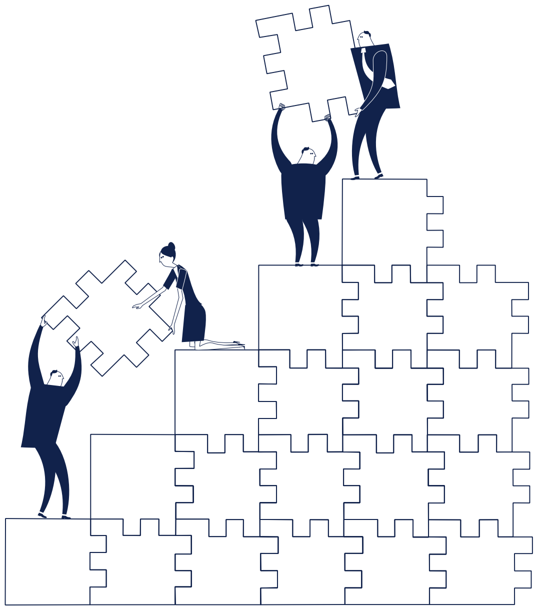
Saïd Business School
Research Papers

July 2015

Labor Flows and the Aggregate Matching Function: A Network-Based Test Using Employer-Employee Matched Records

Omar Guerrero
Saïd Business School, University of Oxford

Eduardo Lopez
Saïd Business School, University of Oxford



Saïd Business School RP 2015-12

Labor flows and the aggregate matching function: a network-based test using employer-employee matched records

Omar A. Guerrero[1,2] and Eduardo López[2,1]

[1]Institute for New Economic Thinking at the Oxford Martin School

[2]CABDyN Complexity Center, Saïd Business School, University of Oxford

Abstract

The assumption of aggregate matching functions in labor markets is tested using a network configuration model for directed multigraphs. We use employer-employee matched records of the universe of employees and firms in Finland and find that aggregate matching functions, even at the level of submarkets, cannot explain the vast majority of the observed patterns of labor flows between firms. Our findings suggest the need for theoretical frameworks that take into account the structure of labor market frictions.

1 Introduction

The paradigm of the aggregate matching function (AMF) is one of the most common assumptions in economics (Mortensen and Pissarides, 1994; Petrongolo and Pissarides, 2001). It provides a reduced way to account for labor market frictions and it naturally integrates into equilibrium models. Despite its parsimony, the assumption of an AMF has been criticized for its lack of micro-foundations and its inability to account for imbalances in labor markets (among other criticisms). On one hand, some studies have addressed the lack of micro-foundations by proposing probabilistic mechanisms to describe the matching process (Montgomery, 1991; Calvó-Armengol, 2004; Stevens, 2007). On the other, labor market imbalances have been studied by partitioning the economy into submarkets, and assigning local AMFs to each one (Shimer, 2007; Sahin et al., 2014). This paper focuses on labor flows across firms, and the role that the AMF plays in explaining them even when considering the economy as a collection of submarkets.

Studying firm-to-firm flows is central to understand labor dynamics because it provides information about who is more likely to find jobs and where. In fact, under the assumption of an AMF there is no reason to expect a significantly high volume of labor flows between two specific firms (even when controlling for their sizes and number of vacancies). We test this hypothesis using employer-employee matched micro-data of the universe of employees in Finland for 20 years. We make use of a well-established method from network science in order to provide a statistical test. Our results show that the AMF does not explain the vast majority of firm-to-firm and firm-to-submarket labor flows.

The rest of the paper is organized as follows. Section 2 introduces the configuration model, commonly used in graph theory to study the formation of random networks. We establish the connection between job search and the configuration model, and present a statistical test. In section 3, we test the hypothesis of the aggregate matching function using employee-employer matched records. Section 4 discusses the results and implications.

2 The Aggregate Matching Function and the Configuration Model

An AMF takes the form $M = f(U, V)$, where U is the number of unemployed individuals in the economy, V is the total number of vacancies, and M is the number of people who are matched to vacancies. It is usually assumed that these M matches are created with homogeneous probability (Pissarides, 2000). A second, less obvious assumption, is that any distribution of vacancies across firms is acceptable. This becomes clear in the wage dispersion literature (Pissarides, 2000), where the AMF allows firm heterogeneity.

To explain the latter point, consider the number of vacancies V_i in firm i. Under the AMF, the number of matches is $M \leq \sum_i^n V_i = V$, where n is the number of firms. In a matching process that involves an AMF, any sequence $\{V_1, V_2, ..., V_n\}$ such that $\sum_i^n V_i = V$ is permissible. The same applies to any sequence $\{M_1, M_2, ..., M_n\}$ such that $\sum_i^n M_i = M$, where M_i denotes the number of matched individuals whose last employer was firm i. We use this idea of firms with different vacancies and job seekers, together with the assumption of homogeneous matching probability, in order to introduce a new test of the AMF.

Consider 4 workers who become separated from firm A and eventually find jobs at different firms. Workers 1 and 2 find jobs in firm B, while workers 3 and 4 become employed by C and D respectively. Each of these flows can be represented as $A \rightarrow B$, $A \rightarrow C$, and $A \rightarrow D$, indicating connections from A to B, C, and D. Since two workers become employees of B, we have two instances of $A \rightarrow B$. This process can be directly mapped into a directed multi-graph $\mathbf{G} := (\mathbf{N}, \mathbf{E})$, defined as an ordered pair with a set of nodes $\mathbf{N}$ and a multiset of edges $\mathbf{E}$. In this context, firms are the nodes in $\mathbf{N}$ and their connections are edges in $\mathbf{E}$. From the previous example, firm A is said to have an out-degree of 4, B has an in-degree of 2, and C and D have in-degrees of 1. This kind of graph has been empirically studied by (Guerrero and Axtell, 2013) under the umbrella of *labor flow networks* and by Schmutte (2014) as *realized mobility networks*. In this paper we use the terms graph and network interchangeably.

Let k_i^{in} and k_i^{out} denote the in and out-degrees of firm i, which are equivalent to their number of matched vacancies V_i and job seekers M_i respectively. Then, we can think of the AMF as a generating mechanism of $\mathbf{G}$, for any degree sequences $\{k_1^{in}, k_2^{in}, ..., k_{|\mathbf{N}|}^{in}\}$ and $\{k_1^{out}, k_2^{out}, ..., k_{|\mathbf{N}|}^{out}\}$ such that $\sum_i^{|\mathbf{N}|} k_i^{in} = \sum_i^{|\mathbf{N}|} k_i^{out} = M$. By randomly matching the job seekers and vacancies of each firm, the AMF generates a random graph $\mathbf{G}$ (Jackson, 2008; Newman, 2010). This process takes into account the fact that some firms receive more workers than others. In fact, this network formation process corresponds a well established framework called the *directed configuration model* (DCM) (Bender and Canfield, 1978; Bollobás and Canfield, 1980; Molloy and Reed, 1995). From this point onwards we will use network terminology.

The DCM hypothesizes that link formation is the result of random matching between nodes. It starts with nodes that possess severed incoming and outgoing edges, also known as *stubs*. The DCM consists of randomly matching each outgoing stub to an incoming stub. When two stubs are matched, an edge is formed. Let $\mathbf{n} \subseteq \mathbf{N}$ denote a subset of nodes, and $\hat{k}_i^{in}$ and $\hat{k}_i^{out}$ the in and out-degrees of firm i generated by the DCM. For a single realization of the DCM, the number of edges from i to any of the firms in $\mathbf{n}$ is

$$\hat{k}^{out}_{i,\mathbf{n}} = \sum_{j\in\mathbf{n}} \sum_{e\in\mathbf{E}} \mathbb{I}(e = i \to j), \tag{1}$$

where $\mathbb{I}$ is an indicator function that takes value 1 when $i \to j$ is an edge of $\mathbf{G}$. If $\mathbf{G}$ has a cumulative degree distribution F, such that

$$\int_0^\infty x dF(x) = \frac{|\mathbf{E}|}{|\mathbf{N}|} = \mu < \infty, \tag{2}$$

a result by (Wilson et al., 2013) shows that, when $|\mathbf{N}| \to \infty$, the random variable

$$\hat{k}^{out}_{i,\mathbf{n}} \sim \text{Binomial}\left(k^{out}_i, \frac{1}{|\mathbf{E}|}\sum_{i\in\mathbf{n}} k^{in}_i\right), \tag{3}$$

where the first parameter corresponds to the number of outgoing stubs of i (the number of trials), and the second corresponds to the probability that an outgoing stub will be matched with an incoming stub from any node in $\mathbf{n}$.

Equation (3) provides the tools to statistically test the significance of individual edges in $\mathbf{G}$. The test consists of comparing $k^{out}_{i,\mathbf{n}}$ against the expectation $\hat{k}^{out}_{i,\mathbf{n}}$ and it takes the from of the p-value of a Binomial test

$$p(i, \mathbf{n}) = \Pr(\hat{k}^{out}_{i,\mathbf{n}} \geq k^{out}_{i,\mathbf{n}}). \tag{4}$$

We can use (4) to test the significance of an edge between a firm and a group of firms (e.g., a submarket), or between two specific firms i and j, in which case the second argument of (3) becomes $k^{in}_i/|\mathbf{E}|$. In the next section, we test both types of methods using comprehensive employer-employee matched records.

3 Testing the Aggregate Matching Function

We perform the test described in (4), by constructing labor flow networks from employer-employee matched data. By measuring the statistical significance of each edge at a 0.001 significance, we count the percentage of edges in which the AMF hypothesis is rejected. We perform this analysis for the annual firm-to-firm flows of every employee in all of Finland.

3.1 Data

We use the Finish Longitudinal Employer-Employee Data (FLEED) Finland (b), which consists of an annual panel of employer-employee matched records of the universe of firms and employees in Finland. The panel is constructed by Statistics Finland from social security records by tracking the association between each worker and each firm at the end of each calendar year. If a worker is not employed, she does not become part of the corresponding cross-section. The result is a panel of 20 years (1988 to 2008) that tracks every firm and every employed individual at the end of each year (approximately 2×10^5 firms and 2×10^6 workers).

FLEED can be merged with other data sets that provide information about companies. We use the Statistics Finland's Business Register Finland (a), which provides information about industrial classification (up to five digits) and geographical location (municipal codes of up to three digits) on an annual basis. The Business Register is built using administrative data from the Tax Administration and through direct inquiries from Statistics Finland to business with more than 20 employees.

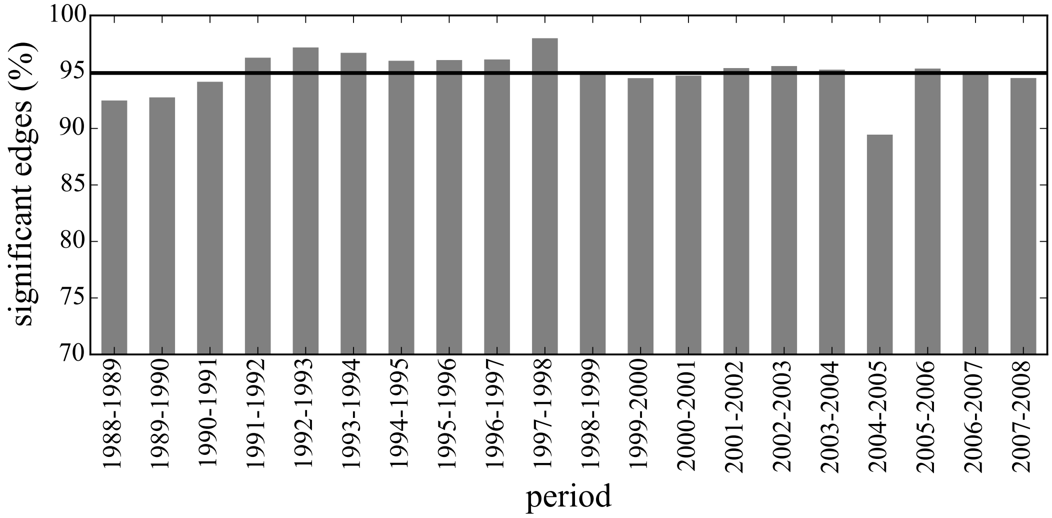
3.2 Results

3.2.1 Entire Economy

We constructed labor flow networks for each cross-section of FLEED and counted the percentage of edges that were statistically significant in each network[1]. Figure 1 shows that the number of edges that are statistically significant are more than 90% each year. This means that the number of labor flows between most pairs of firms are higher than what we would expect under the AMF hypothesis. The average number of significant edges (the horizontal line) is 95%.

[1]We checked that assumption (2) was satisfied. We fitted a Pareto distribution to each labor flow network via MLE and found that, in all cases, the estimated exponents were between 2 and 3 (both exclusive). Given the large size of the networks ($|\mathbf{N}| \approx 2 \times 10^5$) it is reasonable to use the result presented in (3) in order to test the significance of each edge.

Figure 1: Edges Not Explained by the Aggregate Matching Function



A significant edge has a p-value lower than 0.001. The horizontal line indicates the average percentage of significant edges across all years.

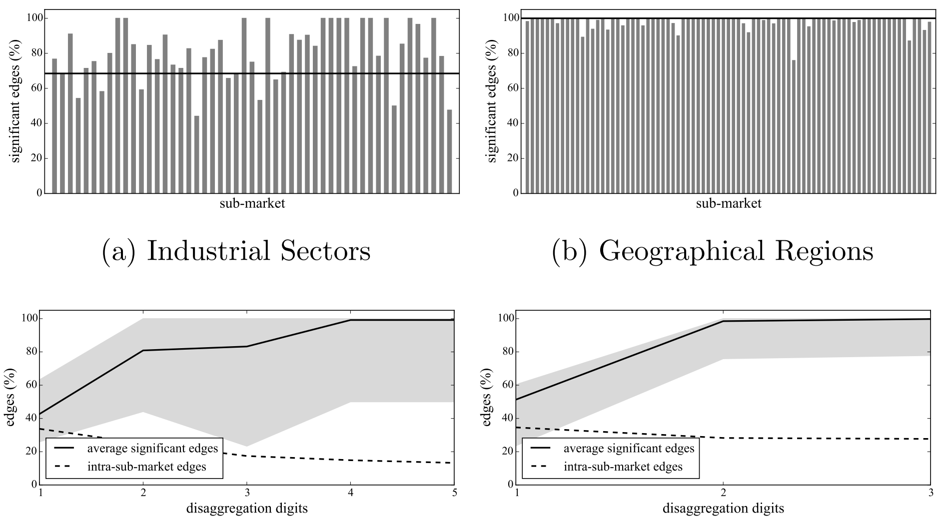
3.2.2 Submarkets

At the level of the entire economy, the AMF can only explain less than 10% of firm-to-firm labor flows. This is partly due to the low likelihood that a specific job seeker and a particular vacancy find each other. This is a common pitfall of thinking in terms of entire economies, so it is not surprising that other tests also reject the AMF (Coles and Petrongolo, 2008). However, it is commonly argued that the AMF is a reasonable assumption when looking into submarkets (Shimer, 2007). Even in that case, we show that local AMFs still do not explain most of these flows.

We merged FLEED with the Business Register and partitioned the data into industries and municipalities. In order to test the significance of an edge, we computed (3), where $\mathbf{n}$ is the submarket to be analyzed. In this way we test the significance of an edge from a firm to any other firm in its own submarket[2].

[2] A related test considers a submarket as a separate network (using only intra-submarket edges). By applying the procedure described in section 3.2.1 to each submarket, we assess the extent to which its own local AMF explains local reallocation patterns. We find that, on average, more than 80% of the edges in submarkets are significant.

Figure 2: Submarket Edges Not Explained by Local AMFs



(a) Industrial Sectors (b) Geographical Regions

(c) Industrial Disaggregation (d) Geographical Disaggregation

The plots correspond to the cross-section of 2006-2007. The analysis for other years yield similar results. The horizontal lines in panels 2a and 2b correspond to the average percentage of significant edges across submarkets. Panels 2c and 2d show the average percentage of significant edges across submarkets (solid line), the gap between the minimum and maximum percentage of significant edges (shaded region), and the percentage of edges contained in submarkets (dashed line).

Figure 2 shows the result of the analysis for a representative cross-section. Panels 2a and 2b correspond to industrial and geographical classifications respectively. They show the percentage of significant edges in each submarket. For industries, on average (horizontal line), the local AMF does not explain nearly 70% of the edges. In the best case, the AMF is rejected in 40% of the edges. The percentage of significant edges is considerably higher in municipalities, being nearly 100% on average.

The results presented in panels 2a and 2b correspond to two-digit disaggregation levels. However, more disaggregated categories can be defined. With more disaggregation, the percentage of significant edges increases because smaller submarkets have a lower $\frac{1}{|\mathbf{E}|}\sum_{i\in\mathbf{n}} k_i^{in}$. Panels 2c and 2d show this pattern. The number of intra-submarket edges as a fraction of all edges drops with disaggregation.

4 Discussion

Using comprehensive employer-employee matched micro-data we have shown that the concept of the aggregate matching function, while useful to model frictional unemployment from an aggregate point of view, is not suited to study labor reallocation. This is important because assuming an aggregate matching process ignores the structure of labor market frictions, which in turn limits our understanding about the role that individual firms play in reallocating labor. Moreover, the structure of frictions provides valuable information about the potential limits of policy when labor mobility is constrained in such ways.

Frictional unemployment is a complex problem because it involves a multiplicity of factors that are specific to firms, workers, and institutions (e.g., mobility costs, social networks, skill specificity etc.). These factors interact in ways that prevent us from disentangling the micro-foundations of how job seekers manage to find jobs. Therefore a different approach is needed in order to study frictional unemployment. Recent advances in network science provide alternatives to study this problem. In particular, we can take advantage of a network representation of frictions. For example, the framework of job search on frictionless networks, developed by Guerrero et al. (2015); López et al. (2015), offers a viable way to overcome some of the limitations of search and matching models. Further developments in this direction could provide new and insightful ways to understand the complexity of labor markets and the implication to policies.

Acknowledgement

We would like to thank Andrew Elliott, Ernesto Carrella, Felix Reed-Tsochas, and Carlos Urzúa for their valuable comments.

This work was supported under the INET at Oxford grant INET12-9001 and by the Oxford Martin School grant LC1213-006.

References

Bender, E. and Canfield, E. (1978). The Asymptotic Number of Labeled Graphs with Given Degree Sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296307.

Bollobás, E. and Canfield, E. (1980). A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1(4):311316.

Calvó-Armengol, A. (2004). Job Contact Networks. *Journal of Economic Theory*, 115(1):191–206.

Coles, M. and Petrongolo, B. (2008). A Test between Stock-Flow Matching and the Random Matching Function Approach. *International Economic Review*, 49(4):11131141.

Finland, S. Business Register.

Finland, S. Finnish Longitudinal Employer-Employee Data.

Guerrero, O. and Axtell, R. (2013). Employment Growth through Labor Flow Networks. *PLoS ONE*, 8(5):e60808.

Guerrero, O., López, E., and Axtell, R. (2015). The Network Composition of Aggregate Unemployment. *Working Paper*.

Jackson, M. (2008). *Social and Economic Networks*. Princeton University Press.

López, E., Guerrero, O., and Axtell, R. (2015). The Network Picture of Labor Flow.

Molloy, M. and Reed, B. (1995). A Critical Point for Random Graphs with a Given Degree Sequence. *Oxford Bulletin of Economics and Statistics*, 6(2-3):161180.

Montgomery, J. (1991). Equilibrium Wage Dispersion and Interindustry Wage Differentials. *The Quarterly Journal of Economics*, 106(1):163–179.

Mortensen, D. and Pissarides, C. (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, 61(3):397–415.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Petrongolo, B. and Pissarides, C. (2001). Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, 39(2):390–431.

Pissarides, C. (2000). *Equilibrium Unemployment Theory*. MIT Press.

Sahin, A., Song, J., Topa, G., and Violante, G. (2014). Mismatch Unemployment. *American Economic Review*, 104(11):3529–64.

Schmutte, I. (2014). Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility. *Labour Economics*, 29:49–61.

Shimer, R. (2007). Mismatch. *American Economic Review*, 97(4):1074–1101.

Stevens, M. (2007). New Microfoundations for the Aggregate Matching Function. *International Economic Review*, 48(3):847–868.

Wilson, J., Bhamidi, S., and Nobel, A. (2013). Measuring the Statistical Significance of Local Connections in Directed Networks. *Neural Information Processing Systems Workshop on Frontiers of Network Analysis: Methods, Models and Applications.*