Measuring Segregation on Small Units: A Partial Identification Analysis
Online Appendix

Xavier D'Haultfœuille* and Roland Rathelot†

April 2016

B Accounting for random unit size and covariates

B.1 Random unit size

We first extend our framework to cases where K is random and takes values in $\{2, ..., \overline{K}\}$. We exclude here units with one individual, for which bounds on segregation indices are trivial. We consider here the agnostic case where p may depend on K; the case where $p \perp\!\!\!\perp K$ is detailed in Section D.4 below. We consider a parameter defined in a similar way as in Assumption 2.1. With random unit sizes, the issue of whether we focus on segregation at the unit level or at the individual level matters. Let F_p^k denote the cdf of p conditional on $K = k$. In the first case, the parameter of interest is defined by

$$\theta_0 = \nu \left(\sum_{k=2}^{\overline{K}} \Pr(K = k) \int h(x, m_{01}) dF_p^k(x), m_{01} \right), \tag{B.1}$$

with, as before $m_{01} = E(p)$. In contrast, in the second case, the parameter of interest satisfies

$$\theta_0 = \nu \left(\sum_{k=2}^{\overline{K}} \frac{k \Pr(K = k)}{E(K)} \int h(x, E(Kp)/E(p)) dF_p^k(x), E(Kp)/E(p) \right). \tag{B.2}$$

*CREST, xavier.dhaultfoeuille@ensae.fr.
†University of Warwick, r.rathelot@warwick.ac.uk.

This expression is therefore the same as above, except that we weight units of size k by k, to express the fact that the parameter is defined at the individual level.

Because the two cases are very similar, we focus on the first one in this section. Let $\eta_0^k = \int h(x, m_{01}) dF_p^k(x)$. Without any joint restriction on $(F_p^k, F_p^{k'})$, the sharp bounds on θ_0 satisfy

$$\{\underline{\theta}_0, \overline{\theta}_0\} = \left\{ \nu\left(\sum_{k=2}^{\overline{K}} \Pr(K=k)\underline{\eta}_0^k, m_{01}\right), \ \nu\left(\sum_{k=2}^{\overline{K}} \Pr(K=k)\overline{\eta}_0^k, m_{01}\right)\right\} \quad \text{(B.3)}$$

where the bounds $\underline{\eta}_0^k$ and $\overline{\eta}_0^k$ on η_0^k can be computed using Theorems 2.1 or 2.2 since $K = k$ is fixed.

We can then estimate these bounds by taking the empirical counterpart of (B.3). First, we estimate $P^k = (\Pr(X=1|K=k), ..., \Pr(X=k|K=k))'$ by constrained maximum likelihood on the subsample of units of size k. Then let $\widehat{m}^k = Q^k \widehat{P}^k$ be the estimator of $m^k = (E(p|K=k), ..., E(p^k|K=k))'$, with Q^k the matrix Q defined in Section 2.2 when $K = k$. Second, for any $F_p^k \in \mathcal{D}_{m^k}^{k+1}$ with support x^k, we may rewrite $\int h(x, m_{01}) dF_p^k(x)$ as a function of x^k, m^k and m_{01} only. We denote this function by $q(x^k, m^k, m_{01})$, and let

$$\overline{\eta}^k(m^k, m_{01}) = \max_{x^k \in \mathcal{S}_{k+1}: V(x^k)^{-1}(1, m^{k\prime})' \geq 0} q(x^k, m^k, m_{01}),$$
$$\underline{\eta}^k(m^k, m_{01}) = \min_{x^k \in \mathcal{S}_{k+1}: V(x^k)^{-1}(1, m^{k\prime})' \geq 0} q(x^k, m^k, m_{01}).$$

Finally, let $\widehat{\Pr}(K=k) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{K_i = k\}$ and $\widehat{m}_{01} = \sum_{k=2}^{\overline{K}} \widehat{\Pr}(K=k)\widehat{m}_1^k$. We estimate the bounds by

$$\begin{aligned} \widehat{\underline{\theta}} &= \nu\left(\textstyle\sum_{k=2}^{\overline{K}} \widehat{\Pr}(K=k)\underline{\eta}^k(\widehat{m}^k, \widehat{m}_{01}), \widehat{m}_{01}\right), \\ \widehat{\overline{\theta}} &= \nu\left(\textstyle\sum_{k=2}^{\overline{K}} \widehat{\Pr}(K=k)\overline{\eta}^k(\widehat{m}^k, \widehat{m}_{01}), \widehat{m}_{01}\right). \end{aligned} \quad \text{(B.4)}$$

Turning to inference, we consider the following bootstrap resampling scheme. First, draw K in its empirical distribution and then draw X conditional on $K = k$ according to the vector of probabilities $\widehat{P}_b^k = Q^k \widehat{m}_b^k$, where $\widehat{m}_b^k$ is defined as $\widehat{m}_b$, for the subsample of units with $K = k$. Then we compute the bootstrap bounds and bootstrap statistics using (B.4). We finally construct the confidence interval $\text{CI}_{1-\alpha}^2$ in exactly the same way as $\text{CI}_{1-\alpha}^1$.

Proposition B.1 *Suppose that Assumption 2.1 holds, ν is C^1 and $\underline{\eta}^k$ and $\overline{\eta}^k$ are differentiable at (m_0^k, m_{01}) for all k in the support of K. Then*

$$\inf_{\theta_0 \in [\underline{\theta}_0, \overline{\theta}_0]} \lim_{n \to \infty} \Pr(\theta_0 \in \ CI_{1-\alpha}^2) = 1 - \alpha.$$

B.2 Including covariates

Now let us consider the inclusion of covariates. Controlling for exogenous characteristics is important because they may constitute an undesirable source of differences in unobserved probabilities across units. For instance, foreigners may be hired more in some sectors of the economy because they have been selected, or they self-selected, on the basis of sector-specific skills. Consider an extreme instance where all firms within each sector would hire foreigners with the same probability. The unconditional index would still be positive if these probabilities differ from one sector to another. The conditional index we consider below, on the other hand, would be zero.

Controlling for characteristics defined at a finer level, such as slots in a geographic area or positions within the firm, may also make the condition $X|K, p \sim B(K, p)$ more credible. To see this, suppose for instance a simple case where $K = 4$ and each firm has two skilled positions and two unskilled ones. Suppose the individuals are hired independently of each other but within each firm, foreigners have a small probability (0.01, say) of being hired on a skilled position, and a higher probability (0.09, say) of being hired on an unskilled one. Then for each firm, the probability of hiring a foreigner on a position drawn randomly is $p = (0.01 + 0.09)/2$. But then $\Pr(X = 4) = (0.01 \times 0.09)^2 \neq p^4$, and $X|K, p \sim B(K, p)$ fails to hold. On the other hand, conditional on the type of position, the binomial mixture model is satisfied.

The discussion above shows that covariates can be defined at the unit level or at the finer level of a slot or a position. We distinguish between these two cases because they lead to different treatments. Let $Z \in \{1, ..., \overline{Z}\}$ denote a characteristic of the unit. Note that we focus here on discrete covariates only; the analysis of continuous covariates raises difficult issues that are beyond the scope of the paper. Suppose that the conditional parameter of interest satisfies $\theta_{0z} = \nu(\int h(x, m_{01z} dF_{p|Z=z}, m_{01z})$, with $m_{01z} = E(p|Z = z)$. For instance, in case of the Duncan and Theil indices, this would amount to considering respectively

$$D_z = E\left[\frac{|p - m_{01Z}|}{2m_{01Z}(1 - m_{01Z})} \,|Z = z\right],$$
$$T_z = 1 - E\left[\frac{p\ln(p)}{m_{01Z}(1 - m_{01Z})}|Z = z\right].$$

The bounds on θ_{0z} can be estimated exactly as previously, focusing on the subsample of units $\{i : Z_i = z\}$. If we are interested in an aggregate parameter

$\theta_{0.} = \sum_{z=1}^{\overline{Z}} \Pr(Z = z)\theta_{0z}$, then bounds on $\theta_{0.}$ can be obtained by a plug-in estimator, and a similar bootstrap procedure as for the random unit size case can be applied to construct valid confidence intervals.

Let us turn to covariates $W \in \{1, ..., \overline{W}\}$ defined at a finer level, such as the type of positions within the firm. We suppose to observe, for each unit and each $w \in \{1, ..., \overline{W}\}$, X_w and K_w, which are respectively the number of individuals belonging to the minority group for positions of type $W = w$ and the number of positions of type $W = w$ (so that $\sum_{w=1}^{\overline{W}} K_w = K$ and $\sum_{w=1}^{\overline{W}} X_w = X$). Let p_w denote the probability of a person belonging to the minority group for positions such as $W = w$. In such a setting, it makes more sense to suppose that the parameter of interest is defined at the position level. Specifically, let

$$\theta_{0w} = \nu\left(\int \frac{k}{E(K_w)} h(x, m_{01w}) dF_{p_w,K_w}(x, k), m_{01w}\right),$$

with $m_{01w} = E(K_w p_w)/E(K_w)$. As above, θ_{0w} is a conditional segregation index, but defined at a position level, so that each unit is weighted by its normalized size $K/E(K_w)$. An aggregated index can then be defined as $\theta_{0.} = \sum_{w=1}^{\overline{W}} \Pr(W = w)\theta_{0w}$. Because

$$\int kh(x, m_{01w}) dF_{p_w,K_w}(x, k) = \sum_{k=1}^{\overline{K}_w} \frac{k \Pr(K_w = k)}{E(K_w)} \int h(x, m_{01w}) dF_{p_w|K_w=k}(x),$$

we can estimate the bounds on $\theta_{0.}$ as in the random unit size case, using X_w and K_w instead of X and K. The confidence intervals can also be obtained similarly.

Computing $\theta_{0.}$ is important from an applied point of view. As we illustrate it in our application, the idea is to compute an index that reflect the residual level of segregation while the contribution of a covariate (*e.g.* the firm's industry, the position's occupation) to segregation is removed. One could make the analogy with linear models, in which covariates are introduced to account for the variance of the outcome. Both Hellerstein and Neumark (2008) and Åslund and Skans (2010) detail the motivation of such an average index reflecting residual segregation.

C Two important tests

C.1 Test of the binomial mixture model

A vector in $\mathcal{M}$ satisfies some restrictions such as variance positivity. Therefore, we may have $Q^{-1}P_0 \notin \mathcal{M}$ if the distribution of X conditional on K and p is not binomial. Such a model is therefore testable. Suppose for instance that $K = 2$ and $P_0 = (0.6, 0.3)'$. This vector would correspond to the vector of raw moments $m_0 = (0.6, 0.3)'$ according to the binomial model. But $0.3 - 0.6^2 < 0$, which violates the restriction that a variance is positive. This implies that such a vector P_0 invalidates the binomial mixture model.

Testing for this restriction is equivalent to testing $P_0 \in \mathcal{P}$. We simply rely, for that purpose, on the following likelihood ratio test statistic

$$LR_n = 2\sum_{k=0}^{K} N_k \ln\left(\frac{N_k}{n\widehat{P}_k}\right).$$

To approximate the distribution of LR_n under the null, we use again the bootstrap. But contrary to what we do in Section 3.2, no correction due to boundary effects is required here. The important point is rather to define a bootstrap distribution that is drawn under the null hypothesis. We thus consider a bootstrap distribution of X defined by the vector of probabilities $\widehat{P}$. Letting LR^*_{2n} denote the bootstrap counterpart of LR_n, we define the critical region of the test by

$$CR_n = \{LR_n > c_{1-\alpha}(LR^*_{2n})\}.$$

Note that when $LR_n = 0$, viz. if $\widetilde{P} \in \mathcal{P}$, we always accept H_0 and it is unnecessary to compute $c_{1-\alpha}(LR^*_{2n})$.

Theorem C.1 *If $P_0 \in \overset{\circ}{\mathcal{P}}$, $\lim_{n\to\infty} \Pr(CR_n) = 0$ with probability one. If $P_0 \in \partial\mathcal{P}$, $\overline{C}_{P_0}$ is a half space and $\alpha < 1/2$, $\lim_{n\to\infty} \Pr(CR_n) = \alpha$ with probability one. Finally, if $P_0 \notin \mathcal{P}$, $\Pr(CR_n) \to 1$ with probability one.*

The theorem shows that the test has asymptotic size equal to α, and is consistent. Of course, this does not mean that we reject the binomial mixture model whenever the true DGP does not satisfy this condition. It may happen that the true DGP is not a binomial mixture model but can be rationalized by such a model.

C.2 A test of constant segregation

We consider here a test that $K \mapsto \theta_0(K)$ is constant over $\mathcal{K}$, where $\theta_0(K)$ denotes the true parameter corresponding to units of size K. Because of partial identification, we cannot test directly for this condition. Instead, we test for an implication of it, namely:

$$H_0: \ \Delta \equiv \min_{K\in\mathcal{K}} \overline{\theta}_0(K) - \max_{K\in\mathcal{K}} \underline{\theta}_0(K) \geq 0, \tag{C.1}$$

where $\underline{\theta}_0(K)$ and $\overline{\theta}_0(K)$ are the sharp lower and upper bounds of $\theta_0(K)$.

We consider the critical region $\{\widehat{\Delta} < c_n\}$, where $\widehat{\Delta}$ is the plug-in estimator of Δ and c_n is chosen such that under the null $\Delta = 0$, $\lim_n P(\widehat{\Delta} < c_n) = \alpha$, where α is the nominal level of the test. The difficulty here is that the distribution of $\widehat{\Delta}$ under the null may be complicated. Moreover, it is known that bootstrap can fail to estimate consistently the asymptotic distribution of statistics involving nondifferentiable functions such as the max or the min operators. We therefore rely on subsampling. Our asymptotic results and standard arguments imply that, whether the null holds or not,

$$\sqrt{n}\left(\min_{K\in\mathcal{K}}\widehat{\overline{\theta}}(K) - \min_{K\in\mathcal{K}}\overline{\theta}_0(K),\ \max_{K\in\mathcal{K}}\widehat{\underline{\theta}}(K) - \max_{K\in\mathcal{K}}\underline{\theta}_0(K)\right) \xrightarrow{d} (\mathcal{Z}_1, \mathcal{Z}_2). \qquad \text{(C.2)}$$

Note that $\mathcal{Z}_1$ and $\mathcal{Z}_2$ may not be normal, either because of the boundary issues or if there exists $K_1 \neq K_2$ such that $\overline{\theta}_0(K_1) = \overline{\theta}_0(K_2) = \min_{K\in\mathcal{K}}\overline{\theta}_0(K)$, say. But in any case, subsampling will provide a consistent estimator of the distribution of $(\mathcal{Z}_1, \mathcal{Z}_2)$. Let us call $(\mathcal{Z}_1^s, \mathcal{Z}_2^s)$ this subsampling distribution. Under the null $\Delta = 0$, $\sqrt{n}\widehat{\Delta} \xrightarrow{d} \mathcal{Z}_1 - \mathcal{Z}_2$. Therefore, the threshold c_n can be chosen as the quantile of order α of the distribution $\mathcal{Z}_1^s - \mathcal{Z}_2^s$.

D Additional results on inference

D.1 A simple test of $\widehat{m} \in \partial\mathcal{M}$

We recall that $\widehat{m} \in \partial\mathcal{M}$ is implied by $\widetilde{P} \notin \mathcal{P}$, and the converse also holds with probability tending to one. In turn, testing for $\widetilde{P} \notin \mathcal{P}$ is equivalent to testing for $Q^{-1}\widetilde{P} \notin \mathcal{M}$. More generally, the issue of whether a given vector μ belongs to the set $\mathcal{M}$ of first K moments of a probability distribution on $[0,1]$ is known as the truncated Hausdorff problem. Several necessary and sufficient conditions have been established for this problem. Proposition D.1 below, which is proved for instance by Krein and Nudel'man (1977, Theorems III.2.3 and III.2.4), provides a characterization that is rather simple to use. It relies on the matrices A_μ, B_μ and C_μ defined in Subsection 2.3.

Proposition D.1 *$\widetilde{P} = Q\mu \in \mathcal{P}$ if and only if:*
- A_μ and $-C_\mu$ are positive when K is even;
- $A_\mu - B_\mu$ and B_μ are positive when K is odd.

D.2 Sufficient conditions for Assumption 3.2

The following lemma shows that under some restrictions on the segregation index, the differentiability requirement of Assumptions 3.2 and 3.3 are satisfied.

Lemma D.2 *Suppose that Assumption 2.1 holds and ν is C^1:*

1. *If h does not depend on m_{01}, $\underline{\theta}$ and $\overline{\theta}$ are directionally differentiable at any $m \in \mathring{\mathcal{M}}$ and differentiable almost everywhere.*
2. *If Assumption 2.2 holds, then $\underline{\theta}$ and $\overline{\theta}$ are differentiable at any $m \in \mathring{\mathcal{M}}$.*

This result relies on a remarkable property of the Chebyshev-Markov moment problem, namely convexity (resp. concavity) of $\min_{F\in\mathcal{D}_m} \int h(x)dF(x)$ (resp. $\max_{F\in\mathcal{D}_m} \int h(x)dF(x)$). Part 1 applies for instance to the Atkinson index for all K, while Part 2 applies to the Theil index, or to the Atkinson index for any odd $K < 50$.

D.3 Uniformly valid but conservative confidence intervals

We consider here confidence intervals that are uniformly valid but also conservative in general. Uniformity is achieved over sets $\mathcal{F}_{u,v}$ defined by

$$\mathcal{F}_{u,v} = \{F \in \mathcal{D} : F(1-u) - F(u) \geq v\}$$

for $0 < u < 1/2$ and $v \in (0,1)$. Restricting F_p to $\mathcal{F}_{u,v}$ may be seen as a reinforcement of Assumption 3.1, since Bernoulli distributions on p are never in $\mathcal{F}_{u,v}$, for any $0 < u < 1/2$ and $v \in (0,1)$. Besides, $F_p \in \cup_{0<u<1/2, v\in(0,1)} \mathcal{F}_{u,v}$ is equivalent to F_p satisfying Assumption 3.1. In this sense, the two conditions become equivalent when u and v tend to zero.

We consider the confidence region on P_0 with asymptotic level $1-\alpha$ defined by

$$I_{1-\alpha} = \{P \in (0,1)^K : (P - \widetilde{P})'\Sigma(P)^{-1}(P - \widetilde{P}) \leq \chi^2_K(1-\alpha)\},$$

where $\chi^2_K(1-\alpha)$ is the $1-\alpha$ quantile of a χ^2_K distribution.[1] Then define

$$\text{CI}^3_{1-\alpha} = \left[\inf_{P\in\mathcal{P}\cap I_{1-\alpha}} \underline{\theta}(Q^{-1}P), \sup_{P\in\mathcal{P}\cap I_{1-\alpha}} \overline{\theta}(Q^{-1}P)\right]. \tag{D.1}$$

[1] An alternative would be to replace $\Sigma(P)$ by $\Sigma(\widetilde{P})$ in $I_{1-\alpha}$. There are two reasons for using $\Sigma(P)$ instead of $\Sigma(\widetilde{P})$. First, in practice $\Sigma(\widetilde{P})$ is often singular because one of the component of $\widetilde{P}$ is zero. Second, it has been shown that, in the case of binomial models (not multinomial as here), the finite sample performances of confidence intervals based on $\Sigma(P)$ are far better than those using $\Sigma(\widetilde{P})$ (Blyth and Still, 1983).

Theorem D.3 *Suppose that Assumptions 2.1 and 3.1 hold. Then, for all* $u \in (0, 1/2)$ *and* $v \in (0, 1)$,

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_{u,v}} \Pr_F([\underline{\theta}_{0F}, \overline{\theta}_{0F}] \subset CI^3_{1-\alpha}) \geq 1-\alpha.$$

This theorem shows actually that $\mathrm{CI}^3_{1-\alpha}$ is uniformly valid for the whole set $[\underline{\theta}_0, \overline{\theta}_0]$, not only for θ_0. This result is obtained under very mild assumptions. Even if these confidence intervals are conservative in general, simulations suggest that they may still be informative, especially when K is small.

D.4 Inference with a random unit size independent of p

The previous bounds are obtained when one is fully agnostic on the dependence between p and K, which is a safe option in the cases in which unit size might be a potential determinant of segregation. However, if one is ready to impose independence between these two variables, we can use units of size $\overline{K}$ to identify the first $\overline{K}$ unconditional moments of p. Actually, the vector of unconditional moments $m_0 = (m_{01}, ..., m_{0\overline{K}})$ is overidentified by

$$\mathbf{P}_0 = \mathbf{Q}m_0, \tag{D.2}$$

where $\mathbf{P}_0$ (resp. $\mathbf{Q}$) stacks vertically the vectors $P_0^k = (\Pr(X = 1|K = k), ..., \Pr(X = k|K = k))$ (resp. the matrices Q^k) for different k. Theorems 2.1 or 2.2 then apply directly with $K = \overline{K}$.

Apart from the accuracy gains due to the overidentification of m_0, the bounds on θ_0 are thus likely to be very close, since they exploit the knowledge of the first $\overline{K}$ moments, with $\overline{K}$ potentially large. In particular, θ_0 is point identified when $\overline{K} = \infty$, because the knowledge of all moments of a distribution on $[0, 1]$ fully characterizes it (see, e.g. Gut, 2005, Theorems 8.1 and 8.3), reducing $\mathcal{D}_{m_0}$ to a singleton. However, to avoid any incorrect inference, the independence assumption should not be used when the overidentification test based on Equation (D.2) is rejected.

We can still estimate $\mathbf{P}_0$ by constrained maximum likelihood, using distributions with at most $\overline{L} + 1$ support points, where $\overline{L}$ is the integer part of $(\overline{K} + 1)/2$. We have to use in this context

$$P_{0i}^k = \Pr(X = i|K = k) = \binom{k}{i} \sum_{j=1}^{\overline{L}+1} y_j x_j^i (1 - x_j)^{k-i}, \quad i \in \{0, ..., k\}, k \in \{2, ..., \overline{K}\}.$$

Given (D.2), we then estimate $m_0 = (m_{01}, ..., m_{0\overline{K}})$ by $\widehat{m} = (\mathbf{Q}'\mathbf{Q})^{-1} \mathbf{Q}\widehat{\mathbf{P}}$. Once more, $\widehat{m} \in \mathcal{M}$ by construction. The bounds can then be estimated as in the case with a single unit size, with K simply replaced by $\overline{K}$.

We can define a bootstrap confidence interval as follows. Letting $\widehat{m}_b = (\widehat{m}_{b1}, ..., \widehat{m}_{b\overline{K}})$ as before, we first draw K in its empirical distribution and then draw X conditional on $K = k$ according to the vector of probabilities $\widehat{P}_b^k = Q^k (\widehat{m}_{b1}, ..., \widehat{m}_{bk})'$. The bootstrap confidence interval can then be obtained using $\text{CI}^1_{1-\alpha}$ as in the case with a single K. We can also test for independence between p and K, for instance with a likelihood ratio test that compares the likelihood under independence with the likelihood obtained without independence. The critical value of this test can be obtained by bootstrapping under the null, as described above.

Finally, an important particular case occurs when some individuals in the unit are unobserved (e.g. survey data). If only $n_K < K$ individuals are sampled from units, then X denotes the number of individuals belonging to the group of interest in this subsample. As previously, X follows, conditional on p and n_K, a binomial distribution $B(n_K, p)$. Hence, the result for a fixed K applies by simply replacing K by n_K. Moreover, it is usually plausible to assume n_K to be independent of p conditional on K. Under this condition, the $\overline{n}_K$ first moments of p conditional on K are identified, $\overline{n}_K$ denoting the maximum of the support of n_K conditional on K. Therefore, we can recover bounds on the segregation index for the whole population using (B.3), by simply changing the unit size K by $\overline{n}_K$.

E Additional Monte Carlo Simulations

In this section, we investigate several other finite sample properties of our estimators, confidence intervals and tests. First, because the bootstrap confidence intervals may not be uniformly consistent, we investigate how the DGP affects its coverage rate. Table 1 displays these coverage rates with three alternative DGP, with K fixed to 6. The first one is discrete: p takes the values 0 and 1/3 with probabilities 0.9 and 0.1 respectively. This DGP was chosen so that $m_0 \in \partial\mathcal{M}$ for all $K \geq 3$, and the first two moments are close to those of our application. The second and third DGP are such that $\Phi^{-1}(p) \sim \mathcal{N}(-2.12, 1.56)$ and $\Phi^{-1}(p) \sim \mathcal{N}(-1.12, 1.56)$, leading to minority proportions higher than in our baseline specification ($E(p) \simeq 0.126$ and 0.273, respectively).

Table 1: Influence of the DGP on inference.

DGP	$[\underline{\theta}_0, \overline{\theta}_0]$	n	$[E(\widehat{\underline{\theta}}), E(\widehat{\overline{\theta}})]$ $(\sigma(\widehat{\underline{\theta}}))$ $(\sigma(\widehat{\overline{\theta}}))$	$CI^1_{0.95}$ Length	$CI^1_{0.95}$ $CR(\theta_0)$
Theil index					
$\Phi^{-1}(p) \sim \mathcal{N}(-2.12, 1.56)$	[0.506, 0.553]	100	[0.520, 0.521] (0.072) (0.072)	0.308	0.968
		1,000	[0.517, 0.538] (0.033) (0.033)	0.147	0.955
		10,000	[0.507, 0.551] (0.009) (0.010)	0.088	0.978
$\Phi^{-1}(p) \sim \mathcal{N}(-1.12, 1.56)$	[0.481, 0.511]	100	[0.491, 0.494] (0.060) (0.061)	0.233	0.945
		1,000	[0.487, 0.504] (0.022) (0.023)	0.100	0.943
		10,000	[0.482, 0.511] (0.007) (0.007)	0.054	1.000
Discrete	[0.677, 0.677]	100	[0.653, 0.653] (0.073) (0.073)	0.436	0.985
		1,000	[0.666, 0.666] (0.028) (0.028)	0.141	0.973
		10,000	[0.673, 0.673] (0.011) (0.011)	0.055	0.983
Duncan index					
$\Phi^{-1}(p) \sim \mathcal{N}(-2.12, 1.56)$	[0.649, 0.740]	100	[0.709, 0.712] (0.065) (0.064)	0.286	0.950
		1,000	[0.688, 0.722] (0.041) (0.033)	0.176	0.973
		10,000	[0.660, 0.738] (0.019) (0.008)	0.144	0.990
$\Phi^{-1}(p) \sim \mathcal{N}(-1.12, 1.56)$	[0.620, 0.677]	100	[0.653, 0.657] (0.060) (0.060)	0.236	0.940
		1,000	[0.637, 0.669] (0.030) (0.022)	0.128	0.968
		10,000	[0.620, 0.675] (0.010) (0.007)	0.088	1.000
Discrete	[0.931, 0.931]	100	[0.904, 0.904] (0.053) (0.053)	0.327	1.000
		1,000	[0.922, 0.922] (0.018) (0.018)	0.102	0.985
		10,000	[0.928, 0.929] (0.006) (0.006)	0.032	0.983

Note: for each DGP and each n, simulations are based on 400 draws of samples. In the discrete DGP, p takes values 0 and 1/3, with probabilities 0.9 and 0.1, leading to $T \simeq 0.677$ and $D \simeq 0.931$. The two other DGP correspond to larger minority proportion than in the baseline specification ($E(p) \simeq 0.126$ and 0.273, respectively). For the first, $T \simeq 0.533$ and $D \simeq 0.707$ while for the second, $T \simeq 0.497$ and $D \simeq 0.658$. In all cases, $K = 6$. $CR(\theta_0)$ denotes the coverage rate of the confidence interval $CI^1_{0.95}$ (i.e. $\Pr(\theta_0 \in CI^1_{1-\alpha})$).

As with our baseline specification, $\text{CI}^1_{0.95}$ is usually conservative. This was expected for the two continuous DGP, but we also observe this pattern with the discrete DGP, for which the asymptotic level is equal to 95%. Monte Carlo simulations (not reported here) reveal that we get closer to 95% for n larger than 10,000. For discrete DGP with higher minority proportions, we also observe levels that are closer to 95% for small or intermediate n.

Finally, Table 2 displays some elements about the performance of the bootstrap test of the binomial model proposed in the previous section. We use the same discrete distribution of p as before. The test performs well in practice, with true levels close to the nominal one except for $n = 100$ and $K = 3$, where it appears to be conservative.

Table 2: Tests of the binomial model: true levels of the bootstrap test (for a nominal level of 5%).

K	n	True level	K	n	True level
3	100	0.020	9	100	0.064
3	1,000	0.045	9	1,000	0.064
3	10,000	0.050	9	10,000	0.058
6	100	0.054	12	100	0.060
6	1,000	0.060	12	1,000	0.049
6	10,000	0.058	12	10,000	0.062

Note: for each (n, K), simulations are based on 2,000 draws. The distribution of p takes values 0 and 1/3 with probability 0.9 and 0.1 respectively.

F Proofs

F.1 Proof of Proposition B.1

The proof is very close to the one of Theorem 3.2 and we only emphasize the differences hereafter. Instead of considering $\widehat{P}_b$ and $\widetilde{P}^*$, we consider respectively the vectors $\widehat{V}_b = \left(\widehat{\Pr}(K=2), ..., \widehat{\Pr}(K=\overline{K}), \widehat{P}^{2\prime}, ..., \widehat{P}^{\overline{K}\prime}\right)'$ and

$$\widetilde{V}^* = \left(\widehat{\Pr}^*(K=2), ..., \widehat{\Pr}^*(K=\overline{K}), \widetilde{P}^{2*\prime}, ..., \widetilde{P}^{\overline{K}*\prime}\right)',$$

where

$$\widetilde{P}^{k*} = \frac{\sum_{i=1}^{n} \mathbb{1}\{K_i^* = k\} I_k(X_i^*)}{\sum_{i=1}^{n} \mathbb{1}\{K_i^* = k\}}$$

and $I_k(x) = (\mathbb{1}\{x = 1\}, ..., \mathbb{1}\{x = k\})'$. The asymptotic distribution of $\widehat{V}_b$ can be established exactly as in Theorem 3.1. Because $\widetilde{P}^{k*}$ is a ratio of averages, asymptotic normality of $\widetilde{V}^*$ follows by the Lindebergh-Feller central limit theorem and the delta method for the bootstrap.

In step 2, we consider, instead of $\widehat{P}^*$, the vector

$$\widehat{V}^* = \left(\widehat{\Pr}^*(K = 2), ..., \widehat{\Pr}^*(K = \overline{K}), \widehat{P}^{2*\prime}, ..., \widehat{P}^{\overline{K}*\prime}\right)',$$

but otherwise the same reasoning as in Theorem 3.2 applies. Regarding Step 3, note that by (B.4) and differentiability of ν, $\underline{\eta}^k$ and $\overline{\eta}^k$, the bounds are differentiable functions of $V = \left(\Pr(K = 2), ..., \Pr(K = \overline{K}), P^{2\prime}, ..., P^{\overline{K}\prime}\right)'$. Finally, Steps 4 and 5 are the same as in Theorem 3.2.

F.2 Proof of Theorem C.1

First, if $P_0 \in \overset{\circ}{\mathcal{P}}$, we have

$$\Pr(C_n) \leq 1 - \Pr(LR_n = 0) = 1 - \Pr(\widetilde{P} \in \mathcal{P}) \to 0.$$

Next, suppose that $P_0 \in \partial\mathcal{P}$, $\overline{C}_{P_0}$ is a half space and $\alpha < 1/2$. As previously, we let hereafter LR_n^* denote the bootstrap counterpart of LR_n, with the underlying vector of probability $\widehat{P}_b$ for X. When $\widetilde{P} \notin \mathcal{P}$, or equivalently $LR_n > 0$, we have $\widehat{P}_b = \widehat{P}$. As a result,

$$LR_{2n}^* | LR_n > 0 \overset{d}{=} LR_n^* | LR_n > 0. \tag{F.1}$$

Thus,

$$\begin{aligned} \Pr(C_n) &= \Pr(LR_n > c_{1-\alpha}(LR_{2n}^*) | LR_n > 0) \Pr(LR_n > 0) \\ &= \Pr(LR_n > c_{1-\alpha}(LR_n^*) | LR_n > 0) \Pr(LR_n > 0) \\ &= \Pr(LR_n > c_{1-\alpha}(LR_n^*)). \end{aligned} \tag{F.2}$$

Now, a similar reasoning as in the first part of the proof of Lemma F.1 yields

$$\begin{aligned} LR_n &= n \left\| \widetilde{P} - \rho(\widetilde{P}) \right\|^2 + o_P\left(n \left\| \widetilde{P} - \rho(\widetilde{P}) \right\|^2\right), \\ LR_n^* &= n \left\| \widetilde{P}^* - \rho(\widetilde{P}^*) \right\|^2 + o_P\left(n \left\| \widetilde{P}^* - \rho(\widetilde{P}^*) \right\|^2\right). \end{aligned}$$

Moreover, by Lemma F.1 and (A.9)

$$\begin{aligned}\sqrt{n}(\rho(\widetilde{P}) - \widetilde{P}) &= \left[\pi_{\overline{C}_{P_0}} - \mathrm{Id}\right]\left(\sqrt{n}(\widetilde{P} - P_0)\right) + o_P(1),\\ \sqrt{n}(\rho(\widetilde{P}^*) - \widetilde{P}^*) &= \left[\pi_{\overline{C}_{P_0}} - \mathrm{Id}\right]\left(\sqrt{n}(\widetilde{P}^* - \widehat{P}_b)\right) + o_P(1).\end{aligned}$$

By the continuous mapping theorem, $LR_n \xrightarrow{d} LR_\infty$, with $LR_\infty = \left\|\pi_{\overline{C}_{P_0}}(\mathcal{Z}) - \mathcal{Z}\right\|^2$. By Part 1 of the proof of Theorem 3.2, we also have $LR_n^* \xrightarrow{d} LR_\infty$. Moreover, because $\overline{C}_{P_0}$ is a half space, there exists $u_0 \in \mathbb{R}^K$ such that $LR_\infty = \max(u_0' Z, 0)$. Moreover, $\Pr(u_0'\mathcal{Z} \leq 0) = 1/2 < 1-\alpha$. Thus, the distribution of LR_∞ is continuous at its quantile $c_{1-\alpha}(LR_\infty)$. As a result, by Theorem 1.2.1 of Politis et al. (1999), and with probability one,

$$\Pr(LR_n > c_{1-\alpha}(LR_n^*)) \to \alpha.$$

The second point of the theorem follows using (F.2).

Finally, suppose that $P_0 \notin \mathcal{P}$. Then $\widetilde{P} \xrightarrow{P} P_0$ and by the continuous mapping theorem, $\widehat{P} = \rho(\widetilde{P}) \xrightarrow{P} \rho(P_0) \neq P_0$. Therefore,

$$\sum_{k=0}^K \widetilde{P}_k \ln(\widetilde{P}_k/\widehat{P}_k) \xrightarrow{P} \sum_{k=0}^K P_{0k} \ln(P_{0k}/\rho_k(P_{0k})) > 0.$$

As a result, LR_n tends to infinity. On the other hand, a similar reasoning as previously shows that the asymptotic distribution LR_{2n}^* is the one of LR_∞. In other words, $c_{1-\alpha}(LR_n^*) = O_P(1)$. This implies that $\Pr(C_n) \to 1$ □

F.3 Proof of Lemma D.2

Let us define $\underline{h}(m) = \min_{F\in\mathcal{D}_m} \int h(x)dF(x)$, and let $\overline{h}$ be defined similarly. We also suppose without loss of generality that $\nu(., m_1)$ is increasing.

1. h is independent of m_{01}. For any $m = (m_1, ..., m_K)' \in \mathcal{M}$,

$$\underline{\theta}(m) = \nu\left(\underline{h}(m), m_{01}\right), \quad \overline{\theta}(m) = \nu\left(\overline{h}(m), m_{01}\right).$$

Now, by Proposition IV.P.4.2 of Krein and Nudel'man (1977), $\underline{h}$ (resp. $\overline{h}$) is convex (resp. concave). Thus, $\underline{h}$ and $\overline{h}$ admit directional derivatives in any $m \in \mathring{\mathcal{M}}$ (see, e.g., Hiriart-Urruty and Lemaréchal, 2001, p. 174) and are differentiable almost everywhere (Hiriart-Urruty and Lemaréchal, 2001, Theorem 4.2.3 of Chapter B). The result follows by the chain rule, since ν is C^1.

2. g **satisfies Assumption 2.2.** We have, for any $m \in \overset{\circ}{\mathcal{M}}$,

$$\{\underline{\theta}(m), \overline{\theta}(m)\} = \left\{\nu\left(\sum_{j=1}^{L+1} \underline{y}_j(m) h(\underline{x}_j(m)), m_1\right), \nu\left(\sum_{j=1}^{L+1} \overline{y}_j(m) h(\overline{x}_j(m)), m_1\right)\right\},$$

where $\underline{x}_j(m)$ (resp. $\overline{x}_j(m)$) are the roots of the polynomial $\underline{P}_m$ (resp. $\overline{P}_m$ and $\underline{y}_j(m)$ is the associated probability. Let us focus hereafter on the lower principal representation. Because

$$(\underline{y}_1(m), ..., \underline{y}_{L+1}(m))' = V(\underline{x}(m))^{-1}(1, m_1, ..., m_L)'$$

and $m \mapsto V(\underline{x}(m))^{-1}$ is smooth, it suffices to prove that $m \mapsto \underline{x}(m)$ is smooth. Now, the coefficients of $\underline{P}_m$ are smooth functions of m. Furthermore, the map from the coefficients of a polynomials to its roots is analytical at all vectors of coefficients such that the corresponding roots are all distinct (see, e.g. Hörmander, 1966), which is the case here. Hence the map $m \mapsto \underline{x}(m)$ is C^1. The result follows by the chain rule, since ν is also C^1 $\square$

F.4 Proof of Theorem D.3

Introduce the function $I(x) = (\mathbb{1}\{x = 1\}, ..., \mathbb{1}\{x = K\})'$ and the vectors $U_i(P_0) = \Sigma(P_0)^{-1/2}(I(X_i) - P_0)$, where $\Sigma(P_0)^{-1/2}$ denotes a square root matrix of $\Sigma(P_0)^{-1}$, and $\overline{U}(P_0) = \frac{1}{n}\sum_{i=1}^n U_i(P_0)$. Hereafter P_0 depends on the cdf F of p but we omit this dependency in the absence of ambiguity. Let $S_{1-\alpha}$ denote the sphere of radius $\sqrt{\chi^2_K(1-\alpha)}$ in $\mathbb{R}^K$. By definition of $\text{CI}^3_{1-\alpha}$,

$$\Pr_F(P_0 \in I_{1-\alpha}) = \Pr_F(\sqrt{n}\,\overline{U}(P_0) \in S_{1-\alpha}). \tag{F.3}$$

We have $E(U_i(P_0)) = 0$ and $V(U_i(P_0)) = I_K$, the identity matrix of size K. As a result, by the multivariate Berry-Esseen bound (see, e.g. Esseen, 1945, p.92),

$$|\Pr_F(\sqrt{n}\,\overline{U}(P_0) \in S_{1-\alpha}) - (1-\alpha)| \leq C(K)\frac{E_F(\sum_{j=1}^K U_{1j}(P_0)^4)}{n^{K/(K+1)}}, \tag{F.4}$$

where $\mathcal{Z} \sim \mathcal{N}(0, I_K)$, $C(K)$ is a constant independent of the distribution of X (and thus p) and $U_{1j}(P_0)$ is the jth component of $U_1(P_0)$. Now, we have

$$E_F\left(\sum_{j=1}^K U_{1j}(P_0)^4\right) \leq E\left[\left(\sum_{j=1}^K U_{1j}(P_0)^2\right)^2\right] = E\left[\|U_1(P_0)\|^4\right].$$

Besides,

$$\|U_1(P_0)\|^2 \leq \frac{\|I(X_1) - P_0\|^2}{\underline{\delta}(P_0)} \leq \frac{K+1}{\underline{\delta}(P_0)},$$

where $\underline{\delta}(P_0)$ is the smallest eigenvalue of $\Sigma(P_0)$. Some algebra show that $\underline{\delta}(P_0) = \min_{i=1}^K P_{0i}$. Thus,

$$E_F\left(\sum_{j=1}^K U_{1j}(P_0)^4\right) \leq \frac{(K+1)^2}{\min_i P_{0i}^2}.$$

Now, for all $i = 1, ..., K$, and all $F \in \mathcal{F}_{u,v}$

$$\begin{aligned}
P_{0i} &= E_F\left[\Pr_F(X = i|p)\right] \\
&= \binom{K}{i} \int_0^1 p^i (1-p)^{K-i} dF(p) \\
&\geq \int_u^{1-u} p^i (1-p)^{K-i} dF(p) \\
&= u^K \int_u^{1-u} \left(\frac{p}{u}\right)^i \left(\frac{1-p}{u}\right)^{K-i} dF(p) \\
&\geq u^K \int_u^{1-u} dF(p) \\
&\geq u^K v,
\end{aligned}$$

where the last inequality follows from $F \in \mathcal{F}_{u,v}$. As a result,

$$\sup_{F \in \mathcal{F}_{u,v}} E_F\left(\sum_{j=1}^K U_{1j}(P_0)^4\right) \leq \frac{(K+1)^2}{u^{2K} v^2}. \quad \text{(F.5)}$$

Combining (F.3), (F.4) and (F.5), we obtain

$$\lim_{n\to\infty} \sup_{F \in \mathcal{F}_{u,v}} |\Pr_F(P_0 \in I_{1-\alpha}) - (1-\alpha)| = 0. \quad \text{(F.6)}$$

Finally, remark that

$$\begin{aligned}
P_0 \in I_{1-\alpha} &\Rightarrow \underline{\theta}(Q^{-1}P_0) \in \underline{\theta}(Q^{-1}I_{1-\alpha}),\ \overline{\theta}(Q^{-1}P_0) \in \overline{\theta}(Q^{-1}I_{1-\alpha}) \\
&\Rightarrow [\underline{\theta}_{0F}, \overline{\theta}_{0F}] \subset \left[\inf_{P \in \mathcal{P} \cap I_{1-\alpha}} \underline{\theta}(Q^{-1}P),\ \sup_{P \in \mathcal{P} \cap I_{1-\alpha}} \overline{\theta}(Q^{-1}P)\right].
\end{aligned}$$

Thus, by definition of $\text{CI}^3_{1-\alpha}$, $\Pr_F([\underline{\theta}_{0F}, \overline{\theta}_{0F}] \subset \text{CI}^3_{1-\alpha}) \geq \Pr_F(P_0 \in I_{1-\alpha})$. The result follows by (F.6) □

F.5 Additional lemmas

In the following, we let, for any vector $x = (x_1, ..., x_K)'$, $x_0 = 1 - \sum_{k=1}^K x_k$. The following lemma is used in the proofs of Theorems 3.1 and 3.2.

Lemma F.1 *Suppose that Assumption 3.1 holds and* $(Q_n)_{n\in\mathbb{N}}$ *is a random sequence such that* $\sqrt{n}(Q_n - P_0) \xrightarrow{d} \mathcal{Z}$ *for some r.v.* $\mathcal{Z}$ *and* $P_0 \in \mathcal{P}$. *For any* $Q = (Q_1, ..., Q_K)$, *let* $\rho(Q) = \arg\min_{R\in\mathcal{P}} \sum_{k=0}^{K} Q_k \ln(Q_k/R_k)$. *Then*

$$\sqrt{n}\,(\rho(Q_n) - P_0)) = \pi_{\overline{C}_{P_0}}(\sqrt{n}(Q_n - P_0)) + o_P(1).$$

Proof: the proof is divided in three steps.

1. Approximation of $\rho(Q_n)$ **by a projection.** First, remark that for any $(x, y) \in \mathbb{R}^K$, $\|x - y\|^2 = \sum_{k=0}^{K}(x_k - y_k)^2/P_{0k}$. We also define $\pi_{\mathcal{P}}(Q) = \arg\min_{R\in\mathcal{P}} \|R - Q\|$. We first prove that, as $Q \to P_0$,

$$\|\rho(Q) - \pi_{\mathcal{P}}(Q)\| = o(\|Q - P_0\|), \tag{F.7}$$

First, $\rho = (\rho_1, ..., \rho_K)$ is continuous by Berge maximum theorem. Thus $||\rho(Q) - Q|| \to 0$ as $Q \to P_0$. Similarly, $|\rho_0(Q) - P_0| \to 0$, with $\rho_0(Q) = 1 - \sum_{k=1}^{K} \rho_k(Q)$. By Assumption 3.1, $P_{0k} > 0$ for $k = 0...K$. Thus, we also have $Q_k > 0$ for Q close enough to P_0. Then, by a Taylor expansion,

$$\begin{aligned}
\sum_{k=0}^{K} Q_k \ln(\rho_k(Q)/Q_k) &= \sum_{k=0}^{K} Q_k \left[\frac{\rho_k(Q) - Q_k}{Q_k} - \frac{(\rho_k(Q) - Q_k)^2}{2Q_k^2} + o\left((\rho_k(Q) - Q_k)^2\right)\right] \\
&= -\frac{1}{2}\sum_{k=0}^{K} \frac{(\rho_k(Q) - Q_k)^2}{Q_k} + o\left(\|\rho(Q) - Q\|^2\right) \\
&= -\frac{1}{2}\|\rho(Q) - Q\|^2 + o\left(\|\rho(Q) - Q\|^2\right),
\end{aligned} \tag{F.8}$$

Similarly, remark that $||\pi_{\mathcal{P}}(Q) - Q|| \le ||P_0 - Q||$ by definition of $\pi_{\mathcal{P}} = (\pi_{1\mathcal{P}}, ..., \pi_{K\mathcal{P}})$. Therefore, by a similar Taylor expansion,

$$\sum_{k=0}^{K} Q_k \ln(\pi_{k\mathcal{P}}(Q)/Q_k) = -\frac{1}{2}\|\pi_{\mathcal{P}}(Q) - Q\|^2 + o\left(\|P_0 - Q\|^2\right), \tag{F.9}$$

where we let $\pi_{0\mathcal{P}}(Q) = 1 - \sum_{k=1}^{K} \pi_{k\mathcal{P}}(Q)$. Now, (F.8) and (F.9) imply, by definition of $\rho(Q)$ and $\pi_{\mathcal{P}}(Q)$,

$$\|\pi_{\mathcal{P}}(Q) - Q\|^2 \le \|\rho(Q) - Q\|^2 \le \|\pi_{\mathcal{P}}(Q) - Q\|^2 + o\left(\|P_0 - Q\|^2 + \|\rho(Q) - Q\|^2\right).$$

This also implies that $\|\rho(Q) - Q\|^2 = O(\|P_0 - Q\|^2)$. Hence,

$$\|\rho(Q) - Q\|^2 - \|\pi_{\mathcal{P}}(Q) - Q\|^2 = o\left(\|P_0 - Q\|^2\right). \tag{F.10}$$

Now, remark that

$$\begin{aligned}
\|\rho(Q) - \pi_{\mathcal{P}}(Q)\|^2 &= \|\rho(Q) - Q\|^2 + \|Q - \pi_{\mathcal{P}}(Q)\|^2 + 2\langle \rho(Q) - Q, Q - \pi_{\mathcal{P}}(Q)\rangle \\
&= \|\rho(Q) - Q\|^2 - \|Q - \pi_{\mathcal{P}}(Q)\|^2 + 2\langle \rho(Q) - \pi_{\mathcal{P}}(Q), Q - \pi_{\mathcal{P}}(Q)\rangle \\
&\leq \|\rho(Q) - Q\|^2 - \|Q - \pi_{\mathcal{P}}(Q)\|^2 ,
\end{aligned}$$

where the inequality follows by the property of the projection onto a convex set. Combined with (F.10), this yields (F.7).

2. Asymptotic distribution of $\pi_{\mathcal{P}}(Q_n)$. Next, we show that

$$\sqrt{n}\,(\pi_{\mathcal{P}}(Q_n) - P_0) = \pi_{\overline{C}_{P_0}}(\sqrt{n}(Q_n - P_0)) + o_P(1). \tag{F.11}$$

Let us consider

$$f_n(h) = \sqrt{n}\left(\pi_{\mathcal{P}}(P_0 + h/\sqrt{n}) - P_0\right), \quad f_\infty(h) = \pi_{\overline{C}_{P_0}}(h).$$

and let $\widetilde{f}_n(h) = (f_n(h), f_\infty(h))$. Let us prove that for all sequence h_n, if a subsequence h_{n_k} tends to h, then

$$\widetilde{f}_{n_k}(h_{n_k}) \to (f_\infty(h), f_\infty(h)). \tag{F.12}$$

The results then follows by applying the extended continuous mapping (see van der Vaart, 2000, Theorem 18.11) to the functions $\widetilde{f}_n$ and the sequence $\sqrt{n}(Q_n - P_0)$.

To prove (F.12), note first that f_∞ is continuous as a projection. Therefore, we just have to prove that $f_{n_k}(h_{n_k}) \to f_\infty(h)$. Now, for all $h \in \mathbb{R}^K$ and as $t \to 0$ (see, e.g., Hiriart-Urruty, 1982),

$$\pi_{\mathcal{P}}(P_0 + th) = P_0 + t\pi_{\overline{C}_{P_0}}(h) + o(t).$$

Taking $t = 1/\sqrt{n}$, this implies that $f_{n_k}(h) - f_\infty(h) = o(1)$. Moreover, as a projection, $\pi_{\mathcal{P}}$ satisfies $\|\pi_{\mathcal{P}}(h') - \pi_{\mathcal{P}}(h)\| \leq \|h' - h\|$ for all (h, h') in $\mathbb{R}^K$. Thus,

$$\begin{aligned}
\|f_{n_k}(h_{n_k}) - f_\infty(h)\| &\leq \|f_{n_k}(h_{n_k}) - f_{n_k}(h)\| + \|f_{n_k}(h) - f_\infty(h)\| \\
&\leq \|h_{n_k} - h\| + o(1) = o(1).
\end{aligned}$$

Hence, (F.12), and thus (F.11), hold.

3. Asymptotic distribution of $\rho(Q_n)$. Finally, let us write

$$\begin{aligned}\sqrt{n}\,(\rho(Q_n)-P_0) &= \sqrt{n}\,(\rho(Q_n)-\pi_{\mathcal{P}}(Q_n))+\sqrt{n}\,(\pi_{\mathcal{P}}(Q_n)-P_0)\\ &= o_P\left(\sqrt{n}(Q_n-P_0)\right)+\pi_{\overline{C}_{P_0}}(\sqrt{n}(Q_n-P_0))+o_P(1).\\ &= \pi_{\overline{C}_{P_0}}(\sqrt{n}(Q_n-P_0))+o_P(1),\end{aligned}$$

where the second equality follows from (F.11) and (F.7) combined with $Q_n-P_0 \xrightarrow{P} 0$ (see, e.g., Lemma 2.12 of van der Vaart, 2000) □

The following lemma is used to establish the consistency of the bounds in the proof of Theorem 3.1.

Lemma F.2 *Suppose that Assumption 2.1 holds. then $\underline{\theta}$ and $\overline{\theta}$ are continuous at any $m \in \mathcal{M}$.*

Proof: we establish the continuity of $\overline{\theta}(.)$ only, the result being similar for $\underline{\theta}(.)$. First, remark that

$$\overline{\theta}(m)=\max_{F\in\mathcal{D}}\widetilde{g}(F,m) \text{ s.t. } \quad F\in G(m),$$

where $\widetilde{g}(F,m)=g(F,m_1)$ and G is the correspondence defined on $\mathcal{M}$ to $\mathcal{D}$ by $G(m)=\mathcal{D}_m$. To show continuity of $\overline{\theta}$, we check that the conditions of the Berge maximum theorem (see, e.g. Carter, 2001, Theorem 2.3) are satisfied. First we show that $\widetilde{g}$ is continuous with respect to the product topology on $\mathcal{D}\times\mathcal{M}$ (we consider the weak topology on $\mathcal{D}$ and the standard topology on $\mathcal{M}$). By an application of the triangular inequality and continuity of h, the function $(F,m)\mapsto \int h(x,m_1)dF(x)$ is continuous. Because ν is continuous by Assumption 2.1, $\widetilde{g}$ is continuous as well.

Second, because any sequence of distributions on $[0;1]$ is uniformly tight, $\mathcal{D}$ is compact for the topology induced by the weak convergence. Because $\mathcal{D}_m$ is closed for the weak convergence and $\mathcal{D}_m\subset\mathcal{D}$, $\mathcal{D}_m$ is also compact. Therefore, G is compact valued. Third, the domain and range of G are compact and the graph of G is $\varphi^{-1}(\{0\})$, with $\varphi(F,m)=\int(x,...,x^K)'dF-m$. Because φ is continuous, the graph of G is closed. As a consequence, G is upper hemicontinuous (see, e.g. Carter, 2001, Exercise 2.107).

Finally, we prove that G is lower hemicontinuous. We have to show that for any (F,m), $F\in G(m)$, and any sequence $m_n\to m$ ($m_n\in\mathcal{M}$), there exists a subsequence $(m_{n_k})_k$ and $F_{n_k} \in G(m_{n_k})$ such that $F_{n_k} \xrightarrow{d} F$. Let $\widetilde{m}_n \in \partial\mathcal{M}$ and $\lambda_n \in [0,1]$ be such that

$$m_n = \lambda_n m + (1-\lambda_n)\widetilde{m}_n.$$

By Theorem 2.1, G is reduced to a singleton on $\partial\mathcal{M}$. Let $\{\widetilde{F}_n\} = G(\widetilde{m}_n)$ and

$$F_n = \lambda_n F + (1-\lambda_n)\widetilde{F}_n.$$

By construction, $F_n \in G(m_n)$. Because $\|m_n - m\| \to 0$ (where $\|.\|$ denotes the euclidian norm), we also have

$$(1-\lambda_n)||m - \widetilde{m}_n|| \to 0.$$

If $\liminf ||m - \widetilde{m}_n|| > 0$, then $\lambda_n \to 1$, implying that $F_n \xrightarrow{d} F$. On the other hand, if $\liminf ||m - \widetilde{m}_n|| = 0$, there exists a subsequence $(m_{n_k})_k$ such that $||m - m_{n_k}|| \to 0$. Moreover, $m \in \partial\mathcal{M}$ because $m_{n_k} \in \partial\mathcal{M}$ and $\partial\mathcal{M}$ is closed. Because G is upper hemicontinuous and single-valued on $\partial\mathcal{M}$, it is continuous on $\partial\mathcal{M}$. As a result, $\widetilde{F}_{n_k} \xrightarrow{d} F$, implying also that $F_{n_k} \xrightarrow{d} F$. Hence, in all cases, we have proved that there exists a subsequence $(m_{n_k})_k$ and $F_{n_k} \in G(m_{n_k})$ such that $F_{n_k} \xrightarrow{d} F$. The result follows $\square$

References

Åslund, O. and O. N. Skans (2010): "Will I See You at Work? Ethnic Workplace Segregation in Sweden, 1985-2002," *Industrial and Labor Relations Review*, 63, 471–493.

Blyth, C. R. and H. A. Still (1983): "Binomial Confidence Intervals," *Journal of the American Statistical Association*, 78, 108–116.

Carter, M. (2001): *Foundations of Mathematical Economics*, MIT Press.

Esseen, C.-G. (1945): "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law," *Acta Mathematica*, 77, 1–125.

Gut, A. (2005): *Probability: A Graduate Course*, Springer-Verlag, New York.

Hellerstein, J. K. and D. Neumark (2008): "Workplace Segregation in the United States: Race, Ethnicity, and Skill," *The Review of Economics and Statistics*, 90, 459–477.

Hiriart-Urruty, J.-B. (1982): "At what points is the projection mapping differentiable?" *The American Mathematical Monthly,*, 89, 456–458.

Hiriart-Urruty, J.-B. and C. Lemaréchal (2001): *Fundamentals of Convex Analysis*, Springer.

Hörmander, L. (1966): *An introduction to complex analysis in several variables*, Van Nostrand, New York.

Krein, M. G. and A. A. Nudel'man (1977): *The Markov Moment Problem and Extremal Problems*, Translations of Mathematical monographs.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*, Springer.

van der Vaart, A. W. (2000): *Asymptotic Statistics*, Cambridge University Press.