Online appendix: Imposing equilibrium restrictions in the estimation of dynamic discrete games

Victor Aguirregabiria*
University of Toronto, CEPR

Mathieu Marcoux†
Université de Montréal, CIREQ, CIRANO

June 21, 2021

Abstract

This online appendix contains: (A) auxiliary theoretical results; (B) proofs of the results from the main text; (C) additional simulation results; and (D) the empirical results from the application.

Keywords: Dynamic discrete games; Nested pseudo-likelihood; Fixed point algorithms; Spectral algorithms; Convergence; Convergence selection bias.

JEL codes: C13; C57; C61; C73.

*150 St. George Street, Toronto, ON, M5S 3G7, Canada, victor.aguirregabiria@utoronto.ca.

†C.P. 6128, succ. Centre-Ville, Montreal, QC, H3C 3J7, Canada, mathieu.marcoux@umontreal.ca.

Appendix A Auxiliary results

One important feature of the sequence of data generating processes defined by $\mathbf{P}_M^0$ is that $\hat{Q}(\boldsymbol{\theta},\mathbf{P}) - Q^0(\boldsymbol{\theta},\mathbf{P}) = O_p\left(M^{-1/2}\right)$ as $M \to \infty$ under $\{\mathbf{P}_M^0 : M \geq 1\}$. This is a direct consequence of a weak law of large number for triangular arrays and $\mathbf{P}_M^0$ being in the $M^{-1/2}$ neighborhood of $\mathbf{P}^0$. To see this, let $\mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right]$ be the expectation of $\hat{Q}(\boldsymbol{\theta},\mathbf{P})$ when the data generating process corresponds to $\mathbf{P}_M^0$, i.e.

$$\mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right] = M^{-1}\sum_{m=1}^{M}\mathrm{E}_{\mathbf{P}_M^0}\left[\ln\left[\Psi\left(\mathbf{y}_m|\mathbf{x}_m,\boldsymbol{\theta},\mathbf{P}\right)\right]\right] \tag{1}$$

$$= \sum_{\mathbf{y},\mathbf{x}}\ln\left[\Psi\left(\mathbf{y}|\mathbf{x},\boldsymbol{\theta},\mathbf{P}\right)\right]P_M^0\left(\mathbf{y}|\,\mathbf{x}\right). \tag{2}$$

By a weak law of large number for triangular arrays, $\hat{Q}(\boldsymbol{\theta},\mathbf{P}) - \mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right] \xrightarrow{p} 0$. For given $(\boldsymbol{\theta},\mathbf{P})$, $\mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right]$ is a function of $\mathbf{P}_M^0$. Then, a first-order Taylor expansion of $\mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right]$ for values of $\mathbf{P}_M^0$ around $\mathbf{P}^0$ gives:

$$\begin{aligned}\mathrm{E}_{\mathbf{P}_M^0}\left[\hat{Q}(\boldsymbol{\theta},\mathbf{P})\right] =& \sum_{\mathbf{y},\mathbf{x}}\ln\left[\Psi\left(\mathbf{y}|\mathbf{x},\boldsymbol{\theta},\mathbf{P}\right)\right]P^0\left(\mathbf{y}|\,\mathbf{x}\right)\\ &+\sum_{\mathbf{y},\mathbf{x}}\ln\left[\Psi\left(\mathbf{y}|\mathbf{x},\boldsymbol{\theta},\mathbf{P}\right)\right]\left(P_M^0\left(\mathbf{y}|\,\mathbf{x}\right) - P^0\left(\mathbf{y}|\,\mathbf{x}\right)\right) + O\left(M^{-1}\right).\end{aligned} \tag{3}$$

Since $\sum_{\mathbf{y},\mathbf{x}}\ln\left[\Psi\left(\mathbf{y}|\mathbf{x},\boldsymbol{\theta},\mathbf{P}\right)\right]P^0\left(\mathbf{y}|\,\mathbf{x}\right) = Q^0(\boldsymbol{\theta},\mathbf{P})$ and $\mathbf{P}_M^0 - \mathbf{P}^0 = \mathbf{c}_{\mathbf{P}}/\sqrt{M}$, we conclude that $\hat{Q}(\boldsymbol{\theta},\mathbf{P}) - Q^0(\boldsymbol{\theta},\mathbf{P}) = O_p\left(M^{-1/2}\right)$ as $M \to \infty$ under $\{\mathbf{P}_M^0 : M \geq 1\}$.

It will also be useful to characterize the asymptotic distribution of $\nabla_{\boldsymbol{\theta}}\hat{Q}_{\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)}$, i.e. the gradient of the pseudo maximum likelihood evaluated at $\boldsymbol{\vartheta}_M^0(\mathbf{P})$ given $\mathbf{P}$. Let $\boldsymbol{\Omega}^0_{M,\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)}$ and $\boldsymbol{\Omega}^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}$ be defined as follows:

$$\boldsymbol{\Omega}^0_{M,\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)} \equiv \mathrm{E}_{\mathbf{P}_M^0}\left[\frac{\partial\ln\left[\Psi\left(\mathbf{y}_m|\,\mathbf{x}_m,\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)\right]}{\partial\boldsymbol{\theta}}\frac{\partial\ln\left[\Psi\left(\mathbf{y}_m|\,\mathbf{x}_m,\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)\right]}{\partial\boldsymbol{\theta}'}\right] \tag{4}$$

$$\boldsymbol{\Omega}^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \equiv \mathrm{E}_{\mathbf{P}^0}\left[\frac{\partial\ln\left[\Psi\left(\mathbf{y}_m|\,\mathbf{x}_m,\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)\right]}{\partial\boldsymbol{\theta}}\frac{\partial\ln\left[\Psi\left(\mathbf{y}_m|\,\mathbf{x}_m,\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)\right]}{\partial\boldsymbol{\theta}'}\right]. \tag{5}$$

Since $\nabla_{\boldsymbol{\theta}}Q^0_{M,\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)} = \mathbf{0}$, a central limit theorem for triangular arrays gives:

$$\sqrt{M}\nabla_{\boldsymbol{\theta}}\hat{Q}_{\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P}\right)} \xrightarrow{d} \mathrm{Normal}\left(\mathbf{0},\boldsymbol{\Omega}^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\right). \tag{6}$$

The following Lemma is used to state the results from the main text.

Lemma A1 (Local asymptotic distribution of the sample NPL mapping). *Let Assumptions 1 and 2 be satisfied and let* $\mathbf{P}_M^0 = \mathbf{P}^0 + \mathbf{c_P}/\sqrt{M}$ *for some unknown constant vector* $\mathbf{c_P}$*. Then, for any* $\mathbf{P} \in \mathcal{P}$*, as* $M \to \infty$ *under* $\{\mathbf{P}_M^0 : M \geq 1\}$*, the pseudo-maximum likelihood estimator* $\hat{\boldsymbol{\vartheta}}(\mathbf{P})$ *has limiting distribution:*

$$\sqrt{M}\left(\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right) \xrightarrow{d} \boldsymbol{\xi}_{\boldsymbol{\vartheta}}(\mathbf{P}) \sim Normal\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}(\mathbf{P})\right)$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}(\mathbf{P}) \equiv \left[\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})}\right]^{-1} \boldsymbol{\Omega}^0_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})} \left[\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})}\right]^{-1}$*. The mapping* $\hat{\boldsymbol{\phi}}(\mathbf{P}) = \mathbf{P} - \hat{\boldsymbol{\varphi}}(\mathbf{P})$ *has a limiting distribution such that, as* $M \to \infty$ *under* $\{\mathbf{P}_M^0 : M \geq 1\}$*:*

$$\sqrt{M}\left(\hat{\boldsymbol{\phi}}(\mathbf{P}) - \boldsymbol{\phi}_M^0(\mathbf{P})\right) = -\sqrt{M}\left(\hat{\boldsymbol{\varphi}}(\mathbf{P}) - \boldsymbol{\varphi}_M^0(\mathbf{P})\right) \xrightarrow{d} \boldsymbol{\xi}_{\boldsymbol{\varphi}}(\mathbf{P}) \sim Normal\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varphi}}(\mathbf{P})\right)$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}}(\mathbf{P}) \equiv \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})} \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}(\mathbf{P}) \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})}$*,* $\boldsymbol{\phi}_M^0(\mathbf{P}) = \mathbf{P} - \boldsymbol{\varphi}_M^0(\mathbf{P})$ *and* $\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}), \mathbf{P}\right)$.

Proof. The first part simply follows from standard pseudo maximum likelihood asymptotic arguments. To see this, consider a stochastic first-order expansion of $\nabla_{\boldsymbol{\theta}} \hat{Q}_{(\hat{\boldsymbol{\vartheta}}(\mathbf{P}),\mathbf{P})} = \mathbf{0}$ around $\boldsymbol{\vartheta}_M^0(\mathbf{P})$:

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} + \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} \left(\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right) + O_p\left(\left\|\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right\|^2\right). \tag{7}$$

Since $\hat{Q}(\boldsymbol{\theta}, \mathbf{P}) - Q^0(\boldsymbol{\theta}, \mathbf{P}) = O_p\left(M^{-1/2}\right)$, we have that:

$$\left(\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} - \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})}\right)\left(\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right) \leq O_p\left(\left\|\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right\|^2\right) \tag{8}$$

By rearranging, we get:

$$\begin{aligned}\sqrt{M}\left(\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right) = &- \left[\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})}\right]^{-1} \sqrt{M} \nabla_{\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} \\ &+ O_p\left(\left\|\hat{\boldsymbol{\vartheta}}(\mathbf{P}) - \boldsymbol{\vartheta}_M^0(\mathbf{P})\right\|^2\right).\end{aligned} \tag{9}$$

The result follows from noting that $\sqrt{M}\nabla_{\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \boldsymbol{\Omega}^0_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})}\right)$ and $\boldsymbol{\theta}_M^0 \to \boldsymbol{\theta}^0$ imply that $\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}_M^0(\mathbf{P}),\mathbf{P})} \to \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Q^0_{(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P})}$. For the second result, notice that, by definition of the NPL mapping:

$$\sqrt{M}\left(\hat{\boldsymbol{\phi}}(\mathbf{P}) - \boldsymbol{\phi}_M^0(\mathbf{P})\right) = -\sqrt{M}\left(\boldsymbol{\Psi}\left(\hat{\boldsymbol{\vartheta}}(\mathbf{P}), \mathbf{P}\right) - \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}), \mathbf{P}\right)\right) \tag{10}$$

which is therefore, given $\mathbf{P}$, a continuous nonlinear transformation of a maximum likelihood estimator. A standard delta method argument combined with $\boldsymbol{\theta}_M^0 \to \boldsymbol{\theta}^0$ gives the multivariate normal distribution of $\boldsymbol{\xi}_{\boldsymbol{\varphi}}(\mathbf{P})$ stated in Lemma A1. □

Appendix B Proofs of results in the main text

B.1 Proof of Lemma 1

(i) Proof of Lemma 1(*A*). For any $\mathbf{P}_M^0$ in the sequence $\{\mathbf{P}_M^0 : M \geq 1\}$, we have that $\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}), \mathbf{P}\right)$ is a continuous mapping from $\mathcal{P}$ to itself. By Brouwer's fixed point theorem, $\mathcal{F}_M^0$ is nonempty for each $\mathbf{P}_M^0$. Therefore, the sequence $\{\mathbf{P}_M^0 : M \geq 1\}$ defines a sequence of nonempty sets $\{\mathcal{F}_M^0 : M \geq 1\}$. Since $\mathbf{P}_M^0$ converges to $\mathbf{P}^0$, we have that $\boldsymbol{\vartheta}_M^0(\mathbf{P}) \to \boldsymbol{\vartheta}^0(\mathbf{P})$, $\boldsymbol{\varphi}_M^0(\mathbf{P}) \to \boldsymbol{\varphi}^0(\mathbf{P})$, and $\mathcal{F}_M^0 \to \mathcal{F}^0$.

Furthermore, every point in the set $\mathcal{F}_M^0$ belongs to a small open ball around a point in set $\mathcal{F}^0$. For a proof of this result, see Aguirregabiria and Mira (2007), pages 46-47, Step 2 in the proof of consistency of the NPL estimator. Therefore, each fixed point $\mathbf{P}_{M*}^0$ in set $\mathcal{F}_M^0$ converges to a well-defined fixed point in the set $\mathcal{F}^0$, that we represent as $\mathbf{P}_*^0$.

(ii) Proof of Lemma 1(*B*). First, we show that $\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\varphi}^0(\mathbf{P}) + \mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P})}/\sqrt{M} + o\left(M^{-1/2}\right)$ for some unknown vector of constants $\mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P})}$ and any $\mathbf{P} \in \mathcal{P}$. Since $\boldsymbol{\vartheta}_M^0(\mathbf{P})$ and $\boldsymbol{\vartheta}^0(\mathbf{P})$ respectively maximize $Q_M^0(\boldsymbol{\theta}, \mathbf{P})$ and $Q^0(\boldsymbol{\theta}, \mathbf{P})$, standard pseudo-likelihood arguments can be used to show that:

$$\boldsymbol{\vartheta}_M^0(\mathbf{P}) = \boldsymbol{\vartheta}^0(\mathbf{P}) - \left[\nabla^2_{\boldsymbol{\theta\theta}} Q^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\right]^{-1} \nabla_{\boldsymbol{\theta}} Q_{M,\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} + O\left(\left\|\boldsymbol{\vartheta}_M^0(\mathbf{P}) - \boldsymbol{\vartheta}^0(\mathbf{P})\right\|^2\right). \quad (11)$$

Let $\mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \equiv \text{diag}\left\{\boldsymbol{\Psi}\left(\boldsymbol{\vartheta}^0(\mathbf{P}), \mathbf{P}\right)\right\}^{-1}$. Then, we have that:

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} Q^0_{M,\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} &= \sum_{\mathbf{x},\mathbf{y}} \frac{P_M^0(\mathbf{y}|\,\mathbf{x})}{\Psi\left(\mathbf{y}|\,\mathbf{x}, \boldsymbol{\vartheta}^0(\mathbf{P}), \mathbf{P}\right)} \left.\frac{\partial \Psi(\mathbf{y}|\,\mathbf{x}, \boldsymbol{\theta}, \mathbf{P})}{\partial \boldsymbol{\theta}}\right|_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \\ &= \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \mathbf{P}_M^0 \end{aligned} \quad (12)$$

and similarly $\nabla_{\boldsymbol{\theta}} Q^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \mathbf{P}^0$. Since $\nabla_{\boldsymbol{\theta}} Q^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} = \mathbf{0}$:

$$\nabla_{\boldsymbol{\theta}} Q_{M,\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \left(\mathbf{P}_M^0 - \mathbf{P}^0\right) = \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \frac{\mathbf{c}_{\mathbf{P}}}{\sqrt{M}}. \quad (13)$$

We can therefore write $\boldsymbol{\vartheta}_M^0(\mathbf{P}) = \boldsymbol{\vartheta}^0(\mathbf{P}) + \mathbf{c}_{\boldsymbol{\vartheta}^0(\mathbf{P})}/\sqrt{M} + o\left(M^{-1/2}\right)$ where:

$$\mathbf{c}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \equiv -\left[\nabla^2_{\boldsymbol{\theta\theta}} Q^0_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\right]^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \mathbf{D}_{\boldsymbol{\vartheta}^0(\mathbf{P})} \mathbf{c}_{\mathbf{P}}. \quad (14)$$

By definition, $\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}), \mathbf{P}\right)$ and $\boldsymbol{\varphi}^0(\mathbf{P}) = \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}^0(\mathbf{P}), \mathbf{P}\right)$. By applying a standard delta method argument, we have:

$$\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\varphi}^0(\mathbf{P}) + \nabla_{\boldsymbol{\theta}}\boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\left(\boldsymbol{\vartheta}_M^0(\mathbf{P}) - \boldsymbol{\vartheta}^0(\mathbf{P})\right) + O\left(\left\|\boldsymbol{\vartheta}_M^0(\mathbf{P}) - \boldsymbol{\vartheta}^0(\mathbf{P})\right\|^2\right) \tag{15}$$

and $\boldsymbol{\varphi}_M^0(\mathbf{P}) = \boldsymbol{\varphi}^0(\mathbf{P}) + \mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P})}/\sqrt{M} + o\left(M^{-1/2}\right)$ with $\mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P})} \equiv \nabla_{\boldsymbol{\theta}}\boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\mathbf{c}_{\boldsymbol{\vartheta}^0(\mathbf{P})}$.

Second, consider a first-order expansion of $\boldsymbol{\varphi}_M^0(\mathbf{P}_{M*}^0)$ around $\mathbf{P}_*^0$:

$$\boldsymbol{\varphi}_M^0\left(\mathbf{P}_{M*}^0\right) = \boldsymbol{\varphi}_M^0\left(\mathbf{P}_*^0\right) + \nabla\boldsymbol{\varphi}_{M,(\mathbf{P}_*^0)}^0\left(\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right) + O\left(\left\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right\|^2\right). \tag{16}$$

Since $\boldsymbol{\varphi}_M^0(\mathbf{P}_{M*}^0) = \mathbf{P}_{M*}^0$ and $\boldsymbol{\varphi}^0(\mathbf{P}_*^0) = \mathbf{P}_*^0$, we can rewrite the previous equation as follows:

$$\begin{aligned}[\mathbf{I} - \nabla\boldsymbol{\varphi}_{(\mathbf{P}_*^0)}^0](\mathbf{P}_{M*}^0 - \mathbf{P}_*^0) = {} & \boldsymbol{\varphi}_M^0\left(\mathbf{P}_*^0\right) - \boldsymbol{\varphi}^0\left(\mathbf{P}_*^0\right) \\ & + [\nabla\boldsymbol{\varphi}_{M,(\mathbf{P}_*^0)}^0 - \nabla\boldsymbol{\varphi}_{(\mathbf{P}_*^0)}^0]\left(\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right) + O\left(\left\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right\|^2\right).\end{aligned} \tag{17}$$

On the left-hand side, the matrix $\mathbf{I} - \nabla\boldsymbol{\varphi}_{(\mathbf{P}_*^0)}^0$ is generically invertible. On the right-hand side, $\boldsymbol{\varphi}_M^0(\mathbf{P}_*^0) - \boldsymbol{\varphi}^0(\mathbf{P}_*^0) = \mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P}_*^0)}/\sqrt{M} + o\left(M^{-1/2}\right)$ and $[\nabla\boldsymbol{\varphi}_{M,(\mathbf{P}_*^0)}^0 - \nabla\boldsymbol{\varphi}_{(\mathbf{P}_*^0)}^0](\mathbf{P}_{M*}^0 - \mathbf{P}_*^0) \leq O\left(\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\|^2\right)$. Therefore, this equation implies that $\mathbf{P}_{M*}^0 = \mathbf{P}_*^0 + \mathbf{c}_{\mathbf{P}}^*/\sqrt{M} + o\left(M^{-1/2}\right)$ where:

$$\mathbf{c}_{\mathbf{P}}^* \equiv [\mathbf{I} - \nabla\boldsymbol{\varphi}_{(\mathbf{P}_*^0)}^0]^{-1}\mathbf{c}_{\boldsymbol{\varphi}(\mathbf{P}_*^0)}. \tag{18}$$

To show $\boldsymbol{\theta}_{M*}^0 = \boldsymbol{\theta}_*^0 + \mathbf{c}_{\boldsymbol{\theta}}^*/\sqrt{M} + o\left(M^{-1/2}\right)$, we start by considering a first-order expansion of $\nabla_{\boldsymbol{\theta}}Q_{M,\left(\boldsymbol{\theta}_{M*}^0,\mathbf{P}_{M*}^0\right)}^0 = \mathbf{0}$ around $\boldsymbol{\theta}_*^0$ and $\mathbf{P}_*^0$:

$$\begin{aligned}\mathbf{0} = {} & \nabla_{\boldsymbol{\theta}}Q_{M,\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0 + \nabla_{\boldsymbol{\theta\theta}}^2 Q_{M,\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0\left(\boldsymbol{\theta}_{M*}^0 - \boldsymbol{\theta}_*^0\right) \\ & + \nabla_{\boldsymbol{\theta}\mathbf{P}}^2 Q_{M,\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0\left(\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right) + O\left(\left\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right\|^2\right).\end{aligned} \tag{19}$$

Notice that $\left(\nabla_{\boldsymbol{\theta\theta}}^2 Q_{M,\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0 - \nabla_{\boldsymbol{\theta\theta}}^2 Q_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0\right)\left(\boldsymbol{\theta}_{M*}^0 - \boldsymbol{\theta}_*^0\right) \leq O\left(\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\|^2\right)$ and that similarly $\left(\nabla_{\boldsymbol{\theta}\mathbf{P}}^2 Q_{M,\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0 - \nabla_{\boldsymbol{\theta}\mathbf{P}}^2 Q_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0\right)\left(\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right) \leq O_p\left(\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\|^2\right)$. Furthermore, one can write:

$$\nabla_{\boldsymbol{\theta}}Q_{M\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0 = \sum_{\mathbf{x},\mathbf{y}}\frac{P_M^0(\mathbf{y}|\,\mathbf{x})}{P_*^0(\mathbf{y}|\,\mathbf{x})}\left.\frac{\partial\Psi(\mathbf{y}|\,\mathbf{x},\boldsymbol{\theta},\mathbf{P})}{\partial\boldsymbol{\theta}}\right|_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)} = \nabla_{\boldsymbol{\theta}}\boldsymbol{\Psi}'_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}\mathbf{D}_*^0\mathbf{P}_M^0 \tag{20}$$

$$\nabla_{\boldsymbol{\theta}}Q_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}^0 = \sum_{\mathbf{x},\mathbf{y}}\frac{P^0(\mathbf{y}|\,\mathbf{x})}{P_*^0(\mathbf{y}|\,\mathbf{x})}\left.\frac{\partial\Psi(\mathbf{y}|\,\mathbf{x},\boldsymbol{\theta},\mathbf{P})}{\partial\boldsymbol{\theta}}\right|_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)} = \nabla_{\boldsymbol{\theta}}\boldsymbol{\Psi}'_{\left(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0\right)}\mathbf{D}_*^0\mathbf{P}^0 = \mathbf{0} \tag{21}$$

where $\mathbf{D}_*^0 \equiv \text{diag}\left\{\mathbf{P}_*^0\right\}^{-1}$. We can therefore write:

$$\begin{aligned}\boldsymbol{\theta}_{M*}^0 =& \boldsymbol{\theta}_*^0 - \left[\nabla^2_{\boldsymbol{\theta\theta}} Q^0_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)}\right]^{-1} \left\{\nabla^2_{\boldsymbol{\theta}\mathbf{P}} Q^0_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)} \left(\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right)\right. \\ & \left. + \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)} \mathbf{D}_*^0 \left(\mathbf{P}_M^0 - \mathbf{P}^0\right)\right\} + O\left(\left\|\mathbf{P}_{M*}^0 - \mathbf{P}_*^0\right\|^2\right)\end{aligned} \tag{22}$$

This last expression leads to $\boldsymbol{\theta}_{M*}^0 = \boldsymbol{\theta}_*^0 + \mathbf{c}_{\boldsymbol{\theta}}^* / \sqrt{M} + o\left(M^{-1/2}\right)$ where:

$$\mathbf{c}_{\boldsymbol{\theta}}^* \equiv -\left[\nabla^2_{\boldsymbol{\theta\theta}} Q^0_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)}\right]^{-1} \left\{\nabla^2_{\boldsymbol{\theta}\mathbf{P}} Q^0_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)} \mathbf{c}_{\mathbf{P}}^* + \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}'_{(\boldsymbol{\theta}_*^0,\mathbf{P}_*^0)} \mathbf{D}_*^0 \mathbf{c}_{\mathbf{P}}\right\}. \tag{23}$$

(iii) Proof of Lemma 1(C). Following Aguirregabiria and Mira (2007, pages 46-47), we can establish that, with probability approaching one, any element $\hat{\mathbf{P}}_*$ in the set $\widehat{\mathcal{F}}$ belongs to a small open ball around an element in the set $\mathcal{F}_M^0$, that we denote as $\mathbf{P}_{M*}^0$. Therefore, under the data generating process $\mathbf{P}_M^0$, we have that $\hat{\mathbf{P}}_* \xrightarrow{p} \mathbf{P}_{M*}^0$.

Using an argument similar as in the proof of Lemma 1(B), but based on a first order expansion of $\hat{\boldsymbol{\phi}}\left(\hat{\mathbf{P}}_*\right) = \mathbf{0}$ around $\mathbf{P}_{M*}^0$:

$$\hat{\mathbf{P}}_* - \mathbf{P}_{M*}^0 = -\left[\nabla \boldsymbol{\phi}^0_{(\mathbf{P}_*^0)}\right]^{-1} \left(\hat{\boldsymbol{\phi}}\left(\mathbf{P}_{M*}^0\right) - \boldsymbol{\phi}_M^0\left(\mathbf{P}_{M*}^0\right)\right) + O_p\left(\left\|\hat{\mathbf{P}}_* - \mathbf{P}_{M*}^0\right\|^2\right). \tag{24}$$

By applying Lemma A1, $\sqrt{M}\left(\hat{\boldsymbol{\phi}}\left(\mathbf{P}_{M*}^0\right) - \boldsymbol{\phi}_M^0\left(\mathbf{P}_{M*}^0\right)\right) \xrightarrow{d} \boldsymbol{\xi}_{\varphi}\left(\mathbf{P}_{M*}^0\right)$. Moreover, since $\mathbf{P}_{M*}^0 \to \mathbf{P}_*^0$, $\boldsymbol{\xi}_{\varphi}\left(\mathbf{P}_{M*}^0\right)$ has the same asymptotic distribution as $\boldsymbol{\xi}_{\varphi}\left(\mathbf{P}_*^0\right)$, i.e. Normal$\left(\mathbf{0}, \boldsymbol{\Sigma}_{\varphi}\left(\mathbf{P}_*^0\right)\right)$. We therefore obtain $\sqrt{M}\left(\hat{\mathbf{P}}_* - \mathbf{P}_{M*}^0\right) \xrightarrow{d} \boldsymbol{\xi}_{\mathbf{P}}\left(\mathbf{P}_*^0\right)$ where $\boldsymbol{\xi}_{\mathbf{P}}\left(\mathbf{P}_*^0\right) = -\left[\nabla \boldsymbol{\phi}^0_{(\mathbf{P}_*^0)}\right]^{-1} \boldsymbol{\xi}_{\varphi}\left(\mathbf{P}_*^0\right)$ is a vector of normal variables with zero means. To derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}_{M*}^0$, consider the following expansion of $\nabla_{\boldsymbol{\theta}} \hat{Q}_{(\hat{\boldsymbol{\theta}}_*,\hat{\mathbf{P}}_*)} = \mathbf{0}$ around $\boldsymbol{\theta}_{M*}^0$ and $\mathbf{P}_{M*}^0$:

$$\begin{aligned}\mathbf{0} =& \nabla_{\boldsymbol{\theta}} \hat{Q}_{(\boldsymbol{\theta}_{M*}^0,\mathbf{P}_{M*}^0)} + \nabla^2_{\boldsymbol{\theta\theta}} \hat{Q}_{(\boldsymbol{\theta}_{M*}^0,\mathbf{P}_{M*}^0)} \left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}_{M*}^0\right) \\ & + \nabla^2_{\boldsymbol{\theta}\mathbf{P}} \hat{Q}_{(\boldsymbol{\theta}_{M*}^0,\mathbf{P}_{M*}^0)} \left(\hat{\mathbf{P}}_* - \mathbf{P}_{M*}^0\right) + O_p\left(\left\|\hat{\mathbf{P}}_* - \mathbf{P}_{M*}^0\right\|^2\right).\end{aligned} \tag{25}$$

Let the empirical measure be denoted by $\mathbb{P}$. More precisely, let:

$$\mathbb{P} = \left\{\mathbb{P}\left(\mathbf{y}|\,\mathbf{x}\right) = \frac{\sum_{m=1}^M \mathbb{1}\left\{\mathbf{y}_m = \mathbf{y}\right\} \mathbb{1}\left\{\mathbf{x}_m = \mathbf{x}\right\}}{\sum_{m=1}^M \mathbb{1}\left\{\mathbf{x}_m = \mathbf{x}\right\}} : \mathbf{y} \in \mathcal{Y}^N, \mathbf{x} \in \mathcal{X}\right\} \tag{26}$$

with the elements of $\mathbb{P}$ ordered in the same way as in $\mathbf{P}$. A triangular array central limit theorem can be used to show that, as $M \to \infty$ under $\left\{\mathbf{P}_M^0 : M \geq 1\right\}$:

$$\sqrt{M}\left(\mathbb{P} - \mathbf{P}_M^0\right) \xrightarrow{d} \boldsymbol{\xi}_{\mathbb{P}} \sim \text{Normal}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbb{P}}\right). \tag{27}$$

The variance-covariance matrix $\mathbf{\Sigma}_{\mathbb{P}}$ is a block diagonal matrix with $\mathbf{x}$-specific blocks corresponding to:

$$\frac{1}{P^0(\mathbf{x})}\left[\text{diag}\left\{\mathbf{P}^0(\mathbf{y}|\,\mathbf{x})\right\} - \mathbf{P}^0(\mathbf{y}|\,\mathbf{x})\,\mathbf{P}^0(\mathbf{y}|\,\mathbf{x})'\right] \tag{28}$$

where $P^0(\mathbf{x})$ is the probability of observing $\mathbf{x}$ under $\mathbf{P}^0$ and covariances between $\mathbb{P}(\mathbf{y}|\,\mathbf{x})$ for different values of $\mathbf{x}$ are 0. Once again using arguments similar as in the proof of 1(*B*), we can write:

$$\begin{aligned}\mathbf{0} =& \nabla_{\boldsymbol{\theta}}\mathbf{\Psi}'_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\mathbf{D}^0_*\left(\mathbb{P}-\mathbf{P}^0_M\right) + \nabla^2_{\boldsymbol{\theta\theta}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_{M*}\right)\\ &+ \nabla^2_{\boldsymbol{\theta}\mathbf{P}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\left(\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right) + O_p\left(\left\|\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right\|^2\right).\end{aligned} \tag{29}$$

Solving for $\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_{M*}$, provided that $\nabla^2_{\boldsymbol{\theta\theta}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}$ is invertible, one gets:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_{M*} =& -\left[\nabla^2_{\boldsymbol{\theta\theta}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\right]^{-1}\left\{\nabla^2_{\boldsymbol{\theta}\mathbf{P}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\left(\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right)\right.\\ &\left.+\nabla_{\boldsymbol{\theta}}\mathbf{\Psi}'_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\mathbf{D}^0_*\left(\mathbb{P}-\mathbf{P}^0_M\right)\right\} + O_p\left(\left\|\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right\|^2\right).\end{aligned} \tag{30}$$

As $M \to \infty$ under $\{\mathbf{P}^0_M : M \geq 1\}$, we have that $\sqrt{M}\left(\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right) \xrightarrow{d} \boldsymbol{\xi}_{\mathbf{P}}(\mathbf{P}^0_*)$. It follows that:

$$\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{M*} - \boldsymbol{\theta}^0_{M*}\right) \xrightarrow{d} \boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{P}^0_*) \tag{31}$$

where $\boldsymbol{\xi}_{\boldsymbol{\theta}}(\mathbf{P}^0_*) = -\left[\nabla^2_{\boldsymbol{\theta\theta}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\right]^{-1}\left[\nabla^2_{\boldsymbol{\theta}\mathbf{P}}Q^0_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\boldsymbol{\xi}_{\mathbf{P}}(\mathbf{P}^0_*) + \nabla_{\boldsymbol{\theta}}\mathbf{\Psi}'_{(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*)}\mathbf{D}^0_*\boldsymbol{\xi}_{\mathbb{P}}\right]$, which follows a mean-zero multivariate distribution since both $\boldsymbol{\xi}_{\mathbf{P}}(\mathbf{P}^0_*)$ and $\boldsymbol{\xi}_{\mathbb{P}}$ follow mean-zero multivariate normal distributions.

(iv) Proof of Lemma 1(*D*). It simply follows from noting that:

$$\hat{\mathbf{P}}_* - \mathbf{P}^0_* = \hat{\mathbf{P}}_* - \mathbf{P}^0_{M*} + \mathbf{P}^0_{M*} - \mathbf{P}^0_*. \tag{32}$$

Lemmas 1(*C*) and 1(*B*) imply that $\hat{\mathbf{P}}_* \xrightarrow{p} \mathbf{P}^0_{M*}$ and $\mathbf{P}^0_{M*} \to \mathbf{P}^0_*$, respectively. It follows that $\hat{\mathbf{P}}_* \xrightarrow{p} \mathbf{P}^0_*$ as required. A similar argument shows that $\hat{\boldsymbol{\theta}}_* \xrightarrow{p} \boldsymbol{\theta}^0_*$.

B.2 Proof of Lemma 2

As shown in the proof of Lemma 1, for any point $\hat{\mathbf{P}}_*$ in the set $\widehat{\mathcal{F}}$ there exists $\mathbf{P}^0_{M*} \in \mathcal{F}^0_M$ and $\mathbf{P}^0_* \in \mathcal{F}^0$ such that, as $M \to \infty$, we have that $\hat{\mathbf{P}}_* \xrightarrow{p} \mathbf{P}^0_{M*}$, and $\mathbf{P}^0_{M*} \to \mathbf{P}^0_*$.

(i) Proof of Lemma 2(*A*). We can write:

$$\rho^0_{M*} - \rho^0_* = \rho^0_{M*} - \rho\left(\nabla\boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)}\right) + \rho\left(\nabla\boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)}\right) - \rho^0_*. \tag{33}$$

A standard delta method argument implies that:

$$\rho^0_{M*} - \rho\left(\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)}\right) = \nabla_{\mathbf{P}} \rho^0_{(\mathbf{P}^0_*)}{}' \left(\mathbf{P}^0_{M*} - \mathbf{P}^0_*\right) + O\left(\left\|\mathbf{P}^0_{M*} - \mathbf{P}^0_*\right\|^2\right) \tag{34}$$

$$= \nabla_{\mathbf{P}} \rho^0_{(\mathbf{P}^0_*)}{}' \frac{\mathbf{c}^*_{\mathbf{P}}}{\sqrt{M}} + o\left(M^{-1/2}\right) \tag{35}$$

where $\nabla_{\mathbf{P}} \rho^0_{(\mathbf{P}^0_*)}$ is the derivative of $\rho\left(\nabla \boldsymbol{\varphi}^0_{(\mathbf{P})}\right)$ with respect to $\mathbf{P}$ evaluated at $\mathbf{P}^0_*$. Another delta method argument implies that:

$$\rho\left(\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)}\right) - \rho^0_* = \text{vec}\left[\nabla \rho^0_{(\mathbf{P}^0_*)}\right]' \text{vec}\left[\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)} - \nabla \boldsymbol{\varphi}^0_{(\mathbf{P}^0_*)}\right] + O_p\left(\left\|\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)} - \nabla \boldsymbol{\varphi}^0_{(\mathbf{P}^0_*)}\right\|^2\right). \tag{36}$$

where $\nabla \rho^0_{(\mathbf{P}^0_*)}$ is the derivative of $\rho\left(\nabla \boldsymbol{\varphi}^0_{(\mathbf{P})}\right)$ with respect to the elements of $\nabla \boldsymbol{\varphi}^0_{(\mathbf{P})}$ evaluated at $\mathbf{P}^0_*$. Let $\nabla^2_j \boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}$ be the derivative of $\nabla \boldsymbol{\varphi}^0_{(\mathbf{P})} = \nabla \boldsymbol{\Psi}\left(\boldsymbol{\vartheta}^0\left(\mathbf{P}\right), \mathbf{P}\right)$ with respect to the j-th element of $\boldsymbol{\theta}$ evaluated at $\left(\boldsymbol{\vartheta}^0\left(\mathbf{P}\right), \mathbf{P}\right)$. Once again using the delta method leads to:

$$\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P})} - \nabla \boldsymbol{\varphi}^0_{(\mathbf{P})} = \sum_{j=1}^{J} \nabla^2_j \boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \left(\vartheta^0_{M,j}\left(\mathbf{P}\right) - \vartheta^0_j\left(\mathbf{P}\right)\right) + O_p\left(\left\|\left(\boldsymbol{\vartheta}^0_M\left(\mathbf{P}\right) - \boldsymbol{\vartheta}^0\left(\mathbf{P}\right)\right)\right\|^2\right). \tag{37}$$

It follows that:

$$\rho\left(\nabla \boldsymbol{\varphi}^0_{M,(\mathbf{P}^0_*)}\right) - \rho^0_* = \text{vec}\left[\nabla \rho^0_{(\mathbf{P}^0_*)}\right]' \text{vec}\left[\sum_{j=1}^{J} \nabla^2_j \boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} \frac{c_{\boldsymbol{\vartheta}^0(\mathbf{P}),j}}{\sqrt{M}}\right] + o\left(M^{-1/2}\right) \tag{38}$$

where $c_{\boldsymbol{\vartheta}^0(\mathbf{P}),j}$ is the j-th element of the vector $\mathbf{c}_{\boldsymbol{\vartheta}^0(\mathbf{P})}$. By using (33), (35) and (38), we have that $\rho^0_{M*} = \rho^0_* + c^*_\rho/\sqrt{M} + o\left(M^{-1/2}\right)$ where:

$$c^*_\rho = \nabla_{\mathbf{P}} \rho^0_{(\mathbf{P}^0_*)}{}' \mathbf{c}^*_{\mathbf{P}} + \text{vec}\left[\nabla \rho^0_{(\mathbf{P}^0_*)}\right]' \text{vec}\left[\sum_{j=1}^{J} \nabla^2_j \boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)} c_{\boldsymbol{\vartheta}^0(\mathbf{P}),j}\right]. \tag{39}$$

(ii) Proof of Lemma 2(B). The random variable $\sqrt{M}\left(\hat{\rho}_* - \rho^0_{M*}\right)$ is equivalent to:

$$\sqrt{M}\left(\hat{\rho}_* - \rho^0_{M*}\right) = \sqrt{M}\left(\hat{\rho}_* - \rho\left(\nabla \hat{\boldsymbol{\varphi}}_{(\mathbf{P}^0_{M*})}\right)\right) + \sqrt{M}\left(\rho\left(\nabla \hat{\boldsymbol{\varphi}}_{(\mathbf{P}^0_{M*})}\right) - \rho^0_{M*}\right). \tag{40}$$

Similarly as in the proof of Lemma 2(A), the result follows from several applications of delta method arguments. First, notice that:

$$\sqrt{M}\left(\hat{\rho}_* - \rho\left(\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)}\right)\right) = \nabla_{\mathbf{P}}\rho^0_{\left(\mathbf{P}^0_*\right)}{}'\sqrt{M}\left(\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right) + O_p\left(\sqrt{M}\left\|\hat{\mathbf{P}}_* - \mathbf{P}^0_{M*}\right\|^2\right). \quad (41)$$

Using Lemma 1(*C*), we have that as $M \to \infty$ under $\{\mathbf{P}^0_M : M \geq 1\}$:

$$\sqrt{M}\left(\hat{\rho}_* - \rho\left(\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)}\right)\right) \xrightarrow{d} \nabla_{\mathbf{P}}\rho^0_{\left(\mathbf{P}^0_*\right)}{}'\,\boldsymbol{\xi}_{\mathbf{P}}\left(\mathbf{P}^0_*\right). \quad (42)$$

Second, we can write:

$$\begin{aligned}\sqrt{M}\left(\rho\left(\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)}\right) - \rho^0_{M*}\right) =& \mathrm{vec}\left[\nabla\rho^0_{\left(\mathbf{P}^0_*\right)}\right]'\mathrm{vec}\left[\sqrt{M}\left(\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)} - \nabla\boldsymbol{\varphi}^0_{M,\left(\mathbf{P}^0_{M*}\right)}\right)\right] \\ &+ O_p\left(\sqrt{M}\left\|\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)} - \nabla\boldsymbol{\varphi}^0_{M,\left(\mathbf{P}^0_{M*}\right)}\right\|^2\right)\end{aligned} \quad (43)$$

where

$$\begin{aligned}\sqrt{M}\left(\nabla\hat{\boldsymbol{\varphi}}_{(\mathbf{P})} - \nabla\boldsymbol{\varphi}^0_{M,(\mathbf{P})}\right) =& \sum_{j=1}^{J}\nabla^2_j\boldsymbol{\Psi}_{\left(\boldsymbol{\vartheta}^0(\mathbf{P}),\mathbf{P}\right)}\sqrt{M}\left(\hat{\vartheta}_j\left(\mathbf{P}\right) - \vartheta^0_{M,j}\left(\mathbf{P}\right)\right) \\ &+ O_p\left(\sqrt{M}\left\|\left(\hat{\boldsymbol{\vartheta}}\left(\mathbf{P}\right) - \boldsymbol{\vartheta}^0_M\left(\mathbf{P}\right)\right)\right\|^2\right).\end{aligned} \quad (44)$$

Then, considering $\mathbf{P} = \mathbf{P}^0_{M*}$ and noting that $\nabla^2_j\boldsymbol{\Psi}_{\left(\boldsymbol{\theta}^0_{M*},\mathbf{P}^0_{M*}\right)} \to \nabla^2_j\boldsymbol{\Psi}_{\left(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*\right)}$, Lemma A1 gives:

$$\sqrt{M}\left(\nabla\hat{\boldsymbol{\varphi}}_{\left(\mathbf{P}^0_{M*}\right)} - \nabla\boldsymbol{\varphi}^0_{M,\left(\mathbf{P}^0_{M*}\right)}\right) \xrightarrow{d} \sum_{j=1}^{J}\nabla^2_j\boldsymbol{\Psi}_{\left(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*\right)}\,\xi_{\boldsymbol{\vartheta}j}\left(\mathbf{P}^0_*\right) \quad (45)$$

where $\xi_{\boldsymbol{\vartheta}j}\left(\mathbf{P}^0_*\right)$ is the j-th element of $\boldsymbol{\xi}_{\boldsymbol{\vartheta}}\left(\mathbf{P}^0_*\right)$.

By combining equations (40), (42) and (45), $\sqrt{M}\left(\hat{\rho}_* - \rho^0_{M*}\right) \xrightarrow{d} \xi_\rho\left(\mathbf{P}^0_*\right)$ where $\xi_\rho\left(\mathbf{P}^0_*\right)$ is equal to:

$$\nabla_{\mathbf{P}}\rho^0_{\left(\mathbf{P}^0_*\right)}{}'\boldsymbol{\xi}_{\mathbf{P}}\left(\mathbf{P}^0_*\right) + \mathrm{vec}\left[\nabla\rho^0_{\left(\mathbf{P}^0_*\right)}\right]'\left[\mathrm{vec}\left[\nabla^2_1\boldsymbol{\Psi}_{\left(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*\right)}\right],\ldots,\mathrm{vec}\left[\nabla^2_J\boldsymbol{\Psi}_{\left(\boldsymbol{\theta}^0_*,\mathbf{P}^0_*\right)}\right]\right]\boldsymbol{\xi}_{\boldsymbol{\vartheta}}\left(\mathbf{P}^0_*\right). \quad (46)$$

Since $\boldsymbol{\xi}_{\mathbf{P}}\left(\mathbf{P}^0_*\right)$ and $\boldsymbol{\xi}_{\boldsymbol{\vartheta}}\left(\mathbf{P}^0_*\right)$ are normally distributed and centered at vectors of 0's, we conclude that $\sqrt{M}\left(\hat{\rho}_* - \rho^0_{M*}\right)$ follows a normal distribution centered at 0. The variance of this distribution is denoted $\sigma^2_{\rho^0_*}$.

(iii) Proof of Lemma 2(*C*). Under $\{\mathbf{P}^0_M : M \geq 1\}$, we have that:

$$\lim_{M\to\infty}\Pr\left(\hat{\rho}_* > \rho^0_*\right) = \Phi\left(-\frac{\lim_{M\to\infty}\sqrt{M}\left[\rho^0_* - \rho^0_{M*}\right]}{\sigma_{\rho^0_*}}\right).$$

Since $\rho^0_{M*} = \rho^0_* + c^*_\rho/\sqrt{M} + o\left(M^{-1/2}\right)$, the probability simplifies to $\Phi\left(c^*_\rho/\sigma_{\rho^0_*}\right)$.

B.3 Proof of Proposition 1

(i) Proof of Proposition 1(*A*). The fact that $\hat{\boldsymbol{\theta}}_{\text{FP}}$ exists if $\min\{\hat{\rho}_{\text{NPL}}, \hat{\rho}_*\} < 1$ directly follows from (16). To derive the limit of the probability $\Pr(E_M) = \Pr(\min\{\hat{\rho}_{\text{NPL}}, \hat{\rho}_*\} < 1)$ as $M \to \infty$, we write this probability as:

$$\Pr\left(\min\left\{\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - \rho^0\right), \sqrt{M}\left(\hat{\rho}_* - \rho^0_*\right) + \sqrt{M}\left(\rho^0_* - \rho^0\right)\right\} < \sqrt{M}\left(1 - \rho^0\right)\right). \tag{47}$$

Notice that $\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - \rho^0_M\right) = \sqrt{M}\left(\hat{\rho}_{\text{NPL}} - \rho^0\right) - c^0_\rho + o(1)$. Since $\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - \rho^0_M\right) \xrightarrow{d} \xi_\rho\left(\mathbf{P}^0\right)$, we can write $\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - \rho^0\right) \to \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho$. Similarly, $\sqrt{M}\left(\hat{\rho}_* - \rho^0_*\right) \to \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho$. Using $\rho^0 = 1$, and defining $\delta\left(\rho^0_*\right) \equiv \lim_{M\to\infty}\sqrt{M}\left(\rho^0_* - 1\right)$, we have that:

$$\lim_{M\to\infty}\Pr(E_M) = \Pr\left(\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho + \delta\left(\rho^0_*\right)\right\} < 0\right). \tag{48}$$

The limiting probability of E_M therefore depends on ρ^0_* via $\delta\left(\rho^0_*\right)$. If $\rho^0_* < 1$, $\delta\left(\rho^0_*\right) = -\infty$ and $\lim_{M\to\infty}\Pr(E_M) = 1$. If $\rho^0_* = 1$, $\delta\left(\rho^0_*\right) = \infty \times 0 = 0$, such that $\lim_{M\to\infty}\Pr(E_M) = \Pr\left(\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho\right\} < 0\right)$. Finally, if $\rho^0_* > 1$, we have that $\delta\left(\rho^0_*\right) = \infty$ and $\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho + \delta\left(\rho^0_*\right)\right\}$ can only be strictly inferior to 0 if $\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0$, i.e. $\lim_{M\to\infty}\Pr(E_M) = \Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right)$.

(ii) Proof of Proposition 1(*B*). Using (16), we can write $\Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}\,\middle|\, E_M\right)$ as:

$$\begin{aligned}\frac{\Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}, E_M\right)}{\Pr(E_M)} =& \frac{\Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{NPL}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}, \hat{\rho}_{\text{NPL}} < 1\right)}{\Pr(E_M)} \\ &+ \frac{\Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}, \hat{\rho}_{\text{NPL}} \geq 1, \hat{\rho}_* < 1\right)}{\Pr(E_M)}.\end{aligned} \tag{49}$$

For $\rho^0 = 1$, this expression is equivalent to:

$$\begin{aligned}&\Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{NPL}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}\,\middle|\,\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - 1\right) < 0\right)\Pr\left(\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - 1\right) < 0\,\middle|\, E_M\right) \\ &+ \Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_M\right) \in \mathcal{B}\,\middle|\,\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - 1\right) \geq 0, \sqrt{M}\left(\hat{\rho}_* - \rho^0_*\right) < \sqrt{M}\left(1 - \rho^0_*\right)\right) \\ &\quad\times\Pr\left(\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - 1\right) \geq 0, \sqrt{M}\left(\hat{\rho}_* - \rho^0_*\right) < \sqrt{M}\left(1 - \rho^0_*\right)\,\middle|\, E_M\right).\end{aligned} \tag{50}$$

Notice that $\sqrt{M}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_M\right) = \sqrt{M}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_{M*}\right) + \sqrt{M}\left(\boldsymbol{\theta}^0_{M*} - \boldsymbol{\theta}^0_M\right)$. As $M \to \infty$, we have that $\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{NPL}} - \boldsymbol{\theta}^0_M\right) \xrightarrow{d} \boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0\right)$ and $\sqrt{M}\left(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}^0_{M*}\right) \xrightarrow{d} \boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0_*\right)$. Moreover, from the proof of Proposition 1(*A*), we have that $\sqrt{M}\left(\hat{\rho}_{\text{NPL}} - 1\right) \to \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho$ and $\sqrt{M}\left(\hat{\rho}_* - \rho^0_*\right) \to$

$\xi_\rho(\mathbf{P}^0_*) + c^*_\rho$. Therefore, as $M \to \infty$, we can write (50) as:

$$\begin{aligned}
&\Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right) \frac{\Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right)}{\lim_{M\to\infty} \Pr\left(E_M\right)} \\
&+ \Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0_*\right) + \lim_{M\to\infty} \sqrt{M}\left(\boldsymbol{\theta}^0_{M*} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho < -\delta\left(\rho^0_*\right)\right) \\
&\quad \times \frac{\Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho < -\delta\left(\rho^0_*\right)\right)}{\lim_{M\to\infty} \Pr\left(E_M\right)}
\end{aligned} \tag{51}$$

where $\lim_{M\to\infty} \Pr(E_M)$ has been derived in Proposition 1(*A*). The limiting distribution $\lim_{M\to\infty} \Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| E_M\right)$ therefore also depends on ρ^0_* via $\delta(\rho^0_*)$. If $\rho^0_* < 1$, we have that $\delta(\rho^0_*) = -\infty$, $\lim_{M\to\infty} \Pr(E_M) = 1$ and $\xi_\rho(\mathbf{P}^0_*) + c^*_\rho < -\delta(\rho^0_*)$ is realized with probability 1, such that $\lim_{M\to\infty} \Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| E_M\right)$ is:

$$\begin{aligned}
&\Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right) \Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right) \\
&+ \Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0_*\right) + \lim_{M\to\infty} \sqrt{M}\left(\boldsymbol{\theta}^0_{M*} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0\right) \Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0\right).
\end{aligned} \tag{52}$$

If $\rho^0_* = 1$, then $\delta(\rho^0_*) = 0$ and $\lim_{M\to\infty} \Pr(E_M) = \Pr\left(\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho\right\} < 0\right)$ such that $\lim_{M\to\infty} \Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| E_M\right)$ is:

$$\begin{aligned}
&\Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right) \frac{\Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right)}{\Pr\left(\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho\right\} < 0\right)} \\
&+ \Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0_*\right) + \lim_{M\to\infty} \sqrt{M}\left(\boldsymbol{\theta}^0_{M*} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho < 0\right) \\
&\quad \times \frac{\Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho < 0\right)}{\Pr\left(\min\left\{\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho\right\} < 0\right)}.
\end{aligned} \tag{53}$$

Finally, if $\rho^0_* > 1$, we have that $\delta(\rho^0_*) = \infty$, $\lim_{M\to\infty} \Pr(E_M) = \Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right)$ and $\Pr\left(\xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho \geq 0, \xi_\rho\left(\mathbf{P}^0_*\right) + c^*_\rho < -\delta\left(\rho^0_*\right)\right) = 0$. As a result, the limiting distribution $\lim_{M\to\infty} \Pr\left(\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{\text{FP}} - \boldsymbol{\theta}^0_M\right) \in \mathcal{B} \middle| E_M\right)$ is:

$$\Pr\left(\boldsymbol{\xi}_{\boldsymbol{\theta}}\left(\mathbf{P}^0\right) \in \mathcal{B} \middle| \xi_\rho\left(\mathbf{P}^0\right) + c^0_\rho < 0\right). \tag{54}$$

Appendix C Additional simulation results

C.1 Summary statistics from simulated data

Table C1 reports summary statistics using simulated data for Experiments I. These statistics are based on simulated firms' decisions in 50000 markets drawn from the ergodic distribution of the state variables. Besides potentially affecting the convergence properties of the NPL algorithm, the value of the strategic interaction parameter has important economic implications. Increasing the value of the strategic interaction parameter considerably reduces the average number of active firms, generating larger reductions in the probabilities of being active for the firms with larger fixed costs. Interestingly, the effect on the number of entry and exits is non-monotonic: these numbers are larger for the "mildly stable" and the "mildly unstable" cases than the other two data generating processes.

Table C2 reports the same competition statistics but for Experiments II, using also 50000 simulated markets.

Table C1: Competition statistics – Experiments I

	Very stable	**Mildly stable**	**Mildly unstable**	**Very unstable**
Number of active firms				
Average	2.7652	1.9939	1.7646	1.2225
Std. dev.	1.6622	1.4320	1.3233	1.0024
AR(1) parameter for number of active firms	0.7070	0.5691	0.5095	0.3519
Average number of entries	0.6917	0.7473	0.7278	0.5492
Average number of exits	0.6933	0.7569	0.7349	0.5558
Average excess turnover	0.4600	0.5110	0.4896	0.2879
Correlation between entries and exits	-0.1743	-0.2225	-0.2141	-0.1854
Probabilities of being active				
Firm 1	0.4993	0.3222	0.2676	0.1239
Firm 2	0.5222	0.3552	0.3032	0.1457
Firm 3	0.5536	0.3975	0.3492	0.1871
Firm 4	0.5797	0.4363	0.3953	0.2689
Firm 5	0.6103	0.4827	0.4493	0.4968

Notes: Statistics computed using $M = 50000$ markets drawn from the ergodic distribution of the state variables. Excess turnover defined as (# entries + # exits) − abs (# entries − # exits).

C.2 Average estimates and standard errors with larger sample

Table C3 reports average estimates and standard errors of the estimators of interest in Experiment I for $M = 5000$.

C.3 Using the frequency estimator as a single set of starting values

In this section, we investigate whether using multiple starting values to initiate the different algorithms studied is necessary for the relative good properties of the spectral approach. To

Table C2: Competition statistics – Experiments II

	Eq (i)	**Eq (ii)**	**Eq (iii)**
Number of active firms			
Average	1.0247	1.1559	1.1621
Std. dev.	0.6027	0.6394	0.6371
AR(1) parameter for number of active firms	0.0160	0.0216	0.0184
Average number of entries	0.3351	0.3331	0.3308
Average number of exits	0.3374	0.3338	0.3302
Average excess turnover	0.0938	0.0567	0.0547
Correlation between entries and exits	-0.2262	-0.2603	-0.2580
Probabilities of being active			
Firm 1	0.7690	0.6009	0.5824
Firm 2	0.2557	0.5550	0.5797

Notes: Statistics computed using $M = 50000$ markets drawn from the ergodic distribution of the state variables. Excess turnover defined as $(\#\text{ entries} + \#\text{ exits}) - \text{abs}(\#\text{ entries} - \#\text{ exits})$.

do so, we compute the estimates one would have obtained in our simulation exercises when considering the frequency count estimator of the conditional choice probabilities as the only set of starting values. We study how this modification affects the average estimates and the convergence rate of the algorithms. More precisely, for each data generating process and each algorithm, we compute the absolute value of the bias estimated by Monte Carlo simulations and we average this absolute bias over the parameters to obtain a scalar measurement. Tables C4 and C5 report the relative average absolute bias, i.e. the computed average absolute bias obtained when using a single set of starting values divided by the one computed from multiple starting values. The relative convergence rates are computed in a similar way: the fraction of Monte Carlo samples which have reached convergence when using a single set of starting values divided by the same fraction obtained when using multiple ones.

From Experiments I's results, using the frequency count estimator as a single set of starting values barely affects the estimated values of the parameters obtained from the spectral algorithm and the spectral solver. The same comment holds for the relaxation algorithm. Some differences are noticeable when looking at the NPL algorithm's results for less stable data generating processes. In particular, using the frequency count estimator as a single set of starting values may reduce the convergence rate and, as a result, alter the estimated bias of converging sequences. Moreover, different starting values may lead to different estimates at $K = 100$ if the sequence of NPL algorithm estimates fail to converge by this number of iterations.

Experiments II's results suggest that using the frequency count estimator as a single set of starting values may severely increase the bias of the estimates, especially for the

Table C3: Simulation results – $M = 5000$, Experiments I

	$\theta_{\text{RS}} = 1$	θ_{RN}	$\theta_{\text{EC}} = 1$	$\theta_{\text{FC},1} = 1.9$	$\theta_{\text{FC},2} = 1.8$	$\theta_{\text{FC},3} = 1.7$	$\theta_{\text{FC},4} = 1.6$	$\theta_{\text{FC},5} = 1.5$
Two-step estimates								
Very stable ($\theta_{\text{RN}} = 1$)	0.8582	0.6470	1.0103	1.8658	1.7711	1.6749	1.5787	1.4808
	(0.0625)	(0.1857)	(0.0362)	(0.0676)	(0.0660)	(0.0650)	(0.0611)	(0.0574)
Mildly stable ($\theta_{\text{RN}} = 2$)	0.7061	0.9286	1.0448	1.8907	1.7799	1.6712	1.5544	1.4374
	(0.0641)	(0.2325)	(0.0341)	(0.0612)	(0.0605)	(0.0563)	(0.0550)	(0.0567)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	0.6673	1.0634	1.0694	1.8909	1.7729	1.6559	1.5326	1.4046
	(0.0638)	(0.2491)	(0.0358)	(0.0634)	(0.0617)	(0.0593)	(0.0588)	(0.0595)
Very unstable ($\theta_{\text{RN}} = 4$)	0.6637	2.0699	1.2045	1.9440	1.8264	1.6964	1.5181	1.2030
	(0.0540)	(0.2830)	(0.0528)	(0.0864)	(0.0820)	(0.0778)	(0.0777)	(0.0845)
Converged $K = 100$ NPL fixed point algorithm estimates								
Very stable ($\theta_{\text{RN}} = 1$)	1.0032	1.0086	1.0007	1.9010	1.8019	1.7016	1.6012	1.5004
	(0.0659)	(0.2052)	(0.0355)	(0.0661)	(0.0664)	(0.0642)	(0.0600)	(0.0582)
Mildly stable ($\theta_{\text{RN}} = 2$)	0.9608	1.8540	1.0162	1.9035	1.7994	1.6997	1.5947	1.4916
	(0.0638)	(0.2294)	(0.0349)	(0.0640)	(0.0653)	(0.0590)	(0.0606)	(0.0612)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	0.9027	2.0002	1.0420	1.9075	1.7972	1.6922	1.5834	1.4706
	(0.0402)	(0.1467)	(0.0321)	(0.0663)	(0.0643)	(0.0653)	(0.0637)	(0.0649)
Very unstable ($\theta_{\text{RN}} = 4$)	–	–	–	–	–	–	–	–
	(–)	(–)	(–)	(–)	(–)	(–)	(–)	(–)
All $K = 100$ NPL fixed point algorithm estimates								
Very stable ($\theta_{\text{RN}} = 1$)	1.0032	1.0086	1.0007	1.9010	1.8019	1.7016	1.6012	1.5004
	(0.0659)	(0.2052)	(0.0355)	(0.0661)	(0.0664)	(0.0642)	(0.0600)	(0.0582)
Mildly stable ($\theta_{\text{RN}} = 2$)	0.9963	1.9875	1.0038	1.9005	1.7990	1.7010	1.5988	1.4994
	(0.0806)	(0.2944)	(0.0387)	(0.0661)	(0.0664)	(0.0622)	(0.0625)	(0.0647)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	0.9696	2.2761	1.0123	1.8991	1.7949	1.6953	1.5947	1.4954
	(0.0508)	(0.1927)	(0.0353)	(0.0699)	(0.0685)	(0.0661)	(0.0666)	(0.0681)
Very unstable ($\theta_{\text{RN}} = 4$)	0.7667	2.6684	1.2213	1.8942	1.7736	1.6647	1.4770	1.2901
	(0.0240)	(0.0603)	(0.0338)	(0.0820)	(0.0791)	(0.0760)	(0.0744)	(0.0678)
All $K = 100$ relaxation algorithm estimates								
Very stable ($\theta_{\text{RN}} = 1$)	1.0129	1.0378	0.9993	1.8993	1.8005	1.7006	1.6005	1.5002
	(0.0687)	(0.2137)	(0.0357)	(0.0665)	(0.0667)	(0.0645)	(0.0603)	(0.0584)
Mildly stable ($\theta_{\text{RN}} = 2$)	1.0437	2.1589	0.9883	1.9065	1.8075	1.7121	1.6132	1.5175
	(0.0960)	(0.3500)	(0.0434)	(0.0674)	(0.0680)	(0.0643)	(0.0652)	(0.0683)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	1.0603	2.6356	0.9749	1.9147	1.8164	1.7223	1.6274	1.5340
	(0.1004)	(0.3943)	(0.0522)	(0.0730)	(0.0728)	(0.0724)	(0.0750)	(0.0800)
Very unstable ($\theta_{\text{RN}} = 4$)	1.0273	4.1534	0.9728	1.9020	1.8023	1.7095	1.6205	1.5462
	(0.0418)	(0.2011)	(0.0487)	(0.0926)	(0.0897)	(0.0869)	(0.0854)	(0.0891)
All $K = 100$ spectral algorithm estimates								
Very stable ($\theta_{\text{RN}} = 1$)	1.0032	1.0086	1.0007	1.9010	1.8019	1.7016	1.6012	1.5004
	(0.0659)	(0.2052)	(0.0355)	(0.0661)	(0.0664)	(0.0642)	(0.0600)	(0.0582)
Mildly stable ($\theta_{\text{RN}} = 2$)	0.9990	1.9970	1.0029	1.9010	1.7996	1.7016	1.5993	1.4997
	(0.0853)	(0.3117)	(0.0399)	(0.0661)	(0.0665)	(0.0624)	(0.0628)	(0.0650)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	1.0035	2.4121	0.9991	1.9049	1.8027	1.7042	1.6037	1.5034
	(0.0886)	(0.3486)	(0.0465)	(0.0708)	(0.0701)	(0.0689)	(0.0706)	(0.0736)
Very unstable ($\theta_{\text{RN}} = 4$)	0.9993	3.9922	1.0025	1.9046	1.8019	1.7036	1.6020	1.5008
	(0.0432)	(0.2132)	(0.0497)	(0.0912)	(0.0882)	(0.0852)	(0.0838)	(0.0880)
Spectral solver estimates								
Very stable ($\theta_{\text{RN}} = 1$)	1.0031	1.0082	1.0008	1.9011	1.8019	1.7017	1.6013	1.5004
	(0.0659)	(0.2052)	(0.0355)	(0.0661)	(0.0664)	(0.0643)	(0.0600)	(0.0583)
Mildly stable ($\theta_{\text{RN}} = 2$)	0.9990	1.9972	1.0029	1.9011	1.7998	1.7017	1.5994	1.4998
	(0.0854)	(0.3122)	(0.0400)	(0.0661)	(0.0665)	(0.0625)	(0.0629)	(0.0651)
Mildly unstable ($\theta_{\text{RN}} = 2.4$)	1.0035	2.4123	0.9991	1.9049	1.8027	1.7043	1.6037	1.5034
	(0.0887)	(0.3490)	(0.0465)	(0.0708)	(0.0701)	(0.0690)	(0.0707)	(0.0737)
Very unstable ($\theta_{\text{RN}} = 4$)	0.9993	3.9918	1.0025	1.9046	1.8018	1.7035	1.6018	1.5006
	(0.0431)	(0.2131)	(0.0498)	(0.0913)	(0.0882)	(0.0851)	(0.0836)	(0.0877)

Notes: Averages and standard errors (in brackets) computed over 500 Monte Carlo samples. The $K = 100$ NPL algorithm is deemed having converged if $\max\left\{\left|\hat{\mathbf{P}}_k - \hat{\mathbf{P}}_{k-1}\right|\right\} < 10^{-5}$ for some $k \leq 100$. The relaxation algorithm never converged by $K = 100$. Since the spectral algorithm almost always converged, the results conditional on convergence are very similar to the ones obtained from all samples and are not reported.

Table C4: Single starting values – Experiments I

	Very stable		**Mildly stable**		**Mildly unstable**		**Very unstable**	
	400	5K	400	5K	400	5K	400	5K
NPL algorithm								
Rel. absolute bias (all)	1.001	1.000	1.275	1.025	1.105	1.036	1.128	1.003
Rel. absolute bias (converged)	1.031	1.000	1.008	1.020	1.010	1.051	1.000	–
Rel. convergence rate	0.998	1.000	0.993	0.989	0.986	0.895	1.000	–
Relaxation algorithm								
Rel. absolute bias (all)	0.980	1.000	0.991	1.000	0.992	1.000	0.997	1.000
Rel. absolute bias (converged)	–	–	–	–	–	–	–	–
Rel. convergence rate	–	–	–	–	–	–	–	–
Spectral algorithm								
Rel. absolute bias (all)	1.000	1.000	1.000	0.999	1.000	1.001	1.000	1.000
Rel. absolute bias (converged)	1.000	1.000	1.000	0.999	1.000	1.001	1.000	1.000
Rel. convergence rate	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spectral solver								
Rel. absolute bias (converged)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Rel. convergence rate	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The Monte Carlo simulation results are computed using a single set of starting values for the choice probabilities (the frequency count estimator) instead of using 5 starting values (including the frequency count estimator). The table reports the average of the parameters' absolute bias and the convergence rates relative to the case with 5 starting values (i.e., the statistics for the single starting value case are divided by their multiple starting values counterpart). The relative absolute bias is computed for converged sequences (i.e., $\max\left\{\left|\hat{\mathbf{P}}_k - \hat{\mathbf{P}}_{k-1}\right|\right\} < 10^{-5}$ for $k \leq 100$) and all sequences at $K = 100$. For the spectral solver, the R's BBsolve default tolerance is used to assess convergence (i.e. the L2 norm of $\hat{\boldsymbol{\phi}}(\mathbf{P})$ being smaller than $\sqrt{\dim(\mathbf{P})} \times 10^{-7}$).

spectral algorithm and solver. This observation suggests that a single starting value may be insufficient to find the NPL fixed point that maximizes the log-likelihood function. It is worth emphasizing that if the increase in the bias is not as striking for the NPL and the relaxation algorithms, it is because these algorithms often fail to deliver the NPL estimator even with multiple starting values.

To summarize, using multiple starting values is necessary for the relative good performance of the spectral algorithm and solver when they are applied to data generating processes featuring multiple fixed points. In that sense, the need for multiple starting values has a similar justification as when maximizing a log-likelihood function that has multiple local maxima.

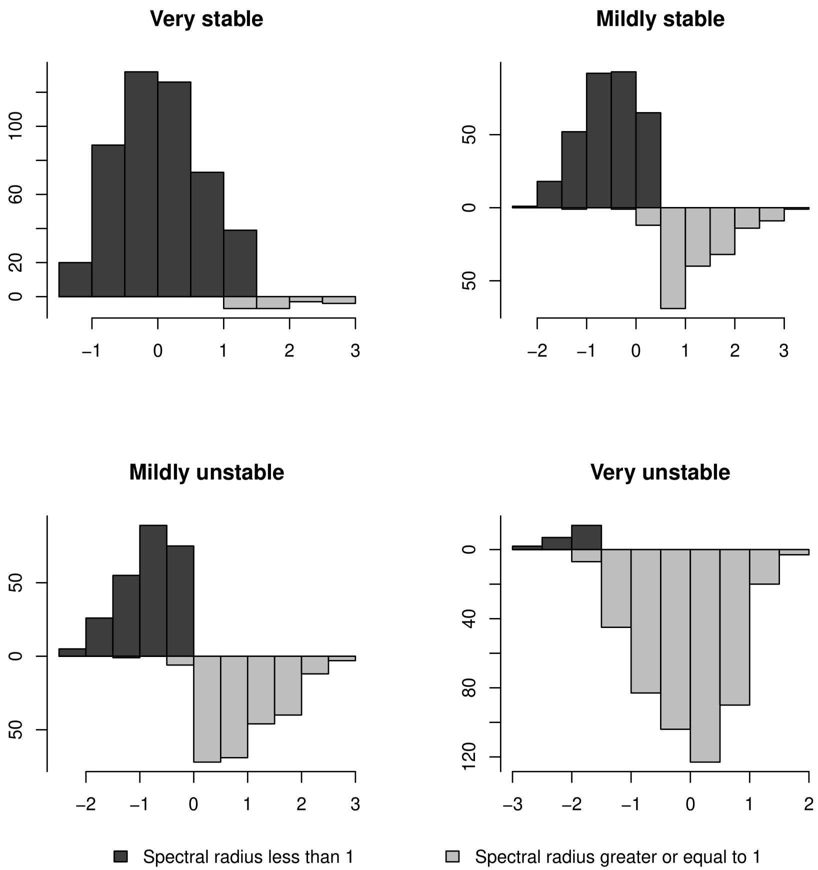
C.4 Split histograms of spectral algorithm estimates

Figures 1 and 2 present the empirical distributions of spectral solver estimates for Experiments I and II, respectively, when we split the Monte Carlo replications in two groups: with $\hat{\rho}_{\text{NPL}} < 1$, and with $\hat{\rho}_{\text{NPL}} \geq 1$.

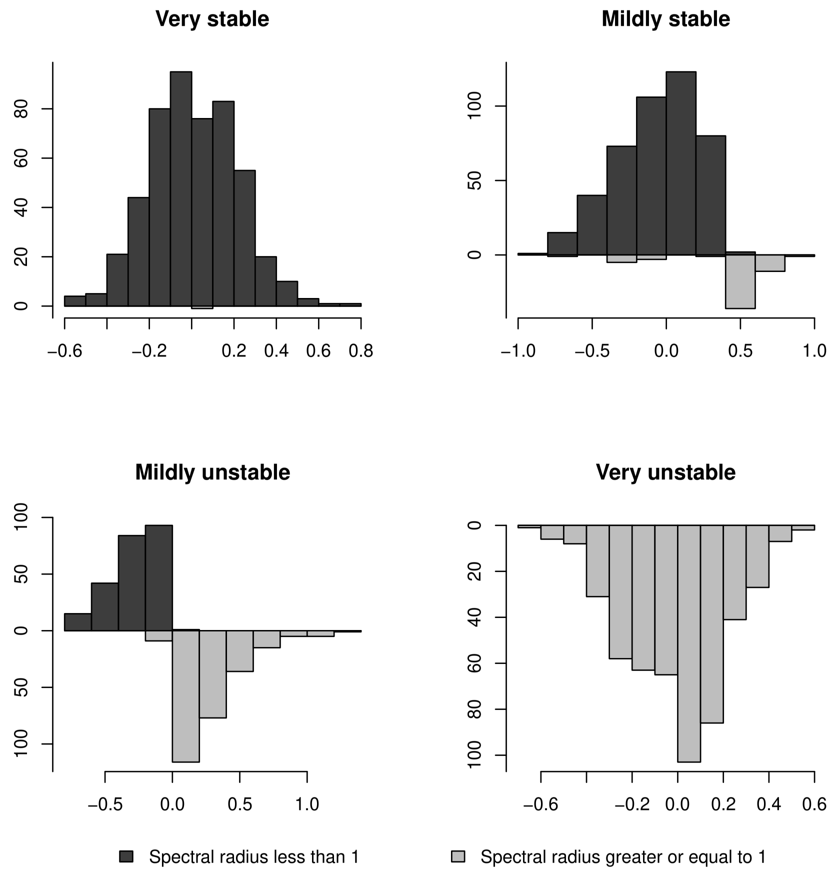
Table C5: Single starting value – Experiments II

	Eq (1)		**Eq (2)**		**Eq (3)**	
	400	5K	400	5K	400	5K
NPL algorithm						
Rel. absolute bias (all)	1.275	1.000	1.013	1.027	1.007	1.003
Rel. absolute bias (converged)	1.275	1.000	1.013	1.027	1.007	1.003
Rel. convergence rate	1.000	1.000	1.000	1.000	1.000	1.000
Relaxation algorithm						
Rel. absolute bias (all)	1.534	1.045	1.023	1.038	1.016	1.015
Rel. absolute bias (converged)	0.987	0.902	1.035	1.033	1.034	1.012
Rel. convergence rate	0.942	0.998	0.922	0.976	0.912	0.982
Spectral algorithm						
Rel. absolute bias (all)	1.174	1.000	4.621	9.450	4.367	20.90
Rel. absolute bias (converged)	1.174	1.000	4.621	9.450	4.367	20.90
Rel. convergence rate	1.000	1.000	1.000	1.000	1.000	1.000
Spectral solver						
Rel. absolute bias (converged)	1.819	1.000	4.192	3.783	1.754	10.74
Rel. convergence rate	1.000	1.000	1.000	0.998	0.998	1.000

Notes: The Monte Carlo simulations are repeated using a single starting value for the choice probabilities (the frequency count estimator) instead of using 100 starting values (including the frequency count estimator and the one-step NPL mapping update). The table reports the average of the parameters' absolute bias and the convergence rates relative to the case with 100 starting values (i.e., the statistics for the single starting value case are divided by their multiple starting values counterpart). The relative absolute bias is computed for converged sequences (i.e., $\max\left\{\left|\hat{\mathbf{P}}_k - \hat{\mathbf{P}}_{k-1}\right|\right\} < 10^{-5}$ for $k \leq 500$) and all sequences at $K = 500$. For the spectral solver, the R's BBsolve default tolerance is used to assess convergence (i.e. the L2 norm of $\hat{\boldsymbol{\phi}}(\mathbf{P})$ being smaller than $\sqrt{\dim(\mathbf{P})} \times 10^{-7}$).



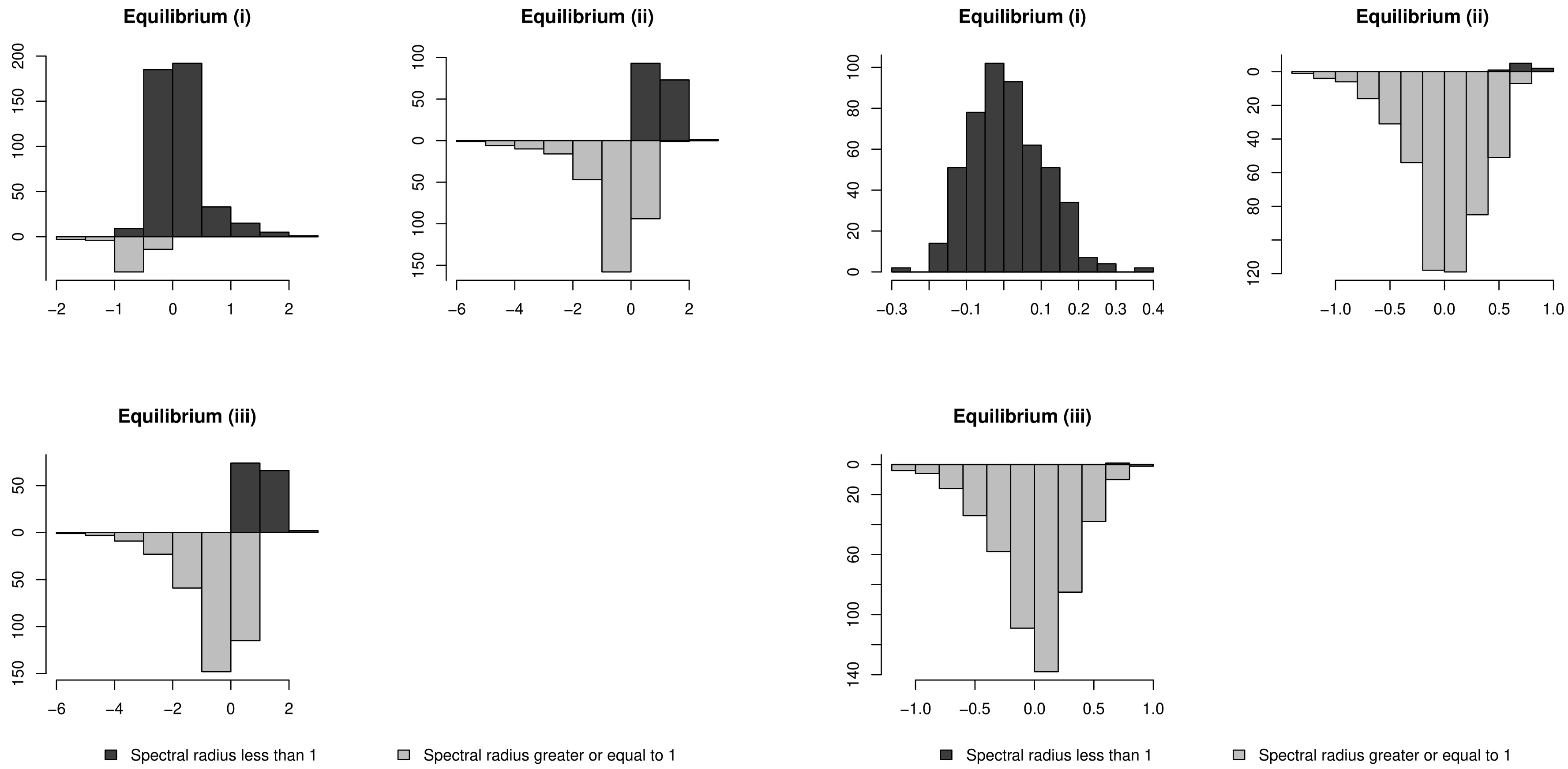
(a) $M = 400$



(b) M=5000

Notes:

Figure 1: Histograms of $\hat{\theta}_{\text{RN}} - \theta^0_{\text{RN}}$ from spectral solver (by value of $\hat{\rho}_{\text{NPL}}$) – Experiments I



(a) $M = 100$

(b) $M = 1000$

Notes:

Figure 2: Histograms of $\hat{\theta}_{\mathrm{C}} - \theta_{\mathrm{C}}^{0}$ from spectral solver (by value of $\hat{\rho}_{\mathrm{NPL}}$)

Appendix D Results from the empirical application

The parameters' estimates and the computation time are reported in Table D6.

Table D6: Results from three different methods across 100 different starting values

	NPL algorithm	**Relaxation algorithm**	**Spectral solver**
		Estimates and standard errors	
$\theta^0_{\text{VP,BK}}$	1.09786***	1.09786***	1.09787***
	(0.21687)	(0.21687)	(0.21687)
$\theta^1_{\text{VP,BK}}$	-0.07653	-0.07653	-0.07654
	(0.07247)	(0.07247)	(0.07247)
$\theta^2_{\text{VP,BK}}$	-0.01297**	-0.01297**	-0.01297**
	(0.00649)	(0.00649)	(0.00649)
$\theta^0_{\text{FC,BK}}$	0.07883**	0.07883**	0.07883**
	(0.03076)	(0.03076)	(0.03076)
$\theta^1_{\text{FC,BK}}$	0.15095***	0.15095***	0.15095***
	(0.02829)	(0.02829)	(0.02829)
$\theta^2_{\text{FC,BK}}$	-0.00547**	-0.00547**	-0.00547**
	(0.00269)	(0.00269)	(0.00269)
$\theta^0_{\text{VP,MD}}$	0.97371***	0.97371***	0.97371***
	(0.30911)	(0.30911)	(0.30911)
$\theta^1_{\text{VP,MD}}$	0.28736***	0.28736***	0.28736***
	(0.09865)	(0.09865)	(0.09865)
$\theta^2_{\text{VP,MD}}$	-0.00738	-0.00738	-0.00738
	(0.00745)	(0.00745)	(0.00745)
$\theta^0_{\text{FC,MD}}$	0.07731***	0.07731***	0.07731***
	(0.02614)	(0.02614)	(0.02614)
$\theta^1_{\text{FC,MD}}$	0.13020***	0.13020***	0.13020***
	(0.01847)	(0.01847)	(0.01847)
$\theta^2_{\text{FC,MD}}$	0.00014	0.00014	0.00014
	(0.00162)	(0.00162)	(0.00162)
Population density	3.94938**	3.94940**	3.94938**
	(1.59148)	(1.59149)	(1.59148)
Income	0.00014	0.00014	0.00014
	(0.00011)	(0.00011)	(0.00011)
Average rent	-0.00003	-0.00003	-0.00003
	(0.00022)	(0.00022)	(0.00022)
Retail taxes	0.00003	0.00003	0.00003
	(0.00003)	(0.00003)	(0.00003)
		Summary statistics of computational time in seconds	
Minimum	159.7	1743	103.0
Median	233.0	3186	194.2
Mean	226.1	2848	193.7
Maximum	261.8	3674	1541

Notes: Significance levels: $*$ is 90%, $**$ is 95% and $***$ is 99%. The estimates are the ones that maximize the log-likelihood function across all 100 vectors of starting values for each method. The computational time is the total time until convergence.

References

Aguirregabiria, V., and P. Mira (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.