A Guide to Estimating Matching Functions in Spatial Models

Giulia Brancaccio*, Myrto Kalouptsidi†, Theodore Papageorgiou‡

Cornell University, Harvard University, CEPR and NBER, McGill University

March 2019

We provide a guide to estimating matching functions in a spatial context. Several interactions in space take place in a decentralized fashion, such as passengers searching for taxis, ships meeting cargo, exporters meeting importers etc. A convenient modeling device to capture these meetings is the matching function, which has been used extensively in labor market settings. However in the spatial context, data availability is often limited to only one side of the market; for instance it is usually hard to find data on the number of passengers searching for a taxi. We discuss an approach to estimating matching functions that allows the researcher to recover the unobserved side of the market with relatively few assumptions. In addition, our approach obtains the matching function non-parametrically, allowing for significantly more flexibility than is commonly assumed. This additional flexibility can be key when deriving welfare and policy implications.

Keywords: matching function, spatial models, transportation, nonparametrics, search frictions, taxis, shipping, geography

*Cornell University, Department of Economics, 464 Uris Hall, Ithaca NY 14850, USA, giulia.brancaccio@cornell.edu

†Harvard University, Department of Economics, Littauer Center 124, Cambridge MA 02138, USA, myrto@fas.harvard.edu

‡McGill University, Department of Economics, 855 Sherbrooke St West, Room 519, Montreal, QC, H3A 2T7, Canada, theodore.papageorgiou@mcgill.ca

1 Introduction

In recent years there has been an increasing focus on the process that governs how agents match with each other in spatial settings. For instance, how do taxis meet passengers? How does a bulk ship meet a cargo owner? How do exporters meet importers? A nascent literature on transportation (Buchholz (2018); Frechette et al. (2018); Brancaccio et al. (2018)- henceforth, BKP) have modeled this process using a matching function, which is a reduced-form approach of modeling who matches with whom in a decentralized market. This function, which has been used extensively in labor market models to study employment outcomes, takes as inputs the stocks of searching agents on either side of the market and dictates the number of matches to be formed. The matching function captures the nature of frictions in our models.

In this paper, borrowing from our previous work, we discuss an estimation method for matching functions in spatial models. In this context, it is often the case that data is very rich on the supply side, but scant on the demand side. For instance, Buchholz (2018) and Frechette et al. (2018) employ rich information on taxi rides, but do not observe hailing passengers. BKP observe rich ship movements, but do not observe searching exporters. Similarly, Eaton et al. (2016) employ data on matches between importers and exporters, but the set of potential importers is not known with certainty.[1]

A second issue with the empirical matching function is the need for flexibility. Indeed, the matching function captures the nature of search frictions in the market in a parsimonious fashion; as such, it may capture a host of different features, such as "information imperfections about potential trading partners, heterogeneities, the absence of perfect insurance markets, slow mobility, congestion from large numbers, and other similar factors" (Petrongolo and Pissarides, 2001). Therefore, a flexible functional form is of utmost importance, as the shape of the matching function is key for welfare and policy implications as shown in Hosios (1990) and more recently in Brancaccio et al. (2019). In addition, the results of any quantitative exploration, such as the ones mentioned above, depend crucially on the matching function specification.

Our approach deals with both having data on only one side of the market, as well as allowing for flexibility. We draw from the literature on nonparametric identification (Matzkin, 2003) and non-separable instrumental variable techniques (e.g. Imbens and Newey, 2009). Roughly, the method leverages (i) an

[1] Arguably, even in labor markets observed information is imperfect; see Lange and Papageorgiou (2018). In all applications discussed, this issue is somewhat reminiscent of the lack of information of potential entrants. In empirical entry models, as only actual entry is observed and one does not observe potential entrants (that did not enter), it is difficult to form estimates of entry probabilities.

invertibility assumption between matches and sellers, (ii) the observed relationship between matches and sellers, (iii) an instrument that shifts the number of sellers, and (iv) a restriction on the matching function that allows us to disentangle monotonic transformations.

In this note, we present the basic estimation approach. Then, we elaborate on a number of practical issues, such as the different restrictions that can be imposed for identification, the instruments and the possibility of parametric restrictions. Finally, we discuss the estimation method in the context of several empirical applications, including taxis, shipping, bike-sharing and importer-exporter matching.

2 Matching Function Estimation

2.1 Setup and Available Data

There are $l = 1, ..., L$ locations or markets. Each location has b_l buyers and s_l sellers. These can be passengers and taxicabs, exporters and ships, customers and trucks, commuters and bikes, importers and exporters, etc. Within each location, the agents in the two sides of the market meet with the goal to contract for a given service (e.g. a ride). The resulting matches from this interaction, m_l, are given by the function,

$$m_{lt} = m_l \left(s_{lt}, b_{lt} \right).$$

The function $m_l(\cdot)$ is continuous and strictly increasing in both s and b; more sellers (buyers) in the market implies that more matches are realized.

Assumption 1. *The function $m_l(\cdot)$ is continuous and strictly increasing in both its arguments.*

The matching function captures the process through which buyers meet and contract with sellers; its focus is on the impact of the mass of each side of the market on the final count of trades.[2] We allow both the distribution of buyers and sellers, as well as the matching function to vary across l; this implies that the geography of different locations affects the interactions between buyers and sellers. Consistent with this, the estimation leverages time series data on the number of matches and sellers within each location. Alternatively, one can assume that the matching function is the same in all locations and exploit the cross-sectional variation across regions, possibly in addition to the time-series one.

Suppose that we observe a time-series dataset for each location l. We focus on the case where the data consists of m_{lt}, s_{lt} for $l = 1, ..., L$ and $t = 1, ..., T_l$, where T_l is the sample size for market l. For instance,

[2] It is possible to account for heterogeneous buyers and sellers by allowing multiple types to enter the matching function.

in BKP, we observe the number of empty ships searching each week t in geographical region l (s_{lt}), as well as the number of ships initiating a loaded trip from each region l in week t (m_{lt}). Geographical regions are sets of ports. In the taxi case, we might observe the number of rides starting at a NYC block (m_{lt}) as well as the number of searching cabs on that block (s_{lt}); we return to this below. Naturally, the procedure can be readily applied to the case where the econometrician observes the number of matches m_{lt} and buyers b_{lt}, but not the number of sellers, s_{lt}.

We want to estimate two sets of unknowns: (i) the functions $m_l(\cdot)$, for all l; (ii) the number of buyers b_{lt} for all l, t. The difficulty here is precisely that we want to estimate *both* unknowns. Indeed, if we had data on the buyers b_{lt}, it would be trivial to obtain the matching function via a flexible approximation. Similarly, if the matching function were known and as it is strictly increasing in both arguments, we could invert it point-wise to obtain the buyers, so that $b_{lt} = m^{-1}(s_{lt}, m_{lt})$.

2.2 Estimation Procedure

We next drop the subscript l for notational convenience. To obtain both unknowns, we borrow from the literature of nonparametric identification and in particular, Matzkin (2003). Assume for now that s_t and b_t are independent,

Assumption 2. *The number of sellers and the number buyers, s_t and b_t, are independent across t.*

We denote the distribution of the number of match realizations over time, conditional on the number of sellers, by $F_{m|s} = \Pr(m_{lt} \leq m | s_{lt} = s)$. Note that, $F_{m|s}$ is observable given the available data and can be estimated with the frequency estimator:

$$\hat{F}_{m|s=s_t}(m_t | s = s_t) = \Pr(m \leq m_t | s = s_t) = \frac{\Pr(m \leq m_t, s = s_t)}{\Pr(s = s_t)} = \frac{\#1\{(m \leq m_t, s = s_t)\}}{\#1\{(s = s_t)\}}$$

where $1\{\cdot\}$ denotes the indicator function and $\#$ denotes the number of observations. To improve efficiency and obtain out of sample values, in practice one may want to use a kernel density estimator (see for instance Fan and Gijbels, 1996 or Pagan and Ullah, 1999).

Let F_b denote the distribution of the number of buyers b_t. At a given point $\{s_t, b_t, m_t\}$ we have:

$$
\begin{aligned}
F_{m|s=s_t}\left(m_t|s=s_t\right) &= \Pr\left(m\left(s,b\right)\leq m_t|s=s_t\right) \\
\text{monotonicity} &= \Pr\left(b\leq m^{-1}\left(s,m_t\right)|s=s_t\right) \\
\text{independence} &= \Pr\left(b\leq m^{-1}\left(s_t,m_t\right)\right) \\
&= F_b\left(b_t\right) \qquad (1)
\end{aligned}
$$

From this relationship the difficulty of recovering both the function $m(\cdot)$ and the buyers becomes clear. Consider the equation:

$$F_{m|s=s_t}\left(m_t|s=s_t\right)=F_b\left(m^{-1}\left(s_t,m_t\right)\right)$$

The left-hand side is known; however, the right-hand side confounds the distribution of buyers F_b and the matching function. Indeed, monotonic transformations of the two are observationally equivalent and we cannot identify both without some restriction (for more details, see Matzkin, 2003).

For ease of exposition, suppose we were willing to assume that the distribution of buyers, F_b, is uniform on [0,1], so that $F_b\left(b_t\right)=b_t$. Then, equation (1) becomes:

$$F_{m|s=s_t}\left(m_t|s=s_t\right)=F_b\left(b_t\right)=b_t$$

and since the left-hand-side is estimated as above, we can recover b_t point-wise. Once b_t is known, we can obtain the matching function directly. Hence, knowledge of the distribution of the unobserved buyers is sufficient to identify both the buyers and the matching function.[3]

The uniform assumption is likely not a good one in our case, because the scale of buyers b_t is important: for instance we require that $b_t\geq m_t$, as there cannot be more matches than there are buyers. Indeed, the matching function must satisfy the following property:

Condition 1. The matching function $m(s,b)$ is bounded above by the minimum of buyers and sellers, i.e.

$$m=m\left(s,b\right)\leq\min\left\{s,b\right\}.$$

What are thus some reasonable restrictions to impose? First, we can consider different distributional

[3] Bajari and Benkard (2005) have applied this technique in the case of demand for PCs. In their case, the unknown function of interest is a hedonic price function and the unobservable is an unobserved product characteristic, such as quality.

assumptions for the buyers.

Assumption 3. *The distribution of buyers is known (possibly up to parameters).*

For instance, we might be willing to assume that F_b is **Poisson** with mean λ_b. In that case, we can interpret the number of buyers b_t as the number of arrivals of buyers at the location every time period (this is the assumption that Buchholz (2018) makes as well when modeling taxi passengers). In that case, (1) becomes:

$$F_{m|s}\left(m|s\right)=F_b\left(b\right)=\exp\left(-\lambda_b\right)\sum_{k=1}^{b}\frac{\lambda_b^k}{k!} \tag{2}$$

If λ_b were known, we could solve this equation for b_t, for all t. However, as in the standard uniform case, the parameter λ_b affects the scale of the estimates of the distribution $\hat{F}_b$ and thus one might not be willing to calibrate it a priori. For instance, one can use Condition 1, which imposes further structure on the matching function, to pin down λ_b. Indeed, one might want to select λ_b to minimize the magnitude of the frictions captured by the matching function m, as measured by $m\left(s,b\right)-\min\left\{s,b\right\}$, in order to not overestimate welfare loss from search frictions.[4]

Instead of distributional assumptions as in Assumption 3, we can instead restrict the matching function. Suppose, for instance, that we are willing to assume that the matching function features **constant returns to scale**. This assumption is commonly used in labor market settings, consistent with the empirical results in that context (see for instance the summary of the estimates in Petrongolo and Pissarides, 2001). Therefore a constant returns to scale matching function is probably a natural starting point here as well, especially if we believe the nature of frictions in the spatial context is similar to those in labor markets (such as information frictions regarding the location of potential buyers and sellers).[5]

Assumption 3′. The matching function exhibits constant returns to scale (CRS), so that $m\left(as,ab\right)=am\left(s,b\right)$ for all $a\geq 0$. Moreover there is a known point $\left\{\bar{m},\bar{s},\bar{b}\right\}$ such that $\bar{m}=m\left(\bar{s},\bar{b}\right)$.

Consider again the main equation (1) at the known point $\left\{\bar{m},\bar{s},\bar{b}\right\}$. Letting $a=b/\bar{b}$, *for all* b,

$$\begin{aligned}F_b\left(a\bar{b}\right)&=F_{m|s=a\bar{s}}\left(m\left(a\bar{s},a\bar{b}\right)|s=a\bar{s}\right)\\&=F_{m|s=a\bar{s}}\left(a\bar{m}|s=a\bar{s}\right).\end{aligned} \tag{3}$$

---

[4]In this case, we choose λ_b to minimize the absolute difference $\sum_{t=1}^{T}\left|m\left(s_t,b_t\right)-\min\left\{s_t,b_t\right\}\right|$ between the potential and realized trades, subject to the constraint that $m\left(s_t,b_t\right)\leq\min\left\{s_t,b_t\right\}$ for all t.

[5]The search literature often regards search frictions as a shortcut for the costly acquisition of information regarding the location of buyers and sellers (see e.g. Mortensen, 1978 and Pissarides, 2000). This interpretation naturally carries forward in the spatial context as well.

We use (3) and vary a to trace out $\hat{F}_b(b)$ for all b, relying on a kernel density estimator for the conditional distribution $\hat{F}_{m|s=a\bar{s}}(a\bar{m}|s=a\bar{s})$ as discussed above. Once the distribution $\hat{F}_b$ is recovered, we obtain the number of buyers b_t from

$$b_t = \hat{F}_b^{-1}\left(\hat{F}_{m|s=s_t}(m_t|s=s_t)\right),$$

and the matching function at *any* point (s,b) from

$$m(s,b) = \hat{F}_{m|s}^{-1}(F_b(b)).$$

The remaining question is how to choose the known point, $\left\{\bar{m},\bar{s},\bar{b}\right\}$. One option is to choose this point to be of the form $1 = m(\bar{s},1)$, so that one buyer is always matched when there are $\bar{s}$ sellers. In order to pin down $\bar{s}$ we can again make use of Condition 1. In particular, we set $\bar{s}$, to be the lowest value such that $m_t \leq b_t$, for all t.

The intuition behind the identification argument is as follows: the observed correlation between s and m informs us on $\partial m(s,b)/\partial s$, since the sensitivity of matches to changes in sellers is observed and s is independent of b by assumption; then, due to homogeneity, this derivative also delivers the derivative $\partial m(s,b)/\partial b$; and once these derivatives are known, integration leads to the matching function, which can be inverted to provide the number of buyers.

In closing, it is worth noting that the estimation can be readily applied to the case of decreasing or increasing returns to scale, with a known degree of homogeneity, $k \neq 1$.[6]

Instruments

Assumption 2 requires that s,b are independent. This assumption may be violated in spatial models where sellers s and buyers b are determined jointly in equilibrium. In this section we show how relax this assumption using an instrument z that shifts the number of sellers exogenously.[7] For instance, in BKP, we use unpredictable weather conditions, in particular ocean wind speed, as an instrument. The intuition

[6] One needs to be cautious of the data range, as homogeneity of degree $k \neq 1$ cannot always be reconciled with the fundamental requirement of Condition 1. Indeed, consider $k > 1$. We have that:

$$m(as,ab) = a^k m(s,b) \leq \min\{as,ab\} = a\min\{s,b\}$$

for all $a > 0$. Therefore, $a^{k-1}m(s,b) \leq \min\{s,b\}$ or $\log a \leq \frac{1}{(k-1)}\log\left(\frac{\min\{s,b\}}{m(s,b)}\right)$, where $k-1 > 0$ and $\log\left(\frac{\min\{s,b\}}{m(s,b)}\right) > 0$ since $\frac{\min\{s,b\}}{m(s,b)} > 1$. Homogeneity requires that the property hold for all $a > 0$, but here we see that a must be constrained for the property $m(s,b) \leq \min\{s,b\}$ to hold. Similar arguments can show that the same problem arises for $k \leq 0$.

[7] Here, we employ non-separable instrumental variables techniques (e.g. Chesher, 2003, Chernozhukov and Hansen, 2005, Chernozhukov et al., 2007 and Imbens and Newey, 2009; for an application of these techniques similar to ours, see Bajari and Benkard, 2005).

for the instrument is that wind affects the speed at which ships travel and therefore exogenously shifts the supply of ships at port. Similarly, in the taxi case, one might exploit the shift change or congestion in distant locations within the city; we return to this below.

Assumption 2′. An instrument z exists, such that $s = h(z, \eta)$ and z is independent of η and b.

Under Assumption 2′, the endogeneity between b and s is driven by the correlation between the shock η and the number of buyers b. In other words, the number of sellers s is independent of the number of buyers b, conditional on the shock η. We exploit this conditional independence to derive a version of equation (1) that does not depend on the independence assumption. In particular,

$$\begin{aligned}
F_{m|s=s_t}(m_t|s = s_t, \eta_t) &= \Pr(m(s,b) \leq m_t | s = s_t, \eta_t) \\
\text{monotonicity} &= \Pr\left(b \leq m^{-1}(s, m_t) | s = s_t, \eta_t\right) \\
\text{cond. independence} &= \Pr\left(b \leq m^{-1}(s_t, m_t) | \eta_t\right) \\
&= F_b(b_t|\eta_t) \qquad (4)
\end{aligned}$$

Equation (4) can be used to recover the unobservables in two steps. In the first step, the relation $s = h(z, \eta)$ is used to recover the realizations of η, $\{\hat{\eta}_t\}_{t=1}^T$. In practice this can be done regressing flexibly the number of sellers s on the instrument, z, and setting $\{\hat{\eta}_t\}_{t=1}^T$ equal to the regression residuals. In the second step, similarly to above, one can recover the unknowns of interest by integrating over η both sides of equation (4).[8]

Parametric Matching Function

Although the nonparametric nature of this procedure has important benefits as already discussed, one may in fact prefer a parametric matching function despite it being more restrictive. This could be either due to some prior knowledge of the form of the matching process, or due to the need to extrapolate substantially in policy analysis, or because of the analytical convenience of the parametric form when used in a structural setting.

We briefly discuss how one might recover both the number of buyers and the matching function parameters in the case of a **Cobb-Douglas matching function** with constant returns to scale. This

[8] It is clear from equation (4) that the procedure does not require integrating out the error term η. However, this may be the best approach for efficiency.

specification is the most commonly used functional form assumption in the matching function estimation literature (Petrongolo and Pissarides, 2001, Elsby et al., 2015).

In particular, let

$$m_{lt} = A_l s_{lt}^{\alpha_l} b_{lt}^{1-\alpha_l}$$

where A_l is a parameter capturing the matching efficiency in region l, while α_l and $1-\alpha_l$ are the elasticities of matches with respect to sellers and buyers respectively. Then we rewrite

$$\begin{aligned} \log(m_{lt}) &= \log(A_l) + (1-\alpha_l)\log b_{lt} + \alpha_l \log s_{lt} \qquad (5) \\ &= \alpha_{l0} + \epsilon_{lt} + \alpha_l \log s_{lt}. \end{aligned}$$

Using an instrument for s_{lt}, we can immediately recover α_l using the above specification. In order to recover the number of buyers as well, b_{lt}, note that since,

$$\alpha_{l0} + \epsilon_{lt} = \log(A_l) + (1-\alpha_l)\log b_{lt},$$

in order to separately identify b_{lt}, from the matching efficiency, A_l, an additional assumption is necessary.[9] Moreover, the Cobb-Douglas specification nicely illustrates that it is not possible to separately identify b_{lt} from the degree of returns to scale α_{l0} of the matching function (see equation (5)).

3 Applications

Shipping:

In BKP we consider the role of transportation costs in world trade. In particular, we divide world ports into 15 regions, which correspond to the L locations. In our setup, meetings between empty ships and

[9] If we normalize $\log(A_l) = K_l$, we can back out the number of buyers:

$$\frac{\alpha_{l0} + \epsilon_{lt} - K_l}{1-\alpha_l} = \log b_{lt}.$$

For instance, similar to the non-parametric specification with CRS discussed above, we can set K_l such that $\log(b_{lt}) > \log(m_{lt})$, or $K_l \leq -(1-\alpha_l)\log(m_{lt}) + \alpha_{l0} + \epsilon_{lt}$. Note that this assumption is equivalent to assuming that the lowest number of buyers necessary to form one match when there is one seller, is

$$\log\left(\bar{b}_l\right) = -\frac{K_l}{1-\alpha_l}.$$

potential exporters are governed by a region-specific matching function. Using satellite AIS (Automatic Identification System) data from exactEarth Ltd we have information on ships' position (longitude and latitude), speed and level of draft (the vertical distance between the waterline and bottom of the ship's hull) at intervals of at most six minutes (see also Adland and Jia, 2016). The draft variable is particular important for our purposes, as it allows us to identify whether ships are loaded or not at any point in time.

Following the notation introduced above, in each region we effectively observe the number of matches, m (loaded ships) and the number of sellers, s (empty ships looking for cargo) and we are looking to recover the region-specific matching function, $m(.)$ and the number of buyers, b (potential exporters). In our baseline specification we assume a CRS matching function, so we do not need to make any assumptions on the distribution of buyers or assume a specific functional form for the matching function. In this setup assuming independence between buyers and sellers (ships and exporters) is not plausible, as the economic forces that affect the number of ships in a region likely also affect the number of exporters. We therefore follow the instrumental variable approach discussed above. In particular, we use unpredictable sea weather shocks that shift the arrival of ships at a port, but do not affect exporters at land.[10] As a robustness, we consider the case where instead of imposing CRS on the matching function we make a functional form assumption on the distribution of buyers (exporters). In particular, we assume that the number of exporters is distributed according to a Poisson distribution. Our results are fairly similar and not surprisingly, the implied returns to scale on the matching function is roughly equal to one, even though we do not impose CRS.

In addition, in Brancaccio et al. (2019) we examine the (constrained) efficiency of decentralized transportation markets and apply our results to the shipping environment described here. We find that allowing for a flexible matching function is key in accurately recovering the welfare implications of search frictions and correctly evaluating optimal policies.

Taxis:

In recent work (e.g. Buchholz, 2018 and Frechette et al., 2018) matching functions have been estimated in the context of NYC taxicabs. The city blocks are divided into a number of different regions corresponding

[10]We divide the sea surrounding each region into zones; for each zone we use information on the wind speed at different distances from the coast and in different directions. To obtain the unpredictable component of weather we run a VAR regression of these weather indicators. The results are robust to the lag structure, as well as estimating jointly for neighboring zones.

to the L locations. Again, each location may have its own matching function.[11] The data (from the Taxi and Limousine Commission) contain every taxi ride realized in the city, including the origin, destination and time of the day. In other words, the data provides the number of matches m_{lt}, but it does not provide a direct measure of the searching taxicabs, s_{lt} (unlike BKP where the AIS data provide both full and empty ships). It is however possible to create a measure of searching taxicabs under certain assumptions.

In particular, we propose the following measure for s_{lt}. Since the entire history of rides is observed, we observe the time between each drop-off and the next pick-up for a given taxi. Therefore we can build a conservative measure of the waiting times using the lag between consecutive drop-offs and pick-ups in the same region. Let p_{lt} denote the match probability for cabs. Then, the wait time is equal to $1/p_{lt}$ and is observed in our data, modulo some assumptions about what drivers do between consecutive trips.[12] Note however that $p_{lt} = m_{lt}/s_{lt}$ and thus we can recover the number of searching cabs from $s_{lt} = m_{lt}/p_{lt}$.

Finally, as the distribution of taxicabs and passengers over space is determined endogenously in equilibrium, the independence assumption between s_{lt} and b_{lt} is likely not valid in this setup and an instrument is required to estimate the matching function. Again, a supply side shifter will suffice; for instance, Frechette et al. (2018) use the drivers' shift change, as well as the lagged number of active cabs.[13] Another potential instrument may involve traffic conditions in distant regions, that affect the supply of taxis in a given region.

Bike-sharing schemes:

Users of bike sharing schemes, which have become quite popular in recent years, also face search frictions: a bike may be available nearby or available bikes may be located elsewhere, but that information is not often readily available. As in the previous applications, it is convenient to capture the matching process using a matching function (see for instance Cao et al., 2018). The available data in this case usually includes when a user finds a bike (matches, m_t), as well as the number of available, unused bikes (s_t), but does not include users that are searching for a bike (b_t). Given the data availability, the procedure outlined above is therefore well-suited to both obtain the unknown side of the market (users searching

[11]For instance, this is the setup in Buchholz (2018). Frechette et al. (2018) handle the matching process somewhat differently. In particular, they assume a specific city-wide matching function which is given by a micro-simulation of driver and passenger behavior; in other words, they assume the matching function is known and recover only the hailing passengers.

[12]For instance, the researcher has to take a stance on whether a long interval between two consecutive trips is search time or a coffee break. However, this assumption is required anyway in the estimation of these models (e.g. Frechette et al. (2018) make such assumptions when modeling search/wait times; Buchholz (2018) also makes such assumptions when modeling taxicab search decisions).

[13]In NYC, practically all shifts change at the same time, leading to a "witching hour" around 5pm with few cabs and many searching passengers. Frechette et al. (2018) argue that the shift change is timed so that the day shift and the night shift have similar returns on average.

for bikes), as well as flexibly recover the matching process. Similarly to other applications, an instrument is needed. As before, conditions in other markets affects the availability of bikes in a given location, providing a plausible instrument.

Exporters and Importers:

The trade literature has recently focused on the implications of search frictions in the formation of matches between importers and exporters (see for instance Eaton et al., 2014 and Eaton et al., 2016). As discussed above, a matching function provides a tractable representation of this process (see for instance Krolikowski and McCallum, 2018). The available data often contains the resulting matches, as well as the identity of the exporter and importer of the match, but does not provide information on who is currently (still) searching. If we were to make additional assumptions on one side of the market, such as for instance that existing exporters are always searching for new importers, then the above procedure can be used to flexibly recover the matching function, as well as the number of searching importers.

4 Conclusion

We outline an estimation approach to matching functions in the context of spatial interactions. This approach is well-suited for several such markets and we discuss a number of potential applications, including taxis, shipping, bike-sharing and import-exporter matching. In addition, we explore several practical issues. In particular, we propose potential instruments in the context of the above applications. Moreover, in the popular case of taxis, where idle taxis are not directly observed, we propose an approximation to the number of searching taxis using the taxi wait time. It should be noted that this list of examples is not exhaustive, and other applications are possible, such as the matching of shipments to trucks.

References

Adland, R. and H. Jia (2016): "Vessel Speed Analytics Using Satellite-Based Ship Position Data," *Proceedings of the 2016 IEEE International Conference on Industrial Engineering and Engineering Management*, 1299–1303.

Bajari, P. and C. L. Benkard (2005): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113, 1239–1276.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2018): "Geography, Search Frictions and Endogenous Trade Costs," *mimeo, Harvard University.*

Brancaccio, G., M. Kalouptsidi, T. Papageorgiou, and N. Rosaia (2019): "Efficiency in Decentralized Transport Markets," *mimeo, Harvard University.*

Buchholz, N. (2018): "Spatial equilibrium, search frictions and efficient regulation in the taxi industry," *mimeo, Princeton University.*

Cao, G., G. Z. Jin, X. Weng, and L.-A. Zhou (2018): "Market Expanding or Market Stealing? Competition with Network Effects in Bike-Sharing," *NBER working paper*, 24938.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Eaton, J., M. Eslava, C. Krizan, M. Kugler, and J. Tybout (2014): "A Search and Learning Model of Export Dynamics," *mimeo, Penn State University.*

Eaton, J., D. Jinkins, J. Tybout, and D. Xu (2016): "Two-sided Search in International Markets," *mimeo, Penn State University.*

Elsby, M. W. L., R. Michaels, and D. Ratner (2015): "The Beveridge Curve: A Survey," *Journal of Economic Literature*, 53, 571–630.

Fan, J. and I. Gijbels (1996): "Local Polynomial Modelling and Its Applications," in *Monographs on Statistics and Applied Probability (Book 66)*, Chapman and Hall.

Frechette, G. R., A. Lizzeri, and T. Salz (2018): "Frictions in a Competitive, Regulated Market Evidence from Taxis," *forthcoming, American Economic Review.*

Hosios, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *The Review of Economic Studies*, 57, 279–298.

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Krolikowski, P. M. and A. H. McCallum (2018): "The Role of Trade Costs in the Surge of Trade Imbalances," *Federal Reserve Bank of Cleveland, Working Paper no. 16-35R.*

Lange, F. and T. Papageorgiou (2018): "Matching Function Estimation Under Unobserved and Endogenous Search Efficacy," *mimeo, McGill University.*

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

Mortensen, D. T. (1978): "Specific Capital and Labor Turnover," *The Bell Journal of Economics*, 9, 572–586.

Pagan, A. and A. Ullah (1999): *Nonparametric Econometrics*, Cambridge, UK: Cambridge University Press.

Petrongolo, B. and C. A. Pissarides (2001): "Looking Into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39, 390–431.

Pissarides, C. A. (2000): *Equilibrium Unemployment Theory*, Cambridge, MA: MIT Press, 2nd ed.