

Bensaid, Bernard; Croiset, Solen; Gary-Bobo, Robert J.

Working Paper

Economies of Density, Team Synergies and Unobserved Heterogeneity: A Study of Home Care Services

CESifo Working Paper, No. 9864

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bensaid, Bernard; Croiset, Solen; Gary-Bobo, Robert J. (2022) : Economies of Density, Team Synergies and Unobserved Heterogeneity: A Study of Home Care Services, CESifo Working Paper, No. 9864, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/263794>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Economies of Density, Team Synergies and Unobserved Heterogeneity: A Study of Home Care Services

Bernard Bensaid, Solen Croiset, Robert Gary-Bobo

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Economies of Density, Team Synergies and Unobserved Heterogeneity: A Study of Home Care Services

Abstract

In the home-care services industry, caregivers drive to visit patients scattered in a district and deliver various care services at the patient's home. We use a unique data set, with a standard panel structure, recording the hours of service and the exact number of miles traveled by each employee. The estimated average transportation cost curves, giving the ratio of miles traveled per hour of service, as a function of hours of service at the patient's home, typically exhibit economies of density. These results suggest that the home-care sector is characterized by excess entry and that a form of franchise-bidding regulation could be appropriate. We study the unobserved heterogeneity of employees and use a quadratic-inlogs, finite-mixture model with latent groups to uncover a finite number of employee types with different cost curves. Some types, but not all, have U-shaped average cost curves yielding an aggregate average transportation cost that is itself U-shaped. Economies of density are mainly due to team synergies operating at the district level. We also show that our model can be interpreted as a test of organizational efficiency based on the Beardwood-Halton-Hammersley (1959) theorem of combinatorial optimization.

JEL-Codes: I110, I180, L220, L230.

Keywords: home care services, elderly, economies of scale, economies of density, transportation costs, Beardwood-Halton-Hammersley theorem, applied econometrics, panel data, unobserved heterogeneity, finite mixture models, entropy criteria.

Bernard Bensaid
Groupe AVEC
France – 94300 Vincennes
bernard.bensaid@avec.fr

Solen Croiset
Université Paris-I Panthéon-Sorbonne
Paris / France
solen.croiset@etu.univ-paris1.fr

Robert Gary-Bobo
Université Paris-I Panthéon-Sorbonne
Paris / France
garybobo@univ-paris1.fr

14 July 2022

The authors acknowledge support from Groupe Avec, a network of Home-Care services operating in France, who facilitated data collection. The data used for the present study have been anonymized and will be available to interested researchers.

1 Introduction

Economies of density are a particular category of economies of scale. These economies are present in many industries in which production takes place in geographical space, or when transportation plays a substantial role. They are characterized by the fact that unit costs decrease when the density of clients increases, keeping the overall scale of operations and input levels fixed. Economies of density may explain why some markets have a tendency towards monopolization, and may carry important consequences in terms of efficient regulation. This notion is a useful concept, not only in Industrial Organization, but also in Urban Economics and Regional Science, because it lies at the heart of agglomeration economies as a basic principle.

In the present paper, we study economies of density in a sector with a very simple and transparent technology: the Home Care Services for the elderly. In this industry, in essence, caregivers drive to visit patients scattered in a district and deliver various care services at the patient's home. To study the technological properties of these services, we use a unique data set, made available to us by a large nonprofit network of Home Care Services located in France. The data, with a standard panel structure, records the hours of service and the exact number of miles traveled by each employee, every month, during several years and in many different districts.¹ With the help of these data we study the technical relationship linking miles traveled to hours of service at the patient's home — a real transportation cost function — and estimate the intensity of economies of density. It is rarely possible to estimate a cost function with the help of completely disaggregated data, allowing separate observations of each employee. As a preliminary step, we use standard regression and panel-data econometric tools, applied to log-linear models, to show the existence of economies of density. Our regression work can be interpreted as a statistical test of organizational efficiency, based on the Beardwood-Halton-Hammersley Theorem (see Beardwood *et al.* (1959)).² According to this view, the data does not exhibit significant inefficiencies in the network's transportation expenditures.

We then go further and estimate non-linear (quadratic-in-logs) models with a finite set of latent types, or latent groups: a finite-mixture model, to capture the unobserved heterogeneity of the network's employees. These models are estimated by the Maximum Likelihood method, using a sequential EM algorithm to produce accurate preliminary estimates. Results show the robust presence of economies of density, taking the form of U-shaped unit cost curves. In other words, the ratio of miles traveled to hours of service at the patient's home substantially declines with the latter hours of service on a sizeable interval of hours. These average cost curves are themselves the weighted average of curves estimated for each latent type of employee taken separately. We show that some of these type-contingent curves, in turn, exhibit economies of density, albeit with various degrees of intensity. The economies of density are mainly generated by district

¹Distance traveled is recorded because each employee's transportation cost is reimbursed by the organization.

²The Beardwood-Halton-Hammersley 1959 theorem is a well-known result in combinatorial optimization. The theorem sheds light on the *Traveling Salesman Problem*. See our discussion below.

effects, that we interpret as *team synergies* operating at the district (*i.e.*, local branch) level.³ The analysis of latent types shows that some types only are responsible for the economies of density, and that these types are concentrated in rural districts.

Finally, we estimate several variants of our model with a number of latent types varying from 1 to 9. Unobserved groups are said to be well-identified, or equivalently, individuals are well-classified, if each observation belongs to a given group with a sufficiently high posterior probability. If this is generally the case, the quality of classification generated by the model is deemed good. We discuss the choice of the appropriate number of latent types, using traditional information criteria (such as Akaike and Bayesian Information Criteria, *i.e.*, AIC and BIC) combined with *entropy* indices, to assess the *quality of classification* produced by the EM and ML estimates. We conclude that, even if 6 or 8 types are optimal according to BIC or AIC, a model with two types only should be selected, for two types is the best compromise if we value well-identified groups.

Our contribution is twofold: the estimation results themselves, to the extent that they entail consequences for regulation policy, and the way these results are established by means of a finite-mixture model. The consequences for public policies and regulation are of course related to the structure of density economies that we uncover. The latter economies, possibly combined with the existence of economies of scope, could justify a shift from the current, relatively free competition regime⁴ to a regime of local regulated monopolies, with a form of franchise bidding, as is often the case, for instance, in the garbage collection services. Given that, in France, the clients of home care services are heavily subsidized, the government has a sizeable stake in a regulatory reform. Indeed, we show that, using our estimated elasticities, the French government could save around a hundred million euros every year.

The methodological originality of our approach is to use a finite-mixture approach with panel data, collected at the level of the individual employee, to analyze the role played by the unobserved heterogeneity of workers and local situations in the estimation of a cost or production function. The statistical model, which is a form of automatic classification method, goes beyond what ordinary methods can achieve by showing that economies of density, arising in less densely populated, rural and peripheral areas, are in fact generated by a specific fraction of the employees only.

Economies of density have been studied in the literature. A pioneering paper, devoted to American airline companies, is due to Caves *et al.* (1984); see also Braeutigam *et al.* (1984). These economies are present in many industries, in particular in postal services, parcel-delivery and garbage-collection services. This notion is crucial to understand the growth of online e-commerce companies like Amazon, but also firms like UPS and Fedex. On *economies of density* at Amazon,

³All other things equal, an isolated employee's travelled distance is proportional to her hours of service at the patient's home.

⁴There exists some regulatory barriers to entry and some licensing rules, though.

see Houde *et al.* (2017). On returns to scale in postal services, see, *e.g.*, Cazals *et al.* (2001)). On garbage collection, see Dubin and Navarro (1988). Economies of density are of course also present in transportation firms: on airlines, see Brueckner and Spiller (1994); on railroad freight transportation, see Bitzan and Keeler (2007); on container shipping, see Xu and Itoh (2018). Economies of density in international transportation may have a strong influence: the agglomeration process in a given country can be influenced by the geography of another country through the channel of trade (more precisely the development and location of ports); see Behrens *et al.* (2006); Mori and Nishikimi (2002). Economies of density are also present in network industries: on electricity distribution, see Roberts (1986); on local telephone networks, see, *e.g.*, Guldmann (1990); on the U.S. water utility industry, see Torres and Morrison-Paul (2006). We find these economies in various personal service industries, see, *e.g.*, Morikawa (2011); in chain stores like Wal-Mart, see Holmes (2011); in the retail banking sector, see, Aguirregabiria *et al.* (2016), and also in the agricultural sector, see Holmes et Lee (2012).

More generally, economies of density are related to the important theme of *agglomeration economies* in Regional and Urban Economics (cf. Combes et Gobillon (2015), Rosenthal et Strange (2020)). The present article shows a pure case of density economies, that is, the direct result of the interaction of a production technology and geographical space, in the presence of random shocks, without any need for the presence of externalities (on this problematic, see Ciccone et Hall (1996)).

The present article can be viewed as a contribution to the study of teams in firms, since economies of density are ultimately generated by a better organization, coordination and team work.⁵ The routing and scheduling of the caregivers and nurses' tours constitute a complicated variant of the *Traveling Salesman Problem*, *i.e.*, the problem of finding the shortest path joining a given number of points on a map. This is a classic and difficult problem in Combinatorial Optimization with direct applications in Home Care and Home Health Care industries.⁶ The recent years have witnessed the development of an important literature on the so-called Home Care Routing and Scheduling Problem in applied Operations Research and Computer Science. This line of research proposes different algorithms to improve the allocation of tasks and the daily tours of nurses and caregivers. For a presentation and further references, see the surveys of Fikar and Hirsch (2017), Cissé *et al.* (2017).

There exists a growing empirical literature on Home Care Services in Health Economics. The recent literature has addressed many questions, but to the best of our knowledge, did not study the problem posed here.⁷

⁵See for instance, among other contributions: on incentives in teams, Bandiera *et al.* (2013); on the team work of nurses, Bartel *et al.* (2014); on emergency departments and moral hazard, Chan (2016); on the effect of team bonuses, Friebel *et al.* (2017); on team-specific human capital (among doctors), Chen (2021); on social skills and teams, Weidman and Deming (2021); on team-specific capital in innovation processes, Jaravel *et al.* (2018).

⁶On the *Traveling Salesman Problem*, and the Beardwood-Halton-Hammersley Theorem, see Applegate *et al.* (2006). The asymptotic behavior of the length of an optimal tour, joining a set of randomly selected points in a given space is the topic of the Beardwood, Halton and Hammersley Theorem.

⁷A recent article by David and Kim (2018), close to ours insofar as it uses personnel data, studies the

In the following, Section 2 presents the intuition for the existence of economies of density in the home care services. Section 3 presents the results obtained with log-linear models, using standard regressions and panel data econometrics. Section 4 presents the model with unobserved heterogeneity and latent groups, discusses ML estimation results and derives the U-shaped average cost curves. Section 5 presents results obtained about the prior and posterior probabilities of latent types and discusses the optimal number of latent groups. Section 6 gathers concluding remarks. A series of appendices provide additional material, estimation results and some additional robustness tests.

2 Economies of Density in the Home Care Services Industry

The technology of home care services is extremely simple. The territory is divided into districts. In each district, there is a team of employees, the caregivers. Employees travel everyday to visit the homes of elderly or disabled persons who need help. The caregivers, mostly women, typically help the patients to accomplish various everyday life tasks: they serve meals, assist ablutions, etc. They spend some time at a patient's home and then drive to visit another patient.

2.1 The Caregivers' Tours

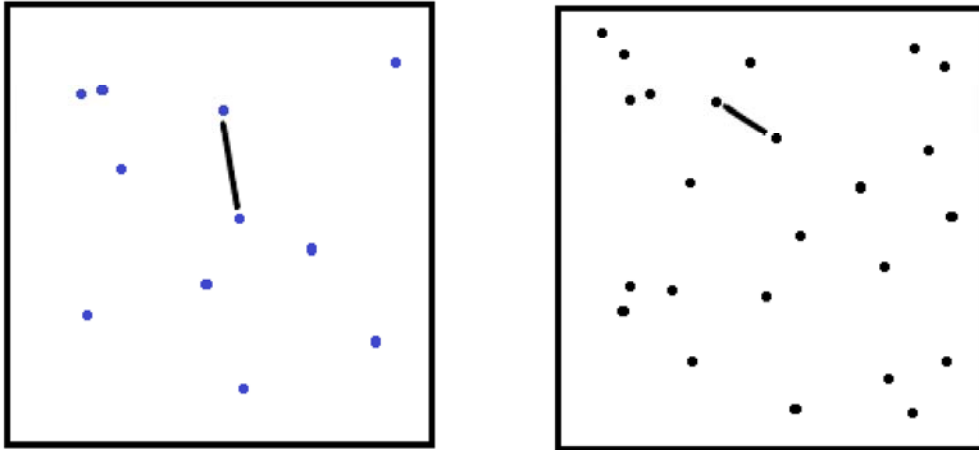
In each district, a manager is in charge of planning the *tours* of a team of employees. The routing and scheduling is frequently revised to adapt to random shocks: employees on sick leave, resignations, temporary absence of patients, changes in the list and addresses of clients, etc.

The district manager tries to reach certain goals as much as possible, for instance: seeks to always send the same caregiver to help a given client; send a person everyday at noon sharp, etc. An optimal organization minimizes the distance traveled by employees, subject to various constraints. To increase flexibility, the home care services have recourse to a combination of full-time, part-time and temporary workers. It follows that the total miles traveled by each employee randomly fluctuates. This variability is of course a blessing for the econometrician, being a source of natural experiments that allows one to identify some key parameters of the firm's cost function.

Let us now provide some intuitions for the reasons why home care services must exhibit economies of density. Consider a portion of geographical space that is served by a given service.

importance of care continuity, estimating the impact of nurse handoffs, due to unplanned employee absences, on hospital readmissions. Forder *et al.* (2014) use a production function approach to explain the home-care patient's quality of life as a function of service intensity. A few papers have used structural microeconomic models to discuss intergenerational family choices concerning formal and informal care (see, e.g., Stabile *et al.* (2006), Byrne *et al.* (2009)). On this theme, see also Bonsang (2009). On the impact of subsidies on recourse to home health care, see Orsini (2010).

Figure 1. Distance Traveled Between Two Clients and Density of Clients



On Figure 1, the left-hand square is the map of some area. On this map, the points are elderly clients. Distances — as the crow flies — between the clients' homes can easily be measured. The solid line segment between two points represents the distance between a client and the nearest next patient. Suppose that we now double the density of clients on the territory. This can be the result of changing for an area with a higher density of seniors, like moving from some small town in the north of France to a region near Nice, on the Riviera, or it can be the result of a merger of several competing home-care services on the same area. The right-hand square has twice as many points to visit. As suggested by the figure, the mean distance between a point and the nearest visit is smaller in the right-hand square. If we double the density of clients, the distance between a client and the nearest stop in the tour is reduced by 30%. The area of the disc centered on a client and including the nearest visit point with a given probability p is divided by two when we double the number of clients. A precise mathematical statement of this result is proposed in Appendix A. It is therefore natural to expect that the ratio of miles traveled to the number of hours of service at the client's home will be smaller in a more densely populated area.

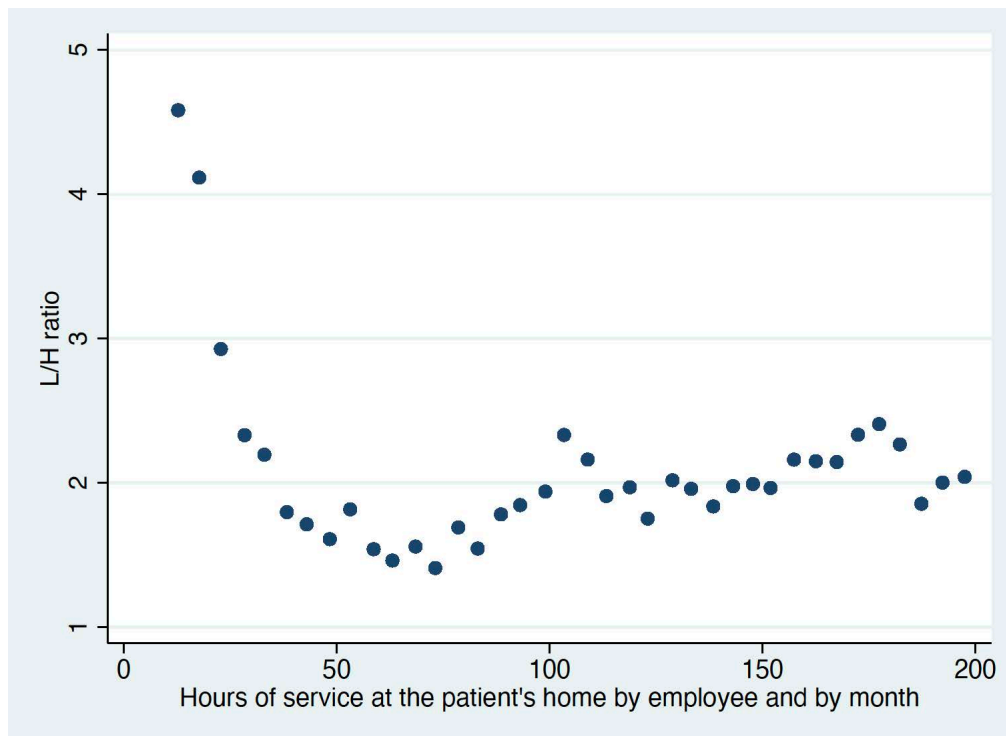
2.2 Kilometers Traveled per Hour of Service at the Patient's Home

Some descriptive statistics will provide a preliminary test for the declining miles-per-hour-of-service property.

2.2.1 Data

The dataset has been constructed thanks to *Groupe Avec*'s personnel management software and records. *Groupe Avec* operates in more than 20 counties⁸. The firm's activity is very much scattered over the French territory. Each county is divided into several districts.⁹ The density of the senior population varies from one district to another, as well as the firm's market share. Indeed, in some locations, there are many competitors. In each district, the management software precisely records the realized hours of service of each employee and the miles traveled between visits. Miles are recorded because they are reimbursed to employees, based on a per-kilometer rate determined by bargaining with trade unions at the national level, *i.e.*, an agreement with unions called a *convention collective*. The time spent driving or traveling between clients by caregivers is also compensated at a wage rate close to the minimum wage.

Figure 2. Kilometers Traveled to Hours of Service at the Patient's Home Ratio as a Function of Hours of Service per Month



Note. To compute the statistics used for Fig. 2, we consider all caregivers and all the months during which they are observed. We partition the monthly hours in bins with a width of 5 hours. In each bin, we compute the mean L/H ratio, that is, the number of kilometers traveled per month divided by the number of hours of service at the patient's home in the same month.

We obtained these data on a monthly basis. This allowed us to build an unbalanced panel. This panel covers 16 counties, divided into 53 branches, themselves including 98 districts, with a total of 3688 caregivers. The first observed month is November 2015 and the last is June 2019.

⁸Counties are the French *départements* here. There are 100 such *départements* in France.

⁹In fact, each county is divided into several branches (*i.e.*, called *agences*), each branch itself includes one or several districts, *i.e.*, local branches called *secteurs*.

Each caregiver is observed during 30 months on average, and the total number of observations is 57 136. We can therefore easily compute the ratio of miles traveled to hours of service, hereafter the L/H ratio, at the level of each individual employee.

Figure 2 plots the ratio of kilometers over hours of service, as a function of an employee’s monthly hours of service at the patient’s home — *i.e.*, not including travel time. Fig. 2 shows the average ratio computed in bins with a 5-hour width. On Fig. 2, the L/H ratio declines from 5 km to about 1.5 km.¹⁰ A ratio that is decreasing with the hours of service indicates the possible presence of economies of density. If we tried to fit a smooth function on Fig. 2, we would find a U-shaped curve. Economies of density exist when the L/H ratio is a decreasing function of hours H. In general, the functional relationship between L/H and H may not be convex and U-shaped. To show the presence of density economies, it would be sufficient to find a decreasing portion of a curve, for instance, a decreasing relationship on a sufficiently wide and relevant interval of hours.

Figure 3. Ratio of Kilometers Traveled per Hour of Service at the Patient’s Home, as a Function of the Density of People Aged 80 and More, by County

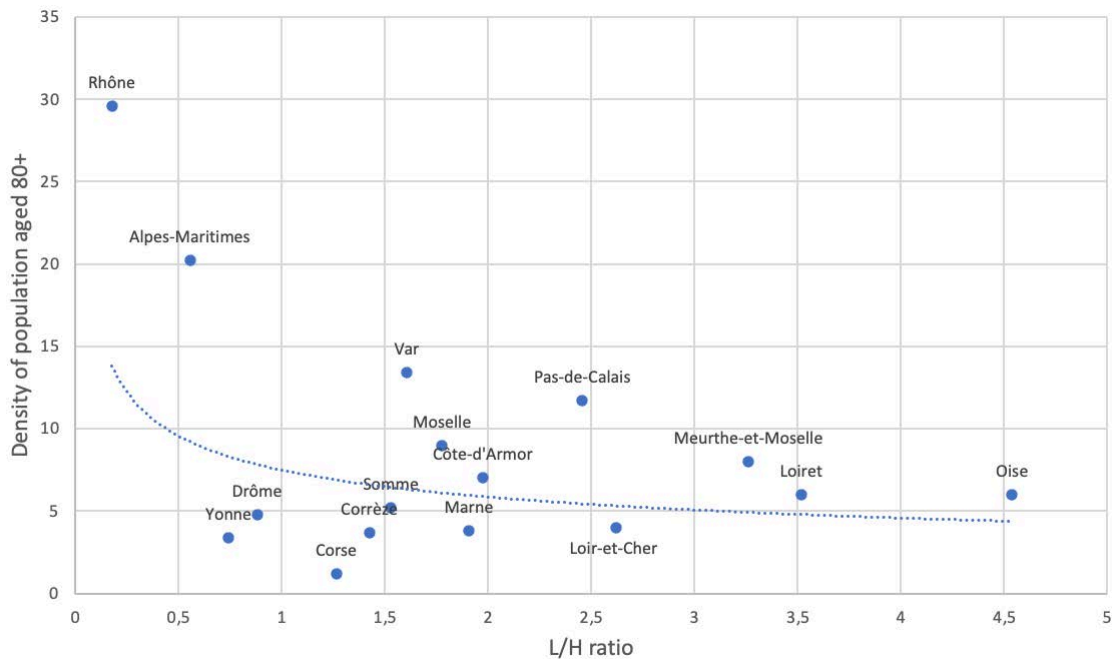


Figure 3 gives a different view: namely, the relationship between the density of persons aged 80 and more in the county and the ratio of kilometers traveled per hour of service at the patient’s home. Fig. 3 exhibits the same declining relationship, but this time, the density is a purely demographic, exogenous factor. Some explanations about French geography will clarify the picture. The Rhone county is dense and includes Lyon, which is a major city and urban cluster. The Alpes-Maritimes includes the cities of Cannes and Nice, and the French Riviera: this county

¹⁰The firm’s personnel is a mixture of full-time and part-time employees (often half or 80% time). Fig. 2 presents averages but, actually, this ratio varies much more. Some employees provide a small number of hours per month. There exists a number of these “small contracts”, used to fill gaps and to ensure a steady service. The L/H ratio can reach 20 km per hour of service for these “small contract” employees.

is not only dense, but also has a high density of retired people. These dense counties have small L/H ratios.

The economies of density are visible, but in a very imperfect way. For instance, Meurthe-et-Moselle, surrounding the city of Nancy, and Moselle, surrounding the city of Metz, are neighboring counties in the east of France. The average densities of the senior populations are about the same in the two counties, but the difference comes from the firm's market share, which is much higher in Moselle (around 33%) than in Meurthe-et-Moselle (around 1%). On Fig. 3, the dotted curve is a suggested hyperbolic function that roughly fits the data.

3 Econometric Analysis of Transportation Costs in the Home Care Services Industry: Log-Linear Models

3.1 Unbalanced Panel Structure. Notations

Our dataset is an unbalanced panel. The employees (*i.e.*, caregivers) are indexed i with $i \in I = \{1, \dots, n\}$. The time periods (months), are indexed by t with $t \in T = \{1, \dots, \tau\}$. Each i is observed during a subset of periods of T denoted T_i . Let τ_i denote the number of periods in T_i .

We also consider the districts. Districts are indexed by s in a set S . Districts form a partition of the set I of employees, since employees do not move from one district to another in the data set. Let A_s denote the subset of employees i in district s . Some districts are not observed in all periods t (because some branches have been recently acquired by the firm). District s is observed during a subset of periods of T denoted T_s . Let τ_s denote the number of periods in T_s . The set of employees i belonging to district s at time t , denoted A_{st} can also vary with t , for easily understandable reasons.¹¹

Fundamentally, we observe a set of pairs (i, t) indexing employees i at time t . We can view each district s as a set of pairs denoted $B_s = \{(i, t) \mid i \in A_{st}, t \in T\}$. The number of pairs (i, t) in district s is denoted n_s . Let also n_{st} denote the number of employees of district s at time t .¹² Conventionally, we set $n_{st} = 0$ if s is not observed at time t . We will also need the average number of employees in district s , that is, $\bar{n}_s = \tau_s^{-1} \sum_t n_{st}$. Let finally N be the total number of observations (i, t) .

Now, for each (i, t) in some district, we observe some variables. Let H_{it} denote employee i 's hours of service in the patients' homes during month t . Let L_{it} denote the kilometers traveled by i during her tours in month t . The corresponding lower-case letters denote the logarithms of these variables. We define $\ell_{it} = \ln(L_{it})$; $h_{it} = \ln(H_{it})$.

¹¹The subset A_s includes all the employees i observed at least once in district s . We have $A_s = \bigcup_{t \in T_s} A_{st}$.

¹²Formally, we have $n_s = \sum_{i \in A_s} \tau_i$. We also have, $n_s = \sum_{t \in T_s} n_{st}$.

3.2 A Basic Model

We start the study with a very simple model. The assumption is that the L/H ratio is a declining function of H . We assume,

$$\frac{L}{H} = \frac{A}{H^\gamma}, \quad (1)$$

where A varies with district and employee characteristics, and where γ measures the importance of density economies. We drop the indices (i, t) here because our hypothesis could be tested at various aggregation levels (see below). When $\gamma = 0$, there are no economies of density and kilometers are proportional to hours of service (that is, $L/H = A$). When $\gamma = 1$, economies of density are very strong, because the traveled kilometers do not depend on hours (that is, $L = A$): kilometers are like a fixed cost, a distance that must be traveled each day or each month, come rain or shine. Economies of density exist as soon as $\gamma > 0$. Taking logarithms, with $\alpha = \ln(A)$, we obtain $\ell = \alpha + (1 - \gamma)h$.

This simple model is also suggested by the Beardwood-Halton-Hammersley Theorem (see Beardwood *et al.* (1959)). According to this mathematical result, under some assumptions, if the caregivers' tours are optimized to minimize their length, then we should find $\gamma \simeq 1/2$, or $L \simeq A\sqrt{H}$ (see Appendix A for technical details, and our discussion below). It so happens that we obtain credible estimates of γ around 0.5 with the basic model.

3.3 Log-Linear Models

Using the notations introduced above, to analyze economies of density in the home care services, we define the intertemporal average of employee i 's log-hours as follows,

$$\bar{h}_i = \sum_{t \in T_i} \frac{h_{it}}{\tau_i}, \quad (2)$$

and with a similar definition for the average log-distance $\bar{\ell}_i$. We also define the district average of log-hours at date t as follows,

$$\bar{h}_{st} = \sum_{i \in A_{st}} \frac{h_{it}}{n_{st}}, \quad (3)$$

with a similar definition for the average $\bar{\ell}_{st}$. Then, we specify the following econometric model, that we shall call the β_1 - β_2 -model,

$$\ell_{it} = \alpha + \beta_1 h_{it} + \beta_2 \bar{h}_{st} + X_{it} \delta + v_{it}. \quad (4)$$

In the above equation, $(\alpha, \beta_1, \beta_2, \delta)$ are parameters; v_{it} is a random error term with a zero mean; X_{it} is a vector of control variables; we also have $s = s(i, t)$, that is, by definition, when $s = s(i, t)$, we have $i \in A_{st}$. This model expresses that economies of density may appear at different levels: $\gamma = 1 - \beta_1$ measures economies of density at the employee level, and we will see below that $\bar{\gamma} = 1 - (\beta_1 + \beta_2)$ measures economies of density at the aggregate district level.

The \bar{h}_{st} term captures specific effects of the district that are supposed to depend on the average level of activity in district s . We can also add district dummies if needed in X_{it} .

If we take the average of equation (4), summing over all $i \in A_{st}$ with weights $1/n_{st}$, we easily obtain,

$$\bar{\ell}_{st} = \alpha + (\beta_1 + \beta_2)\bar{h}_{st} + \bar{X}_{st}\delta + \bar{v}_{st}, \quad (5)$$

with obvious notations for the mean of errors in district s at time t , denoted \bar{v}_{st} and the corresponding mean of controls \bar{X}_{st} . Equation (5) shows that $\beta_1 + \beta_2$ measures the economies of density at the level of district aggregates. It is clearly possible to exploit the variability of \bar{h}_{st} to estimate the model $\bar{\ell}_{st} = \alpha + \beta_3\bar{h}_{st} + \bar{X}_{st}\delta + \bar{v}_{st}$. It will be possible to check that we have $\hat{\beta}_3 \simeq \hat{\beta}_1 + \hat{\beta}_2$. Equation (5) can typically be estimated by a fixed effects, panel data method, defining districts as groups. In fact this model will yield the most credible results, revealing a positive and significant γ at the district level, that we interpret as a team synergy. Finally, if we sum over all observations (i, t) in B_s with an equal weight, we define,

$$\bar{h}_s = \frac{1}{n_s} \sum_{(i,t) \in B_s} h_{it}, \quad (6)$$

with a similar definition for the district average $\bar{\ell}_s$, and we have $\bar{\ell}_s = \alpha + \beta_3\bar{h}_s + \bar{X}_s\delta + \bar{v}_s$, using obvious notations.

3.4 Log-linear Models: Estimation Results

We assume that random errors have the following structure: $v_{it} = u_i + \epsilon_{it}$ where ϵ_{it} is a random variable independent of other variables and u_i is an individual effect, *i.e.*, a fixed effect. Then, β_1 , β_2 and β_3 are identified and can be estimated by means of a *within* estimator. We will see that the endogeneity biases are limited with our data. Table 1 gives descriptive statistics for the main variables used in the present article.

3.4.1 Preliminary Results: OLS on Pooled Data and Fixed-Effect Estimates

To start the analysis, we estimated the simplest model, that is, $\ell_{it} = \alpha + \beta h_{it} + u_{it}$, called *Model A*, by means of OLS on pooled data, with the possible addition of controls. The parameter of interest is $\gamma = 1 - \beta$, where β is the coefficient of log-hours. Results for the simplest model are given in Appendix B. In this appendix, column (1) in Table 8 shows that, in essence, $\hat{\beta} \simeq 1$. This means that returns to scale are constant at the individual employee level. Yet, it may be that the endogeneity of hours causes an upward bias. Columns (2) and (3) of Table 8 show that adding controls does not change the result, even if, as in column (3), we add the complete set of *district dummies*. Economies of density may be hidden by caregiver population heterogeneity. Adding some flexibility changes the picture a little bit (see Appendix B), but we do not find economies of density at the level of individual employees. This preliminary test naturally leads to the estimation of Model β_1 - β_2 , as specified by Equation (4).

Table 1: SUMMARY STATISTICS

	Mean	Standard Dev.	Median	Observations*
Kilometers Traveled (L_{it})	210.78	193.97	158.06	57,136
Hours per month (H_{it})	110.04	45.78	115	57,136
Part-Time Dummy	0.225	0.418	0	12,903
80%-Time Dummy	0.487	0.500	0	27,810
Full-Time Dummy	0.287	0.452	0	16,423
Rural District Dummy	0.271	0.445	0	15,516
Urban District Dummy	0.360	0.480	0	20,580
Competition	15.84	15.08	15	57,136
Density of 80 years old ⁺	7.64	3.21	9.00	57,136
Speed	34.93	13.03	35.12	57,088

Note: The density of 80 years old and more is measured in inhabitants per square kilometer (in the county), using French Census data, INSEE 2015. *Competition* is the number of competitors of the firm established in the main town of district s or in its suburbs (in 2019). *Speed* is defined for each (i, t) as the ratio L_{it}/Θ_{it} , in kilometers per hour, where Θ_{it} is the time spent driving between visits. *Density* is the logarithm of the density of population aged 80 and more in the county (French Census, INSEE 2015). Rural and urban districts are defined by the French National Statistics Institute, *i.e.*, INSEE, based on Census data. (*) The Observations column is the number of available observations in the data set for the continuous or count variables; it is the number of observations with a value equal to 1 for dummies.

Table 2 gives the estimation results of Model β_1 - β_2 , obtained by OLS and the fixed-effects, *within* estimator. Estimated β_1 coefficients are significant, precise and stable across different specifications, around 1.45. In Table 2, column (1) has no controls and no district dummies; column (2) shows that adding a full set of district dummies does not change the β_1 coefficient in a significant way; column (3) has controls but we still find $\beta_1 \simeq 1.40$. In all columns, the district effect, that is, the coefficient β_2 of the average hours per month in the district, *i.e.*, \bar{h}_{st} , is significant and negative; column (4) shows that the district and individual effects, resp. β_2 and β_1 are still significant with the *within* estimator, but the magnitudes are somewhat smaller.¹³

The most remarkable property of Table 2 is that column (2) gives $\hat{\beta}_1^{ols} + \hat{\beta}_2^{ols} \simeq 1.451 - 0.867 = 0.584$, column (4) yields $\hat{\beta}_1^{fe} + \hat{\beta}_2^{fe} = 1.109 - 0.522 = 0.587$, in principle, the most reliable estimate. This is not by chance, and we will see below that $\hat{\beta}_3 \simeq \hat{\beta}_2 + \hat{\beta}_1$ when we directly estimate the model at the level of district aggregates. We know that $1 - \beta_1 - \beta_2$ is supposed to measure economies of density at the aggregate district level. If this is true, we have $\gamma = 1 - \beta_1 - \beta_2 \simeq 0.42$.

¹³In column (4), the constant is the average of individual fixed effects.

Table 2: TRAVELED KILOMETERS. ESTIMATION OF MODEL β_1 - β_2

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	FE
h_{it}	1.447*** (0.020)	1.451*** (0.020)	1.404*** (0.021)	1.109*** (0.018)
\bar{h}_{st}	-0.795*** (0.032)	-0.867*** (0.044)	-0.767*** (0.042)	-0.522*** (0.028)
Constant	1.828*** (0.168)	1.261*** (0.252)	2.457** (0.696)	2.128*** (0.137)
Speed	. (.)	. (.)	0.031*** (0.006)	. (.)
Density of 80+ years old	. (.)	. (.)	-0.542*** (0.161)	. (.)
Competition	. (.)	. (.)	0.080** (0.026)	. (.)
Time (month)	. (.)	. (.)	0.002* (0.001)	. (.)
District dummies	NO	YES	YES	NO
Observations	56,878	56,878	56,830	56,878
R^2	0.268	0.488	0.553	.
R^2 within				0.232
R^2 between				0.258
R^2 overall				0.267
F	2789.78***	383.67***	453.94***	1842.08***

Note: Significance is indicated by: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates of columns (1)-(3) have been obtained by OLS on pooled data. The dependent variable is ℓ_{it} . Column (4) gives the within-employee, fixed-effects estimator. Standard deviations are robust. The *Time* variable is the month, ranging from $t = 1$ to $t = 44$, between November 2015 and June 2019. Variable *Competition* is the number of competitors of the firm in district s . *Speed* is kilometers traveled in the month divided by time spent driving. *Density* is the log of the density of population aged 80 and more in the county.

We conclude that our preliminary results show the presence of significant economies of density, but these economies are generated at the district level, within a specific and relatively stable team of workers who must coordinate efforts to achieve an efficient way of servicing a local group of patients. It follows that the most likely explanation is that we observe a team-synergy effect: when employees work more together in a given district, they are more efficient in the use of transportation: they drive less per hour of service. This effect may be generated by the added

flexibility of a team in which employees are working longer hours on average per day or per month. The effect may also be generated in a larger team: we will study this effect later.¹⁴

3.4.2 Evaluation of potential transportation cost economies

These economies of density matter for public policy. Using the above results we can estimate the transportation-cost economies that could be achieved by a shift to a franchise bidding regime with local monopolies. More precisely, an increase in the density of clients leads to a substantial drop in these costs. It follows that to exploit the economies of density, public authorities could in principle divide the territory into regions, and auction the right to run a single, local monopoly Home-Care service in each of these regions — *i.e.*, create a franchise bidding regime. With the help of the model, we can compute a simple simulation: If each local branch of the company under study was granted the exclusive right to operate in its county, how much could the network save in terms of caregivers' transportation costs per hour of service at the patient's home?

Assume that each local branch of the company becomes the unique provider on its territory, or district, and inherits all of the patients: the market share of each branch becomes 100%. As a consequence, the number of hours of service at the patient's home will increase and the number of kilometers per hour of service will drop. Table 3 shows the results of this simulation exercise. The middle columns of Table 3 give the percentage decrease of transportation costs per hour of service for various initial market shares, from 1/10th to 1/2, and for two values of γ , namely 0.5, that we believe is likely to be the right value, and 0.25, a conservative estimate. We see that the decrease in costs is sizable, even for the small value of γ . For instance, with an initial market share of 1/4th, the expected gains are between 30% and 50%.¹⁵

At this point, it is possible to estimate the transportation costs saved in the entire country, if everywhere we changed for a system of exclusive territory concessions. The right-hand column of Table 3 gives the estimated gains for France in an entire year, assuming that the initial market share is everywhere equal to that indicated in the left-hand column. To obtain this result we multiply a few figures. There are 768,837 seniors receiving public subsidies for Home Care services.¹⁶ A subsidized patient of Home Care services consumes 269.4 hours of service by year on average.¹⁷ On average, these hours generate a traveled distance very close to 2 km per hour

¹⁴But the results may not necessarily be caused by the fact that employees work harder when they work together more, in a given district. We remain agnostic about moral hazard in teams here: in fact, the team effect could be purely technological.

¹⁵By assumption, the local branch of district s jumps from P_s , a value smaller than 1, to 1, that is 100%. The hours then increase by $x_s\%$, so that the market share becomes 100%, that is, $(1 + x_s)P_s = 1$. Next, the kilometers per hour of service in district s , denoted k_s , is given by the estimated model, predicting that $k_s = A_s/H_s^\gamma$, where A_s is a constant characterizing district s and H_s is total hours in the district. The upward jump in hours causes a change in the ratio of kilometers to hours that drops from k_s to k'_s . The relative variation of the ratio can then be written $(k'_s/k_s) - 1 = (1/(1 + x_s))^\gamma - 1 = P_s^\gamma - 1$. The square root of the market share minus one is thus the percentage decrease in transportation costs per hour of service if $\gamma = 1/2$.

¹⁶More precisely, the APA, *allocation personnalisée d'autonomie*, in 2017, according to official figures of the Ministry of Public Health, DREES.

¹⁷According to data provided by *Groupe Avec*, 2015-2019.

of service. Finally, each of these kilometers costs 0.6746 euros on average, taking the form of travel cost reimbursements and paid working time.¹⁸ Multiplying these figures we obtain an estimate of the total transportation cost of Home Care services in France during a given year, namely, in euros,

$$279,289,064 = 768,837 \times (269.4) \times 2 \times (0.6746).$$

A 50% reduction of this total amounts to 140 million euros par year.

Table 3: ESTIMATION OF THE REDUCTION IN TRANSPORTATION COSTS, PER YEAR, IN FRANCE, AFTER A MOVE TO EXCLUSIVE TERRITORIES

Initial market share	% Drop in transportation costs if $\gamma = 0.5$	% Drop in transportation costs if $\gamma = 0.25$	Gains if $\gamma = 0.5$ (million euros)
1/10	-68%	-43%	190,48
1/8	-64%	-40%	179,28
1/6	-59%	-36%	165,27
1/5	-55%	-33%	154,07
1/4	-50%	-29%	140,06
1/3	-42%	-24%	117,65
1/2	-29%	-16%	81,23

Note: If the initial market share of the provider is $1/n$ and if $\gamma = 1/2$, then, the percentage drop in transportation costs per hour of service is $\sqrt{1/n} - 1$. More generally, this percentage drop can be expressed as $(1/n)^\gamma - 1$, and n can be interpreted as the number of competitors initially sharing the market.

To increase the density of clients on a given area, an obvious possibility is thus to end the division of the market in this area. Some available data, and anecdotal evidence suggest that the current French Home-Care Services market generates excess entry. Market concentration is very low in many French counties (*i.e.*, *départements*).¹⁹ This is indeed causing market instability: in the recent years, the sector has witnessed an unprecedented wave of bankruptcies, and the recent pandemic only made things worse. This happened a few years after governmental decisions partially liberalized the sector and eased the entry of new for-profit companies in the market. These factual elements show that economies of density, and more generally economies of scale, are important to understand the Home Care industry, in spite of the fact that this activity does not require large amounts of capital.

¹⁸Again according to *Avec* figures in the period 2015-2019.

¹⁹There exists a website listing all the licensed Home Care services in France, l'*Annuaire sanitaire et social* (cf. sanitaire-social.com). This website shows 237 services in the county surrounding Bordeaux (*i.e.*, the *Gironde*), with 108 structures for Bordeaux itself (in 2020). Even if we take into account that some of these businesses are not independent, because they are local branches of larger networks, and the possibility that some firms do not serve the entire market, the degree of market fragmentation seems very high.

3.4.3 Fixed Effects Estimation Using District-Level Aggregates

We now aggregate hours and kilometers at the district level. Even if the endogeneity of hours does not seem to be a big concern, we estimated the log-linear model given by Equation (5) by means of a fixed-effects estimator at the district level. We also directly controlled for district effects with the help of district dummies. In addition, it is natural to estimate weighted regressions in which observations (s, t) are weighted by the number of employees of district s at time t , denoted n_{st} . Another possibility is to use the average number of observations in district s , denoted \bar{n}_s , as regression weights — we tried both ways.

The regressions presented in Table 4 use the variability of average log-hours per employee \bar{h}_{st} , and kilometers per employee $\bar{\ell}_{st}$, across the time dimension, in each district s . This variability is large enough to yield precise estimates of β_3 . Economies of density appear at the district level; they are again stable and significant, with $1 - \beta = \gamma \simeq 0.5$. The estimates of Table 4, with $\beta_3 < 1$, also reveal the operation of *team synergies* at the district level. Districts with more intense activity have significantly smaller L/H ratios.

In the first three columns of Table 4, the idea is to use weighted OLS and to control for district effects, either by means of the set of district dummies or by means of the continuous variable \bar{h}_s . On Table 4, in the regression of column (1), district effects are controlled by \bar{h}_s and we find $\hat{\beta}_3 = 0.502$; in column (2), district effects are controlled by means of district dummies, and the estimation of β is very similar, we have $\hat{\beta}_3 = 0.498$; column (3) is just the same as column (1) but districts are declared as clusters: the β_3 estimate remains very significant. The last three columns treat the fixed effect problem by differencing. In principle, these treatments take care of the bulk of a possible endogeneity-of-hours problem. Differences with the time average purge the results from the possible correlation of district fixed effects with district log-hours, and the value of $\hat{\beta}_3$ must be the same as that of column (2). Column (4) is a simple OLS regression of $(\bar{\ell}_{st} - \bar{\ell}_s)$ on $(\bar{h}_{st} - \bar{h}_s)$, weighted by n_{st} . Column (5) uses a weighted fixed-effect, within estimator, but the Econometrics software uses \bar{n}_s as a weight: this explains the difference in the estimated value of β , which remains close to 0.5 anyway. Finally, column (6) of Table 4 displays the weighted first-differences estimator of β , somewhat lower than 0.5 — thus pointing towards stronger economies of density. So, it could be that the real value of β lies somewhere between 0.4 and 0.5.

On Table 4, the identification of regressions (2), (4), (5) and (6) is correct if (with obvious notations) $\bar{v}_{st} = \bar{u}_s + \bar{\epsilon}_{st}$, that is, if the district effect \bar{u}_s can be eliminated by differencing. These results suggest that $\hat{\beta} = 0.498 \simeq 0.5$. If we compute the regressions of Table 4 without weighting, we obtain the same qualitative results — that is, economies of density are significant — but the estimated β_3 becomes close to 0.6.²⁰

²⁰This seems to indicate that smaller districts have smaller economies of density. The unweighted results are available upon request.

Table 4: KILOMETERS TRAVELED. DISTRICT LEVEL. WEIGHTED REGRESSIONS

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS Clustered	Fixed Effects [†]	Fixed Effects	First Differences
\bar{h}_s	0.774*** (0.110)	.	0.774 (0.635)	.	.	.
\bar{h}_{st}	.	0.498*** (0.046)	.	.	0.515*** (0.051)	.
$\bar{h}_{st} - \bar{h}_s$	0.502*** (0.097)	.	0.502*** (0.051)	0.498*** (0.046)	.	.
$\Delta \bar{h}_{st}$	0.429*** (0.057)
Constant	1.265** (0.513)	1.641*** (0.217)	1.265 (2.963)	0.001 (0.004)	2.465*** (0.237)	-0.000 (0.005)
District Dummies	NO	YES	NO	NO	NO	NO
Observations	3,117	3,117	3,117	3,117	3,117	3,065
Groups	.	.	98	.	98	.
R^2	0.027	0.891	0.027	0.074	.	0.072
F	37.50***	327.91***	47.48***	117.98***	100.86***	57.16***
R^2 within	0.087	.
R^2 between	0.052	.
R^2 overall	0.024	.
σ_u	0.719	.
σ_ϵ	0.214	.
ρ	0.918	.

Note: Significance is indicated by: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates have been obtained with data aggregated at the (district, month) level. Variable t denotes the month, numbered from $t = 1$ to $t = 44$, between November 2015 and June 2019. All regressions are weighted. For columns (1) to (4) and (6) observations are weighted by the number of employees of district s in month t , that is n_{st} . The regression of column (5) is weighted by the mean number of employees in district s , *i.e.*, \bar{n}_s . The dependent variable is $\bar{\ell}_{st}$ in columns(1)-(3) and $\Delta \bar{\ell}_{st} = \bar{\ell}_{st} - \bar{\ell}_{s,t-1}$ for column (6). Robust standard deviations are provided for the coefficients of column (1). For the regression of column (2), the district dummies are the only controls; standard deviations are again robust (White correction). In column (3), districts s are defined as clusters. [†]The dependent variable of column (4) is $(\bar{\ell}_{st} - \bar{\ell}_s)$, and the explanatory variable is $(\bar{h}_{st} - \bar{h}_s)$. Column (5) is estimated by the fixed-effects within estimator where districts s are the groups; the constant here can be interpreted as the average of fixed effects ; the values of σ_u and σ_ϵ are the standard deviations of the district fixed effects and of the error term ϵ , respectively; the ρ statistic is the percentage of variance due to district effects. Column (6) is estimated by OLS with robust standard deviations; $\Delta \bar{h}_{st} = \bar{h}_{st} - \bar{h}_{s,t-1}$ is the explanatory variable; the first observations in each district are lost. The F statistic tests the overall significance of the model.

In Appendix C (Table 9), we report the results obtained when travel time is used instead of kilometers traveled. The results are very similar. In particular, when the β_1 - β_2 Model is estimated with travel time as the dependent variable, we still find, with a fixed-effects, within employee estimator, $\hat{\beta}_1 + \hat{\beta}_2 = 1.047 - 0.485 = 0.562$ (see column (4) of Table 9). This shows the robustness of our result.

3.5 Relationship with the Beardwood-Halton-Hammersley Theorem

The combinatorial optimization literature provides us with results that shed light on the analysis presented here. Coming back to Fig. 1 above, assume that there are ν points (x_1, \dots, x_ν) drawn at random, independently and identically distributed on the unit square $[0, 1]^2$. Let $L(x_1, \dots, x_\nu)$ denote the minimal length in the set of all tours joining the ν points, where the distance between the points is the usual Euclidean distance. The BHH Theorem (*i.e.*, Beardwood, Halton and Hammersley (1959)) states that if the points x_i are i.i.d and distributed on a territory, denoted X , with a nonzero area, and included in the plane, then, there exists a positive constant ρ such that, with probability 1,

$$\frac{L(x_1, \dots, x_\nu)}{\sqrt{\nu}} \rightarrow \rho \quad \text{as the number of visits } \nu \rightarrow \infty.$$

In our case, the constant ρ depends on the district X , and on the distribution of x_i . But, the crucial point is the square-root property: *almost surely*, the minimal length of a tour²¹ with ν visits is asymptotically close to $\rho\sqrt{\nu}$. The average care time spent per visit, excluding the time spent driving, denoted V , is a random variable. Assume that $V = V_0 e^{-2\epsilon}$ where ϵ is a random shock with a zero mean and $V_0 = v(s)$ is a function of district s . We have $H = \nu V_0 e^{-2\epsilon}$. If the tours are efficiently organized, as a consequence of the BHH Theorem, for ν large enough, we have, $L \simeq \rho e^\epsilon \sqrt{H/V_0}$. Taking logarithms, we find the model,

$$\ell = \alpha + \frac{h}{2} + \epsilon,$$

where $\alpha = \ln(\rho) - (1/2) \ln(V_0)$ also depends on district characteristics.

Empirically, we can take H to be the individual hours during a day tour, but the number of visits of this tour is probably not large enough. So we can aggregate the hours of service of an employee during month t or even during all periods. If the visit schedules are efficient, this is like pasting efficient tours together, yielding a long efficient tour. We can also aggregate the tour lengths and the hours of different employees in a given district s . Again, pasting together the efficient tours of individuals sharing a given district may yield a long efficient tour of the district team. It is then justified to run the regressions of Table 4 and to expect $\hat{\beta} \simeq 1/2$. As a robustness check, we ran regressions with the total sum of hours and kilometers in districts, that is $\ln(\sum_{i \in A_{st}} H_{it})$ and $\ln(\sum_{i \in A_{st}} L_{it})$. We also regressed the individual aggregate $\ln(\sum_{t \in T_i} H_{it})$

²¹For more detailed explanations, see Appendix A.

on $\ln\left(\sum_{t \in T_i} L_{it}\right)$. If the coefficient of the log-sum of hours is close to one half, we do not detect inefficiencies in the scheduling of tours. Table 5 reports the results of this type of regression.

Table 5: KILOMETERS TRAVELED: LOG-LINEAR REGRESSIONS WITH TOTAL HOURS AND KM

	(1)	(2)	(3)	(4)
	OLS	OLS	FE	FE
Dependent Variable:	$\ln(\sum_{t \in T_i} L_{it})$	$\ln(\sum_{i \in A_{st}} L_{it})$	$\ln(\sum_{i \in A_{st}} L_{it})$	$\ln(\sum_{i \in A_{st}} L_{it})$
$\ln(\sum_{t \in T_i} H_{it})$	0.987*** (0.016)	.	.	.
$\ln(\sum_{i \in A_{st}} H_{it})$.	0.438*** (0.035)	0.699*** (0.070)	0.561*** (0.090)
Constant	-0.504* (0.195)	4.130*** (0.274)	2.709*** (0.524)	3.897*** (0.699)
District Dummies	YES	YES	NO	NO
Observations	3,687	3,117	3,117	3,117
Weights	1	n_{st}	1	\bar{n}_s
R^2	0.831	0.950	.	.
F	141.39***	820.18***	98.18***	38.77***

Note: Significance is indicated by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 3,687 is the number of employees in the employee base and 3,117 is the number of observed (district, month), i.e., (s, t) pairs. The regression of column (1) is not weighted. The regression of column (2) is weighted with n_{st} . Columns (3) and (4) report fixed-effects, *within* estimates. Column (4) reports the results of a weighted FE regression with weights \bar{n}_s .

The results of Table 5 are striking. In the regression of column (1), we use the total hours and kilometers aggregated at the employee level, and, in essence, we find constant returns to scale. In contrast, if we aggregate at the district level, and control for district effects, we find economies of density. Column (2) shows that $\beta \simeq 0.44$, more or less as predicted by the BHH Theorem. The result is confirmed by the weighted fixed-effect estimate of column (4), yielding $\hat{\beta}_{fe} = 0.56$. The regression of column (3) is not weighted. The BHH assumptions are not rejected by the data.²² The BHH Theorem is by definition true if the assumptions on points x_i and X are true. Tours may not be modeled as loops in a subset of the plane; or finally, tours are not efficiently organized. Our regressions are a joint test of these assumptions. Appendix A recalls that the BHH Theorem is valid in three dimensions (points are then randomly drawn in a cube). In the latter case, we expect $\beta = 2/3$, not far from the unweighted estimate of column (3). The regressions of Table 5 do not seem to reject the efficiency or near efficiency of tours, within a

²²The ordinary F -test of $H_0: \beta = .5$ in regression (2) of Table 4 yields a p -value of 0.96; the same test of β in regression (5), Table 4, yields a p -value of 0.76; finally, the same test of β in regression (4), Table 5, yields a p -value of 0.51.

district. Besides this, we confirm that team synergies seem to operate in districts, leading to smaller L/H ratios in more active districts.

4 U-shaped Curves and Unobserved Heterogeneity

The results that we obtained above with the log-linear models may be viewed as a preliminary step. We suspect that there is a substantial amount of unobserved heterogeneity in the data. If the true model is not linear in logs at the individual or at the district level, and if, in addition, the β parameter varies with observable and unobservable characteristics of employees and districts, as suggested by Table 8, it is fully justified to estimate more flexible forms of our model. We will add flexibility by assuming that the model is quadratic in log-hours, and by assuming the existence of a finite set of latent types. In other words, we will now estimate variants of a finite-mixture version of a quadratic-in-logs model. We will show that the latter models typically generate U-shaped average transportation cost curves and uncover the unobserved heterogeneity of workers.

Let K denote the number of latent employee types in the population. Types, or groups, are indexed by $k = 1, \dots, K$. The fact that quadratic terms are significant justifies the use of a quadratic model. Next, a difficulty with finite mixture models is the choice of the number of groups K . We use a combination of criteria, including usual likelihood-based information criteria. AIC and BIC, that is, Akaike and Bayesian Information Criteria have a tendency to select too many latent types; in other words, they lead to overfitting. With our data, the BIC is minimized when $K = 8$. We therefore combine the usual criteria with entropy-based criteria, to assess the quality of classification of employees by our model. We will see that two types provide a good representation of the underlying heterogeneity of workers. When $K > 2$, the quality of classification declines — *i.e.*, entropy rises. When $K > 2$, we find that the marginal gain of an additional type in terms of the log-likelihood is not worth the cost of the added complexity, and all the more since individuals are not well identified as members of a given group.

4.1 Modeling Unobserved Heterogeneity with Latent Types

Let K be the number of unobserved types, indexed by k . We assume that employee i , at date t , belongs to group k with prior probability p_k . Each group can be described by a quadratic model. The simplest model, called Model A, is specified as follows. If i belongs to group k at time t , we assume,

$$\ell_{it} = \alpha_k + \delta_k h_{it} + \zeta_k h_{it}^2 + \epsilon_{ikt}. \quad (7)$$

It happens that this model is too simple, insofar as it doesn't explicitly allow for local, *i.e.*, district-based, team effects. Yet, Model A yields significant values of ζ_k . The analysis of Model A is presented in Appendix E: in spite of its simplicity, this model generates U-shaped curves

giving L_{it}/H_{it} as a function of H_{it} .

We will mainly focus on the following richer specification, called Model B. If i belongs to group k at time t , we assume,

$$\ell_{it} = \alpha_k + \beta_k \bar{h}_{st} + \nu_k \bar{h}_{st}^2 + \delta_k h_{it} + \rho_k (1/n_{st}) + \epsilon_{ikt}, \quad (8)$$

where $\epsilon_{ikt} \sim \mathcal{N}(0, \sigma_k^2)$, and $(\alpha_k, \beta_k, \nu_k, \delta_k, \rho_k)$ are parameters that depend on group k . We assume that the ϵ_{ikt} are i.i.d random variables, independent of hours conditional on group k . The groups $k = 1, \dots, K$ are therefore taking care of the possible endogeneity problem that would arise in a model assuming that parameters are independent of k . The normality assumption improves tractability but the reader must keep in mind that any distribution of ℓ_{it} can be approximated by a mixture of normal distributions, to any desired degree of precision. Model B can capture district-based team effects through \bar{h}_{st} , \bar{h}_{st}^2 and $1/n_{st}$. The average hours in district s capture an “intensive” form of the economies of density, while $1/n_{st}$ captures an “extensive” form (if $\rho_k > 0$).

4.1.1 Interpretation of the Models

To understand the meaning of Model B more completely, define first the log-ratio of kilometers to hours in district s at time t as $\bar{\kappa}_{st} = \bar{\ell}_{st} - \bar{h}_{st}$. We then aggregate equation (8) over all observations $i \in A_{st}$, expressing the ratio $\bar{\kappa}_{st}$, and dropping index k to lighten notation, we easily obtain,

$$\bar{\kappa}_{st} = \alpha + (\beta + \delta - 1)\bar{h}_{st} + \nu \bar{h}_{st}^2 + \rho(1/n_{st}) + \bar{\epsilon}_{st}, \quad (9)$$

We expect that $(\beta + \delta - 1)$ and ν will provide a convex curve for $\exp(\bar{\kappa})$, typically with $\beta + \delta - 1 < 0$ and $\nu > 0$, but to show the presence of density economies, it is sufficient to find a curve that is decreasing on a relevant interval of hours — say, between 0 and 200. Note that $\exp(\bar{\kappa}_{st})$ is just the geometric mean of the ratio H_{it}/L_{it} for $i \in A_{st}$, and $\bar{\kappa}_{st} = n_{st}^{-1} \ln [\prod_{i \in A_{st}} (L_{it}/H_{it})]$. The intuition for the use of n_{st}^{-1} as a right-hand side variable is straightforward.²³ Economies of density in the “extensive” sense exist if $\rho > 0$ because a larger team implies a smaller average L/H ratio.

To draw the average L/H curves generated by Model B, we compute the expectation of $\exp(\bar{\kappa}_{st})$, as given by equation (9). Dropping index st to lighten notation, we have,

$$\mathbb{E} \left(\exp(\bar{\kappa}) \mid \bar{h}, k, n \right) = \exp \left\{ \alpha_k + (\beta_k + \delta_k - 1)\bar{h} + \nu_k \bar{h}^2 + \left(\rho_k + \frac{\sigma_k^2}{2} \right) \frac{1}{n} \right\}, \quad (10)$$

where n stands for team size n_{st} .²⁴ We obtain the average transportation cost curve by taking

²³Equation (9) means that when n_{st} increases, given a fixed average value of hours \bar{h}_{st} , then, $\rho > 0$ implies that $\bar{\ell}_{st}$ has to adjust downward: the average distance travelled by each worker must decrease.

²⁴We find (10) because $\bar{\kappa} \sim \mathcal{N}(\mu, n^{-1}\sigma_k^2)$, conditional on \bar{h} , k and n . Since μ is a function of \bar{h} and $1/n$, it follows that $\mathbb{E}(e^{\bar{\kappa}} \mid \bar{h}, k, n) = \exp(\mu + \sigma_k^2/(2n))$.

the expectation of (10) over k , that is,

$$\mathbb{E} \left(\exp(\bar{\kappa}) \mid \bar{h}, n \right) = \sum_k p_k \mathbb{E} \left(\exp(\bar{\kappa}) \mid \bar{h}, k, n \right), \quad (11)$$

where p_k is the frequency of group k . The curves (10) and (11) can be computed with the help of the estimated values of parameters, including σ_k and p_k .

4.1.2 Estimation Method

These models can be estimated by a straightforward likelihood maximization routine, provided that good preliminary estimates are available. The preliminary estimates are obtained by means of an EM algorithm. We will present the ML, and some EM estimates of the model. In essence, the log-likelihood is maximized when a particular weighted regression is estimated with weights equal to the posterior probabilities that each observation (i, t) belongs to group k . This posterior probability, denoted p_{itk} is derived from the likelihood function by means of Bayes' Theorem. This particular EM algorithm is described in Appendix D. The same method has been justified and advocated by various authors (see, *e.g.*, Arcidiacono and Jones (2003), Bonhomme and Robin (2009)); there are also various applications in other works.²⁵

Let Λ_{itk} be the contribution to likelihood of observation (i, t) , conditional on type k and the explanatory variables. We have,

$$\Lambda_{itk} = \left(\frac{1}{\sigma_k} \right) f \left(\frac{\epsilon_{itk}}{\sigma_k} \right), \quad (12)$$

where ϵ_{itk} is given by (8) and $f(x) = (\sqrt{2\pi})^{-1} e^{-x^2/2}$. The contribution to likelihood of observation (i, t) is therefore,

$$\Lambda_{it} = \sum_{k=1}^K p_k \Lambda_{itk}.$$

The likelihood Λ is the product of the Λ_{it} contributions over all $t \in T_i$ and all i . The log-likelihood is therefore,

$$\mathcal{L} = \ln \Lambda = \sum_{i=1}^n \sum_{t \in T_i} \ln \left(\sum_{k=1}^K p_k \Lambda_{itk} \right). \quad (13)$$

To compute the posterior probabilities of an individual's type, we apply Bayes' rule. We obtain,

$$p_{itk} = \Pr(k \mid i, t, \ell_{it}, h_{it}) = \frac{p_k \Lambda_{itk}}{\sum_{j=1}^K p_j \Lambda_{itj}}. \quad (14)$$

These *individual posterior probabilities* are computed with the help of contributions to likelihood, evaluated at the estimated values of the model parameters. These probabilities tell us how likely it is that individual i belongs to group k at time t . If for each i , the individual posterior

²⁵See, for instance, Gary-Bobo *et al.* (2016), Cassagneau-Francis *et al.* (2021). On finite mixtures in general, see McLachlan and Peel (2000), Bouveyron *et al.* (2019). We use straightforward MLE here but there exists other methods, see *e.g.*, Bonhomme and Manresa (2015), Su, Shi and Phillips (2016).

probability is close to 1 for some k , individuals are very well classified. If, on the contrary, all individual posterior probabilities are near $1/K$, the system of latent groups is completely ineffective at characterizing individuals: every i is represented by a uniform lottery on a set of virtual types.

4.2 Estimation Results

The results obtained for Model A are presented in Appendix E, for the sake of completeness. We focus here on Model B. We present the results obtained with $K = 1, 2, 3$. We discuss below the reasons why we think that $K = 2$, *i.e.*, two types only, is the best choice. But we estimated the model up to $K = 9$. The results obtained with $K = 6$ are presented in Appendix G for the sake of completeness.

To estimate the prior probabilities p_k , we use a classic parametrization making sure that all probabilities remain in $[0, 1]$ and sum to one. To this end, we define the vector of parameters (r_1, \dots, r_k) such that

$$p_k = \frac{e^{r_k}}{\sum_{j=1}^K e^{r_j}}, \quad (15)$$

and we impose the constraint $r_1 = 0$. The likelihood is maximized with respect to r , without any constraints.

4.2.1 Model B

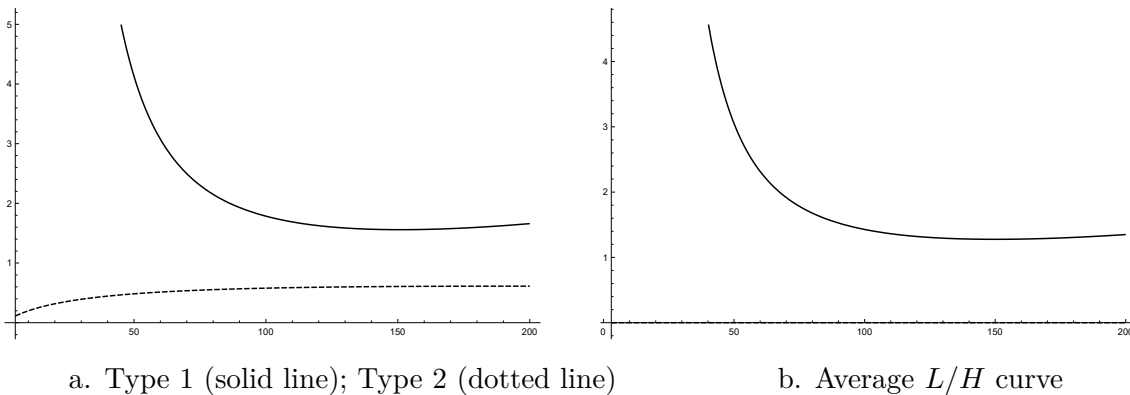
The ML estimation results of Model B are reported in Table 6. Each column of Table 6 gives the parameters of equation (8) in the (K, k) case, *i.e.*, for a number of types $K = 1, 2, 3$ and $k \leq K$. At first glance, we see that all model parameters are precisely estimated for $(K, k) = (2, 1)$, *i.e.*, the first type in a model with 2 types. Column (2, 1) shows that the estimated $\bar{\kappa}$ -ratio curve will be convex with respect to \bar{h}_{st} , since $\beta_k + \delta_k - 1 < 0$ and $\nu_k > 0$. When $K = 2$, type $k = 2$ has some nonsignificant parameters and will yield a flat curve. When $K = 3$, the results seem to indicate that type 2 is very flat and that type 1 and 3 have significant quadratic terms but with opposite signs. The prior probabilities of types are always non-negligible. Yet the probability of types is estimated with good precision only in the case of two types (columns (2, 1), (2, 2)). A huge improvement of the likelihood is achieved when we move from $K = 1$ to $K = 2$, but the gains of a third type are limited. We will see below that the quality of classification is maximal with 2 types. To understand the meaning of these results, we will draw the average L/H curves, as given by Equations (10) and (11), with the estimated values of the model's parameters in Table 6.

Using the estimated values of parameters given by Table 6, Figure 4 gives the plots of (10) for $k = 1$ (solid line) and $k = 2$ (dotted line), on the left-hand panel, and the average curve (11), on the right-hand panel. To draw these pictures, $(1/n)$ is fixed at its median value, that is

$(1/n) = 0.043$ (the median value of n_{st} is equal to 23).

Figure 4. Model B

L/H Curves Conditional on Type k and Average *L/H* Curve, for $K = 2$.



Type 2, a minority, has a low ratio, below 1 km per hour of service, and type 2’s *L/H* curve reaches a minimum at $H = 150$ hours per month (the solid line on the left panel of Fig. 4). If we take the average of the two curves we find a convex, U-shaped curve, as shown by Fig. 4b. We conclude that economies of density exist, and that they seem to be exhausted as soon as employees reach an average number of hours per month around 150, but there are employees (*i.e.*, Type 2) for which economies of density do not exist, or are negligible in practice.

We now briefly discuss the estimated model with three types (columns $(3, k)$ on Table 6). The pictures of *L/H* curves are given by Figure 5. On Fig. 5a, we see that Type 3, a minority with a probability of 16%, has a quasi-concave (in fact log-concave) curve and a very low *L/H* ratio (below .25 km per hour of service). Type 3, which is a kind of offshoot of Type 2, is not very useful. But when $K = 3$, Type 1 is still present and exhibits clear economies of density. Type 1’s curve is still U-shaped and convex and reaches a minimum at $H = 187$. It follows that, in practice, Type 1 is still driving the overall economies of density in the aggregate. Fig. 5b shows that the average *L/H* curve is U-shaped. In the case of Model B, we conclude that economies of density not only exist, but also play a major role. A possible explanation for the “anomalous” quasi-concave look of Types 2 and 3 is that these types capture some employees that happen to live very close to their patients. Further scrutiny shows that these individuals are thinly spread almost everywhere, but mainly concentrated in a dozen of districts. To select the real-life type-2 individuals, we select the subsample of observations (i, t) such that $p_{it2} > 0.85$. We then observe that the districts s in which more than 10% of the employees are identified as type 2 in this sense are some big cities, like Lyon and Nice, but also other Riviera cities like Cannes, Menton and Antibes, as well as rural places or districts centered around small towns: like Sens (in Burgundy) and, curiously, many districts in Corsica (Bastia, Borgo, Corte, Ile-Rousse and a village called Saint Florent). This means that type-2 employees are not necessarily operating in high-density urban districts. We will learn more about this below.

Table 6: TRAVELED KILOMETERS
 MAXIMUM LIKELIHOOD ESTIMATION OF MODEL B.

(K, k)	(1,1)	(2,1)	(2,2)	(3,1)	(3,2)	(3,3)
\bar{h}_{st}	-0.168 (0.600)	-8.125*** (0.812)	0.071 (0.915)	-8.666*** (1.088)	-0.375 (0.670)	5.908** (2.081)
\bar{h}_{st}^2	-0.067 (0.066)	0.798*** (0.088)	-0.127 (0.102)	0.830*** (0.119)	-0.023 (0.074)	-0.786** (0.229)
h_{it}	1.460*** (0.010)	1.118*** (0.010)	2.268*** (0.033)	0.983*** (0.012)	1.734*** (0.025)	2.616*** (0.052)
$1/n_{st}$	-0.447*** (0.096)	0.522*** (0.105)	-1.468*** (0.233)	3.220*** (0.189)	-0.173 (0.106)	-1.857*** (0.329)
Constant	0.323 (1.376)	20.495*** (1.876)	-3.987 (2.062)	23.076*** (2.497)	-1.068 (1.513)	-18.901*** (4.774)
σ_k	1.057*** (0.003)	0.688*** (0.004)	1.211*** (0.009)	0.590*** (0.006)	0.679*** (0.010)	1.262*** (0.015)
r_k	. (.)	0 (0)	-0.866*** (0.034)	0 (0)	0.069 (0.045)	-0.886*** (0.061)
p_k	1	0.704	0.296	0.403	0.431	0.166
Log-Lik	-84,254.79	-79,507.32	-79,507.32	-78,522.24	-78,522.24	-78,522.24
Observ.	57,130	57,130	57,130	57,130	57,130	57,130

Note : Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model B. The dependent variable is always ℓ_{it} . Estimates have been obtained by Maximum Likelihood. Column (1,1) has $K = 1$ (or no group); Column (2,k) gives the coefficients of the model for type $k = 1, 2$ with $K = 2$ and Column (3,k) gives the coefficients of the model for type $k = 1, 2, 3$ when $K = 3$. The σ_k line gives the estimated standard deviations of the error term for each type.

Until now, we explored the “intensive” economies of density, but in addition, there are extensive economies of density driven by the number of employees n_{st} . Consider Model B with two types, *i.e.*, $K = 2$. Figure 6 gives 3-dimensional representations of the L/H surface as a function of H and n when $K = 2$, based on the results of Table 6 again. It is very clear, on Figure 6, that only Type 1 exhibits both kinds of economies of density. Type 2, dwarfed by Type 1 because of much smaller L/H ratios, has a surface that looks like a textbook production function with two factors.

Figure 5. Model B

L/H Curves Conditional on Type k and Average *L/H* Curve, for $K = 3$.

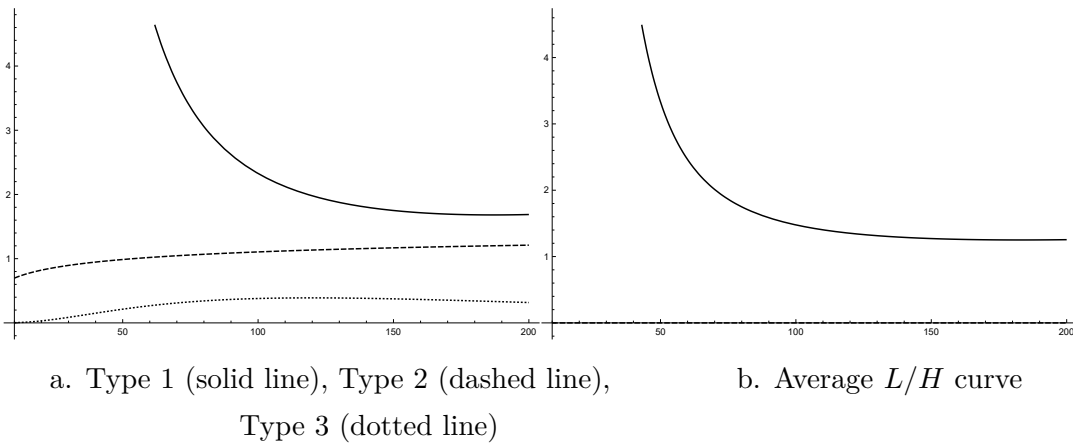
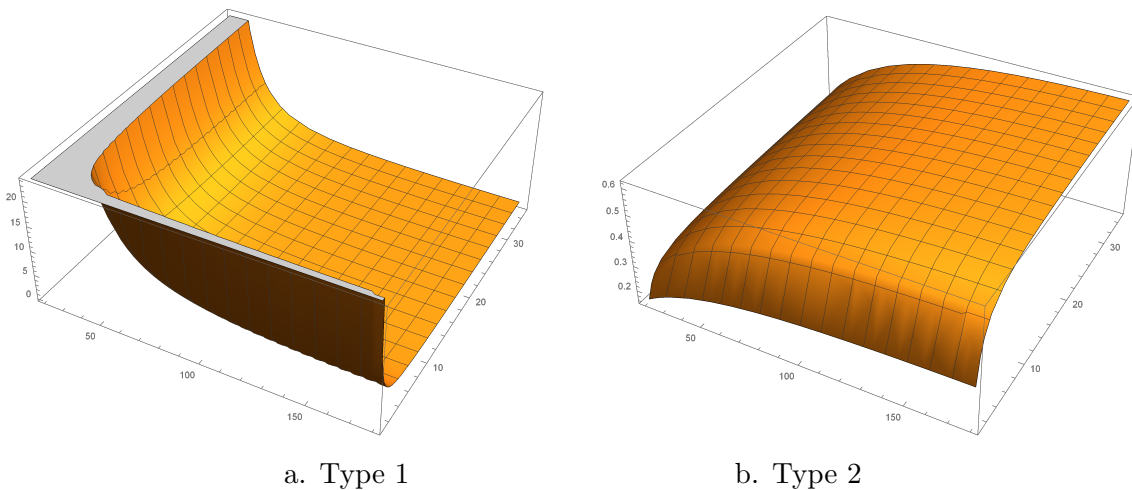


Figure 6. Model B. 3D Plots of *L/H* Surfaces Conditional on Type k , for $K = 2$



4.3 Robustness

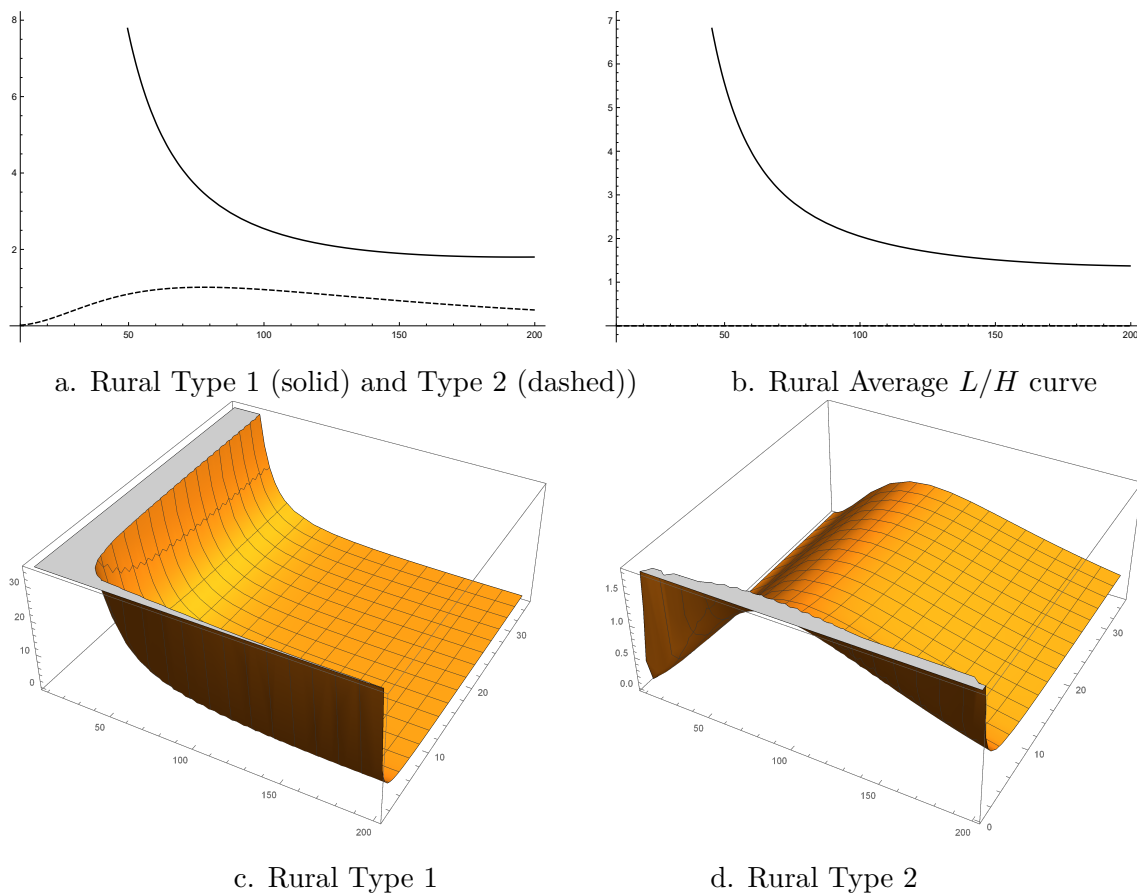
4.3.1 Model B in Urban and Rural Districts

The system of types is not simply replicating the urban/rural divide. To check the robustness of our findings, we re-estimated Model B with indicators of two subsamples, the urban and rural districts, interacting with the parameters of Model B. The urban or rural denomination is not

our choice, but the official classification of the National Institute of Statistics, *i.e.*, *INSEE*). The full results are presented in Appendix F, in Tables 13 and 14, for Urban and Rural districts respectively. We discuss here the average L/H curves that are computed with the help of subsample estimates.²⁶

Figure 7 gives the L/H curves of the 2-types version of Model B in the rural districts. Fig. 7a shows that Type 1, the majority type with a frequency of nearly 70%, has a typical, convex U-shaped curve. Type 1 exhibits economies of density. Type 2 has a much smaller L/H ratio and a quasi-concave curve, with rather negligible economies of density. Figure 7b gives the average curve, which is U-shaped and convex. We know that economies of density are due to the presence of Type 1.

Figure 7. Model B. L/H Curves and 3D Surfaces, Conditional on Type k and Average L/H Curve, for $K = 2$, in the Rural Districts



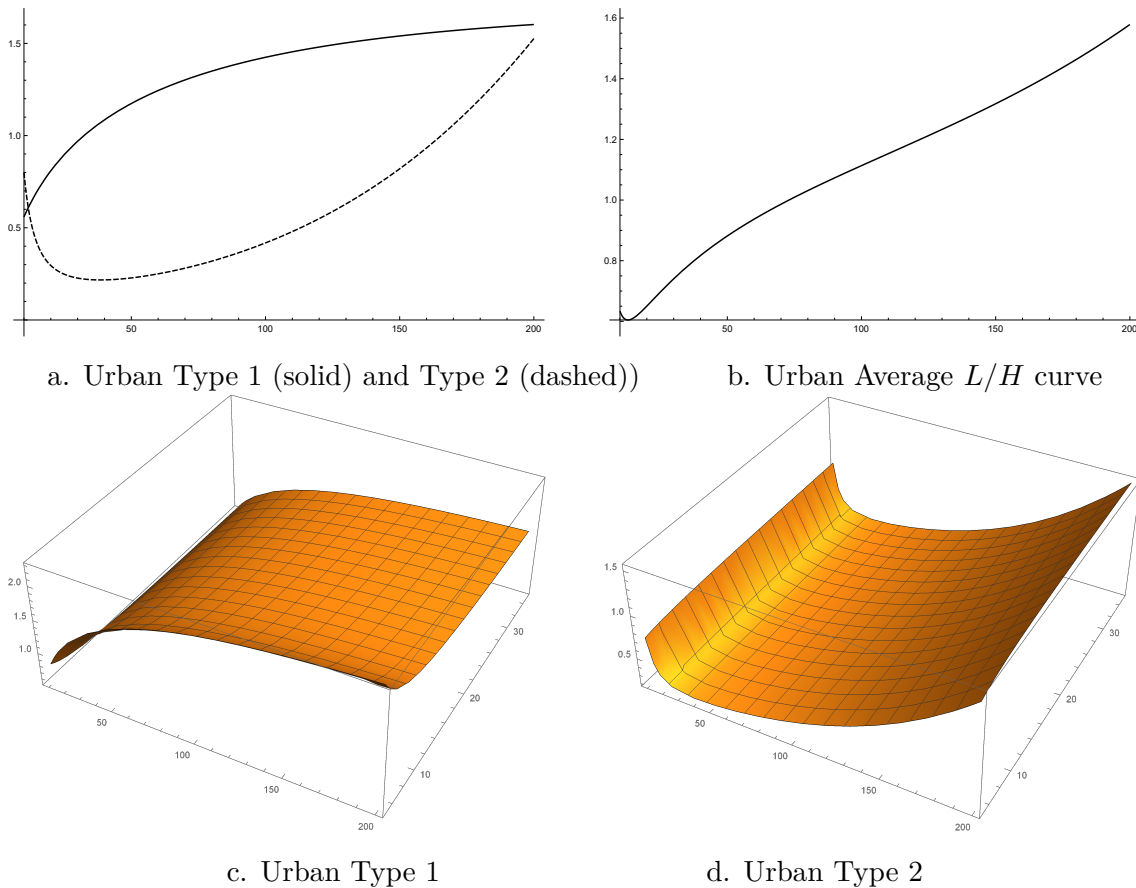
Figures 7c and 7d, show that economies of density are also extensive for both types, the curves being decreasing with the size of the team n . The results are very different in urban districts. Figure 8 depicts the L/H curves of the 2-type version of Model B, in urban districts. Figure 8a shows that Type 2 has a convex and U-shaped L/H curve but Type 1 exhibits decreasing

²⁶Model B is re-estimated by ML, adding interactions of all variables with dummies indicating the Rural or Urban classification of each district.

returns on the relevant range of hours. Since Type 1 is more frequent, Figure 8b shows that the average curve is increasing on the relevant range. Figures 8c and 8d, give the 3D surfaces, that is, L/H as a function of H and n , for Type 1 and Type 2 respectively: the extensive economies of density are negligible. To sum up, in the urban districts, economies of density play a minor role.

Some interesting conclusions can be drawn from this exercise. The finding that there are “convex” and “quasi-concave” types seems to be robust. The quasi-concave types are more frequent in urban districts while the convex, U-shaped types are more frequent in rural districts. The concave types typically have a low L/H ratio (less than 1.5 km per hour of service): these types do not spend much time driving. The economies of density observed in the aggregate are driven by the convex U-shaped types. Besides, the low L/H , quasi-concave types are not confined to urban districts. We are not surprised to see that the kilometers consumed per hour of service are below 1.5 in urban districts, because the density of patients in geographical space is high. But there are concave types with a low L/H ratio in rural districts too: presumably, these employees stay in the same village.

Figure 8. Model B. L/H Curves and 3D Surfaces, Conditional on Type k and Average L/H Curve, for $K = 2$, in the Urban Districts



4.3.2 Model B with 6 Types

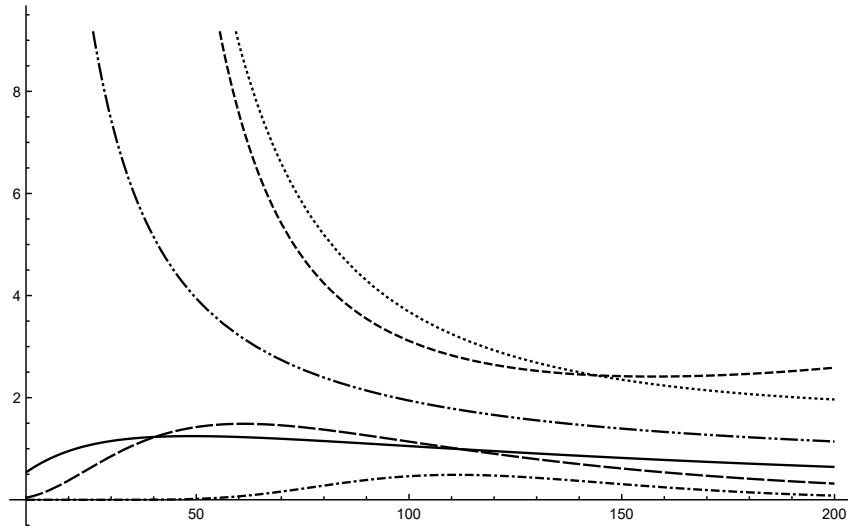
Our general conclusion resists conditioning on observable characteristics like rural or urban districts, but the distinction between district categories may be inadequate for the study of our problem. A legitimate approach is then to increase flexibility by increasing the number of types and let the likelihood maximization algorithm speak. We estimated Model B with 6 types by means of the sequential EM algorithm (the full results are displayed in Appendix G, Table 15 and Table 16), and the results confirm our general conclusion that economies of density are present in the home care industry, and more precisely, mostly present in rural districts.

Figure 9 displays the L/H curves computed with the estimated parameters of a 6-type version of Model B, estimated by means of the sequential EM algorithm described in Appendix D, and in the rural districts.²⁷ Figure 9a plots the six curves simultaneously. Three types out of six, *i.e.*, Types 2, 3 and 6, exhibit decreasing and convex L/H curves. Types 1, 4 and 5 have quasi-concave hump-shaped curves, with a decreasing portion, but they do not consume many kilometers: their L/H ratios are below 1.5 km per hour of service. Figure 9b shows the quasi-concave curves of types 1, 4 and 5 separately.

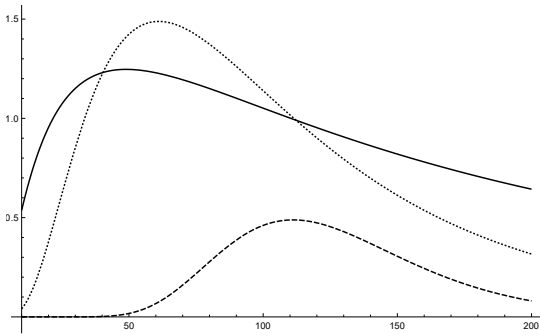
Fig. 9c depicts the L/H curves of types 2, 3 and 6 separately. Types 2, 3 and 6 together amount to 65% of the employees in rural districts. So we find more or less the same proportion of convex types as in the model with $K = 2$ (*i.e.*, Table 13, column (2,1)). Figure 9d gives the average L/H curve of the 6-type model. This curve has the typical convex and decreasing shape characterizing economies of density on the relevant range; it is essentially the same as before, when $K = 2$, but expressed as an average of 6 curves instead of 2 or 3. We conclude that Model B exhibits economies of density with two, three or six types. Our finding is therefore robust. The question is now to choose the appropriate value of K . It may be that using 6 types is just a form of overfitting.

²⁷The curves are drawn with $1/n$ fixed at its median value.

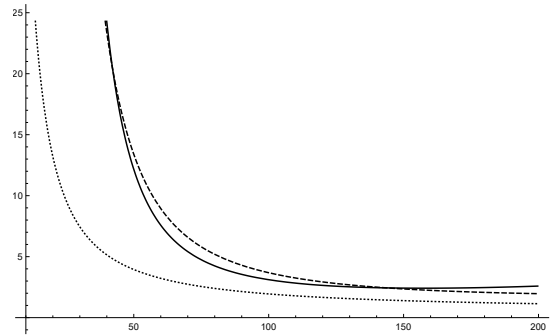
Figure 9. Model B. L/H Curves Conditional on Type k and Average Curve, for $K = 6$, in Rural Districts



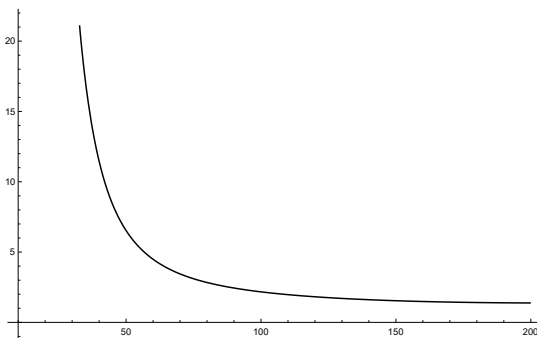
a. Type 1 (solid), Type 2 (dashed), Type 3 (dotted), Type 4 (dash-dot), Type 5 (long -), Type 6 (dash-dot-dot).



b. Type 1 (solid), Type 4 (dashed), Type 5 (dotted).



c. Type 2 (solid), Type 3 (dashed), Type 6 (dotted).



d. Average L/H curve over 6 types

5 Choice of the Number of Types K . Quality of Classification

We estimated several versions of the model. How should we choose K , the number of types? When $K = 6$, the model fits the data better than a model with only two types, but are the 6 types interpretable? In other words, is the type of each individual well-identified? Do all types correspond to real-world individuals? For instance, we know that many low-ratio employees live near the French Riviera or in Corsica. To answer these questions, we will first study the empirical distribution of estimated posterior probabilities \hat{p}_{itk} for each type k . We will then study a battery of criteria: the classic information criteria, and measures of *entropy*.

5.1 Empirical Distributions of Posterior Probabilities p_{itk} .

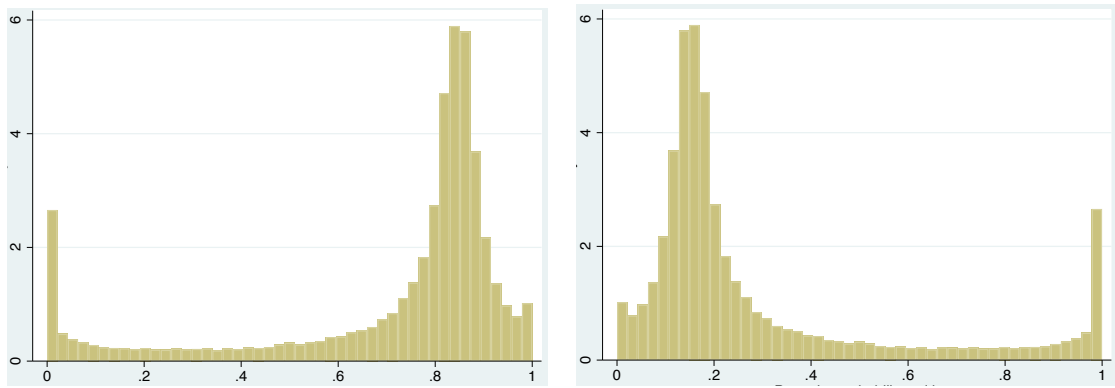
If the *probability distribution of the posterior probabilities* \hat{p}_{itk} has only masses close to zero and one, then, type k is well-identified (or well-classified). We know which employees belong or not to group k . Figure 10 below shows the empirical distributions of p_{it1} and p_{it2} for Model B with $K = 2$. Fig. 10a exhibits a probability mass between 0.75 and 1 and Fig 10b is a mirror image of Fig. 10a because $p_{it2} = 1 - p_{it1}$. Type 2 is visibly a minority. These pictures are typical of a case in which individual types are (relatively) well-identified. In other words, for $K = 2$, the partitioning of observations associated with Model B is excellent.

Figure 11 shows the equivalent results for the 3-type version of Model B. Fig. 11a shows that type 1 is easy to recognize (and it is relatively easy to point which observations do not belong to group 1 with a probability greater than .8, but the other two groups are difficult to distinguish, as shown by Fig. 11b and Fig. 11c. Many (i, t) s have a probability around 0.45 or .5 of belonging to group 2 and a probability around 0.5 of belonging to group 3.²⁸ It seems that types 2 and 3 cannot easily be distinguished from other types. The classification is much fuzzier when $K = 3$, as compared to $K = 2$.

Figure 12 displays the densities of the \hat{p}_{itk} , $k = 1, \dots, 6$ in the six-type version of Model B. Visual inspection of Figure 12 shows that the types of the 6-type model are difficult to recognize. Yet some types are better identified than others. Understandably, the quasi-concave types 1, 4 and 5 are relatively easy to separate from others because they have markedly different L/H ratios. This is because we see masses at and near zero for these types. We know who's not a Type 1, 4 or 5. But they constitute only a minority. The convex types, that is, 2, 3 and 6, are typically not well identified (the distributions of p_{it2} , p_{it3} and p_{it6} have substantial masses around 0.3). Again, we know only relatively well who's not a Type 2, 3 or 6. Visual inspection (Fig. 9) shows that types 2 and 3 are difficult to distinguish. The essence of the underlying structure seems to be uncovered with only two types.

²⁸There are masses at point 0 on Figures 11b and 11c, but these masses only indicate that we know that many observations *do not* belong to group 2 (or 3) with a high probability.

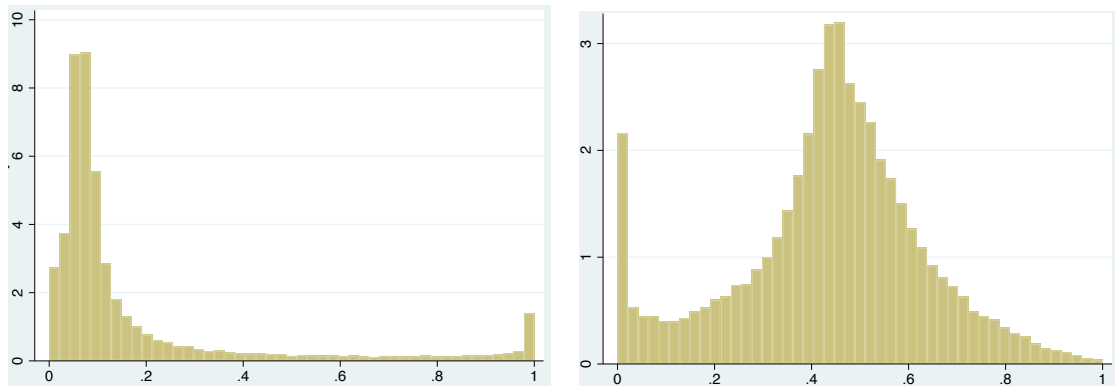
Figure 10. Model B. Distribution of Posterior Probabilities p_{itk} , with $K = 2$



a. Distribution of p_{it1} (Type 1)

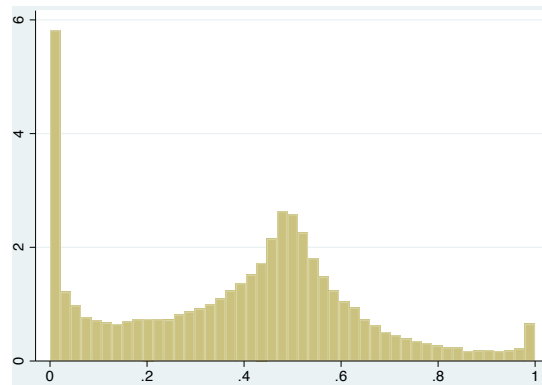
b. Distribution of p_{it2} (Type 2).

Figure 11. Model B. Distribution of Posterior Probabilities p_{itk} , with $K = 3$



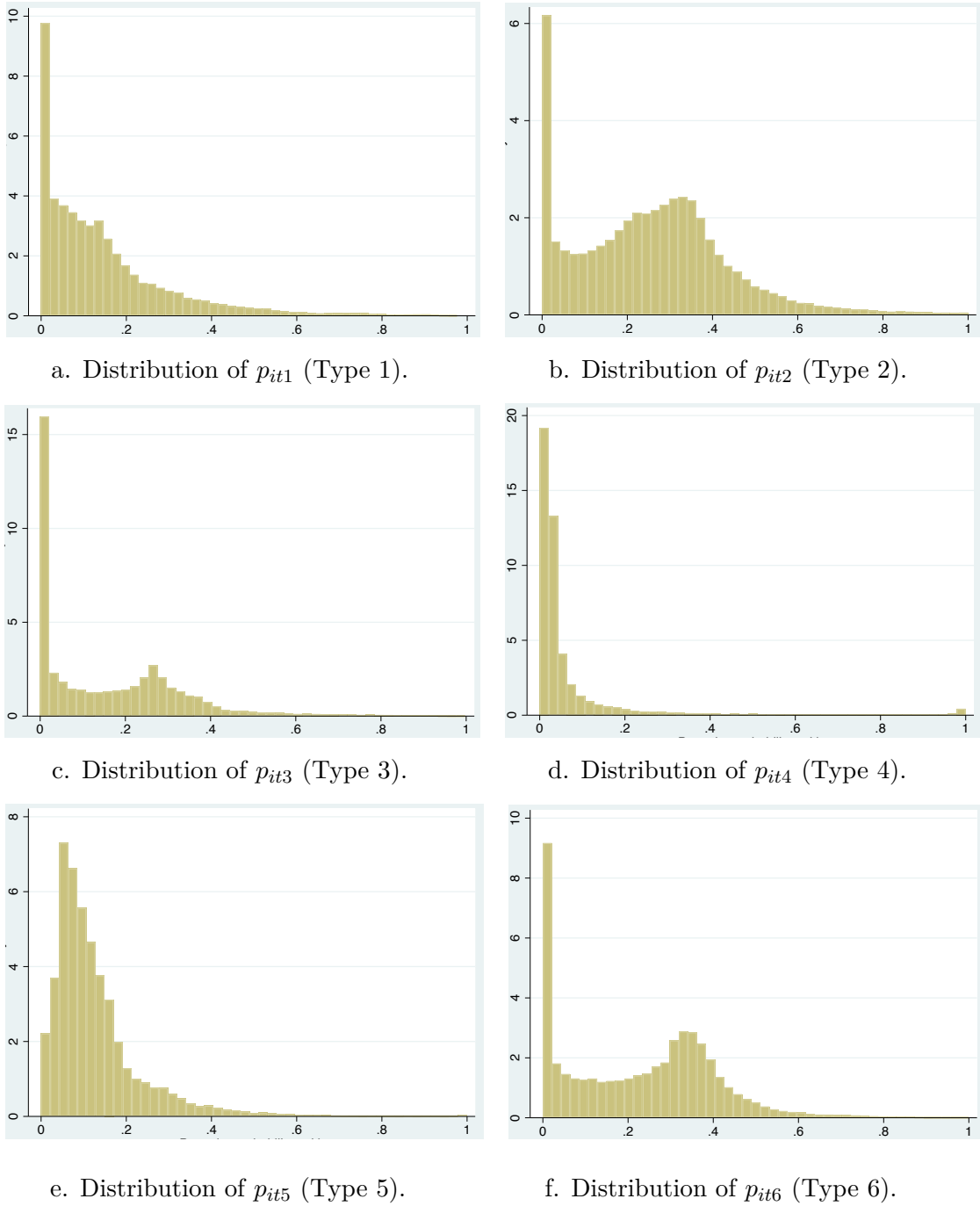
a. Distribution of p_{it1} (Type 1)

b. Distribution of p_{it2} (Type 2).



c. Distribution of p_{it3} (Type 3).

Figure 12. Model B. Distribution of Posterior Probabilities p_{itk} , with $K = 6$



5.2 Information and Entropy Criteria

When it comes to the choice of the number of types K , there exists a tension between Information and Entropy criteria. First of all, the log-likelihood of the model with K types, denoted $\mathcal{L}(K)$, is typically increasing and concave: an additional type will always lead to some improvement of $\mathcal{L}(K)$, but with decreasing marginal values. Next, the Akaike and Bayesian Information Criteria (*i.e.*, AIC and BIC) penalize the number of parameters of the model (see Akaike (1974), Schwarz

(1978)), and both criteria will in principle reach a minimum for some value of K .

But AIC and BIC are not a very good guide in the context of finite mixture models. AIC tends to overestimate the correct number of components (AIC pushes towards over-fitting). BIC corrects for these difficulties but tends to underestimate K .²⁹ Celeux and Soromenho (1996) have proposed a simple choice criterion based on the notion of entropy called the *Normalized Entropy Criterion*, or NEC (see Appendix H for some details).

In our context, entropy \mathcal{E} must be defined as follows,

$$\mathcal{E}(K) = - \sum_{i=1}^n \sum_{t \in T_i} \sum_{k=1}^K \hat{p}_{itk} \ln(\hat{p}_{itk}). \quad (16)$$

It is easy to check that $\mathcal{E}(1) = 0$ and $0 \leq \mathcal{E}(K) \leq N \ln(K)$, where N is the number of observations (i, t) . The entropy is maximal when $p_{itk} = 1/K$ for all k and all (i, t) . Entropy is maximal when types cannot be distinguished because any observation can belong to every group with the same probability $1/K$. Entropy is minimal (equal to zero) when partitioning is perfect. Indeed, if for all (i, t) , there exists a type $k = k(i, t)$ such that $p_{itk} = 1$, then, $\mathcal{E}(K) = 0$. We can divide entropy by its maximum value to obtain an index taking values in $[0, 1]$. Define $\mathfrak{E}(K) = (N \ln(K))^{-1} \mathcal{E}(K)$. This index should be small to ensure a good classification.

To define the NEC, we consider the gains, in terms of the Log-Likelihood, with respect to $K = 1$, that is $\mathcal{L}(K) - \mathcal{L}(1)$. Entropy is now divided by this gain. NEC is defined as follows, for $K > 1$,

$$\text{NEC}(K) = \frac{\mathcal{E}(K)}{\mathcal{L}(K) - \mathcal{L}(1)}. \quad (17)$$

Another simple criterion that measures the quality of classification is the Average Hirschman-Herfindahl Index. This index is defined as follows,

$$\text{AHHI}(K) = \frac{1}{N} \sum_{i=1}^n \sum_{t \in T_i} \sum_{k=1}^K \hat{p}_{itk}^2 \quad (18)$$

Note that AHHI is equal to 1 if all observations (i, t) are perfectly classified (*i.e.*, there exists $k = k(i, t)$ such that $p_{itk} = 1$). In addition we have, $1/K \leq \text{AHHI}(K) \leq 1$. It follows that the lower bound of AHHI is decreasing with K .³⁰ A normalized index can be constructed as follows. For $K > 1$, define $\mathfrak{H}(K) = (K * \text{AHHI} - 1)/(K - 1)$. We have $0 \leq \mathfrak{H}(K) \leq 1$. AHHI and $\mathfrak{H}(K)$ may increase with K (as we will see below); if these indices drop, this is because the quality of classification deteriorates when K increases.

²⁹For references on these problems and the discussion of other information criteria, see Celeux and Soromenho (1996).

³⁰On the use of AHHI in classification problems, see Windham and Cutler (1992).

Table 7: MODEL B. INFORMATION AND ENTROPY CRITERIA. CHOICE OF K

K	1	2	3	4	5	6	7	8	9
$L(K)$	-82,468	-77,013	-75,819	-75,133	-74,718	-74,497	-74,321	-74,143	-74,118
BIC	165,112	154,388	152,186	151,001	150,357	150,100	149,935	149,765	149,902
AIC	164,698	154,093	151,738	150,401	149,605	149,196	148,880	148,557	148,542
$\mathcal{E}(K)$	0	24,132	43,992	51,263	64,836	73,234	76,956	82,018	88,938
NEC	.	4.424	6.616	6.989	8.366	9.187	9.446	9.852	10.651
$\mathfrak{E}(K)$	0	0.609	0.701	0.723	0.705	0.715	0.692	0.690	0.708
AHHI	1	0.733	0.529	0.487	0.397	0.345	0.330	0.306	0.273
$\mathfrak{H}(K)$.	0.466	0.293	0.316	0.246	0.214	0.218	0.206	0.182

Note: Model B has been estimated by the EM algorithm repeatedly with values of K ranging from $K = 1$ to $K = 9$. Model B has been estimated, adding interactions of the K types with the indicators of three subsamples: the Urban, Peri-Urban and Rural districts (that partition the dataset). Parameters therefore vary not only with type, but also with the three types of district. $L(K)$ is the estimated Log-Likelihood with K types. BIC is the Bayesian Information Criterion. AIC is Akaike’s Information Criterion. $\mathcal{E}(K)$ is entropy as defined above. NEC is Celeux and Soromenho’s Normalized Entropy Criterion. $\mathfrak{E}(K)$ is just $\mathcal{E}(K)$ divided by its maximal value, $N \ln(K)$. AHHI is the Average Hirschman-Herfindahl Index defined in the text. $\mathfrak{H}(K)$ is the normalized Herfindahl index defined in the text. In the case of Model B; BIC seems to reach a minimum for $K = 8$; AIC never reaches a minimum between $K = 1$ and $K = 9$; NEC is minimal for $K = 2$. Most of the gains in terms of $L(K)$ are achieved with $K = 2$ or $K = 3$.

We can now discuss the optimal value of K with the help of the various criteria. First we look at increments in the log-likelihood. We observe on Table 7 that a large part of the task of improving the model with a mixture is achieved with only two types.

Table 7 displays information, entropy and Herfindahl criteria for Model B from $K = 1$ to $K = 9$.³¹ First, we find that the likelihood increment $\mathcal{L}(2) - \mathcal{L}(1)$ is the largest among all the increments $\mathcal{L}(K + 1) - \mathcal{L}(K)$. At the same time, AIC is never minimized (in the range $K = 1, \dots, 8$), but it is very flat for $K > 3$, while BIC seems to be minimized at $K = 8$, and NEC and $\mathfrak{E}(K)$ are minimized at $K = 2$. Finally, Herfindahl indices are maximized at $K = 2$. Model B provides a case in which the entropy and information criteria disagree. Note that $\mathfrak{E}(K)$ and $\mathfrak{H}(K)$ are nonmonotonic. (See Appendix E for another instance of nonmonotonic Herfindahl indices). We conclude that $K = 2$ is the most reasonable choice.

The usual information criteria, likelihood, AIC and BIC, are imperfect guides to choose the number of types. To make our final choice, we combined the information criteria with two other

³¹To compute Table 7, Model B has been estimated with added interactions. More precisely, we added the interactions of types with three indicators of urban, rural and peri-urban districts (the latter being just districts that are neither rural nor urban). This version of the Model is more flexible.

requirements: (i), a model with precisely estimated coefficients and (ii), a model with sufficiently well-separated types. Entropy and Herfindahl indices measure this quality of classification: both types of indices indicate that $K = 2$, for model B, is the best compromise.

6 Conclusion

We used a unique data set with a panel structure, reporting the activity of a network of Home Care Services. The data provide monthly observations of the hours of service at the patients' homes and of the kilometers traveled, for each employee, during several years and in many districts scattered on the French territory. We first estimated log-linear models of a real transportation cost function that exhibit substantial economies of density. We discussed the connection of our results with the Bearwood-Halton-Hammersley Theorem. The latter theorem characterizes the asymptotic behavior of the minimal length of a tour visiting a number of points randomly scattered on a plane. We show that our results do not obviously reject the mathematical model, predicting that the minimal distance is asymptotic to the square root of the number of visited points. The results strongly suggest that economies of density are the result of team or group synergies operating at the district (*i.e.*, local branch) level.

We then investigated the data further to take the unobserved heterogeneity of employees into account. A finite mixture, quadratic-in-logs model, estimated by Maximum Likelihood, yielded improved results and we uncovered a set of latent groups. Each latent group of employees is characterized by its own average transportation cost curve. The finite-mixture model generates an average cost curve that is typically U-shaped for a majority of individuals. We discussed the best choice of the number of latent groups, using, not only standard information criteria such as AIC and BIC, but also entropy indices, to better measure the quality of classification generated by the model. We conclude that a model with 2 types is enough to capture the essential result. The average transportation cost function is the weighted average of the cost functions of two latent groups of employees: a majority group with a convex decreasing, *i.e.*, U-shaped average cost curve, and a minority group with a much smaller consumption of kilometers and a quasi-concave curve. These results confirm the existence of substantial economies of density in the Home Care Industry, and hence carry possible consequences in terms of regulation, because free competition inflates transportation costs in an industry that has a local natural monopoly structure.

7 References

Arcidiacono, Peter and John B. Jones (2003), "Finite Mixture Distributions, Sequential Likelihood, and the EM Algorithm", *Econometrica*, 71(3), 933-946

Aguirregabiria, Victor, Clark, Robert and Hui Wang (2016), “Diversification of Geographic Risk in Retail Bank Networks: Evidence from Bank Expansion after the Riegle-Neal Act”, *Rand Journal of Economics*, 47(3), Fall, p 529-572.

Akaike, Hirotugu (1974), “A New Look at the Statistical Identification Model”, *IEEE Transactions on Automatic Control*, 19, p 716-723.

Applegate, David L., Bixby, Robert E., Chvatál, Vašek and William J. Cook (2006), *The Traveling Salesman Problem: A Computational Study*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, New Jersey.

Arlotto, Alessandro and J. Michael Steele (2016), “Beardwood-Halton-Hammersley Theorem for Stationary Ergodic Sequences: A Counterexample”, *Annals of Applied Probability*, 26(4), 2141-2168.

Bandiera, Oriana, Barankay, Iwan and Imran Rasul (2013), “Team Incentives: Evidence from a Firm Level Experiment”, *Journal of the European Economic Association*, 11(5), 1079–1114, doi.org/10.1111/jeea.12028.

Bartel, Ann P., Nancy D. Beaulieu, Ciaran S. Phibbs, and Patricia W. Stone (2014), “Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector.” *American Economic Journal: Applied Economics*, 6(2), 231-59.

Beardwood, Jillian, Halton, John H. and John M. Hammersley (1959), “The Shortest Path Through Many Points”, *Proceedings of the Cambridge Philosophical Society*, 55, 299-327.

Behrens, Kristian, Gaigné, Carl, Ottaviano, Gianmarco I.P. and Jacques-François Thisse (2006), “How Density Economies in International Transportation Link the Internal Geography of Trading Partners”, *Journal of Urban Economics*, 60, 248-263.

Bitzan, John D. and Theodore E. Keeler (2007), “Economies of Density and Regulatory Change in the U.S. Railroad Freight Industry”, *Journal of Law and Economics*, 50(1), p 157-180.

Bonhomme, Stéphane and Jean-Marc Robin (2009), “Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000”, *Review of Economic Studies*, 76(1), 63-92.

Bonhomme, Stéphane and Elena Manresa (2015), “Grouped Patterns of Heterogeneity in Panel Data.” *Econometrica*, 83, 1147-1184, doi.org/10.3982/ECTA11319

Bonsang, Eric (2009), “Does Informal Care from Children to their Elderly Parents Substitute for Formal Care in Europe?”, *Journal of Health Economics*, 28, 143-154.

Bouveyron, Charles, Celeux, Gilles, Murphy, T. Brendan and Adrian E. Raftery (2019), *Model-Based Clustering and Classification for Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Braeutigam, Ronald R., Andrew F. Daughety and Mark A. Turnquist (1984), “A Firm Specific Analysis of Economies of Density in the U.S. Railroad Industry”, *Journal of Industrial Economics*, 33(1), 3-20.

Brueckner, Jan K. and Pablo T. Spiller (1994), “Economies of Traffic Density in the Deregulated Airline Industry”, *Journal of Law and Economics*, 37(2) , p 379-415.

Byrne, David, Goeree, Michelle S., Hiedemann, Bridget and Steven Stern (2009), “Formal Home Health Care, Informal Care, and Family Decision Making”, *International Economic Review*, 50(4), 1205-1242.

Caves, Douglas W., Christensen, Laurits R., and Michael W. Tretheway (1984), “Economies of Density versus Economies of Scale. Why Trunk and Local Service Airline Costs Differ”, *Rand Journal of Economics*, 15(4), Winter, p 471-489.

Cazals, Catherine, Duchemin, Pascale, Florens, Jean-Pierre, Roy, Bernard, and Olivier Vialanex (2001), “An Econometric Study of Cost Elasticity in the activities of Post-Office Counters”, in: M.A Crew and P.R. Kleindorfer, editors, *Postal and Delivery Services : Pricing, Productivity, Regulation and Strategy*, p 161-171.

Celeux, Gilles and Gilda Soromenho (1996), “An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model”, *Journal of Classification*, 13, 195-212.

Chan, David C. (2016), “Teamwork and Moral Hazard: Evidence from the Emergency Department”, *Journal of Political Economy*, 124(3), 734-770.

Chen, Yiqun (2021), “Team-Specific Human Capital and Team Performance: Evidence from Doctors.” *American Economic Review*, 111(12), 3923-3962, doi: 10.1257/aer.20201238

Ciccone, Antonio, and Robert E. Hall (1996), “Productivity and the Density of Economic Activity”, *American Economic Review*, 86(1), p 54-70.

Cissé, Mohamed, Yalçındag, Semih, Kergosien, Yannick, Sahin, Evren, Lenté Christophe and Andrea Matta (2017), “OR Problems Related to Home Health Care: A Review of Relevant Routing and Scheduling Problems”, *Operations Research for Health Care*, 13-14, p 1-22.

Combes Pierre-Philippe and Laurent Gobillon (2015), “The Empirics of Agglomeration Economies”, Chapter 5 in: Duranton, Henderson and Strange, editors, *Handbook of Regional and Urban Economics*, vol 5A, Elsevier, Amsterdam.

David, Guy and Kunhee Lucy Kim (2018), “The Effect of Workforce Assignment on Performance: Evidence from Home Health Care”, *Journal of Health Economics*, 50, 26-45.

Dubin, Jeffrey A. and Peter Navarro (1988), “How Markets for Impure Public Goods Organize: The Case of Household Refuse Collection”, *Journal of Law, Economics and Organization*, 4(2), 217-241.

Fikar, Christian and Patrick Hirsch (2017), “Home Health Care Routing and Scheduling: A Review”, *Computers and Operations Research*, 77, p 86-95.

Forder, Julien, Malley, Juliette, Towers, Ann-Marie and Ann Netten (2014), “Using Cost-Effectiveness Estimates from Survey Data to Guide Commissioning: An Application to Home Care”, *Health Economics*, 23, 979-992.

Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov (2017), “Team Incentives and Performance: Evidence from a Retail Chain.” *American Economic Review*, 107(8), 2168-2203, doi: 10.1257/aer.20160788

Gary-Bobo, Robert, Goussé, Marion and Jean-Marc Robin (2016), "Grade Retention and Unobserved Heterogeneity", *Quantitative Economics, Journal of the Econometric Society*, vol. 7(3), p 781-820.

Guldmann, Jean-Michel (1990), “Economies of Scale and Density in Local Telephone Networks”, *Regional Science and Urban Economics*, 20, 521-535.

Holmes, Thomas J. (2011), “The Diffusion of Wal-Mart and Economies of Density”, *Econometrica*, 79(1), p 253-302.

Holmes Thomas J. and Sanghoon Lee (2012), “Economies of Density versus Natural Advantage: Crop Choice on the Back Forty”, *Review of Economics and Statistics*, 94(1), p 1-19.

Houde, Jean-François, Newberry, Peter and Katja Seim (2017), “Economies of Density in E-Commerce: A Study of Amazon’s Fulfillment Center Network”, NBER WP no 23361, NBER, Cambridge, Massachusetts.

Jaravel, Xavier, Neviana Petkova and Alex Bell (2018), “Team-Specific Capital and Innovation.” *American Economic Review*, 108(4-5), 1034-73,doi: 10.1257/aer.20151184

McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, Wiley Series in Probability and Statistics, J. Wiley, New York.

Mori, Tomoya and Koji Nishikimi (2002), “Economies of Transport Density and Industrial Agglomeration”, *Regional Science and Urban Economics*, 32, 167-200.

Morikawa, Masayuki (2011),“Economies of Density and Productivity in Service Industries: An Analysis of Personal Service Industries Based on Establishment Level Data”, *Review of Economics and Statistics*, 93(1), p 179-192.

Orsini, Chiara (2010), “Changing the Way the Elderly Live: Evidence from the Home Health Care Market in the United States”, *Journal of Public Economics*, 94, 142-152.

Roberts, Mark J. (1986),“Economies of Density and Size in the Production and Delivery of Electric Power”, *Land Economics*, 62(4), p 378-387.

Rosenthal Stuart S. and William C. Strange (2020), “How Close is Close? The Spatial Reach of Agglomeration Economies”, *Journal of Economic Perspectives*, 34(3), p 27-49.

Schwarz, Gideon (1978), “Estimating the Dimension of a Model”, *Annals of Statistics*, 6, 461-464.

Stabile, Mark, Laporte, Audrey and Peter C. Coyte (2006), “Household Responses to Public Home Care Programs”, *Journal of Health Economics*, 25, 674-701.

Steinerberger, Stefan (2015), “New Bounds for the Traveling Salesman Constant”, *Advances in Applied Probability*, 47, 27-36.

Su, Liangjun, Shi, Zhentao and Peter C. B. Phillips (2016), “Identifying Latent Structures in Panel Data”. *Econometrica*, 84, 2215-2264, doi.org/10.3982/ECTA12560

Torres, Marcelo and Catherine J. Morrison Paul (2006), “Driving Forces for Consolidation or Fragmentation of the US Water Utility Industry: A Cost Function Approach with Endogenous Output”, *Journal of Urban Economics*, 59, 104-120.

Windham, M. P. and A. Cutler (1992), “Information Ratios for Validating Cluster Analyses”, *Journal of the American Statistical Association*, 87, 1188-1192.

Weidmann, Ben and David J. Deming, (2021), “Team Players: How Social Skills Improve Team Performance.” *Econometrica*, 89, 2637-2657, doi.org/10.3982/ECTA18461

Xu, Hangtian and Hidezaku Itoh (2018), “Density Economies and Transport Geography: Evidence from the Container Shipping Industry”, *Journal of Urban Economics*, 105, 121-132.

A Appendix: Density, Distance and the Beardwood-Halton-Hammersley Theorem

A.1 Density of Clients and Miles Traveled

To rationalize Figure 1, consider the unit square $[0, 1]^2$. Let n denote the number of patients. Let (x_1, \dots, x_n) denote the array of points representing the clients' locations, and define the set of locations $X_n = \{x_1, \dots, x_n\}$, with each $x_i \in [0, 1]^2$, $i = 1, \dots, n$. We assume that the points x_i are independent, identically distributed (i.i.d), uniformly distributed random variables drawn in $[0, 1]^2$. Let now $y(x)$ denote the nearest neighbor of patient x in the unit square (in the sense of Euclidean distance). Let $z = \|x - y(x)\|$, be the distance between x and the nearest neighbor. Suppose that x is in the interior of the square. Then, the probability that $y(x)$ is located out of a closed disc of center x and radius r can be written,

$$\Pr(z > r) = \prod_{y \in X_n \setminus \{x\}} \Pr(\|x - y\| > r) = \left(1 - \pi r^2\right)^{n-1},$$

provided that $r > 0$ is small enough to ensure that $\pi r^2 < 1$ and that the disc is entirely included in the square. We then immediately derive the probability of finding the nearest neighbor in the disc with center x and radius r , denoted $F_n(x, r)$, that is,

$$F_n(x, r) = \Pr(z \leq r) = 1 - \left(1 - \pi r^2\right)^{n-1}.$$

It is possible to show that the expected value of z goes to 0 when n tends towards infinity. Let us now fix p , the probability that the nearest neighbor is located at a distance smaller than r , that is, fix $F_n(x, r) = p$. Next, compute the radius r' ensuring that this probability remains equal to p when the number of clients is doubled, that is, $p = F_{2n}(x, r')$. We must have $(n - 1) \ln[1 - \pi r^2] = (2n - 1) \ln[1 - \pi (r')^2]$. If n is large enough, using the approximation $\ln(1 + u) \simeq u$ for small values of u , we easily find,

$$\frac{1}{2} \simeq \frac{n - 1}{2n - 1} \simeq \left(\frac{r'}{r}\right)^2.$$

The ratio of r' to r is approximately,

$$\frac{r'}{r} \simeq \sqrt{\frac{1}{2}} \simeq 0.707.$$

Hence, the distance to the nearest neighbor is 30% smaller when the number of patients is doubled. Let us denote $r = r_n(p)$ the radius of the disc in which the nearest neighbor is located with probability p . Then, we see that the area of the disc with radius r' , that is, $\pi (r')^2 = \pi r_{2n}^2(p)$ is near one half of the area of the disc with radius $r = r_n(p)$, i.e, $\pi r_n^2(p)$, when the number of clients grows from n to $2n$.

A.2 The Theorem of Beardwood, Halton and Hammersley

Consider now *tours*, defined as paths joining all points in $X_n = \{x_1, \dots, x_n\}$. The distance between two points is the Euclidean distance. Let $L(x_1, \dots, x_n)$ denote the length of the smallest tour joining all points in X_n . Then we can state the following result.

Theorem (Beardwood, Halton, Hammersley, 1959) Let x_1, x_2, \dots, x_n be independent and identically distributed, uniformly distributed random variables in $[0, 1]^2$. There exists a constant ρ such that, with probability 1,

$$\lim_{n \rightarrow \infty} \frac{L(x_1, \dots, x_n)}{\sqrt{n}} = \rho.$$

For a proof of this result, see Beardwood *et al.* (1959). More recent developments and further references will be found in Steinerberger (2015) and Arlotto and Steele (2016). The exact value of the constant ρ is still unknown, but various arguments have shown that $\rho \simeq 0.712$. A more general statement of the above result considers i.i.d variables, randomly distributed with respect to some probability distribution on the unit square of the plane (not necessarily a continuous distribution without mass points). Let $f(x)$ denote the density of the absolutely continuous part of this probability distribution, and denote coordinates $x = (y, z)$, then we have,

$$\lim_{n \rightarrow \infty} \frac{L(x_1, \dots, x_n)}{\sqrt{n}} = \rho \int_0^1 \int_0^1 f(y, z)^{1/2} dy dz.$$

A particular case is of course the uniform distribution on the unit square. In this case, since $f = 1$, the double integral above is the area of the square and boils down to 1. Suppose that X is a bounded subset of the plane with positive area (i.e., positive Lebesgue measure), and let $a(X)$ denote the area of X , then, there exists a constant ρ such that, with probability 1,

$$\lim_{n \rightarrow \infty} \frac{L(x_1, \dots, x_n)}{\sqrt{n}} = \rho \sqrt{a(X)}.$$

A difficulty comes from the fact that the i.i.d condition in the statement of the Beardwood-Halton-Hammersley Theorem cannot be weakened without caution. Some weaker forms of the independence condition lead to a weakened statement saying that the *expectation* of L converges towards $\rho\sqrt{n}$ as $n \rightarrow \infty$ (cf. Arlotto and Steele (2016)).

Finally, it is interesting to note that the BHH Theorem is also valid in spaces of higher dimension. Some districts being mountainous, it could be useful to consider the shortest tour visiting points in a cube. If $x_i \in [0, 1]^3$, we can state that there exists a constant ρ_3 such that, with probability 1,

$$\lim_{n \rightarrow \infty} \frac{L(x_1, \dots, x_n)}{n^{2/3}} = \rho_3.$$

B Appendix: Ordinary Least Squares

Table 8: KILOMETERS TRAVELED: OLS ON POOLED DATA

	(1)	(2)	(3)	(4)	(5)	(6*)
h_{it}	1.013*** (0.025)	1.050*** (0.025)	1.049*** (0.025)	. (.)	. (.)	. (.)
$h_{it} * \text{Part time}$. (.)	. (.)	. (.)	0.972*** (0.040)	0.973*** (0.041)	0.973*** (0.052)
$h_{it} * 80\% \text{ time}$. (.)	. (.)	. (.)	1.046*** (0.022)	1.113*** (0.018)	1.113*** (0.027)
$h_{it} * \text{Full time}$. (.)	. (.)	. (.)	1.127*** (0.037)	1.255*** (0.033)	1.255*** (0.045)
80%	0.470*** (0.018)	0.391*** (0.018)	0.392*** (0.018)	0.146 (0.197)	-0.222 (0.192)	-0.222 (0.262)
Full time	0.663*** (0.023)	0.641*** (0.024)	0.642*** (0.024)	-0.070 (0.250)	-0.692** (0.237)	-0.692* (0.307)
Constant	-0.259* (0.104)	-2.519*** (0.129)	0.259 (0.646)	-0.093 (0.168)	0.601 (0.662)	0.601 (1.784)
Speed	. (.)	0.031*** (0.006)	0.031*** (0.006)	. (.)	0.031*** (0.006)	0.031* (0.015)
Density of 80+ years old	. (.)	. (.)	-0.545** (0.160)	. (.)	-0.553** (0.160)	-0.553 (0.477)
Competition	. (.)	. (.)	0.079** (0.026)	. (.)	0.080** (0.026)	0.080 (0.078)
Time (in months)	. (.)	. (.)	0.002*** (0.001)	. (.)	0.002** (0.001)	0.002 (0.002)
District dummies	NO	YES	YES	NO	YES	YES
Observations	56,878	56,830	56,830	56,878	56,830	56,830
R ²	0.280	0.568	0.568	0.280	0.568	0.568
F	4530.44	589.26	589.51	3284.13	616.55	92.41

Note: Significance is indicated by: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable is always the log-km ℓ_{it} . Estimates are obtained by OLS on pooled data. (*) Equation (6) is estimated by clustering at the employee i level; there are 3686 clusters. *Time* is just calendar time t , ranging from $t = 1$ to $t = 44$ (between Novembre 2015 and June 2019). Variables $h_{it} * \text{Full Time}$, $h_{it} * 80\% \text{ Time}$ and $h_{it} * \text{Part Time}$ are interactions of log-hours h with the indicators of an interval of average hours. *Competition* is the number of competitors of the firm in district s . *Speed* is the ratio of kilometers to hours spent traveling. *Density* is the log of the density of population aged 80 and more in the county.

For the models of columns (4), (5) and (6) in Table 8, we defined three sub-samples, partitioning the set of employees, based on the value of an employee's average monthly hours. The three bins are called *Full-Time*, *80%-Time* and *Part-Time*. There is in fact a good deal of variation of hours in each of the bins, defined as wide intervals. When log-hours and the constant are interacted with the bin-dummies, column (4) of Table 8 shows that β is slightly smaller than one for part-time employees, with respectively $\hat{\beta} \simeq 0.97$ and that $\hat{\beta} \simeq 1.13$ for full-time employees. At the same time, the constant of full-time employees is much lower than the constant of the part-time ones (they consume much less kilometers). Adding controls and a complete set of district dummies does not change the result, in essence. Column (6) of Table 8 shows that clustering the data at the employee level does not change the key result, that is, estimates of β are still significant and around one.

C Appendix: Time Spent Travelling instead of Kilometers

Table 9: TIME SPENT TRAVELLING

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	FE	OLS	FE
h_{it}	1.327*** (0.019)	1.361*** (0.018)	1.362*** (0.018)	1.047*** (0.017)	.	.
\bar{h}_{st}	-0.282*** (0.023)	-0.758*** (0.034)	-0.751*** (0.034)	-0.485*** (0.025)	0.350*** (0.039)	0.383*** (0.045)
Constant	-3.495*** (0.127)	-1.955*** (0.195)	0.279 (0.302)	-1.275*** (0.122)	-0.656*** (0.182)	-0.432* (0.206)
Speed	.	.	-0.596*** (0.038)	.	.	.
Density of 80 ⁺	.	.	-0.076 (0.055)	.	.	.
Competition	.	.	0.004 (0.009)	.	.	.
Time (month)	.	.	0.004*** (0.000)	.	.	.
District dummies	NO	YES	YES	NO	YES	NO
Observations	57,091	57,091	57,091	57,091	3,100	3,100
R^2	0.313	0.482	0.485	.	0.833	.
R^2 within	.	.	.	0.247	.	0.043
R^2 between	.	.	.	0.302	.	0.095
R^2 overall	.	.	.	0.310	.	0.048
σ_u	.	.	.	0.920	.	0.562
σ_e	.	.	.	0.513	.	0.228
ρ	.	.	.	0.762	.	0.859

Note: Significance is indicated by: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. In columns 1-4, the dependent variable is the log-time spent travelling $\theta_{it} = \ln(\Theta_{it})$, for the employees with a nonzero reported time. The dependent variable is the average of travel time at time t in district s , $\bar{\theta}_{st}$ for columns (5) and (6). Estimates are obtained by OLS on pooled data (columns 1,2,3,5) or the fixed-effects, within estimator (columns 4 and 6). *Time* is just calendar time t , ranging from $t = 1$ to $t = 44$ (between November 2015 and June 2019). *Competition* is the number of competitors of the firm in district s . *Speed* is the ratio of kilometers to hours spent traveling. *Density* is the log of the density of population aged 80 and more in the county. Columns (4) and (6) report the fixed-effects within estimates where districts s are the groups; the constant here can be interpreted as the average of fixed effects; the values of σ_u and σ_e are the standard deviations of the district fixed effects and of the error term ϵ , respectively; the ρ statistic is the percentage of variance due to district effects. In column (5) the regression is weighted by \bar{n}_{st} and in column (6) it is weighted by \bar{n}_s .

D Appendix: Estimation Method; EM Algorithm

D.1 Score of a Linear Model

Let us consider the following linear model. If observation (i, t) belongs to group k , we specify

$$\ell_{it} = a_k + b_k h_{it} + c_k \bar{h}_{st} + d_k \bar{h}_{st}^2 + g_k n_{st}^{-1} + \epsilon_{itk}, \quad (19)$$

where $\ell_{itk} = \ell_{it}$, $h_{itk} = h_{it}$ for all k , etc, and where $(a_k, b_k, c_k, d_k, g_k)$ are parameters, and $s = s(i, t)$, as defined above. Let us now write the score with respect to b_k . The formula for other parameters (a_k, c_k, \dots) is very similar to $\partial \mathcal{L} / \partial b_k = 0$. To simplify notation, denote $\sum_{(i,t)} = \sum_{i=1}^n \sum_{t \in T_i}$. We have,

$$\frac{\partial \mathcal{L}}{\partial b_k} = \sum_{(i,t)} \left(\frac{p_k \Lambda_{itk}}{\sum_{j=1}^K p_j \Lambda_{itj}} \right) \left(\frac{\epsilon_{itk}}{\sigma_k} \right) h_{it} = 0, \quad (20)$$

where ϵ_{itk} is defined by equation (19). Using Bayes' rule (14), equation (20) can be equivalently rewritten,

$$\sum_{(i,t)} p_{itk} \epsilon_{itk} h_{it} = 0. \quad (21)$$

Equation (21) is just a normal equation for weighted OLS with weights p_{itk} . If we write the score for another parameter than b_k , we obtain,

$$\sum_{(i,t)} p_{itk} \epsilon_{itk} x_{it} = 0. \quad (22)$$

When x_{it} takes the values, $x_{it} = 1$, $x_{it} = \bar{h}_{it}$, $x_{it} = \bar{h}_{st}$, $x_{it} = \bar{h}_{st}^2$ and $x_{it} = 1/n_{st}$, we obtain a different normal equation for each of the model's coefficients.

To derive the estimator of σ_k^2 , we just compute $\partial \mathcal{L} / \partial \sigma_k = 0$, that is,

$$\frac{\partial \mathcal{L}}{\partial \sigma_k} = \sum_{(i,t)} p_{itk} \left(\frac{\epsilon_{itk}^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right) = 0, \quad (23)$$

and we easily derive from (23),

$$\sigma_k^2 = \frac{\sum_{(i,t)} p_{itk} \epsilon_{itk}^2}{\sum_{(i,t)} p_{itk}}, \quad (24)$$

the formula of a weighted estimator for the variance of error terms, based on squared residuals.

Finally, we write the score for the prior probabilities p_k , with $k = 2, \dots, K$, under the constraint $p_1 = 1 - p_2 - \dots - p_K$. We easily obtain, for $k > 1$,

$$\frac{\partial \mathcal{L}}{\partial p_k} = \sum_{(i,t)} \left(\frac{p_{itk}}{p_k} - \frac{p_{it1}}{p_1} \right) = 0. \quad (25)$$

Next, we multiply equation (25) by p_k for each k and sum these equations over k . We obtain, $Np_1 = \sum_{(i,t)} p_{it1}$, where $N = \sum_i \tau_i$, and $\tau_i = |T_i|$. Finally, using (25), we derive the ML estimator of p_k , that is, for all k ,

$$p_k = N^{-1} \sum_{(i,t)} p_{itk}. \quad (26)$$

D.2 Iterative Estimation Method (E-M) Based on a Weighted Regression

The above derivation of the score shows that our linear model can be estimated by an iterative method using a weighted linear regression with weights p_{itk} . The steps are as follows.

1°) Choose initial values for p_k and p_{itk} .

2°) With K types, create a database where all lines are replicated (K copies of each line (i, t)). Let $\theta_k(it\kappa)$ be the indicator of type k . We add the K indicators as control variables. Each copy (i, t, k) of (i, t) has $\theta_k(itk) = 1$ and $\theta_k(it\kappa) = 0$ for $\kappa \neq k$. We consider the interactions of indicators θ_k with the explanatory variables x_{it} .

3°) Run a weighted OLS regression with weights p_{itk} , and estimate

$$\begin{aligned} \ell_{it\kappa} = & \sum_k a_k \theta_k(it\kappa) + \sum_k b_k \theta_k(it\kappa) h_{it} + \sum_k c_k \theta_k(it\kappa) \bar{h}_{st} \\ & + \sum_k d_k \theta_k(it\kappa) \bar{h}_{st}^2 + \sum_k g_k \theta_k(it\kappa) n_{st}^{-1} + \epsilon_{it\kappa}, \end{aligned} \quad (27)$$

where $\ell_{it\kappa} = \ell_{it}$ for all $\kappa = 1, \dots, K$.

4°) Probabilities are updated with the help of the residuals ϵ_{itk} of the weighted regression (27) and the weighted estimator of σ_k^2 (given by (24)): compute the updated value of contributions to likelihood Λ_{itk} (given by (12)). Probabilities p_{itk} can then be updated using Bayes' rule (14). We then obtain the updated weights p'_{itk} . Prior probabilities p_k are also updated using (26).

5°) Iterate, going back to 3°, until the convergence criterion $\sum_k \sum_{(i,t)} (p'_{itk} - p_{itk})^2 < 10^{-m}$ is met. Store the last estimated parameters and the last estimated probabilities.

6°) Use the results as preliminary estimates to run a standard likelihood maximization routine. The ML algorithm will then yield the standard deviations of all estimated parameters, including the r_k 's, from which \hat{p}_k 's are derived using Equation (15).

E Appendix: Estimation Results of Model A

Table 10: MAXIMUM LIKELIHOOD ESTIMATION OF MODEL A, WITH $K = 1, 2, 3$

	$K = 1$	$K = 2$		$K = 3$		
	(1,1)	(2,1)	(2,2)	(3,1)	(3,2)	(3,3)
h_{it}	-0.609*** (0.074)	-0.139* (0.068)	0.487 (0.255)	1.670*** (0.099)	-1.142*** (0.327)	-2.376*** (0.152)
h_{it}^2	0.239*** (0.009)	0.147*** (0.008)	0.181*** (0.029)	-0.108*** (0.013)	0.389*** (0.038)	0.480*** (0.019)
Constant	2.484*** (0.153)	2.628*** (0.138)	-2.246*** (0.557)	0.073 (0.192)	0.435 (0.696)	5.357*** (0.287)
σ_k	1.058*** (0.003)	0.718*** (0.005)	1.270*** (0.011)	0.624*** (0.006)	1.287*** (0.016)	0.660*** (0.009)
r_k	. (.)	0 (0)	-1.101*** (0.038)	0 (0)	-0.937*** (0.065)	0.000 (0.063)
p_k	1	0.750	0.250	0.418	0.163	0.418
Log-Likelihood	-84,266.527	-79,929.198		-78,966.303		
Observations	57,130	57,130		57,130		

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model A. The dependent variable is always ℓ_{it} . Estimates have been obtained by Maximum Likelihood. Column 1 has $K = 1$ (or no group); Column (2,k) gives the coefficients of the model for type $k = 1, 2$ with $K = 2$ and Column (3,k) gives the coefficients of the model for type $k = 1, 2, 3$ when $K = 3$. The σ_k line gives the estimated values of the standard deviation.

To illustrate the predictions of Model A, we draw the L/H curves with the estimated values of parameters in the case of three types. The expression of the curve is very simple in this case. We have

$$\mathbb{E}[\exp(\kappa) | h, k] = \exp \left[\alpha_k + (\delta_k - 1)h + \zeta_k h^2 + (1/2)\sigma_k^2 \right].$$

Figure E1a displays the L/H curves for the three types and Fig. E1b gives the average L/H curve. Model A yields results that are roughly similar to Model B. In particular, the ML estimates with $K = 3$ yield a U-shaped curve (Type 3) a hump-shaped curve (Type 1) and a increasing curve with small ratios (Type 2). The weighted average of the three curves yields the expected U-shaped curve. Results obtained with $K = 2$ and $K = 4$ (see below) are similar.

Table 12 displays information, entropy and Herfindahl criteria from $K = 1$ to $K = 7$. To

compute Table 12, Model A has been estimated with added interactions. More precisely, we added interactions of types with three indicators of urban, rural and peri-urban districts (the latter are just districts that are neither rural nor urban). This version of the Model is more flexible and yields similar results. We again find that the likelihood increment $\mathcal{L}(2) - \mathcal{L}(1)$ is the largest among the $\mathcal{L}(K + 1) - \mathcal{L}(K)$. At the same time, AIC is minimized at $K = 6$, NEC is minimized at $K = 2$ and $\mathfrak{E}(K)$ is minimized at $K = 4$. Finally, Herfindahl indices are maximized at $K = 2$ (with “bumps” at $K = 4$). We have found a case in which the criteria disagree. Note that AHHI and $\mathfrak{H}(K)$ are nonmonotonic. We conclude that $K = 2$ seems to be the most reasonable choice.

Figure E1. Model A. L/H Curves Conditional on Type k and Average L/H Curve, for $K = 3$

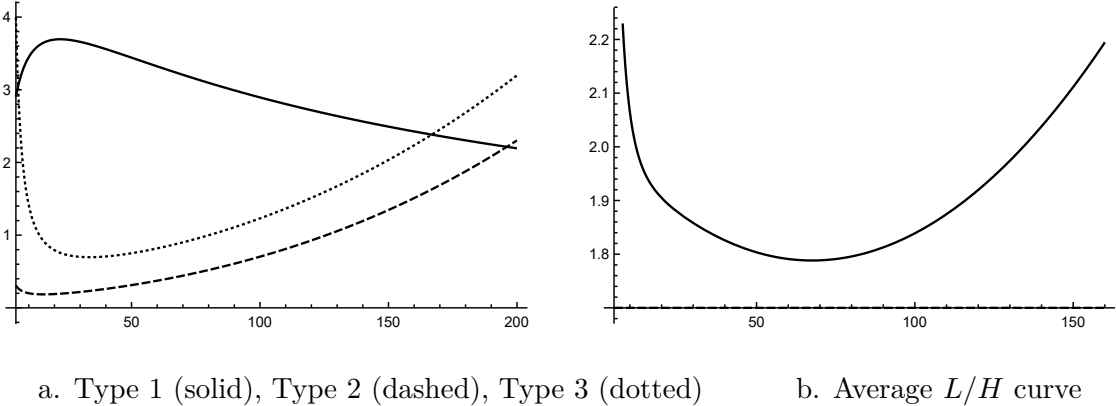


Figure E2. Model A. EM Estimation. L/H Curves Conditional on Type k and Average L/H Curve, for $K = 4$

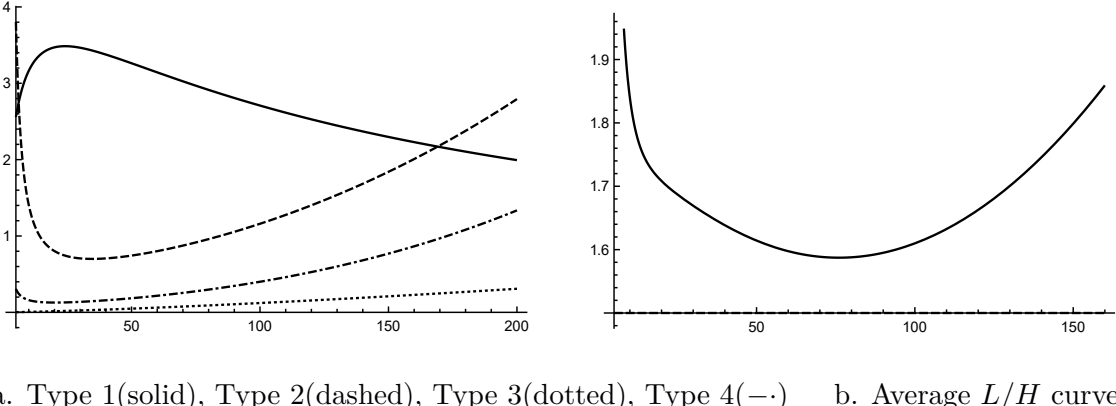


Table 11: EM ESTIMATION OF MODEL A WITH $K = 4$

(K, k)	(4,1)	(4,2)	(4,3)	(4,4)
h_{it}	1.796*** (0.040)	-2.199*** (0.044)	2.271*** (0.354)	-1.690*** (0.061)
h_{it}^2	-0.125*** (0.005)	0.451*** (0.005)	0.007 (0.041)	0.447*** (0.007)
Constant	-0.096 (0.084)	5.240*** (0.123)	-8.510*** (0.761)	1.854*** (0.153)
σ_k	0.155	0.152	0.809	0.279
p_k	0.391	0.406	0.005	0.198
Log-Likelihood	-78,901.07			
Observations	57,130			
R^2	0.621			

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model A. The dependent variable is always ℓ_{it} . Column (K, k) gives the coefficients of the model for type $k = 1, 2, 3, 4$ with $K = 4$. The σ_k line gives the estimated values of the standard deviations of error terms. Estimates have been obtained by means of an EM algorithm, described in Appendix B. The standard deviations of coefficients are those of a weighted least squares regression in which each observation (i, t) is replicated K times and the weights are the posterior probabilities p_{itk} . These standard deviations underestimate the true ones, but typically not enough to invalidate significance, due to the large number of observations. The p_k s are estimated by taking the arithmetic average of the p_{itk} summed over (i, t) for each type k . The posterior probabilities p_{itk} are estimated using Bayes' rule iteratively.

Table 12: MODEL A: INFORMATION AND ENTROPY CRITERIA. CHOICE OF K

K	1	2	3	4	5	6	7
$L(K)$	-82,150.62	-77,022.56	-76,679.12	-75,988.17	-75,342.42	-75,186.77	-75,182.32
BIC	164,410.77	154,275.14	153,708.74	152,447.33	151,276.31	151,085.50	150,364.64
AIC	164,321.24	154,087.14	153,422.28	152,062.41	150,792.94	150,503.69	150,516.84
$\mathcal{E}(K)$	0	22,231.14	42,668.87	42,829.05	60,359.18	69,871.3	76,063.41
NEC	.	4.391	8.380	6.879	8.816	10.117	10.850
$\mathfrak{E}(K)$	0	0.561	0.680	0.541	0.656	0.682	0.684
AHHI	1	0.761	0.541	0.543	0.431	0.370	0.333
$\mathfrak{H}(K)$.	0.522	0.312	0.390	0.288	0.245	0.221

Note: Model A has been estimated by the EM algorithm repeatedly with values of K ranging from $K = 1$ to $K = 7$. Model A has been estimated on the entire *employee base*, adding interactions of the K types with the indicators of three subsamples: the Urban, Peri-Urban and Rural districts (that partition the dataset). Parameters therefore vary not only with type, but also with the three types of district. $L(K)$ is the estimated Log-Likelihood with K types. BIC is the Bayesian Information Criterion. AIC is Akaike's Information Criterion. $\mathcal{E}(K)$ is entropy as defined above. NEC is Celeux and Soromenho's Normalized Entropy Criterion. $\mathfrak{E}(K)$ is just $\mathcal{E}(K)$ divided by its maximal value, $N \ln(K)$. AHHI is the Average Hirschman-Herfindahl Index defined in the text. $\mathfrak{H}(K)$ is the normalized Herfindahl index defined in the text. In the case of Model A; AIC is minimal for $K = 6$; NEC is minimal for $K = 2$; $\mathfrak{E}(K)$ is minimal for $K = 4$.

F Appendix: Robustness. Estimations of Model B in Rural and Urban Subsamples

Table 13: TRAVELED KILOMETERS
ESTIMATION OF MODEL B BY ML IN RURAL AREAS

(K, k)	(1,1)	(2,1)	(2,2)
\bar{h}_{st}	-3.484** (1.177)	-8.342*** (0.952)	7.332** (2.587)
\bar{h}_{st}^2	0.239 (0.127)	0.788*** (0.103)	-0.997*** (0.281)
h_{it}	1.465*** (0.019)	1.040*** (0.017)	2.349*** (0.053)
$1/n_{st}$	1.203*** (0.183)	2.975*** (0.330)	1.005*** (0.258)
Constant	9.383 (2.722)	22.311*** (2.204)	-18.963*** (5.974)
σ_k	1.021*** (0.003)	0.636*** (0.004)	1.175*** (0.008)
r_k	. (.)	. (.)	-0.804*** (0.031)
p_k	1	0.691	0.309
Observations	57,130		
Log-Likelihood	-82,269.24	-76,782.39	-76,782.39

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model B. The dependent variable is always ℓ_{it} . Estimates have been obtained by Maximum Likelihood, adding interactions with a dummy for urban and for rural districts. Column (1,1) has $K = 1$ (or no group); Column (2,k) gives the coefficients of the model for types $k = 1, 2$ with $K = 2$.

Table 14: TRAVELED KILOMETERS
ESTIMATION OF MODEL B BY ML IN URBAN AREAS

(K, k)	(1,1)	(2,1)	(2,2)
\bar{h}_{st}	-3.612*** (0.964)	0.753 (0.878)	-6.482** (2.478)
\bar{h}_{st}^2	0.402*** (0.105)	-0.079 (0.095)	0.718** (0.269)
h_{it}	1.513*** (0.016)	1.199*** (0.015)	2.240*** (0.051)
$1/n_{st}$	-0.542*** (0.128)	2.134*** (0.202)	-1.343*** (0.213)
Constant*	-3.671 (3.511)	-24.771*** (2.933)	27.033** (8.261)
σ_k	1.021*** (0.003)	0.636*** (0.004)	1.175*** (0.008)
r_k	. (.)	. (.)	-0.804*** (0.031)
p_k	1	0.691	0.309
Observations	57,130		
Log-Likelihood	-82,269.24	-76,782.39	-76,782.39

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model B. The dependent variable is always ℓ_{it} . Estimates have been obtained by Maximum Likelihood adding interactions with dummy variables indicating rural and urban districts. Column 1 has $K = 1$ (or no group); Column (2.k) gives the coefficients of the model for type $k = 1, 2$ with $K = 2$. (*)The rural districts are the reference; it follows that to obtain the constant of urban districts, the constant of Table 13 should be added to the constant of the present Table 14.

G Appendix: Estimation Results with K=6.

Table 15: TRAVELED KILOMETERS

ESTIMATION OF MODEL B IN RURAL DISTRICTS, BY MEANS OF THE EM ALGORITHM, WHEN $K = 6$

(K, k)	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)
\bar{h}_{st}	1.122 (0.877)	-12.126*** (0.810)	-7.906*** (0.573)	46.825*** (4.631)	8.441*** (1.361)	-3.079*** (0.589)
\bar{h}_{st}^2	-0.334*** (0.095)	1.227*** (0.088)	0.690*** (0.062)	-5.227*** (0.499)	-1.100*** (0.148)	0.184** (0.064)
h_{it}	2.477*** (0.018)	0.709** (0.021)	1.166*** (0.010)	3.421*** (0.084)	1.606*** (0.022)	1.490*** (0.010)
$1/n_{st}$	7.209*** (0.188)	-1.251*** (0.105)	12.886*** (0.233)	1.147*** (0.306)	0.820*** (0.134)	6.216*** (0.122)
Constant*	-5.169* (2.037)	37.514*** (2.766)	26.906*** (2.431)	-111.632*** (10.981)	-13.092*** (3.739)	13.570*** (2.453)
p_k	0.149	0.255	0.161	0.078	0.128	0.229
Observations	57,130					

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model B. The dependent variable is always ℓ_{it} . Estimates have been obtained by the sequential EM algorithm, adding interactions with a dummy for urban and for rural districts. There are 6 types ($K = 6$); Column (6,k) gives the coefficients of the model for types $k = 1, \dots, 6$. (*) To read the line of constants: note that type 1 in rural districts is the reference, so that the constant of column (6,1) should be added to the constants of other columns.

Table 16: TRAVELED KILOMETERS

ESTIMATION OF MODEL B IN URBAN DISTRICTS, BY MEANS OF THE EM ALGORITHM, WHEN $K = 6$

(K, k)	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)
\bar{h}_{st}	13.868*** (1.302)	10.481*** (0.923)	15.295*** (0.880)	-25.085*** (4.815)	-12.532*** (1.673)	2.453*** (0.362)
\bar{h}_{st}^2	-1.551*** (0.141)	-1.163*** (0.099)	-1.765*** (0.094)	2.683*** (0.523)	1.517*** (0.181)	-0.167*** (0.039)
h_{it}	3.218*** (0.019)	2.278*** (0.012)	1.141*** (0.009)	2.767*** (0.061)	0.345*** (0.023)	1.564*** (0.010)
$1/n_{st}$	-5.354** (0.151)	-1.400*** (0.079)	-0.518*** (0.072)	0.110 (0.212)	6.110*** (0.225)	6.213*** (0.123)
Constant*	-36.482*** (3.637)	-24.301*** (2.956)	-27.450*** (2.886)	53.848*** (11.292)	32.978*** (4.359)	-5.161* (2.202)
p_k	0.149	0.255	0.161	0.078	0.128	0.229
Observations	57,130					

Note: Significance is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Model B. The dependent variable is always ℓ_{it} . Estimates have been obtained by the sequential EM algorithm, adding interactions with a dummy for urban and for rural districts. There are 6 types ($K = 6$); Column (6,k) gives the coefficients of the model for types $k = 1, \dots, 6$. (*) To read the line of constants: note that type 1 in rural districts is the reference, so that the constant of column (6,1) in Table 15 should be added to the constants of the present Table 16.

H Appendix: The Normalized Entropy Criterion

Let the maximized Log-Likelihood be denoted

$$\hat{\mathcal{L}}(K) = \sum_{(i,t)} \ln \left[\sum_k \hat{p}_k \hat{\Lambda}_{itk} \right].$$

Using the notation introduced above and (14), we can break the maximized log-likelihood as follows,

$$\hat{\mathcal{L}}(K) = \hat{\mathcal{C}}(K) + \hat{\mathcal{E}}(K), \quad (28)$$

where $\hat{\mathcal{E}}(K)$ is entropy, as defined by (16), but computed with the estimated values \hat{p}_{itk} , and

$$\hat{\mathcal{C}}(K) = \sum_{(i,t)} \sum_{k=1}^K \hat{p}_{itk} \ln \left(\hat{p}_k \hat{\Lambda}_{itk} \right). \quad (29)$$

When sorting is perfect, that is, if for each (i, t) , there exists $k = k(i, t)$ such that $p_{itk} = 1$, then entropy is nil and $\hat{\mathcal{L}}(K) = \hat{\mathcal{C}}(K)$. Now, given that $\mathcal{E}(1) = 0$, we have $\hat{\mathcal{L}}(1) = \hat{\mathcal{C}}(1)$. Given that likelihood increases with K , we can write, for $K > 1$,

$$1 = \frac{\hat{\mathcal{C}}(K) - \hat{\mathcal{C}}(1)}{\hat{\mathcal{L}}(K) - \hat{\mathcal{L}}(1)} + \frac{\hat{\mathcal{E}}(K)}{\hat{\mathcal{L}}(K) - \hat{\mathcal{L}}(1)}, \quad (30)$$

and we define NEC as $\hat{\mathcal{E}}(K)/(\hat{\mathcal{L}}(K) - \hat{\mathcal{L}}(1))$. The criterion must be minimized. Note that NEC may be greater than 1 since the first term on the right-hand side of (30) may be negative. See Celeux and Soromenho (1996) for details.