Econometrica Supplementary Material

SUPPLEMENT TO "THE PRODUCTIVITY ADVANTAGES OF LARGE CITIES: DISTINGUISHING AGGLOMERATION FROM FIRM SELECTION": ADDITIONAL APPENDICES

(*Econometrica*, Vol. 80, No. 6, November 2012, 2543–2594)

BY PIERRE-PHILIPPE COMBES, GILLES DURANTON, LAURENT GOBILLON, DIEGO PUGA, AND SÉBASTIEN ROUX

This document contains a set of appendices with supplemental material. Appendix E extends the model to introduce worker mobility, consumption amenities, and urban crowding costs. Appendix F derives asymptotic properties of the estimator. Appendix G explains how we compute the minimization criterion to estimate the values of the parameters. Appendix H provides further details on the data. Appendix I explains how we implement alternative approaches to estimate TFP. Finally, Appendix J provides sector-level estimates using urban areas as spatial units.

APPENDIX E: LABOR MOBILITY, URBAN CROWDING COSTS, AND CONSUMPTION AMENITIES

IN THIS SECTION, we extend the model to introduce worker mobility, consumption amenities, and urban crowding costs in the spirit of Henderson (1974) and Roback (1982). Introducing worker mobility makes city sizes endogenous and equalizes equilibrium utility. Consumption amenities provide a reason for size heterogeneity across cities. Urban crowding costs (which include commuting and housing costs) provide a dispersion force that explains why workers do not end up all concentrating in a single city in equilibrium.

Workers are now freely mobile within and across cities, and their utility function is extended to incorporate amenities. For a worker in city i, utility is given by

$$V_i = U_i + B_i, \tag{E.1}$$

where U_i is the subutility derived from the consumption of differentiated products and from the consumption of the numéraire good. It is defined just as in equation (1) from the main text. The second term in equation (E.1), B_i, is the level of amenities (or quality of life) in this city. This simple parameterization for amenities is fairly standard and our imposition of additive separability is mostly for convenience.[1]

Next, we incorporate urban crowding costs through a simple version of the monocentric city model (Alonso (1964)). Production in each city takes place at a single point. Surrounding each city's center, there is a line with residences of constant unit length owned by absentee landlords. A resident living at a distance x from the city center incurs a cost of commuting to and from work of

[1]Minimally, we would only require V_i to be quasiconcave so that the associated expenditure function is well behaved.

 DOI: 10.3982/ECTA8442

$\varsigma(2x)^\rho$. Land rent at the city edge (i.e., the rental price of land in the best alternative use) is normalized to zero. The possibility of arbitrage across residential locations together with fixed unit lot size ensures that at the residential equilibrium, the city is symmetric with its edges located at a distance $N_i/2$ of the center (where N_i is total population in city i), and that the sum of commuting cost and land rent expenditures is the same for all residents and is equal to the commuting cost of those residents farthest away from the center ςN_i^ρ.

For simplicity, we keep a simple version of agglomeration economies where $D_i = 1$ for all cities, so that the effective labor supplied by an individual worker in city i is $a(N_i + \delta \sum_{j \neq i} N_j) = e^{A_i}$, while retaining all other aspects of our model. Indirect utility for a worker in city i can then be expressed as

$$W_i = B_i + e^{A_i} + \mathrm{CS}_i - \varsigma N_i^\rho, \tag{E.2}$$

where

$$\mathrm{CS}_i = \frac{\omega_i}{2(\gamma + \eta \omega_i)}(\alpha - P_i)^2 + \frac{\omega_i}{\gamma}\sigma_{Pi}^2 \tag{E.3}$$

is consumer surplus (CS) from consumption of the differentiated products and the numéraire good, which depends only on the number of product varieties available locally, ω_i, and on the mean, P_i, and variance, σ_{Pi}^2, of their prices (Melitz and Ottaviano (2008)). Free worker mobility must equate indirect utility across cities to some common level $\overline{W}$, so that $W_i = \overline{W}\ \forall i$.

Substituting (E.2), (E.3), the definition of the mean and variance of local prices, and the pricing equation (5) into the equality $W_i = \overline{W}$ yields, for each of the I cities, an equation relating its population, N_i, and the unit cost cutoff for all I cities, $\bar{h}_j$ for $j = 1, \dots, I$. These I equations can be solved simultaneously with the I free entry conditions (6) for N_j and $\bar{h}_j$ for $j = 1, \dots, I$ as a function of $\overline{W}$. Provided the urban crowding cost parameter ρ is large enough, so that urban crowding costs eventually dominate agglomeration benefits, there is a unique stable solution for population in each city for given $\overline{W}$. Then, conditional on which potential locations for cities are populated, the population constraint (i.e., the equality of the sum of population in all cities to aggregate population) determines $\overline{W}$. Finally, to determine which cities are populated, one must specify a mechanism for city formation. The simplest such mechanism is to allow the absentee landlords to operate as competitive profit-maximizing city developers as in Henderson (1974). In this case, the population constraint determines a minimum level of amenities below which cities are not populated. Cities with amenities above that threshold are inhabited by the level of population N_i^* that maximizes W_i in (E.2) for the level of amenities B_i of that city. This is such that $\partial N_i^*/\partial B_i > 0$, so that cities with greater amenities are larger in size. If $A > 0$ (i.e., if $\partial A_i/\partial N_i > 0$), this larger city goes together with higher nominal earnings for workers due to stronger agglomeration economies. If $S > 0$ (i.e., if $\partial S_i/\partial N_i > 0$), the larger city size also goes together with higher consumer surplus because consumers in larger cities enjoy greater variety of differentiated products at more favorable prices. On the other hand, larger cities have the disadvantage of higher costs associated with housing and commuting, and, in equilibrium, city sizes adjust so that the net advantages and disadvantages of larger cities exactly balance out against the value of the amenities they provide.

For our purposes, the main point to be drawn here is that our theoretical Proposition 1 in the main text still holds in this extended version of the model, since it only relies on equations (6), (7), and (10), and these are still satisfied.[2]

APPENDIX F: ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

In this section, we establish some asymptotic results for the vector of estimated parameters $\hat{\theta} = (\hat{A}, \hat{D}, \hat{S})$, which is chosen in a set of values denoted Φ. These results draw from Gobillon and Roux (2010), who studied a similar (but more general) setting. The asymptotic properties of the estimated parameters can be established using the same line of argument as Carrasco and Florens (2000), who studied the generalized method of moments estimator when there is a continuum of moments. Consistency can be proved using Theorem 1 of Carrasco and Florens (2000) under standard assumptions.

To ease the exposition, we index areas by $j = 1$ and $i = 2$. The vector of estimated parameters verifies

$$\hat{\theta} = \arg\min_{\theta \in \Phi} \|B_n \hat{m}_\theta\|,$$

where B_n is a (possibly random) sequence of bounded linear operators and

$$\hat{m}_\theta(u) = \hat{\lambda}_2\big(r_S(u)\big) - D\hat{\lambda}_1\big(S + (1 - S)r_S(u)\big) - A$$

is the empirical counterpart of $m_\theta(u)$ as given by equation (21), $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are some estimators of the quantile functions, and $r_S(u) = \max(0, \frac{-S}{1-S}) + [1 - \max(0, \frac{-S}{1-S})]u$. We begin by making the following assumptions:

ASSUMPTION F.1: *The function $\tilde{F}(\cdot)$ is three times differentiable with a continuous derivative. Its support is a bounded interval.*

ASSUMPTION F.2: *The set of admissible parameters Φ is compact.*

ASSUMPTION F.3: *The equation $m_\theta = 0$ has a unique interior solution θ_0 within Φ.*

[2]Provided, of course, the stability condition for the city population equilibrium holds.

ASSUMPTION F.4: *The estimators of the quantile functions are of the form*

$$\hat{\lambda}_k(u) = \int_0^1 \lambda_k^s(v)\, d_v K_{n_k}(u, v), \quad k \in \{1, 2\},$$

where $\lambda_k^s(\cdot)$ *is the sample quantile function of area* k, n_k *is the number of observations in area* k *(with* $n_1 + n_2 = n$*), and* $K_n(\cdot, \cdot)$ *is a general kernel function verifying one of the following alternative subassumptions*:

F.4.1 *We have* $K_n(u, v) = \frac{(u_i - u)\delta_{u_{i-1}}(v) + (u - u_{i-1})\delta_{u_i}(v)}{u_i - u_{i-1}}$ *for* $u \in [u_{i-1}, u_i)$, *where* $u_i = \frac{i}{n}$, $i \in \{0, \dots, n\}$, *and* $\delta_u(\cdot)$ *is a Dirac mass in* u.

F.4.2 *We have* $d_v K_n(u, v) = \frac{1}{\varphi_n} b(\frac{v-u}{\varphi_n})\, dv$, *where* $b(\cdot)$ *is a density and* $\varphi_n \to 0$ *as* $n \to +\infty$.

Assumptions F.1 and F.2 are standard regularity conditions. Assumption F.1 ensures that the quantile functions take bounded values. In practice, the compact set of parameters invoked in Assumption F.2 is of the form $[-M_A, M_A] \times [0, M_D] \times [-1 + \epsilon, 1 - \epsilon]$ with M_A and M_D positive and large, and ϵ very small. Assumption F.3 is an identification condition stating that there is a unique interior value θ_0 within Φ for which the continuum of equalities $m_{\theta_0}(u) = 0$, $u \in [0, 1]$ holds. Assumption F.4 states that the quantile estimators are either a linear interpolation between the sample quantiles computed at the observed ranks or some differential kernel estimators with vanishing bandwidth. These two types of estimators are presented in the nested specification proposed by Cheng and Parzen (1997) that makes use of the general kernel function $K_n(\cdot, \cdot)$. Assumptions F.1 and F.4 are enough to obtain continuity and uniform consistency of the quantile estimators:

LEMMA F.1: *Under Assumptions* F.1 *and* F.4, *the functions* $\hat{\lambda}_1(\cdot)$ *and* $\hat{\lambda}_2(\cdot)$ *are uniformly continuous over* $[0, 1]$, *and we have* $\hat{\lambda}_k(u) \to \lambda_k(u)$ *almost surely uniformly for all* $u \in [0, 1]$ *when* $n_k \to +\infty$, *where* $k \in \{1, 2\}$.

The proof for the interpolated sample quantiles is given by van der Vaart (2000). The proof for the differential kernel estimators can be found in Bae and Kim (2004).

Lemma F.1 ensures that the function $(\theta, u) \to \hat{m}_\theta(u)$ is uniformly continuous. Note that this would not be the case if the estimators of the quantile functions were discontinuous at values $u_{ki} = \frac{i}{n_k}$ for $i \in \{1, \dots, n_k - 1\}$. Lemma F.1 also ensures that $\hat{m}_\theta(u) \to m_\theta(u)$ almost surely uniformly for all (θ, u) as the sample quantiles converge to the true quantiles almost surely uniformly for all u.

We also make some assumptions regarding the linear operator of the minimization program. In these assumptions, we refer to the kernel of an intregral operator. The kernel $\bar{\ell}(\cdot, \cdot)$ of an integral operator L is a two-dimensional function such that $(Lf)(x) = \int_0^1 \bar{\ell}(x, u) f(u)\, du$.

ASSUMPTION F.5: *There is a* (*nonrandom*) *bounded linear operator* B *that may depend on* θ_0 *but not on* θ *such that* $Bm_{\theta_0} = 0 \Longrightarrow m_{\theta_0} = 0$ *and such that the kernel of* $B_n^* B_n$, $\ell_n(\cdot, \cdot)$ *converges to the kernel of* $B^* B$, $\ell(\cdot, \cdot)$ *in the sense that* $\int_0^1 \int_0^1 |\ell_n(u, v) - \ell(u, v)| \, du \, dv \to 0$ *almost surely*.

This assumption ensures that the sequence of bounded linear operators B_n converges (in a specific way) to B. We can prove the following lemma, which is used to establish the consistency of the estimated parameters.

LEMMA F.2: *Under Assumptions* F.1, F.2, F.3, F.4, *and* F.5, $\|B_n \hat{m}_\theta\|$ *is a continuous function of* θ *and* $\|B_n \hat{m}_\theta\| \to \|Bm_\theta\|$ *almost surely uniformly for all* θ *in the set of admissible parameters when* $n_k \to +\infty$, $k \in \{1, 2\}$.

PROOF: This proof is drawn from Gobillon and Roux (2010), Appendix 1. We have

$$\|B_n \hat{m}_\theta\| = \int_0^1 \int_0^1 \ell_n(u, v) \hat{m}_\theta(u) \hat{m}_\theta(v) \, du \, dv.$$

The function $(\theta, u) \to \hat{m}_\theta(u)$ is uniformly continuous and so is $(\theta, u, v) \to \hat{m}_\theta(u)\hat{m}_\theta(v)$. This yields that $\|B_n \hat{m}_\theta\|$ is uniformly continuous. We also have

$$\begin{aligned} \|B_n \hat{m}_\theta\| - \|Bm_\theta\| &= \langle \hat{m}_\theta, B_n^* B_n \hat{m}_\theta \rangle - \|Bm_\theta\| \\ &= \langle \hat{m}_\theta, (B_n^* B_n - B^* B) \hat{m}_\theta \rangle + \|B\hat{m}_\theta\| - \|Bm_\theta\|. \end{aligned} \tag{F.1}$$

Furthermore,

$$\begin{aligned} &\left| \langle \hat{m}_\theta, (B_n^* B_n - B^* B) \hat{m}_\theta \rangle \right| \\ &\quad \le \sup_{u, v \in [0,1]^2} \left| \hat{m}_\theta(u) \hat{m}_\theta(v) \right| \int_0^1 \int_0^1 \left| \ell_n(u, v) - \ell(u, v) \right| du \, dv. \end{aligned}$$

Since the quantiles take their values in an interval that is bounded, the function $|\hat{m}_\theta(u)\hat{m}_\theta(v)|$ is uniformly bounded in (θ, u, v) and n. Using Assumption F.5, we get that $|\langle \hat{m}_\theta, (B_n^* B_n - B^* B)\hat{m}_\theta \rangle| \to 0$ almost surely uniformly for all θ in Φ. We also have

$$\|B\hat{m}_\theta\| - \|Bm_\theta\| = \langle \hat{m}_\theta - m_\theta, B^* B \hat{m}_\theta \rangle + \langle m_\theta, B^* B(\hat{m}_\theta - m_\theta) \rangle, \tag{F.2}$$

where

$$\begin{aligned} \langle \hat{m}_\theta - m_\theta, B^* B \hat{m}_\theta \rangle &= \langle \hat{m}_\theta, B^* B(\hat{m}_\theta - m_\theta) \rangle \\ &= \int_0^1 \int_0^1 [\hat{m}_\theta(u) - m_\theta(u)] \hat{m}_\theta(v) \ell(u, v) \, du \, dv. \end{aligned}$$

Hence

$$\begin{aligned}&\left|\langle\hat{m}_\theta, B^*B(\hat{m}_\theta - m_\theta)\rangle\right|\\&\quad\le \sup_u\left|\hat{m}_\theta(u) - m_\theta(u)\right|\sup_u\left|\hat{m}_\theta(u)\right|\int_0^1\int_0^1\left|\ell(u,v)\right|du\,dv.\end{aligned}$$

We have $\int_0^1\int_0^1|\ell(u,v)|\,du\,dv < +\infty$. This is because for $|\ell(u,v)| > 1$, we have $|\ell(u,v)| \le \ell(u,v)^2$ and $\int_0^1\int_0^1 \ell(u,v)^2\,du\,dv < +\infty$ as B is bounded. Also, $\sup_u|\hat{m}_\theta(u)|$ is bounded uniformly for all $(\theta, n) \in \Phi \times N$ since the quantiles are bounded according to Assumption F.1. Finally, we have $\sup_u|\hat{m}_\theta(u) - m_\theta(u)| \to 0$ almost surely uniformly for all $\theta \in \Phi$ because of Assumption F.4. We then get $|\langle\hat{m}_\theta, B^*B(\hat{m}_\theta - m_\theta)\rangle| \to 0$ almost surely uniformly for all $\theta \in \Phi$ and then, from equation (F.2), $\|B\hat{m}_\theta\| \to \|Bm_\theta\|$ almost surely for all $\theta \in \Phi$. Using equation (F.1), we finally obtain $\|B_n\hat{m}_\theta\| \to \|Bm_\theta\|$ almost surely for all $\theta \in \Phi$. *Q.E.D.*

PROPOSITION F.1: *Under Assumptions F.1–F.5, we have $\hat{\theta} \to \theta_0$ almost surely when $n_k \to +\infty$, $k \in \{1, 2\}$.*

This proposition follows from Lemma F.1 and Lemma F.2, and its proof is a direct application of Lemma 2.2 in White (1980, p. 736).

We now turn to the asymptotic distribution of the estimated parameters. We adapt Theorem 2 of Carrasco and Florens (2000) on asymptotic normality to our setting. We make an additional assumption.

ASSUMPTION F.6: *The function $\frac{\partial K_n}{\partial v}(u,v)$ is differentiable in u.*

This ensures that $\hat{m}_\theta(\cdot)$ is differentiable in u when $S \neq 0$. This property will be used when making a Taylor expansion of the function $\hat{m}_\theta(\cdot)$. Note that Assumption F.6 is verified for differentiable kernel estimators defined in F.4.2, but not for the interpolated sample quantile estimators defined in F.4.1, since they are not differentiable at points $u_{ki} = i/n_k$ for $i \in \{1, \dots, n_k - 1\}$. Hence, we restrict our attention to differentiable kernel estimators. Also, the differentiability of $\hat{m}_\theta(\cdot)$ is not granted when $S = 0$, as the function $r_S(\cdot)$ is not differentiable for that value of the truncation parameter. Assumptions F.1, F.4.2, and F.6 ensure the convergence of the estimated quantile functions to Brownian bridges uniformly on any closed interval in $(0, 1)$.

LEMMA F.3: *Under Assumptions F.1, F.4.2, and F.6, we have*

$$\sup_{u\in[\underline{u},\overline{u}]}\left|\sqrt{n_k}\big(\hat{\lambda}_k(u) - \lambda_k(u)\big) - \lambda_k'(u)Y_k^{n_k}(u)\right| \xrightarrow[n_k\to\infty]{p} 0$$

for $k \in \{1, 2\}$ and any interval $[\underline{u}, \overline{u}] \subset (0, 1)$, where $(Y_k^{n_k}(u))_{n_k}$ is a sequence of Brownian bridges.

PROOF: We apply the theorem of Cheng and Parzen (1997). They showed the convergence of the estimated quantile functions to Brownian bridges under some assumptions regarding the quantile functions and assumptions about the function $K_n(\cdot,\cdot)$. Their conditions on the quantile functions are met under the additional assumption that $\lambda_k(\cdot)$, $k\in\{1,2\}$, are three times differentiable. In our setting, this is the case when $\tilde{F}(\cdot)$ is three time differentiable, which is granted by Assumption F.1. Their conditions on the function $K_n(\cdot,\cdot)$ are met when this function is a differential kernel with a vanishing bandwidth, which is granted by Assumption F.4.2. *Q.E.D.*

We now show that under our assumptions, our set of estimated equalities $\hat{m}_{\theta_0}$ is asymptotically normal when the number of observations in each of the two areas goes to infinity at the same speed as n.

ASSUMPTION F.7: *We have* $\lim_{n\to\infty}\frac{n_k}{n}=\omega_k>0$, $k\in\{1,2\}$.

LEMMA F.4: *Under Assumptions* F.1, F.2, F.3, F.4.2, F.5, F.6, *and* F.7, *when* $S_0\neq 0$, *where* S_0 *is the truncation parameter in* $\theta_0=(A_0,D_0,S_0)$, $\sqrt{n}\hat{m}_{\theta_0}(\cdot)\xrightarrow{d} N(0,\tilde{L})$ *in distribution over any interval* $[\underline{u},\overline{u}]\subset(0,1)$, *where* $\tilde{L}$ *is a covariance operator with kernel* $\tilde{\ell}_{\theta_0}(\cdot,\cdot)$ *such that*

$$\begin{aligned}\tilde{\ell}_{\theta_0}(u,v)&=\lambda_2'\big(r_{S_0}(u)\big)\lambda_2'\big(r_{S_0}(v)\big)\\&\quad\times\left[\frac{1}{\omega_2}C_{r_{S_0}}(u,v)+\frac{1}{\omega_1}\frac{1}{(1-S_0)^2}C_{S_0+(1-S_0)r_{S_0}(u)}(u,v)\right],\end{aligned}\tag{F.3}$$

where $C_h(u,v)=h(u)\wedge h(v)-h(u)h(v)$.

PROOF: From Lemma F.3, we get

$$\sqrt{n_2}\big(\hat{\lambda}_2(u)-\lambda_2(u)\big)=\lambda_2'(u)\big[Y_{n_2}^2(u)+e_{n_2}^2(u)\big],\tag{F.4}$$

$$\sqrt{n_1}\big(D\hat{\lambda}_1(u)-D\lambda_1(u)\big)=D\lambda_1'(u)\big[Y_{n_1}^1(u)+e_{n_1}^1(u)\big],\tag{F.5}$$

where $\lim_{n_k\to\infty}\sup_{u\in[\underline{u},\overline{u}]}|e_{n_k}^k(u)|\xrightarrow{p}0$, $k\in\{1,2\}$. Applying equation (F.4) in $r_S(u)$ and equation (F.5) in $\tilde{r}_S(u)=S+(1-S)r_S(u)$, we get

$$\begin{aligned}\hat{m}_\theta(u)&=\frac{\lambda_2'(r_S(u))}{\sqrt{n_2}}\big[Y_{n_2}^2\big(r_S(u)\big)+e_{n_2}^2\big(r_S(u)\big)\big]\\&\quad-D\frac{\lambda_1'(\tilde{r}_S(u))}{\sqrt{n_1}}\big[Y_{n_1}^1\big(\tilde{r}_S(u)\big)+e_{n_1}^1\big(\tilde{r}_S(u)\big)\big].\end{aligned}$$

Deriving the equality $m_{\theta_0}(u)=0$ with respect to u, we obtain

$$\lambda_2'\big(r_{S_0}(u)\big)=D_0(1-S_0)\lambda_1'\big(\tilde{r}_{S_0}(u)\big).$$

Using these last two equations, we get

$$\hat{m}_{\theta_0}(u) = \lambda_2'\big(r_{S_0}(u)\big)\left[\frac{1}{\sqrt{n_2}}Y_{n_2}^2\big(r_{S_0}(u)\big) - \frac{1}{\sqrt{n_1}}\frac{1}{1-S_0}Y_{n_1}^1\big(\tilde{r}_{S_0}(u)\big)\right] + \lambda_2'\big(r_{S_0}(u)\big)\left[\frac{1}{\sqrt{n_2}}e_{n_2}^2\big(r_{S_0}(u)\big) - \frac{1}{\sqrt{n_1}}\frac{1}{1-S_0}e_{n_1}^1\big(\tilde{r}_{S_0}(u)\big)\right].$$

Defining $e_{n_2,n_1}(v) = \sqrt{\frac{n}{n_2}}e_{n_2}^2(r_{S_0}(u)) - \sqrt{\frac{n}{n_1}}\frac{1}{1-S_0}e_{n_1}^1(\tilde{r}_{S_0}(u))$, we deduce from the properties of e_{n_2} and e_{n_1} that $\lim_{n\to\infty}\sup_{v\in U}|e_{n_2,n_1}(v)| \xrightarrow{p} 0$. From this, we deduce that $\sqrt{n}\hat{m}_{\theta_0}(\cdot)$ converges in distribution to a normal process whose covariance function is denoted $\tilde{\ell}_{\theta_0}(\cdot,\cdot)$. We have

$$\tilde{\ell}_{\theta_0}(u,v) = \lim_{n\to\infty}\operatorname{cov}\big(\sqrt{n}\hat{m}_{\theta_0}(u), \sqrt{n}\hat{m}_{\theta_0}(v)\big),$$

where

$$\begin{aligned}
&\operatorname{cov}\big(\sqrt{n}\hat{m}_{\theta_0}(u), \sqrt{n}\hat{m}_{\theta_0}(v)\big) \\
&\quad= \operatorname{cov}\Bigg(\lambda_2'\big(r_{S_0}(u)\big)\left[\sqrt{\frac{n}{n_2}}Y_{n_2}^2\big(r_{S_0}(u)\big) - \sqrt{\frac{n}{n_1}}\frac{1}{1-S_0}Y_{n_1}^1\big(\tilde{r}_{S_0}(u)\big)\right], \\
&\qquad \lambda_2'\big(r_{S_0}(v)\big)\left[\sqrt{\frac{n}{n_2}}Y_{n_2}^2\big(r_{S_0}(v)\big) - \sqrt{\frac{n}{n_1}}\frac{1}{1-S_0}Y_{n_1}^1\big(\tilde{r}_{S_0}(v)\big)\right]\Bigg) \\
&\quad= \lambda_2'\big(r_{S_0}(u)\big)\lambda_2'\big(r_{S_0}(v)\big)\left[\frac{n}{n_2}\operatorname{cov}\big(Y_{n_2}^2\big(r_{S_0}(u)\big), Y_{n_2}^2\big(r_{S_0}(v)\big)\big)\right. \\
&\qquad \left. + \frac{n}{n_1}\frac{1}{(1-S_0)^2}\operatorname{cov}\big(Y_{n_1}^1\big(\tilde{r}_{S_0}(u)\big), Y_{n_1}^1\big(\tilde{r}_{S_0}(v)\big)\big)\right] \\
&\quad= \lambda_2'\big(r_{S_0}(u)\big)\lambda_2'\big(r_{S_0}(v)\big)\left[\frac{n}{n_2}C_{r_{S_0}}(u,v) + \frac{n}{n_1}\frac{1}{(1-S_0)^2}C_{\tilde{r}_{S_0}}(u,v)\right],
\end{aligned}$$

with $C_h(u,v) = h(u)\wedge h(v) - h(u)h(v)$ for a given function h. Hence

$$\tilde{\ell}_{\theta_0}(u,v) = \lambda_2'\big(r_{S_0}(u)\big)\lambda_2'\big(r_{S_0}(v)\big) \times \left[\frac{1}{\omega_2}C_{r_{S_0}}(u,v) + \frac{1}{\omega_1}\frac{1}{(1-S_0)^2}C_{\tilde{r}_{S_0}}(u,v)\right].$$

Q.E.D.

The expression of the kernel involves $C_h(\cdot,\cdot)$, which is the covariance function of a Brownian bridge when h is the identity function ($h = I_d$). For any other h, we have $C_h(u,v) = C_{I_d}(h(u), h(v))$: the covariance function of the Brownian bridge is evaluated at the arguments once they have been transformed by h. In equation (F.3), $C_h(\cdot,\cdot)$ is evaluated for functions h corresponding to two rank transformations resulting from the selection process.

Lemma F.4 is finally used in the application of Theorem 2 of Carrasco and Florens (2000) to obtain the asymptotic distribution of the estimated parameters:

PROPOSITION F.2: *Under Assumptions* F.1, F.2, F.3, F.4.2, F.5, F.6, *and* F.7, *the asymptotic distribution of* $\hat{\theta}$ *is given by*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, V),$$

with

$$V = \left\| B\frac{\partial m_{\theta_0}}{\partial \theta'} \right\|^{-2} \left\langle B\frac{\partial m_{\theta_0}}{\partial \theta'}, B\tilde{L}B^* B\frac{\partial m_{\theta_0}}{\partial \theta'} \right\rangle \left\| B\frac{\partial m_{\theta_0}}{\partial \theta'} \right\|^{-2}.$$

For the proof, see Carrasco and Florens (2000).

APPENDIX G: IMPLEMENTATION OF THE MINIMIZATION CRITERION

In this section, we explain how we compute the minimization criterion of equation (25), which is used to estimate the values of the parameters.

First note that the data consist of a set of log productivities in large cities (indexed by i) and in small cities (indexed by j), ranked in ascending order and denoted Φ_i and Φ_j, respectively. From these data, for any θ, we need to be able to evaluate $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ at any ranks $u \in [0, 1]$ to compute $M(\theta) = \int_0^1 [\hat{m}_\theta(u)]^2\, du + \int_0^1 [\hat{\tilde{m}}_\theta(u)]^2\, du$. For that purpose, we construct some estimators $\hat{\lambda}_i(u)$ and $\hat{\lambda}_j(u)$ of the quantiles $\lambda_i(u)$ and $\lambda_j(u)$. Focusing on large cities (replace i with j for small cities), we start from the set of log productivities

$$\Phi_i = \left[\phi_i(0), \dots, \phi_i(E_i - 1)\right]',$$

where E_i is the number of establishments in i and $\phi_i(0) < \cdots < \phi_i(E_i - 1)$. We can construct the sample quantiles at the observed ranks as $\hat{\lambda}_i(\frac{k}{E_i}) = \phi_i(k)$ for $k \in \{0, \dots, E_i - 1\}$. For any other rank $u \in (0, 1)$, the estimators of the quantiles are recovered by linear interpolation,

$$\hat{\lambda}_i(u) = \left(k_i^* + 1 - uE_i\right)\hat{\lambda}_i\left(\frac{k_i^*}{E_i}\right) + \left(uE_i - k_i^*\right)\hat{\lambda}_i\left(\frac{k_i^* + 1}{E_i}\right), \tag{G.1}$$

where $k_i^* = \lfloor uE_i \rfloor$ and $\lfloor \cdot \rfloor$ denotes the integer part. From equation (G.1) and the corresponding expression for j, we can use the empirical counterparts of equations (21) and (24),

$$\hat{m}_\theta(u) = \hat{\lambda}_i\left(r_S(u)\right) - D\hat{\lambda}_j\left(S + (1 - S)r_S(u)\right) - A,$$

$$\hat{\tilde{m}}_\theta(u) = \hat{\lambda}_j\left(\tilde{r}_S(u)\right) - \frac{1}{D}\hat{\lambda}_i\left(\frac{\tilde{r}_S(u) - S}{1 - S}\right) + \frac{A}{D},$$

to compute $\hat{m}_\theta(u)$ and $\hat{\hat{m}}_\theta(u)$ at any rank u and for any θ. We then consider $K = 1001$ ranks evenly distributed over the interval [0, 1]. These ranks are denoted u_k, $k \in \{0, \dots, K\}$, with $u_0 = 0$ and $u_K = 1$. We approximate the two subcriteria using the formulas

$$\int_0^1 \left[\hat{m}_\theta(u)\right]^2 du \approx \frac{1}{2}\sum_{k=1}^{K}\left\{\left[\hat{m}_\theta(u_k)\right]^2 + \left[\hat{m}_\theta(u_{k-1})\right]^2\right\}(u_k - u_{k-1}),$$

$$\int_0^1 \left[\hat{\hat{m}}_\theta(u)\right]^2 du \approx \frac{1}{2}\sum_{k=1}^{K}\left\{\left[\hat{\hat{m}}_\theta(u_k)\right]^2 + \left[\hat{\hat{m}}_\theta(u_{k-1})\right]^2\right\}(u_k - u_{k-1}).$$

The estimated parameters $\hat{\theta}$ are those that minimize the sum of these two quantities.

APPENDIX H: Further Description of the Data

SIREN (Système d'Identification du Répertoire des ENtreprises)

These annual data contain the following information for all registered active establishments in France: establishment identifier, year, legal status, municipality identifier, municipality identifier for the headquarter, and four-digit sector identifier. We note that establishments in the finance and real estate sectors are not included. Over 1994–2002, the data set contains 27,282,570 observations, including 3,074,401 in 2000. A significant share of these observations corresponds to establishments with no salaried worker.

DADS (Déclarations Annuelles de Données Sociales)

These annual data contain the following information for all establishments with at least one salaried worker in France during the year: establishment identifier, firm identifier, year, legal status, four-digit sector identifier, total working days, total working hours, total labor costs, and total wages.[3] We note that the last three variables are also available by skill group (see below for the definition of skill groups).

Over 1994–2002, this data set contains 14,535,717 observations, including 1,693,312 in 2000. These numbers of establishments are smaller than for SIREN because only establishments with at least one salaried worker are included here.

When merging DADS with SIREN, we end up with 11,183,561 observations at the establishment–year level including 1,298,954 for 2000. The decrease in

[3]In France, total wages and total labor costs differ because employers need to pay various taxes and contributions over and above the wages paid to the employees. The most important among these are social security and pension contributions.

the number of observations mostly comes from the absence of the finance and insurance sectors in SIREN.

BRN–RSI (Bénéfices Réels Normaux and Régime Simplifié d'Imposition)

These annual data result from the merge of the BRN and RSI data. All French firms must report balance sheet information either in the "standard" manner (larger firms), which appears in the BRN, or in a "simplified" way (small firms), which appears in the RSI. These data contain the following information for all registered firms in France: firm identifier, year, two-digit industry identifier, number of full time equivalent workers, total revenue, value added, operating profit (*excédent brut d'exploitation*), total wages, social security and pension contributions, value of tangible assets, and value of total assets (including intangible assets). Asset values and the shares of wages and capital in value added were computed by Boutin and Quantin (2008), but only up to 2002. Therefore, we restrict our attention to the 1994–2002 period.

Over 1994–2002, this data set contains 16,023,214 observations, including 705,785 firms in the BRN data and 1,185,522 firms in the RSI data in 2000.

An additional decrease in the number of observations happens when merging DADS–SIREN with BRN–RSI. It occurs because firms that cease their operations often do not make any report for their last year of activity and thus are not present in the BRN–RSI data. We end up with 1,704,415 firms and 2,352,898 establishments observed at least once during the study period, including 1,136,479 establishments for 2000.

Further Data Restrictions

We restrict our attention to firms in continental France (thus excluding Corsica and overseas territories) in all manufacturing sectors and in business services, with the exception of finance and insurance given their specific reporting rules. We also exclude distribution and consumer services from our main estimations. The assignment of a specific location to distribution (which involves moving goods across locations) is difficult and the estimation of a production function in consumer services is more problematic. This leads to a data set, that includes 363,001 firms and 503,475 establishments. For multiestablishment firms, we aggregate establishments at the firm–geographical unit level. This leaves us with 430,237 establishments. We only select establishments in the same industry as their firm and delete firms with establishments in more than 20 employment areas, as they create mass points in the data. This leaves us with 350,291 firms and 367,241 establishments (including 339,223 monoestablishment firms). We retain information on all establishments from all firms with six employees or more and finally end up with data on 148,705 firms and 166,086 establishments (including 137,014 monoestablishment firms) observed at least once during the study period. We also report results for firms with between one and five employees in Table 6 of the main text.

To sum up, for each firm and each year between 1994 and 2002, we know the firm's value added, the value of its capital, and its sector of activity. For each establishment within each firm, we know its location and the number of hours worked by its employees by skill level.[4]

To obtain reliable estimates of A, D, and S from firm-level TFP, we need to exclude extreme outliers. We thus trim the 1 percent of observations with the highest TFP values and the 1 percent of observations with the lowest TFP values in each city size class, and end up with 162,765 establishments (98 percent of 166,086) in the estimations that combine all establishments from all sectors (such as the bottom panel of Table III) and 134,275 establishments (98 percent of 137,013) in the majority of estimations that focus on monoestablishment firms (such as Tables I and II).

Definition of Skill Groups

We now explain how the three skill groups (low, intermediate, and high) are defined. For white-collar workers, we follow Burnod and Chenu (2001) since there is no official classification.

The low-skill group includes low-skill blue collars (in craft, manufacturing, and agriculture) and low-skill white collars (sales clerk, employees in personal services). (In the French standard occupational classification, the following two-digit occupations are included: 55, *employés de commerce*; 56, *personnels des services directs aux particuliers*; 67, *ouvriers non qualifiés de type industriel*; 68, *ouvriers non qualifiés de type artisanal*; and 69, *ouvriers agricoles*.)

The intermediate-skill group includes high-skill blue collars (in craft, manufacturing, handling, and transport), taxi drivers, and intermediate-skill white collars (administrative employees). (In the French standard occupational classification, the following two-digit occupations are included: 54, *employés administratifs d'entreprise*; 62, *ouvriers qualifiés de type industriel*; 63, *ouvriers qualifiés de type artisanal*; 64, *chauffeurs*; and 65, *ouvriers qualifiés de la manutention, du magasinage et du transport*.)

The high-skill group includes managers (in craft, manufacturing or sales), executive and knowledge workers (doctors, lawyers, executives, professors, scientists, engineers), intermediate professions (primary teachers, intermediate professions in health, social work, administration and sales firms, religious, technicians, foremen). (In the French standard occupational classification, the following two-digit occupations are included: 21, *chefs d'entreprise artisanale*; 22, *chefs d'entreprise industrielle ou commerciale de moins de* 10 *salariés*; 23,

[4]The merged data set contains much more information than is usually available. For instance, U.S.-based research relies either on sectoral surveys or on 5-year censuses for which value added is difficult to compute. We instead have exhaustive annual data. We also have information on the number of hours worked by skill level instead of total employment as is often the case.

chefs d'entreprise industrielle ou commerciale de 10 *salariés et plus*; 31, *professionnels de la santé et avocats*; 33, *cadres de la fonction publique*; 34, *professeurs, professions scientifiques*; 35, *professions de l'information, des arts et des spectacles*; 37, *cadres administratifs et commerciaux d'entreprises*; 38, *ingénieurs et cadres techniques d'entreprises*; 42, *instituteurs et assimilés*; 43, *professions intermédiaires de la santé et du travail social*; 46, *professions intermédiaires administratives et commerciales des entreprises*; 47, *techniciens*; and 48, *contremaîtres, agents de maîtrise*.)

APPENDIX I: IMPLEMENTATION OF ALTERNATIVE APPROACHES TO PRODUCTIVITY

Olley–Pakes

In this section, we present three alternative approaches to TFP estimation. The first is the methodology proposed by Olley and Pakes (1996) to account for the endogeneity of production factors when estimating the parameters of equation (34). These authors considered that the residual ϕ_t can be decomposed into an unobserved factor φ_t, which is potentially correlated with labor and capital, and an uncorrelated error term η_t such that $\phi_t = \varphi_t + \eta_t$. They supposed that the unobserved factor φ_t can be rewritten as its projection on its lag and an innovation: $\varphi_t = \kappa(\varphi_{t-1}) + \xi_t$. They also made the crucial assumption that capital investment at time t depends on the capital stock and the unobserved factor φ_t: $I_t = i_t(k_t, \varphi_t)$. The function i_t is supposed to be strictly increasing in the unobserved factor. It can be inverted such that $\varphi_t = f_t(k_t, I_t)$. Equation (28) can then be rewritten as

$$\ln(V_t) = \beta_2 \ln(l_t) + \sum_{s=1}^{3} \sigma_s l_{s,t} + \Psi_t(k_t, I_t) + \eta_t, \tag{I.1}$$

where the auxiliary function Ψ_t is defined as

$$\Psi_t(k_t, I_t) = \beta_{0,t} + \beta_1 \ln(k_t) + f_t(k_t, I_t). \tag{I.2}$$

Equation (I.1) can be estimated with OLS after $\Psi_t(k_t, I_t)$ has been replaced with a third-order polynomial crossing k_t, I_t, and year dummies. This allows recover of some estimators of the labor and skill share coefficients ($\hat{\beta}_2$ and $\hat{\sigma}_s$), as well as the auxiliary function ($\hat{\Psi}_t$). It is then possible to construct the variable

$$v_t = \ln(V_t) - \hat{\beta}_2 \ln(l_t) - \sum_{s=1}^{3} \hat{\sigma}_s l_{s,t}. \tag{I.3}$$

From equation (I.2), the lagged value of the unobserved factor φ_{t-1} can be approximated by $\hat{\Psi}_{t-1}(k_{t-1}, I_{t-1}) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1})$. Using equations (I.1)–

(I.3), and the projection of the unobserved factor on its lag, the value-added equation then becomes

$$v_t = \beta_{0,t} + \beta_1 \ln(k_t) + \kappa\big(\hat{\Psi}_{t-1}(k_{t-1}, I_{t-1}) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1})\big) + \vartheta_t, \tag{I.4}$$

where ϑ_t is a random error. The function $\kappa(\cdot)$ is approximated by a third-order polynomial and equation (I.4) is estimated with nonlinear least squares. We thus recover some estimators of the year dummies ($\hat{\beta}_{0,t}$) and the capital coefficients ($\hat{\beta}_1$). An estimator of ϕ_t is then given by $\hat{\phi}_t = v_t - \hat{\beta}_{0,t} - \hat{\beta}_1 \ln(k_t)$.

Although the Olley–Pakes method allows us to control for simultaneity, it has some drawbacks. In particular, we need to construct investment from the data: $I_t = k_t - k_{t-1}$. Since investment enters lagged into equation (I.4), we must observe firms for at least three consecutive years to compute their TFP with this method. Other observations must be dropped. Furthermore, the investment equation $I_t = i_t(k_t, \varphi_t)$ can be inverted only if $I_t > 0$. Hence, we can keep only observations for which $I_t > 0$. This double selection may introduce a bias, for instance, if (i) there is greater "churning" (i.e., entry and exits) in denser areas and (ii) age and investment affect productivity positively. Then more establishments with a low productivity may be dropped in high density areas. In turn, this may increase the measured difference in local productivity between areas of low and high density. Reestimating OLS TFP on the same sample of firms used for Olley–Pakes shows that this is, fortunately, not the case on French data.

Levinsohn–Petrin

We also implement the approach proposed by Levinsohn and Petrin (2003). Its main difference with Olley and Pakes (1996) is that the quantity of inputs is used to account for the unobservables instead of investment. The unobserved factor is then rewritten as $\varphi_t = f_t(k_t, Ic_t)$, where Ic_t is the consumption of inputs. Otherwise, the estimation procedure remains the same. However, we lose fewer observations, since the use of materials instead of investment means we need to observe firms for 2 consecutive years instead of 3.

Cost Shares

Alternatively, a TFP measure can be constructed using cost shares as estimates of the labor and capital coefficients in equation (28). The costs of labor and capital were evaluated by Boutin and Quantin (2008) for each cell defined by the three-digit industry, the year, and the number of firm employees (less than 5, 5–20, 20–50, 100, more than 100). The share of capital (resp. labor) in these costs is denoted $\hat{\beta}_{1,t}$ (resp. $\hat{\beta}_{2,t}$). Implicitly, we assume constant returns to scale as we have $\hat{\beta}_{1,t} + \hat{\beta}_{2,t} = 1$. The predicted value added based on capital and labor is $\ln V_t^p = \hat{\beta}_{1,t} \ln(k_t) + \hat{\beta}_{2,t} \ln(l_t)$. The following specification can then be estimated with OLS:

$$\ln(V_t) - \ln V_t^p = \beta_{0,t} + \sum_{s=1}^{3} \sigma_s l_{s,t} + \tilde{\phi}_t.$$

Denoting $\hat{\beta}_{0,t}^c$ and $\hat{\sigma}_s^c$ the estimated coefficients, the TFP measure is given by

$$\hat{\tilde{\phi}}_t = \ln(V_t) - \ln V_t^p - \hat{\beta}_{0,t} - \sum_{s=1}^{3} \hat{\sigma}_s l_{s,t}.$$

For all methods, the TFP of a firm is the firm-level average of yearly TFP over the period 1994–2002. The TFP estimates we recover with these four approaches are highly correlated. The correlation between OLS TFP and Olley–Pakes TFP is 0.73. The correlation between OLS TFP and Levinsohn–Petrin TFP is 0.85. The correlation between OLS TFP and cost-share TFP is 0.93. Unsurprisingly, these alternative methods to estimate TFP give results that are qualitatively similar for A, D, and S at the sector level.

APPENDIX J: ESTIMATIONS FOR URBAN AREAS

TABLE J.I

CITIES WITH POPULATION > 200,000 VERSUS POPULATION < 200,000[a]

	OLS, Monoestablishments				
	$\hat{A}$	$\hat{D}$	$\hat{S}$	R^2	Obs.
Sector	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.062 (0.004)*	0.966 (0.020)	0.003 (0.002)	0.951	21,187
Apparel, leather	0.041 (0.010)*	1.392 (0.053)*	0.009 (0.006)	0.987	5711
Publishing, printing, recorded media	0.173 (0.008)*	1.324 (0.053)*	−0.001 (0.004)	0.986	8991
Pharmaceuticals, perfumes, soap	0.039 (0.054)	1.210 (0.139)	−0.007 (0.056)	0.885	1014
Domestic appliances, furniture	0.118 (0.011)*	1.217 (0.050)*	0.007 (0.006)	0.992	6172
Motor vehicles	0.076 (0.035)*	1.291 (0.179)	0.003 (0.035)	0.818	1410

(*Continues*)

TABLE J.I—*Continued*

	OLS, Monoestablishments				
	$\hat{A}$	$\hat{D}$	$\hat{S}$	R^2	Obs.
Sector	(1)	(2)	(3)	(4)	(5)
Ships, aircraft, railroad equipment	0.097 (0.035)*	1.140 (0.202)	−0.005 (0.039)	0.798	966
Machinery	0.076 (0.005)*	1.057 (0.027)*	−0.004 (0.004)	0.984	14,084
Electric and electronic equipment	0.079 (0.009)*	1.022 (0.045)	−0.003 (0.005)	0.957	5550
Building materials, glass products	0.068 (0.014)*	1.077 (0.061)	0.001 (0.010)	0.933	3048
Textiles	0.050 (0.015)*	1.101 (0.055)	0.001 (0.007)	0.935	3273
Wood, paper	0.087 (0.010)*	1.103 (0.041)*	−0.002 (0.005)	0.992	5629
Chemicals, rubber, plastics	0.075 (0.010)*	1.048 (0.041)	0.003 (0.005)	0.969	5119
Basic metals, metal products	0.074 (0.005)*	1.056 (0.024)*	0.000 (0.002)	0.997	13,911
Electric and electronic components	0.079 (0.024)*	1.000 (0.080)	0.002 (0.042)	0.944	2485
Consultancy, advertising, business services	0.190 (0.016)*	1.101 (0.030)*	−0.006 (0.024)	0.976	35,738
All sectors	0.087 (0.002)*	1.241 (0.009)*	0.000 (0.001)	0.998	134,275

[a]The asterisk denotes that $\hat{A}$ and $\hat{S}$ are significantly different from 0 at 5% and $\hat{D}$ is significantly different from 1 at 5%.

REFERENCES

ALONSO, W. (1964): *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press. [1]

BAE, J., AND S. KIM (2004): "The Uniform Consistency of the Sample Kernel Quantile Process," *Bulletin of the Korean Mathematical Society*, 41 (3), 403–412. [4]

BOUTIN, X., AND S. QUANTIN (2008): "Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises (1984–2002)," *Économie et Statistique*, 413 (1), 47–64. [11, 14]

BURNOD, G., AND A. CHENU (2001): "Employés qualifiés et non-qualifiés: Une proposition d'aménagement de la nomenclature des catégories socioprofessionnelles," *Travail et Emploi*, 86, 87–105. [12]

CARRASCO, M., AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16 (6), 797–834. [3,6,9]

CHENG, C., AND E. PARZEN (1997): “Unified Estimators of Smooth Quantile and Quantile Density Functions,” *Journal of Statistical Planning and Inference*, 59 (2), 297–307. [4, 7]

GOBILLON, L., AND S. ROUX (2010): “Quantile-Based Inference of Parametric Transformations Between Two Distributions,” Unpublished Manuscript, CREST-INSEE. [3,5]

HENDERSON, J. V. (1974): “The Sizes and Types of Cities,” *American Economic Review*, 64 (4), 640–656. [1,2]

LEVINSOHN, J., AND A. PETRIN (2003): “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 70 (2), 317–342. [14]

MELITZ, M., AND G. I. P. OTTAVIANO (2008): “Market Size, Trade and Productivity,” *Review of Economic Studies*, 75 (1), 295–316. [2]

OLLEY, G. S., AND A. PAKES (1996): “The Dynamics of Productivity in the Telecommunication Equipment Industry,” *Econometrica*, 64 (6), 1263–1297. [13,14]

ROBACK, J. (1982): “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90 (6), 1257–1278. [1]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*. New York: Cambridge University Press. [4]

WHITE, H. (1980): “Nonlinear Regression on Cross-Section Data,” *Econometrica*, 48 (3), 721–746. [6]

Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, 2 Rue de la Charité, 13236 Marseille Cedex 02, France; ppcombes@univmed.fr; http://www.vcharite.univ-mrs.fr/pp/combes/,

Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA and CEPR; duranton@wharton.upenn.edu; https://realestate.wharton.upenn.edu/profile/21470/,

Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris Cedex 20, France and PSE, CREST, CEPR, and IZA; laurent.gobillon@ined.fr; http://laurent.gobillon.free.fr/,

Centro de Estudios Monetarios y Financieros (CEMFI), Casado del Alisal 5, 28014 Madrid, Spain and CEPR; diego.puga@cemfi.es; http://diegopuga.org,

and

CREST (INSEE), 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France and PSE; roux@ensae.fr; http://www.crest.fr/ses.php?user=3049.

Manuscript received February, 2009; final revision received February, 2012.