

Additional Materials B: Insights from text analysis

This appendix describes the text analysis that has been performed on the “purpose of the payment field” (see Section “Dataset Description” in the paper for more information) of each wire transfer transaction. The outcomes of such experiment have been briefly presented in Section “Insights from text analysis” of the paper.

In order to dig deeper in the dynamics underlying the patterns we discussed in the paper, we analysed also the *purpose of the payment* field, that is edited by the payer when the wire transfer is settled. As an example, we extracted this information from the wire transfers settled during January 2016, and we reported in Fig. B1 the most recurrent words for each kind of sender-receiver pair. To be noted that:

- (a) for natural persons paying natural persons the most common word is “affitto”, the Italian word for rent;
- (b) for legal persons paying natural persons, the two most recurrent words are “stipendio” and “pensione”, namely salary and pension;
- (c) for natural persons paying legal persons, the most recurring words apart “gennaio” (January) are “quota”, “pagamento”, “retta” and “saldo”, meaning respectively share, payment, fee and settlement;
- (d) finally, for payments from legal persons to legal persons, the most recurrent words are “saldo” (settlement) and “fattura” (invoice). It should be noted that the *settlement* and *invoice* labels are too wide to give further insights on the reasons behind the payments. Please notice that this observations should be complemented with the analyses we described in the paper in Section “Inequalities and specialisations at a regional basis”, where we aimed to better understand interactions dynamics among legal persons using business sectors information, and also to let emerge the inner complexity of supply chains.

If we consider pairs of words instead of single words as in Fig. B2, we have that:

- (a) “rents”, “condominium fees”, and “housing expenses” are clearly the dominant reasons for a transfer from natural persons to natural persons;

- (b) it is much clearer that “salaries”, and “retirement pensions” are the most frequently reported reasons for payments from legal persons to natural persons, way ahead of “insurance settlements” and other kind of transfers;
- (c) it emerges a smoother distribution of reasons of payments from natural persons to legal persons. However, it is clear that customers use wire transfers to pay generic services after an “invoice” as been issued, as well as “membership fees”, “taxes”, “school fees”, etc.;
- (d) also the reasons of payments from legal persons to legal persons are represented by a smooth distribution of bi-grams frequency. Moreover, the general interpretation is harder than the other cases, because the information is still quite noisy. Nevertheless, it is striking that the most common bi-gram unequivocally refers to a multinational gambling company operating in Italy, whose name has not been unfold by Intesa Sanpaolo for privacy reasons. This over representation of a single company is due to the very frequent and regular payments from this company to each of the sales points of tickets every Monday, and also because the purpose of the payment stated is a fixed string for every wire transfers. The second and third most used word pairs were “saldo contabile” and “ripiamento saldo”, both referring to accounting balance operations.

We also calculated the tri-grams frequencies distributions for all the four types of interactions between account owners. However, such analysis does not add any additional insights w.r.t. what we learnt from the the previous one; moreover, it does not help to mitigate the noise we found while exploring the legal to legal persons interactions. As a consequence, we need to adopt some different analytical tools to better understand what happens when legal persons are involved on both sides of the wire transfers. This is one of the main points addressed by the bow tie analysis that is presented in the paper.

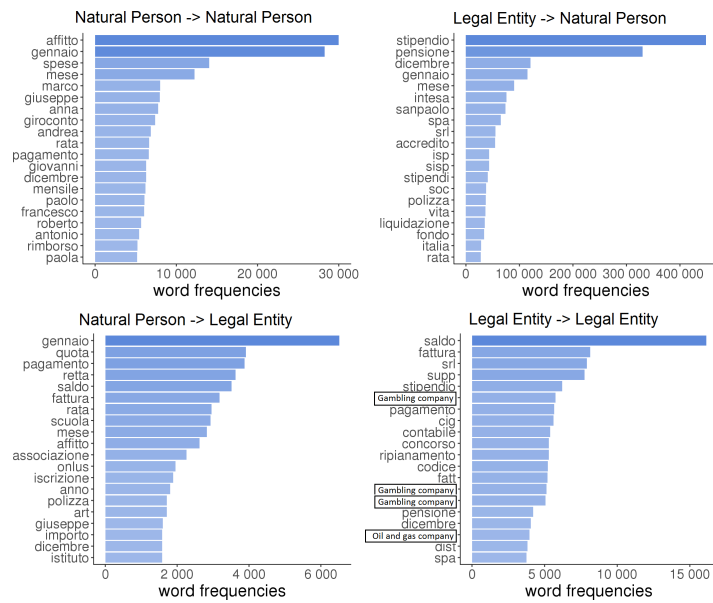


Figure B1: Word frequencies distribution: top 20 most frequent single words in the purpose of payment fields in wire transfers settled during January 2016 (we found very similar distributions for other months).

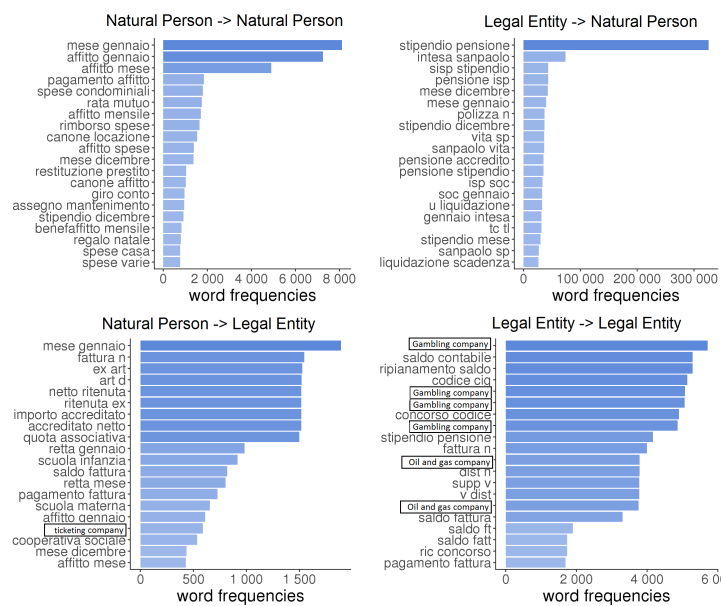


Figure B2: Bi-gram frequencies distribution: top 20 most frequent words pairs in the purpose of payment fields in wire transfers settled during January 2016 (we found very similar distributions for other months).