, December 2025

Motivation

- Computation matters:
  - **Full solution** (Miller (1984), Pakes (1986), Rust (1987))
  - **CCP estimators** (Hotz and Miller (1993), Hotz et al. (1994))
- CCP requirements:
  - Use ΔVF $\Leftrightarrow$ CCPs, transition probabilities etc.
  - Consistent estimation of CCPs and transition probabilities.
- **Two-step estimators** ($\subseteq$ CCP) use a *consistent nonparametric CCP estimator* without *iteration.* Benefits of iterations include bias reduction and efficient estimation (Aguirregabiria and Mira (2002)), and accommodation of unobserved heterogeneity (Arcidiacono and Miller (2011)).

Two-step estimation

Why want to use two-step methods?

- Computationally light – e.g., closed-form estimation that guarantees global solution (Sanches, Silva and Srisuma (2016)).
- Consistent estimation of games when data come from a single equilibrium while iterative methods cause problems (Pesendorfer and Schmidt-Dengler (2010), Kasahara and Shimotsu (2012)).

To apply the two-step approach, one needs an initial consistent nonparametric CCP estimator, which appears infeasible with unobserved heterogeneity.

Our goal is to make two-step methods feasible with unobserved heterogeneity.

What We Do

- Clustering of CCPs to **estimate** latent types. Then, any standard two-step estimation method can be applied.
- Provide weak conditions where oracle efficient estimation of θ_0 is possible.
- We allow for:
  - heterogeneity over time (macro common shocks) and individual/market types (heterogeneity in payoffs, transition law, and multiple equilibria).
  - unknown number of latent types and different types of heterogeneity (e.g., payoff het+multiple equilibria etc).
- Illustrate with Sanches et al. (2016) estimator in a MC under different heterogeneity settings.

Related Literature

- **Single-agent model estimation:** Rust (1987), Hotz and Miller (1993), Hotz et al. (1994), Aguirregabiria and Mira (2002).
- **Dynamic game estimation:** Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008), Srisuma (2013), Sanches, Silva, and Srisuma (2016).
- **With unobserved heterogeneity:** Kasahara and Shimotsu (2008), Aguirregabiria and Mira (2007), Arcidiacono and Miller (2011), Hu and Shum (2012), Kaloupstidi, Scott, and Souza-Rodrigues (2021)
- **K-means clustering:** Pollard (1981,1982), Bonhomme and Manresa (2015), Bonhomme, Lamadon, and Manresa (2022).

Plan of Talk

- Modelling Assumption
- Clustering CCPs
- Estimating Number of Types
- Simulations
- Conclude

Modelling Assumption

Model Assumptions

Assumption M'

1. $|\mathcal{A} \times \mathcal{X}| < \infty$ and $\mathcal{E} = \mathbb{R}^{|A+1|}$;
2. $u(a, x, \varepsilon) = \theta^{\top} \pi(a, x) + \varepsilon(a)$;
3. $P(x', \varepsilon' | x, \varepsilon, a) = Q(\varepsilon') \, G(x' | x, a)$ and ;
4. Q *is known (cf. Norets and Tang (2014))*;
5. β *is known (cf. Komarova et al. (2018), Abbring and Daljord (2020))*.

Model Assumptions with Unobserved Heterogeneity

Assumption M

1. $|\mathcal{A} \times \mathcal{X} \times \mathcal{C} \times \mathcal{D}| < \infty$ and $\mathcal{E} = \mathbb{R}^{|A+1|}$;
2. $u_c(a, x, d, \varepsilon) = \theta_c{}^{\top} \pi_c(a, x, d) + \varepsilon(a)$;
3. $P_c(x', d', \varepsilon' | x, d, \varepsilon, a) = Q_c(\varepsilon') H(d'|d) G_c(x'|x, a, d)$ and ;
4. Q_c *is known (cf. Norets and Tang (2014))*;
5. β_c *is known (cf. Komarova et al. (2018), Abbring and Daljord (2020))*.

Estimation

(Sanches et al. (2016)) For each c, under a rank condition:

$$\theta_c = (\mathbf{Z}_c^\top \mathbf{Z}_c)^{-1} \mathbf{Z}_c^\top \mathbf{Y}_c,$$

where $(\mathbf{Z}_c, \mathbf{Y}_c)$ are continuous in CCPs and (H, G_c).
Suppose we have $\{(a_{mt}, x_{mt}, d_t)\}_{t=1, m:c_m=c}^{T}$. The infeasible empirical estimator gives:

$$\widetilde{\theta}_c = (\widetilde{\mathbf{Z}}_c^\top \widetilde{\mathbf{Z}}_c)^{-1} \widetilde{\mathbf{Z}}_c^\top \widetilde{\mathbf{Y}}_c.$$

We cluster CCPs to obtain $\left\{\left(a_{mt}, x_{mt}, \hat{d}_t\right)\right\}_{t=1, m:\hat{c}_m=c}^{T}$ and

$$\widehat{\theta}_c = (\widehat{\mathbf{Z}}_c^\top \widehat{\mathbf{Z}}_c)^{-1} \widehat{\mathbf{Z}}_c^\top \widehat{\mathbf{Y}}_c,$$

that is **oracle efficient** under weak conditions.

Clustering CCPs

Time Invariant States

Suppose $\mathcal{A} = \{0, \ldots, A\}$, $\mathcal{X} = \{1, \ldots, X\}$, $\mathcal{C} = \{1, \ldots, C\}$ with A, X, C known.

- Given $\{(a_{mt}, x_{mt})\}_{m=1,t=1}^{M,T}$, estimate $\mathfrak{q}^0_{c_m} = \Pr[a_{mt} = a | x_{mt} = x, c_m]$ by

$$\widehat{q}_{mT}(a|x) = \frac{\sum_{t=1}^{T} \mathbf{1}[a_{mt} = a, x_{mt} = x]}{\sum_{t=1}^{T} \mathbf{1}[x_{mt} = x]}.$$

- Let $\widehat{q}_{mT}$ be $(\widehat{q}_{mT}(a|x))_{(a,x)\in\mathcal{A}\backslash\{0\}\times\mathcal{X}}$. Choose $\mathbf{q} = (\mathfrak{q}_c)_{c\in\mathcal{C}} \in [0,1]^{AXC}$ to minimize:

$$\begin{aligned} \widehat{W}_{\mathcal{C}}(\mathbf{q}) &= \frac{1}{M}\sum_{m=1}^{M} \phi_{\mathcal{C}}(\widehat{q}_{mT}, \mathbf{q}), \text{ where} \\ \phi_{\mathcal{C}}(q, \mathbf{q}) &= \min_{c\in\mathcal{C}} \|q - \mathfrak{q}_c\|^2. \end{aligned}$$

Time Invariant States

Let $\mathfrak{q}_c^0 = \left(\Pr\left[a_{mt} = a | x_{mt} = x, c_m = c\right]\right)_{(a,x) \in \mathcal{A} \setminus \{0\} \times \mathcal{X}}$.

Assumption C.

(i) $\{c_m\}_{m=1}^M$ are i.i.d. variables drawn from some distribution γ such that $\gamma(c) > 0$ for all $c \in \mathcal{C}$.

(ii) For each m, conditional on c_m, $\{(a_{mt}, x_{mt})\}_{t \geq 1}$ is stationary and geometric $\alpha-$mixing with common conditional distribution $P_{A,X|C}(\cdot|c_m)$ such that $P_{X,C}(x, c) > 0$ all $(x, c) \in \mathcal{X} \times \mathcal{C}$.

(iii) For all $c \neq c'$, $\left\|\mathfrak{q}_c^0 - \mathfrak{q}_{c'}^0\right\| \geq \epsilon_{\mathcal{C}} > 0$.

Time Invariant States

Lemma C1. Under Assumption C and $\log(M) = o(T)$,
$\sup_{m \leq M} \left\| \widehat{q}_{mT} - \mathfrak{q}^0_{c_m} \right\| = o_p(1)$.

Let $(\widehat{\mathfrak{q}}_c)_{c \in \mathcal{C}} \in \arg\min_{\mathbf{q}} \widehat{W_{\mathcal{C}}}(\mathbf{q})$.

Lemma C2. Under Assumption C,
$\min_{\sigma} \max_c \left\| \widehat{\mathfrak{q}}_c - \mathfrak{q}^0_{\sigma(c)} \right\| = o_p(1)$ as $M, T \to \infty$.

Theorem C. Under Assumption C and $\log(M) = o(T)$, then
$\frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\left[\widehat{c}_m \neq c_m\right] = o_p\left(M^{-\rho}\right)$ for all $\rho > 0$.

Time Invariant States

Corollary C. For any bounded sequence of random variables $\{z_m\}_{m=1}^{M}$, under Assumption C and $\log(M) = o(T)$:
$\frac{1}{M}\sum_{m=1}^{M} z_m \mathbf{1}[\widehat{c}_m = c] - \frac{1}{M}\sum_{m=1}^{M} z_m \mathbf{1}[c_m = c] = o_p(M^{-\rho})$ for all $c, \rho > 0$.

Proposition C. Under Assumption C and $\log(M) = o(T)$, the feasible frequency estimators, via k-means, for $\Pr[c_m = c]$, $\Pr[x_{mt} = x | c_m = c]$, $\Pr[a_{mt} = a | x_{mt} = x, c_m = c]$ etc are oracle efficient.

Time Varying States

Let $\mathcal{D} = \{1, \ldots, D\}$.

- Estimate $\mathfrak{r}^0_{d_t} = \Pr\left[a_{mt} = a | x_{mt} = x, d_t\right]$ by

$$\widehat{r}_{Mt}\left(a|x\right) = \frac{\sum_{m=1}^{M} \mathbf{1}\left[a_{mt} = a, x_{mt} = x\right]}{\sum_{m=1}^{M} \mathbf{1}\left[x_{mt} = x\right]}.$$

- Let $\widehat{r}_{Mt}$ be $\left(\widehat{r}_{Mt}\left(a|x\right)\right)_{(a,x)\in\mathcal{A}\backslash\{0\}\times\mathcal{X}}$, choose $\mathbf{r} = \left(\mathfrak{r}_d\right)_{d\in\mathcal{D}} \in \left[0,1\right]^{AXD}$ to minimize:

$$\begin{aligned}
\widehat{W}_{\mathcal{D}}\left(\mathbf{r}\right) &= \frac{1}{T}\sum_{t=1}^{T}\phi_{\mathcal{D}}\left(\widehat{r}_{Mt}, \mathbf{r}\right), \text{ where} \\
\phi_{\mathcal{D}}\left(r, \mathbf{r}\right) &\equiv \min_{d\in\mathcal{D}} \left\| r - \mathfrak{r}_d \right\|^2.
\end{aligned}$$

Time Varying States

Let $\mathfrak{r}_d^0 = \left(\Pr\left[a_{mt} = a | x_{mt} = x, d_t = d\right]\right)_{(a,x)\in\mathcal{A}\setminus\{0\}\times\mathcal{X}}$.

Assumption D.

(i) $\{d_t\}_{t\geq 1}$ is stationary and geometric $\alpha-$mixing, whose stationary distribution δ satisfies $\delta(d) > 0$ for all $d \in \mathcal{D}$.

(ii) For each t, conditional on d_t, $\{(a_{mt}, x_{mt})\}_{m=1}^{M}$ are i.i.d. with common conditional distribution $P_{A,X|D}(\cdot|d_t)$ such that $P_{X,D}(x,d) > 0$ all $(x,d) \in \mathcal{X}\times\mathcal{D}$.

(iii) For all $d \neq d'$, $\left\|\mathfrak{r}_d^0 - \mathfrak{r}_{d'}^0\right\| \geq \epsilon_{\mathcal{D}} > 0$.

Time Varying States

Let $\widehat{d}_t = \arg\min_{d \in \mathcal{D}} \left\| \widehat{r}_{Mt} - \widehat{\mathfrak{r}}_d \right\|$.

Theorem D. Under Assumption D and $\log(T) = o(M)$, then $\frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\left[\widehat{d}_t \neq d_t\right] = o_p\left(T^{-\rho}\right)$ for all $\rho > 0$.

Corollary D. For any bounded sequence of random variables $\{z_t\}_{t=1}^{T}$, under Assumption D and $\log(T) = o(M)$:
(i) $\frac{1}{T} \sum_{t=1}^{T} z_t \mathbf{1}\left[\widehat{d}_t = d\right] - \frac{1}{T} \sum_{t=1}^{T} z_t \mathbf{1}\left[d_t = d\right] = o_p\left(T^{-\rho}\right)$ for $d, \rho > 0$;
(ii) $\frac{1}{T} \sum_{t=1}^{T} z_t \mathbf{1}\left[\widehat{d}_{t+1} = d', \widehat{d}_t = d\right] - \frac{1}{T} \sum_{t=1}^{T} z_t \mathbf{1}\left[d_{t+1} = d', d_t = d\right]) = o_p\left(T^{-\rho}\right)$ for all $d, d', \rho > 0$.

Proposition D. Under Assumption D and $\log(T) = o(M)$, the feasible frequency estimators, via k-means, for $\Pr\left[d_t = d\right]$, $\Pr\left[d_{t+1} = d' | d_t = d,\right]$, $\Pr\left[a_{mt} = a | x_{mt} = x, d_t = d\right]$ etc are oracle efficient.

More Generally

Assumption CD.
(i) $\{c_m\}_{m=1}^{M}$ are i.i.d. variables drawn from some distribution γ such that $\gamma(c) > 0$ for all $c \in \mathcal{C}$.
(ii) $\{d_t\}_{t\geq 1}$ is stationary and geometric $\alpha-$mixing, whose stationary distribution δ satisfies that $\delta(d) > 0$ for all $d \in \mathcal{D}$. Moreover, the σ-fields generated by $\{c_m\}_{m=1}^{M}$ and $\{d_t\}_{t=1}^{T}$ are independent.
(iii) For each t, conditional on $\{c_m\}_{m=1}^{M}$ and d_t, $\{(a_{mt}, x_{mt})\}_{m=1}^{M}$ is independent and drawn from conditional distribution $P_{A,X|C,D}(\cdot|c_m, d_t)$ such that $P_{X,C,D}(x, c, d) > 0$ all $(x, c, d) \in \mathcal{X} \times \mathcal{C} \times \mathcal{D}$.
(iv) For each m, conditional on c_m, $\{(a_{mt}, x_{mt}, d_t)\}_{t\geq 1}$ is stationary and geometric $\alpha-$mixing.
(v) For all $c \neq c'$, $\left\|\mathfrak{q}_c^0 - \mathfrak{q}_{c'}^0\right\| \geq \epsilon_{\mathcal{C}} > 0$.
(vi) For all $d \neq d'$, $\left\|\mathfrak{r}_d^0 - \mathfrak{r}_{d'}^0\right\| \geq \epsilon_{\mathcal{D}} > 0$.

More Generally

Theorem CD. Under Assumption CD and $\log(M) = o(T)$ and $\log(T) = o(M)$, then $\frac{1}{M}\sum_{m=1}^{M} \mathbf{1}[\widehat{c}_m \neq c_m] = o_p(M^{-\rho})$ and $\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\left[\widehat{d}_t \neq d_t\right] = o_p(T^{-\rho})$ for all $\rho > 0$.

Analogous results to Corollaries C and D hold.

Proposition CD. Under Assumption CD and $\log(M) = o(T)$ and $\log(T) = o(M)$, the feasible frequency (via k-means) estimators for $\Pr[c_m = c]$, $\Pr[d_t = d]$, $\Pr[d_{t+1} = d' | d_t = d]$, $\Pr[a_{mt} = a | x_{mt} = x, c_m = c, d_t = d]$ etc are oracle efficient.

Estimating Number of Latent Types

Number of Latent Types

We only observe $\{\widehat{q}_{mT}\}_{m=1}^{M}$.

Let:

$$\begin{aligned}\widehat{\mathrm{W}}\left(n_{\mathcal{C}}\right) &= \inf_{\mathbf{q}} \frac{1}{M} \sum_{m=1}^{M} \min_{c \in \{1,\ldots,n_{\mathcal{C}}\}} \left\| \widehat{q}_{mT} - \mathfrak{q}_c \right\|^2, \\ \widehat{\mathcal{J}}_{\mathrm{W}}\left(n_{\mathcal{C}}\right) &= \widehat{\mathrm{W}}\left(n_{\mathcal{C}}\right) + \lambda_M n_{\mathcal{C}}.\end{aligned}$$

Theorem NC1. Under Assumption C, $\log(M) = o(T)$, and λ_M is a positive sequence that satisfies $\sqrt{\frac{\log(M)}{T}} = o(\lambda_M)$, then $\widehat{n}_{\mathcal{C}} := \arg\min_{n_{\mathcal{C}} \leq N_{\mathcal{C}}} \widehat{\mathcal{J}}_{\mathrm{W}}\left(n_{\mathcal{C}}\right) \xrightarrow{p} C$ as $M, T \to \infty$.

Number of Equilibria

- Suppose C in (the game version of) Assumption C satisfies $C = \sum_{j=1}^{J} K_j$ with $K_j = 1 + k_j$ such that
  - J denotes the number of $\{\theta_j\}_{j=1}^{J}$ that generated the data;
  - for each j, $k_j > 0$ means there are $1 + k_j$ equilibria being played.
- Cluster $\left\{\widehat{\theta}_c\right\}_{c=1}^{\widehat{n}_{\mathcal{C}}}$ to identify J and k_j:

$$
\begin{aligned}
\widehat{\mathrm{Q}}\left(n_\Theta\right) &= \inf_{\left(\mathfrak{t}_{c'}\right)_{c'=1,\ldots,n}} \sum_{c=1}^{\widehat{n}_{\mathcal{C}}} \min_{c' \in \left\{1,\ldots,n_\Theta\right\}} \left\|\widehat{\theta}_c - \mathfrak{t}_{c'}\right\|^2 \\
\widehat{\mathcal{I}}_{\mathrm{Q}}\left(n_\Theta\right) &= \widehat{\mathrm{Q}}\left(n_\Theta\right) + \lambda_M n_\Theta .
\end{aligned}
$$

Theorem NC2. Under Assumption C, $\log\left(M\right) = o\left(T\right)$, and λ_M is a positive sequence that satisfies $\sqrt{\frac{1}{M}} = o\left(\lambda_M\right)$, then $\widehat{n}_\Theta := \arg\min_{n_\Theta \leq \widehat{n}_{\mathcal{C}}} \widehat{\mathcal{I}}_{\mathrm{Q}}\left(n_\Theta\right) \xrightarrow{p} J$ as $M, T \to \infty$.

Monte Carlo Simulation

Simulations

Based on an entry game in Pesendorfer and Schmidt-Dengler (2008) where in each market m at time t:

- $a_{imt} \in \{0,1\}$ and $x_{mt} = (a_{1mt-1}, a_{2mt-1})$
- Firm $1'$s period payoffs is

$$\begin{aligned}\theta^{\intercal}\pi_1(a_{mt}, x_{mt}) &= a_{1mt}(1-a_{2mt})\mu_1 + \mu_2 a_{1mt}a_{2mt} \\ &\quad + a_{1t}(1-a_{1mt-1})F + (1-a_{1mt})a_{1mt-1}W.\end{aligned}$$

- Set $\theta = (\mu_1, \mu_2, F, W) = (1.2, -1.2, -0.2, 0.1)$
- The game has 3 equilibria.
- Estimation is done based on SSS's OLS.

Simulations

We consider the following types of UH:

1. (c_m only) There are θ_1 and θ_2 depending on market types.
2. (d_t only) There are additive time fixed effects in π_1.
3. (c_m only) Just θ_1 but different market can play different equilibria.
4. (c_m only) θ_1 and θ_2 + multiple equilibria.
5. (both c_m and d_t) are in progress.

Performance varies depending on (T, M) and how well separated the equilium CCPs are.

	Eq-1		Eq-2		Eq-3	(Sym)
States	p_i	p_{-i}	p_i	p_{-i}	p_i	p_{-i}
$(0,0)$	0.733	0.276	0.615	0.528	0.576	0.576
$(0,1)$	0.613	0.420	0.312	0.840	0.305	0.842
$(1,0)$	0.800	0.223	0.831	0.303	0.842	0.305
$(1,1)$	0.752	0.294	0.606	0.578	0.595	0.595

Table: Equilibria under θ_1. The CCPs are conditional on states $x_t = (a_{1t-1}, a_{2t-1})$.

	Sym Eq	
States	p_i	p_{-i}
$(0,0)$	0.526	0.526
$(0,1)$	0.718	0.322
$(1,0)$	0.322	0.718
$(1,1)$	0.516	0.516

Table: Equilibria under θ_2. The CCPs are conditional on states $x_t = (a_{1t-1}, a_{2t-1})$.

Simulation Results

Case 1 - Market Heterogeneity

(M, T)	Estimator	F_1	μ_{11}	μ_{12}	F_2	μ_{21}	μ_{22}	A-MSE	C-E%
$(500, 10)$	Naive	0.054	0.339	0.604	-	-	-	64.45	-
		(0.054)	(0.371)	(0.337)	-	-	-	-	-
	Feasible	0.236	0.336	0.483	0.215	0.534	0.227	55.50	8.59
		(0.232)	(0.417)	(0.561)	(0.141)	(0.248)	(0.170)	-	-
	Infeasible	0.095	0.074	0.111	0.082	0.138	0.082	4.68	-
		(0.119)	(0.090)	(0.137)	(0.103)	(0.172)	(0.102)	-	-
$(500, 50)$	Naive	0.043	0.195	0.533	-	-	-	37.91	-
		(0.026)	(0.195)	(0.179)	-	-	-	-	-
	Feasible	0.041	0.031	0.047	0.040	0.066	0.039	0.94	0.12
		(0.051)	(0.038)	(0.058)	(0.049)	(0.082)	(0.048)	-	-
	Infeasible	0.040	0.031	0.046	0.039	0.066	0.039	0.94	-
		(0.051)	(0.038)	(0.058)	(0.048)	(0.082)	(0.048)	-	-
$(500, 100)$	Naive	0.042	0.160	0.524	-	-	-	33.69	-
		(0.018)	(0.147)	(0.134)	-	-	-	-	-
	Feasible	0.028	0.022	0.032	0.026	0.044	0.026	0.44	0.00
		(0.035)	(0.027)	(0.040)	(0.032)	(0.055)	(0.032)	-	-
	Infeasible	0.028	0.021	0.027	0.026	0.043	0.026	0.42	-
		(0.035)	(0.026)	(0.040)	(0.032)	(0.054)	(0.032)	-	-
$(1000, 10)$	Naive	0.047	0.266	0.558	-	-	-	47.73	-
		(0.042)	(0.275)	(0.251)	-	-	-	-	-
	Feasible	0.214	0.262	0.354	0.213	0.468	0.180	35.50	8.99
		(0.184)	(0.322)	(0.444)	(0.097)	(0.160)	(0.106)	-	-
	Infeasible	0.069	0.055	0.080	0.062	0.107	0.060	2.60	-
		(0.088)	(0.068)	(0.101)	(0.078)	(0.132)	(0.076)	-	-
$(1000, 50)$	Naive	0.041	0.155	0.520	-	-	-	32.35	-
		(0.018)	(0.131)	(0.121)	-	-	-	-	-
	Feasible	0.028	0.022	0.033	0.025	0.042	0.025	0.43	0.12
		(0.036)	(0.028)	(0.041)	(0.037)	(0.182)	(0.124)	-	-
	Infeasible	0.029	0.022	0.024	0.025	0.043	0.025	0.44	-
		(0.036)	(0.028)	(0.042)	(0.032)	(0.054)	(0.032)	-	-
$(1000, 100)$	Naive	0.041	0.140	0.518	-	-	-	30.66	-
		(0.013)	(0.099)	(0.092)	-	-	-	-	-
	Feasible	0.021	0.016	0.024	0.018	0.031	0.019	0.23	0.00
		(0.026)	(0.020)	(0.030)	(0.023)	(0.039)	(0.024)	-	-
	Infeasible	0.021	0.016	0.023	0.018	0.030	0.018	0.22	-
		(0.026)	(0.020)	(0.029)	(0.023)	(0.039)	(0.023)	-	-

Table 4: Monte Carlo estimates of the model under Case 1. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

Case 2 - Time Heterogeneity

(M,T)	Estimator	F	μ_1	μ_2	ω_2	A-MSE	C-E%
$(500,10)$	Naive	0.066	0.104	0.127	-	4.60	-
		(0.083)	(0.084)	(0.105)	-	-	-
	Feasible	0.081	0.088	0.111	0.065	4.46	15.71
		(0.103)	(0.112)	(0.131)	(0.081)	-	-
	Infeasible	0.070	0.068	0.089	0.047	2.79	-
		(0.089)	(0.082)	(0.110)	(0.060)	-	-
$(500,50)$	Naive	0.030	0.096	0.106	-	2.52	-
		(0.037)	(0.037)	(0.044)	-	-	-
	Feasible	0.032	0.035	0.042	0.025	0.63	6.13
		(0.039)	(0.042)	(0.051)	(0.034)	-	-
	Infeasible	0.030	0.027	0.037	0.019	0.46	-
		(0.037)	(0.033)	(0.045)	(0.024)	-	-
$(500,100)$	Naive	0.020	0.099	0.104	-	2.30	-
		(0.026)	(0.026)	(0.032)	-	-	-
	Feasible	0.021	0.025	0.027	0.018	0.30	4.33
		(0.027)	(0.030)	(0.035)	(0.020)	-	-
	Infeasible	0.021	0.019	0.025	0.014	0.23	-
		(0.026)	(0.023)	(0.032)	(0.017)	-	-
$(1000,10)$	Naive	0.049	0.103	0.110	-	3.49	-
		(0.061)	(0.068)	(0.079)	-	-	-
	Feasible	0.053	0.056	0.071	0.042	1.78	7.62
		(0.066)	(0.071)	(0.089)	(0.057)	-	-
	Infeasible	0.052	0.048	0.063	0.035	1.40	-
		(0.064)	(0.060)	(0.078)	(0.043)	-	-
$(1000,50)$	Naive	0.020	0.101	0.103	-	2.32	-
		(0.024)	(0.028)	(0.033)	-	-	-
	Feasible	0.020	0.020	0.025	0.014	0.22	2.32
		(0.025)	(0.025)	(0.031)	(0.018)	-	-
	Infeasible	0.020	0.019	0.024	0.014	0.11	-
		(0.025)	(0.023)	(0.030)	(0.018)	-	-
$(1000,100)$	Naive	0.014	0.099	0.102	-	2.15	-
		(0.018)	(0.020)	(0.023)	-	-	-
	Feasible	0.014	0.014	0.018	0.010	0.11	1.76
		(0.018)	(0.017)	(0.023)	(0.012)	-	-
	Infeasible	0.014	0.013	0.018	0.010	0.11	-
		(0.018)	(0.016)	(0.022)	(0.013)	-	-

Table 5: Monte Carlo estimates of the mean bias and standard deviation (in brackets) of different estimator for θ with Symmetric Equilibrium. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

Case 3 - Multiple Equilibria

(M,T)	Estimator	F_1	μ_{11}	μ_{12}	F_2	μ_{21}	μ_{22}	F_3	μ_{31}	μ_{32}	A-MSE	C-E%
$(500, 10)$	Naive	0.074	0.121	0.084	-	-	-	-	-	-	3.64	-
		(0.072)	(0.054)	(0.090)	-	-	-	-	-	-	-	-
	Feasible	0.309	1.068	1.066	0.316	0.631	0.678	0.275	0.626	0.922	198.14	8.11
		(0.204)	(0.422)	(0.540)	(0.252)	(0.467)	(0.523)	(0.358)	(0.736)	(0.573)	-	-
	Infeasible	0.069	0.058	0.082	0.109	0.090	0.129	0.116	0.089	0.136	4.76	-
		(0.086)	(0.072)	(0.105)	(0.139)	(0.111)	(0.160)	(0.144)	(0.109)	(0.167)	-	-
$(500, 50)$	Naive	0.056	0.116	0.052	-	-	-	-	-	-	2.18	-
		(0.032)	(0.024)	(0.039)	-	-	-	-	-	-	-	-
	Feasible	0.077	0.074	0.091	0.227	0.189	0.233	0.350	0.200	0.272	19.01	6.42
		(0.129)	(0.069)	(0.190)	(0.196)	(0.124)	(0.226)	(0.346)	(0.325)	(0.211)	-	-
	Infeasible	0.029	0.024	0.037	0.052	0.040	0.059	0.050	0.038	0.058	0.93	-
		(0.037)	(0.030)	(0.046)	(0.065)	(0.049)	(0.074)	(0.063)	(0.047)	(0.073)	-	-
$(500, 100)$	Naive	0.055	0.117	0.050	-	-	-	-	-	-	2.03	-
		(0.020)	(0.016)	(0.025)	-	-	-	-	-	-	-	-
	Feasible	0.034	0.076	0.046	0.118	0.100	0.191	0.121	0.076	0.084	5.52	16.90
		(0.042)	(0.091)	(0.057)	(0.104)	(0.089)	(0.129)	(0.150)	(0.104)	(0.094)	-	-
	Infeasible	0.020	0.017	0.025	0.034	0.026	0.039	0.035	0.027	0.039	0.43	-
		(0.026)	(0.022)	(0.032)	(0.043)	(0.033)	(0.049)	(0.044)	(0.033)	(0.049)	-	-
$(1000, 10)$	Naive	0.063	0.118	0.067	-	-	-	-	-	-	2.78	-
		(0.052)	(0.039)	(0.064)	-	-	-	-	-	-	-	-
	Feasible	0.314	1.167	1.205	0.298	0.586	0.609	0.239	0.617	0.893	205.84	7.98
		(0.156)	(0.413)	(0.522)	(0.192)	(0.414)	(0.450)	(0.316)	(0.724)	(0.548)		
	Infeasible	0.049	0.041	0.058	0.083	0.065	0.098	0.084	0.067	0.096	2.52	-
		(0.061)	(0.050)	(0.074)	(0.103)	(0.080)	(0.121)	(0.104)	(0.082)	(0.118)	-	-
$(1000, 50)$	Naive	0.056	0.117	0.050	-	-	-	-	-	-	2.09	-
		(0.022)	(0.028)	(0.017)	-	-	-	-	-	-	-	-
	Feasible	0.082	0.095	0.107	0.194	0.166	0.224	0.303	0.188	0.333	18.68	6.42
		(0.145)	(0.081)	(0.228)	(0.208)	(0.124)	(0.264)	(0.342)	(0.334)	(0.209)	-	-
	Infeasible	0.020	0.017	0.025	0.036	0.027	0.041	0.036	0.028	0.043	0.47	-
		(0.025)	(0.021)	(0.032)	(0.045)	(0.034)	(0.052)	(0.046)	(0.035)	(0.053)	-	-
$(1000, 100)$	Naive	0.056	0.117	0.050	-	-	-	-	-	-	2.01	-
		(0.016)	(0.012)	(0.020)	-	-	-	-	-	-	-	-
	Feasible	0.023	0.044	0.033	0.122	0.098	0.154	0.137	0.087	0.090	4.87	10.21
		(0.028)	(0.058)	(0.043)	(0.100)	(0.079)	(0.116)	(0.146)	(0.100)	(0.092)	-	-
	Infeasible	0.015	0.013	0.019	0.024	0.019	0.028	0.025	0.019	0.028	0.22	-
		(0.019)	(0.016)	(0.023)	(0.030)	(0.024)	(0.035)	(0.032)	(0.024)	(0.036)	-	-

Table 4: Monte Carlo estimates of the model under θ_1 equilibria. A-MSE and C-E% respectively are the sum of mean square errors (scaled by 100) and percentage of clustering error.

(M, T)	Estimator	F_1	μ_{11}	μ_{12}	F_3	μ_{31}	μ_{32}	A-MSE	C-E%
$(500, 10)$	Naive	0.082	0.159	0.093	-	-	-	5.01	-
		(0.068)	(0.051)	(0.085)	-	-	-	-	-
	Feasible	0.195	0.629	0.600	0.205	0.513	0.694	109.11	5.93
		(0.184)	(0.505)	(0.557)	(0.249)	(0.322)	(0.431)	-	-
	Infeasible	0.057	0.048	0.070	0.095	0.074	0.111	2.89	-
		(0.072)	(0.059)	(0.087)	(0.119)	(0.090)	(0.137)	-	-
$(500, 50)$	Naive	0.074	0.156	0.076	-	-	-	3.82	-
		(0.029)	(0.022)	(0.036)	-	-	-	-	-
	Feasible	0.039	0.053	0.037	0.055	0.043	0.062	1.12	2.33
		(0.039)	(0.043)	(0.043)	(0.069)	(0.053)	(0.076)	-	-
	Infeasible	0.024	0.020	0.030	0.040	0.031	0.046	0.52	-
		(0.030)	(0.025)	(0.037)	(0.051)	(0.038)	(0.058)	-	-
$(500, 100)$	Naive	0.073	0.156	0.075	-	-	-	3.65	-
		(0.019)	(0.014)	(0.024)	-	-	-	-	-
	Feasible	0.018	0.016	0.021	0.029	0.021	0.032	0.26	0.55
		(0.022)	(0.019)	(0.027)	(0.036)	(0.027)	(0.040)	-	-
	Infeasible	0.016	0.014	0.020	0.028	0.032	0.021	0.25	-
		(0.021)	(0.017)	(0.026)	(0.035)	(0.026)	(0.040)	-	-
$(1000, 10)$	Naive	0.075	0.158	0.081	-	-	-	4.31	-
		(0.049)	(0.036)	(0.061)	-	-	-	-	-
	Feasible	0.171	0.596	0.566	0.189	0.500	0.672	102.00	5.20
		(0.149)	(0.506)	(0.566)	(0.221)	(0.257)	(0.368)	-	-
	Infeasible	0.040	0.033	0.050	0.069	0.055	0.080	1.53	-
		(0.050)	(0.041)	(0.061)	(0.088)	(0.068)	(0.101)	-	-
$(1000, 50)$	Naive	0.073	0.157	0.074	-	-	-	3.67	-
		(0.020)	(0.014)	(0.026)	-	-	-	-	-
	Feasible	0.029	0.064	0.030	0.052	0.043	0.061	1.07	2.29
		(0.032)	(0.041)	(0.033)	(0.060)	(0.045)	(0.068)	-	-
	Infeasible	0.016	0.014	0.021	0.029	0.022	0.034	0.26	-
		(0.021)	(0.018)	(0.027)	(0.036)	(0.028)	(0.042)	-	-
$(1000, 100)$	Naive	0.072	0.156	0.074	-	-	-	3.56	-
		(0.014)	(0.011)	(0.018)	-	-	-	-	-
	Feasible	0.013	0.013	0.016	0.021	0.016	0.024	0.15	0.55
		(0.015)	(0.014)	(0.019)	(0.026)	(0.020)	(0.030)	-	-
	Infeasible	0.011	0.010	0.015	0.021	0.016	0.023	0.13	-
		(0.014)	(0.012)	(0.018)	(0.026)	(0.020)	(0.029)	-	-

Table XX: Monte Carlo estimates of the model under Equilibrium 1 and 3 generated by $theta_1$. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

(M,T)	Estimator	F_1	μ_{11}	μ_{12}	F_2	μ_{21}	μ_{22}	A-MSE	C-E%
$(500, 10)$	Naive	0.075	0.137	0.086	-	-	-	4.03	-
		(0.064)	(0.049)	(0.080)	-	-	-	-	-
	Feasible	0.196	0.703	0.667	0.191	0.555	0.763	127.30	5.58
		(0.178)	(0.525)	(0.582)	(0.230)	(0.320)	(0.432)	-	-
	Infeasible	0.055	0.048	0.068	0.091	0.073	0.109	2.79	-
		(0.069)	(0.060)	(0.085)	(0.115)	(0.091)	(0.135)	-	-
$(500, 50)$	Naive	0.063	0.131	0.067	-	-	-	2.81	-
		(0.029)	(0.021)	(0.036)	-	-	-	-	-
	Feasible	0.041	0.050	0.036	0.057	0.040	0.061	1.09	2.55
		(0.042)	(0.040)	(0.043)	(0.070)	(0.050)	(0.074)	-	-
	Infeasible	0.025	0.020	0.031	0.041	0.031	0.047	0.53	-
		(0.031)	(0.025)	(0.038)	(0.051)	(0.039)	(0.059)	-	-
$(500, 100)$	Naive	0.064	0.133	0.067	-	-	-	2.74	-
		(0.019)	(0.015)	(0.024)	-	-	-	-	-
	Feasible	0.018	0.017	0.021	0.029	0.022	0.033	0.28	0.76
		(0.022)	(0.020)	(0.027)	(0.037)	(0.028)	(0.042)	-	-
	Infeasible	0.016	0.014	0.020	0.029	0.022	0.033	0.25	-
		(0.021)	(0.017)	(0.025)	(0.036)	(0.027)	(0.041)	-	-
$(1000, 10)$	Naive	0.067	0.135	0.075	-	-	-	3.34	-
		(0.047)	(0.035)	(0.058)	-	-	-	-	-
	Feasible	0.180	0.650	0.619	0.176	0.543	0.743	118.39	4.66
		(0.139)	(0.517)	(0.585)	(0.204)	(0.252)	(0.361)	-	-
	Infeasible	0.041	0.032	0.047	0.067	0.051	0.079	1.42	-
		(0.051)	(0.040)	(0.059)	(0.084)	(0.063)	(0.098)	-	-
$(1000, 50)$	Naive	0.063	0.133	0.067	-	-	-	2.73	-
		(0.020)	(0.015)	(0.024)	-	-	-	-	-
	Feasible	0.032	0.062	0.028	0.054	0.038	0.058	0.97	2.20
		(0.035)	(0.040)	(0.032)	(0.063)	(0.041)	(0.062)	-	-
	Infeasible	0.017	0.014	0.020	0.029	0.022	0.033	0.26	-
		(0.021)	(0.018)	(0.025)	(0.037)	(0.028)	(0.042)	-	-
$(1000, 100)$	Naive	0.064	0.133	0.067	-	-	-	2.68	-
		(0.014)	(0.011)	(0.018)	-	-	-	-	-
	Feasible	0.015	0.015	0.016	0.021	0.016	0.024	0.16	0.76
		(0.017)	(0.015)	(0.020)	(0.026)	(0.020)	(0.031)	-	-
	Infeasible	0.012	0.010	0.015	0.020	0.016	0.023	0.13	-
		(0.015)	(0.013)	(0.019)	(0.025)	(0.020)	(0.030)	-	-

Table XX: Monte Carlo estimates of the model under Equilibrium 1 and 2 generated by $theta_1$. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

(M,T)	Estimator	F_2	μ_{21}	μ_{22}	F_3	μ_{31}	μ_{32}	A-MSE	C-E%
$(500,10)$	Naive	0.066	0.051	0.076	-	-	-	2.00	-
		(0.082)	(0.064)	(0.095)	-	-	-	-	-
	Feasible	0.325	0.666	0.686	0.262	0.418	0.574	97.80	7.82
		(0.195)	(0.204)	(0.306)	(0.306)	(0.342)	(0.440)	-	-
	Infeasible	0.091	0.073	0.109	0.095	0.074	0.111	4.08	-
		(0.115)	(0.091)	(0.135)	(0.119)	(0.090)	(0.137)	-	-
$(500,50)$	Naive	0.030	0.023	0.035	-	-	-	0.42	-
		(0.038)	(0.029)	(0.044)	-	-	-	-	-
	Feasible	0.158	0.113	0.130	0.193	0.123	0.116	9.10	4.21
		(0.121)	(0.094)	(0.114)	(0.158)	(0.102)	(0.115)	-	-
	Infeasible	0.041	0.031	0.047	0.040	0.031	0.046	0.75	-
		(0.051)	(0.039)	(0.059)	(0.051)	(0.038)	(0.058)	-	-
$(500,100)$	Naive	0.020	0.015	0.023	-	-	-	0.18	-
		(0.025)	(0.019)	(0.028)	-	-	-	-	-
	Feasible	0.129	0.093	0.109	0.158	0.099	0.106	6.48	3.44
		(0.103)	(0.078)	(0.095)	(0.136)	(0.084)	(0.101)	-	-
	Infeasible	0.029	0.022	0.033	0.028	0.021	0.032	0.36	-
		(0.036)	(0.027)	(0.041)	(0.035)	(0.026)	(0.040)	-	-
$(1000,10)$	Naive	0.049	0.038	0.057	-	-	-	1.10	-
		(0.061)	(0.047)	(0.071)	-	-	-	-	-
	Feasible	0.321	0.671	0.700	0.236	0.377	0.509	88.15	7.67
		(0.127)	(0.126)	(0.193)	(0.263)	(0.291)	(0.376)	-	-
	Infeasible	0.067	0.051	0.079	0.069	0.055	0.080	2.17	-
		(0.084)	(0.063)	(0.098)	(0.088)	(0.068)	(0.101)	-	-
$(1000,50)$	Naive	0.021	0.016	0.023	-	-	-	0.20	-
		(0.026)	(0.020)	(0.030)	-	-	-	-	-
	Feasible	0.122	0.085	0.098	0.149	0.093	0.094	5.96	3.44
		(0.111)	(0.083)	(0.097)	(0.142)	(0.091)	(0.101)	-	-
	Infeasible	0.029	0.022	0.033	0.029	0.022	0.034	0.38	-
		(0.037)	(0.028)	(0.042)	(0.036)	(0.028)	(0.042)	-	-
$(1000,100)$	Naive	0.015	0.011	0.017	-	-	-	0.10	-
		(0.018)	(0.014)	(0.021)	-	-	-	-	-
	Feasible	0.010	0.070	0.082	0.120	0.077	0.083	4.21	2.46
		(0.094)	(0.068)	(0.082)	(0.118)	(0.076)	(0.087)	-	-
	Infeasible	0.020	0.016	0.023	0.025	0.020	0.030	0.19	-
		(0.025)	(0.020)	(0.030)	(0.026)	(0.020)	(0.029)	-	-

Table XX: Monte Carlo estimates of the model under Equilibrium 2 and 3 generated by $theta_1$. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

Simulations: Time Heterogeneity with multiple ω's

(M,T)	Estimator	F	μ_1	μ_2	ω_2	ω_3	A-MSE
$(500,50)$	Naive	0.027	0.200	0.206	-	-	8.84
		(0.034)	(0.047)	(0.048)	-	-	-
	Feasible	0.033	0.046	0.056	0.056	0.043	1.02
		(0.040)	(0.058)	(0.063)	(0.072)	(0.051)	-
	Infeasible	0.028	0.032	0.036	0.025	0.024	0.48
		(0.034)	(0.038)	(0.045)	(0.031)	(0.031)	-
$(500,100)$	Naive	0.020	0.200	0.205	-	-	8.52
		(0.025)	(0.034)	(0.036)	-	-	-
	Feasible	0.023	0.034	0.038	0.040	0.032	0.50
		(0.028)	(0.043)	(0.045)	(0.057)	(0.040)	-
	Infeasible	0.020	0.022	0.027	0.016	0.017	0.25
		(0.025)	(0.027)	(0.033)	(0.021)	(0.021)	-
$(500,200)$	Naive	0.014	0.200	0.204	-	-	8.29
		(0.017)	(0.024)	(0.024)	-	-	-
	Feasible	0.015	0.023	0.023	0.022	0.022	0.20
		(0.018)	(0.027)	(0.029)	(0.036)	(0.027)	-
	Infeasible	0.014	0.015	0.019	0.012	0.012	0.12
		(0.017)	(0.018)	(0.023)	(0.015)	(0.015)	-
$(100,50)$	Naive	0.021	0.199	0.206	-	-	8.59
		(0.027)	(0.040)	(0.041)	-	-	-
	Feasible	0.021	0.026	0.029	0.020	0.020	0.32
		(0.026)	(0.032)	(0.037)	(0.026)	(0.026)	-
	Infeasible	0.021	0.023	0.028	0.018	0.017	0.27
		(0.026)	(0.028)	(0.035)	(0.022)	(0.022)	-
$(1000,100)$	Naive	0.014	0.199	0.204	-	-	8.31
		(0.018)	(0.027)	(0.028)	-	-	-
	Feasible	0.014	0.017	0.019	0.012	0.013	0.13
		(0.017)	(0.021)	(0.024)	(0.016)	(0.017)	-
	Infeasible	0.014	0.015	0.018	0.012	0.012	0.11
		(0.017)	(0.019)	(0.022)	(0.015)	(0.015)	-
$(1000,200)$	Naive	0.010	0.201	0.205	-	-	8.34
		(0.012)	(0.020)	(0.019)	-	-	-
	Feasible	0.010	0.012	0.013	0.008	0.009	0.06
		(0.012)	(0.014)	(0.017)	(0.010)	(0.011)	-
	Infeasible	0.009	0.011	0.013	0.008	0.008	0.06
		(0.012)	(0.013)	(0.016)	(0.010)	(0.010)	-

Table 8: Monte Carlo estimates of the mean bias and standard deviation (in brackets) of different estimator for θ with Symmetric Equilibrium. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

(M,T)	Estimator	F	μ_1	μ_2	ω_2	ω_3	ω_4	A-MSE
$(500,50)$	Naive	0.028	0.304	0.308	-	-	-	19.57
		(0.035)	(0.061)	(0.056)	-	-	-	-
	Feasible	0.035	0.057	0.068	0.063	0.067	0.049	1.50
		(0.042)	(0.072)	(0.073)	(0.078)	(0.078)	(0.058)	-
	Infeasible	0.028	0.036	0.041	0.028	0.029	0.029	0.58
		(0.035)	(0.043)	(0.050)	(0.036)	(0.037)	(0.037)	-
$(500,100)$	Naive	0.021	0.301	0.307	-	-	-	18.93
		(0.026)	(0.044)	(0.041)	-	-	-	-
	Feasible	0.023	0.039	0.047	0.056	0.055	0.037	0.69
		(0.028)	(0.049)	(0.052)	(0.073)	(0.065)	(0.043)	-
	Infeasible	0.019	0.025	0.028	0.019	0.019	0.020	0.27
		(0.024)	(0.030)	(0.034)	(0.024)	(0.025)	(0.025)	-
$(500,200)$	Naive	0.014	0.301	0.306	-	-	-	18.63
		(0.018)	(0.030)	(0.027)	-	-	-	-
	Feasible	0.016	0.030	0.032	0.045	0.044	0.028	0.36
		(0.019)	(0.036)	(0.038)	(0.062)	(0.052)	(0.034)	-
	Infeasible	0.013	0.017	0.020	0.014	0.014	0.014	0.13
		(0.016)	(0.021)	(0.024)	(0.017)	(0.017)	(0.018)	-
$(1000,50)$	Naive	0.022	0.301	0.307	-	-	-	19.12
		(0.028)	(0.053)	(0.050)	-	-	-	-
	Feasible	0.022	0.033	0.035	0.027	0.029	0.027	0.45
		(0.027)	(0.041)	(0.044)	(0.039)	(0.040)	(0.035)	-
	Infeasible	0.021	0.026	0.030	0.020	0.021	0.021	0.32
		(0.026)	(0.033)	(0.037)	(0.025)	(0.026)	(0.026)	-
$(1000,100)$	Naive	0.015	0.301	0.306	-	-	-	18.73
		(0.019)	(0.037)	(0.034)	-	-	-	-
	Feasible	0.014	0.019	0.021	0.015	0.016	0.016	0.16
		(0.017)	(0.024)	(0.026)	(0.019)	(0.021)	(0.021)	-
	Infeasible	0.014	0.017	0.019	0.013	0.014	0.015	0.13
		(0.017)	(0.021)	(0.024)	(0.017)	(0.017)	(0.018)	-
$(1000,200)$	Naive	0.010	0.302	0.308	-	-	-	18.76
		(0.013)	(0.027)	(0.023)	-	-	-	-
	Feasible	0.009	0.013	0.014	0.010	0.011	0.011	0.07
		(0.012)	(0.016)	(0.018)	(0.013)	(0.014)	(0.013)	-
	Infeasible	0.009	0.011	0.014	0.010	0.010	0.009	0.06
		(0.012)	(0.014)	(0.017)	(0.012)	(0.012)	(0.012)	-

Table 9: Monte Carlo estimates of the mean bias and standard deviation (in brackets) of different estimator for θ. A-MSE and C-E% respectively are the sum of mean square errors of all estimators (scaled by 100) and percentage of clustering error.

Handles unobserved states by dropping states

(M, T)	Mean					Median				
	C					C				
	0,5	1	1,5	2	2,5	0,5	1	1,5	2	2,5
(500, 25)	5,4	4,2	3,3	2,7	2,7	6	5	4	3	3
(500, 50)	5,3	3,0	3,0	2,9	2,8	6	3	3	3	3
(500, 100)	4,8	3,5	3,0	3,0	3,0	5	4	3	3	3
(1000, 25)	5,9	4,6	3,7	2,9	2,9	6	5	4	3	3
(1000, 50)	4,8	3,0	2,9	2,8	2,6	5	3	3	3	3
(1000, 100)	4,4	3,3	3,0	3,0	3,0	4	3	3	3	3

		Probability of Optimal Cluster K =k being selected					
(M, T)	C	K					
		1	2	3	4	5	6
(500, 25)	0,5	0,0%	0,1%	3,9%	22,3%	1,7%	72,0%
	1	0,0%	8,7%	20,2%	19,1%	50,8%	1,2%
	1,5	0,0%	27,1%	20,0%	52,9%	0,0%	0,0%
	2	0,0%	28,9%	69,0%	2,1%	0,0%	0,0%
	2,5	1,6%	27,3%	71,1%	0,0%	0,0%	0,0%
(500, 50)	0,5	0,0%	0,0%	14,4%	6,5%	16,4%	62,7%
	1	0,0%	0,1%	96,6%	3,3%	0,0%	0,0%
	1,5	0,0%	2,7%	97,3%	0,0%	0,0%	0,0%
	2	0,0%	13,0%	87,0%	0,0%	0,0%	0,0%
	2,5	0,0%	24,9%	75,1%	0,0%	0,0%	0,0%
(500, 100)	0,5	0,0%	0,0%	3,1%	41,6%	24,5%	30,8%
	1	0,0%	0,0%	48,8%	50,9%	0,3%	0,0%
	1,5	0,0%	0,0%	97,5%	2,5%	0,0%	0,0%
	2	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%
	2,5	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%
(1000, 25)	0,5	0,0%	0,0%	0,4%	5,2%	0,0%	94,4%
	1	0,0%	1,0%	4,6%	30,6%	63,7%	0,1%
	1,5	0,0%	5,6%	21,6%	72,8%	0,0%	0,0%
	2	0,0%	5,6%	94,4%	0,0%	0,0%	0,0%
	2,5	0,3%	5,3%	94,4%	0,0%	0,0%	0,0%
(1000, 50)	0,5	0,0%	0,7%	29,7%	5,4%	17,1%	47,1%
	1	0,0%	0,7%	96,9%	2,4%	0,0%	0,0%
	1,5	0,0%	7,9%	92,1%	0,0%	0,0%	0,0%
	2	0,0%	22,6%	77,4%	0,0%	0,0%	0,0%
	2,5	0,0%	41,4%	58,6%	0,0%	0,0%	0,0%
(1000, 100)	0,5	0,0%	0,0%	3,6%	66,1%	18,5%	11,8%
	1	0,0%	0,0%	70,2%	29,8%	0,0%	0,0%
	1,5	0,0%	0,0%	99,9%	0,1%	0,0%	0,0%
	2	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%
	2,5	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%

(M, T)	Mean					Median				
	C					C				
	0,5	1	1,5	2	2,5	0,5	1	1,5	2	2,5
(500, 25)	2,8	2,6	2,4	2,3	2,2	3	3	2	2	2
(500, 50)	2,5	2,2	2,1	2,1	2,0	2	2	2	2	2
(500, 100)	2,2	2,1	2,0	2,0	2,0	2	2	2	2	2
(1000, 25)	2,7	2,4	2,3	2,2	2,1	3	2	2	2	2
(1000, 50)	2,3	2,1	2,0	2,0	2,0	2	2	2	2	2
(1000, 100)	2,1	2,0	2,0	2,0	2,0	2	2	2	2	2

Probability of Optimal Number of Theta					
(M, T)	C	Theta			
		1	2	3	4
(500, 25)	0,5	0,0%	20,7%	79,3%	0,0%
	1	0,0%	43,6%	56,4%	0,0%
	1,5	0,0%	59,5%	40,5%	0,0%
	2	0,0%	70,7%	29,3%	0,0%
	2,5	0,0%	78,5%	21,5%	0,0%
(500, 50)	0,5	0,0%	51,6%	46,6%	1,8%
	1	0,0%	77,5%	21,6%	0,9%
	1,5	0,0%	87,7%	11,9%	0,4%
	2	0,0%	93,1%	6,8%	0,1%
	2,5	0,0%	95,4%	4,6%	0,0%
(500, 100)	0,5	0,0%	77,6%	22,4%	0,0%
	1	0,0%	93,3%	6,7%	0,0%
	1,5	0,0%	97,9%	2,1%	0,0%
	2	0,0%	99,5%	0,5%	0,0%
	2,5	0,0%	100,0%	0,0%	0,0%
(1000, 25)	0,5	0,0%	29,7%	70,3%	0,0%
	1	0,0%	59,0%	41,0%	0,0%
	1,5	0,0%	73,4%	26,6%	0,0%
	2	0,0%	84,0%	16,0%	0,0%
	2,5	0,0%	90,1%	9,9%	0,0%
(1000, 50)	0,5	0,0%	70,6%	29,3%	0,1%
	1	0,0%	89,2%	10,7%	0,1%
	1,5	0,0%	95,9%	4,0%	0,1%
	2	0,0%	98,8%	1,1%	0,1%
	2,5	0,0%	99,3%	0,7%	0,0%
(1000, 100)	0,5	0,0%	90,9%	9,1%	0,0%
	1	0,0%	98,9%	1,1%	0,0%
	1,5	0,0%	99,9%	0,1%	0,0%
	2	0,0%	100,0%	0,0%	0,0%
	2,5	0,0%	100,0%	0,0%	0,0%

Conclusion

Concluding Remarks

- Propose to estimate dynamic models with unobserved heterogeneity by hard-classification of latent types to preserves all benefits of the two-step method.
- We can handle: macro common shocks, payoff heterogeneity, transition heterogeneity, and multiple equilibria.
- Performs well in simulations depending on sample size and separation of CCPs.
- Idea readily extends to ordered discrete games (e.g., Gowrisankaran, Lucarelli, Schmidt-Dengler, and Town (2018) and Gowrisankaran and Schmidt-Dengler (2025)).
- Extension to continuous observables is also possible, e.g., for continuous a_{mt} (Bajari, Benkard and Levin (2007), Srisuma (2013)) and/or with continuous x_{mt} (Srisuma and Linton (2012), Buchholz, Shum and Xu (2021)).

The End - THANK YOU!