

Constructing a Supplier-Buyer Network: Imported Goods Flow Tracking in the Tax Transaction Records in Armenia

Navasardyan Zaruhi H.

PHD candidate, Mathematical Modeling of Economy, ASUE (Yerevan, RA)

<https://orcid.org/0009-0003-1660-7173>

zara.navasardyan@gmail.com

UDC: 330.4; EDN: MBTZYS; JEL: H2, D23;

DOI: 10.58587/18292437-2024.2-111

Keywords: Taxpayer networks, Natural language processing, Network analysis, Import tracking, Product flow traceability, Supply chain, Transactional documents

Մատակարար-Գնորդ ցանցի կառուցում. ներմուծվող ապրանքների հետազոտելիության ուսումնասիրություն Հայաստանում հարկային փաստաթղթերի միջոցով

Նավասարդյան Չարուհի Հ.

Հայտնի, Տնտեսության մաթեմատիկական մոդելավորում, ՀՊՏՀ (Երևան, ՀՀ)

Ամփոփագիր. Վերջին տարիներին հարկ վճարողների ցանցերը հետազոտության արդիական թեմաներից են՝ հաշվի առնելով ժամանակակից տեխնոլոգիական լուծումների շարունակարար աճող կիրառելիությունը և հարկ վճարողների վարքագծի թաքնված փոփոխարարությունների կամ օրինաչափությունների բացահայտման կարևորությունը հարկային մարմինների արդյունավետ գործունեության համար: Այս ուսումնասիրության մեջ հարկ վճարողների ցանցերը վիզուալիզացվում և վերլուծվում են ցանցային վերլուծության տեխնիկայի կիրառմամբ և մի քանի ապրանքի օրինակով առաջարկվում է մատակարարման շղթայում արտադրանքի հետազոտելիության ընդհանրացված մեթոդաբանություն: Ըստ էության, հարկ վճարողների ցանցերը արտացոլում են հարկատուների միջև կապերի համակարգը՝ ընդգծելով նրանց միջև ապրանքների, ծառայությունների և ֆինանսական գործարքների հոսքը: Նախորդ ուսումնասիրություններում առաջարկվում են հարկային փաստաթղթերում ներկայացված ապրանքների նկարագրությունների դասակարգման խնդրի լուծումներ: Բնական լեզվի մշակման ժամանակակից ալգորիթմները կիրառվում են արտադրանքի դասակարգման համար՝ հաշվի առնելով միայն արտադրանքի նկարագրությունը հայերեն լեզվով և սահմանափակ տեղեկատվությամբ: Մշակված ալգորիթմը օգտագործվել է որպես ելակետ՝ պարզելու համար, թե որ հարկատուներն են ներգրավված ընտրված ապրանքի առևտրում և վերջնարդյունքում՝ ցանցի կառուցման համար: Ընտրվել է ապրանքների խումբ ներմուծողից մինչև վերջնական սպառող մատակարարման շղթան ձևավորող հարկ վճարողների ցանցի կառուցման համար: Ճշգրտված գնի միջին մարժայով՝ Հայաստան ներկրված բրնձի քանակությունը հայտնաբերվել և բացահայտվել է վերջնական սպառողին վաճառքը արտացոլող՝ մատակարարման շղթայի վերջին հանգույցներում և փաստաթղթավորված հարկային անդորրագրերում: Ուսումնասիրության ընթացքում դիտարկվել են նաև այնպիսի գործոններ, ինչպիսիք են տեղական արտադրությունը, և առաջարկվել են լուծումներ՝ ներառելով առևտրի ենթակա ապրանքների գների գոտիներ՝ տեղական արտադրությունը գտելու համար:

Հանգուցարաններ՝ հարկ վճարողների ցանցեր, բնական լեզվի մշակում, ցանցային վերլուծություն, ներմուծման վերահսկողություն, արտադրանքի հոսքի հետազոտելիություն, մատակարարման շղթա, հարկային փաստաթղթեր

Создание сети поставщиков-покупателей: отслеживание потока импортных товаров в учете налоговых операций в Армении

Навасардян Заруи О.

Соискатель, Математическое моделирование экономики, АГЭУ (Ереван, РА)

Аннотация. В последние годы сети налогоплательщиков являются одной из основных тем исследований, учитывая растущую применимость современных технологических решений и растущую важность для налоговых органов выявления скрытых взаимосвязей или закономерностей в поведении налогоплательщиков. В этом исследовании сети налогоплательщиков визуализируются и анализируются с применением методов сетевого анализа, а также предлагается обобщаемая методология отслеживания продукции по цепочке поставок. По сути, сети налогоплательщиков отображают сеть связей между налогоплательщиками, подчеркивая поток товаров, услуг и финансовых операций между ними. Предыдущие исследования посвящены вопросу классификации продуктов из налоговых документов. Современные методы обработки естественного языка были применены для категоризации продуктов, учитывая только описание продукта на армянском языке и ограниченную информацию. Разработанный алгоритм использовался в качестве отправной точки для

определения того, какие налогоплательщики участвовали в торговле конкретным товаром и, следовательно, для построения сети. Набор продуктов был выбран для построения сети налогоплательщиков, формирующей цепочку поставок от импортеров до конечных потребителей. С учетом наценки количество импортированного в Армению риса было обнаружено и идентифицировано на последних узлах цепочки поставок, проданных конечному потребителю, и зафиксировано в налоговых квитанциях. В ходе анализа также учитываются такие факторы, как местное производство, и предлагаются решения, включающие ценовые фильтры на товары, подлежащие торговле, для фильтрации местного производства.

Ключевые слова: сети налогоплательщиков, обработка естественного языка, сетевой анализ, отслеживание импорта, отслеживание потока продукции, цепочка поставок, транзакционные документы

Introduction

Efficient tax collection is a cornerstone of any functioning economy, and tracking mechanisms are of a high importance in this domain. In the context of Armenia, where imported goods form a significant portion of the market, understanding the intricate flow of these goods from importers to final consumers is paramount. This process, often referred to as "good flow tracking," describes the intricate web of transactions within the supply chain, facilitating not only fiscal accountability but also enhancing economic transparency and governance.

In the setting of this research, taxpayer networks are defined as the connections formed through pairwise interactions between taxpayers. These interactions represent the relationships between entities involved in economic transactions, such as buying and selling goods or services. Moreover, the emergence of interconnected networks among taxpayers amplifies the importance of effective tracking mechanisms. Understanding and leveraging these networks gains vital importance for tax authorities to adapt to the evolving landscape of economic activities and react respectively from both functional and legal perspectives.

This paper, delves into the intricacies of constructing a Supplier-Buyer Network within the context of Armenia's taxation system. By harnessing the power of NLP, data analytics and network analysis techniques, the interconnected relationships among taxpayers are visualized.

The main contribution of the research is constructing product flow tracking networks from import to final consumer. The significance of the research contribution is amplified with the suggested evaluation approach which compares the final consumer sales and import amount for the same product where final sales is adjusted by an average markup and other external factors are considered.

Literature review

Customs authorities in different countries have widely adopted analytical technologies, employing a mix of big data, data analytics, artificial intelligence, and machine learning. Even those not currently utilizing these tools have intentions to incorporate them in the future. Most authorities recognize

significant advantages from these advanced technologies, especially in enhancing risk management, profiling, fraud detection, and improving tax collection practices [8, p. 7-9].

Many researchers acknowledge the importance of the network analysis in the taxation system, given that taxpayer tend to act more and more in a collaborative manner [4, p. 1-3] and traditional approaches are not efficient anymore [3].

Baghdasaryan et al. [2, p. 16-18] showcase the importance of the taxpayer buyer-seller network and provide evidence of network effect on real life data. They show that the company historical audit information can be replaced by its network information with no harm on the accuracy of the fraud prediction model.

In their study, Alexopoulos et al. highlight the importance of the analysis of taxpayer networks as many hidden insights may arise when analyzing the interactions rather than taxpayers individually [1, p.2-3].

Recent studies have shown that traditional auditing methods struggle with the complexities of taxpayer transaction relationships and to address this, researchers from China suggest a novel framework based on fused transactional network representations integrating transaction network topology and taxpayer attributes to enhance tax evasion detection. They test the methodology on real-world tax dataset and the suggested methods ends up providing superior results to those of the state-of-the-art approaches [7, p.241].

In a related study, researchers construct a framework analyzing the connections and trade-offs between taxes and traditional cross-border supply chain modeling features [5, p. 20-46].

Another research suggests employing data resampling techniques alongside the Random Forest ML algorithm to create a framework that identifies typical patterns and highlights features distinguishing transactions involving strategic goods from the larger context of international trade flows [6].

Methodology

To construct a supplier-buyer network for tracking imported goods within Armenia's tax transaction records, a multi-step approach was employed.

Firstly, building on previous research addressing product classification from transactional tax documents, modern Natural Language Processing (NLP) techniques were applied to categorize products based on Armenian language descriptions. The resulting algorithm representing a combination of several similarity scores provided a basis for identifying taxpayers involved in specific goods trade, serving as the cornerstone for constructing the taxpayer network. A selected set of products was then chosen and the supply chain networks for those products were constructed and visualized. These networks reveal the interconnections among taxpayers, illustrating the flow of goods, services, and financial transactions between them. Finally, the methodology compares the imported amount of the product with the internal trade records and based on various typology of supply chain, the research suggests estimates of tax revenues forgone due to utilization of untaxed or less taxed supply chain options by the taxpayers.

The experiment focused on two key products: milk and rice, selected for specific reasons. Firstly,

milk was chosen due to its shorter cycle compared to rice. Secondly, milk is extensively produced within Armenia, whereas rice is predominantly imported. Thirdly, there exists a relatively limited number of importers for these products, facilitating the preliminary validation of the obtained network and analysis.

The supply chain graph below illustrates the journey of goods from import to final consumer, traversing through several key nodes. At the outset, goods are imported into the country, initiating the supply chain process. The importer sells them to a supplier. The supplier, acting as an intermediary, facilitates the distribution of goods to various companies or business entities, which themselves can be retailers or suppliers again. Finally, the goods reach their ultimate destination at the final consumer. This graph succinctly captures the interconnected flow of goods through the supply chain, highlighting the pivotal roles played by importers, suppliers, companies, and ultimately, the end consumers.

Figure 1. *An example of a supply chain from a single importer to consumer.*



Results

The network is constructed based on invoice and tax-receipt data. It is worth mentioning that in the invoices the seller has a legal opportunity to not disclose the buyer. So, wherever the buyer information is provided the network goes deeper while when it is missing or when the products are sold with tax receipts it stops.

Milk is equally imported and produced locally. Given the fact that exact differentiation between imported goods and local production is not possible as there is no identifier, we apply a price filter to separate local production. The minimum invoice price of the first level transaction in the supply chain + markup (30%) was used as threshold for filtering.

The chart below shows the supply chain a product, e.g. milk passes from import to final suppliers.

The construction of the network involves multiple layers that correspond to successive trade transactions, each revealing crucial insights into the trajectory of imported goods. These layers extend until the point where no further transactions with invoices occur, indicating that the imported goods have reached their final destination, typically the consumer market.

Within each layer of the network, three distinct nodes are identified:

- The primary node serves as the core track, capturing the initial acquisition of goods by the buyer, complete with comprehensive transactional details. This node allows tracking of subsequent steps in the supply chain.
- The secondary node denotes sales transactions recorded without complete buyer information on the invoice. This absence of critical buyer data disrupts the flow of the network, rendering it untraceable beyond this point. It is important to highlight that legally the buyer information is not a requested field for some predefined cases.
- The final node provides insights into sales transactions documented with tax receipts, signaling the culmination of the supply chain as goods are sold to end consumers through retail outlets. These nodes signify the consumption of goods by the population, effectively marking the conclusion of the supply chain journey.

Overall, the total amount of milk consumption can be calculated as the sum of the sales with tax receipt in each layer of the network and the sales with invoice where the buyer is not identified so total import (X) should be equal to the sum of F2-F6

. We should consider that the sales invoice with missing buyer identification may not in fact be directly consumed, thus we need to add some markup there to simulate the reality. To see how successful the identification of the given product along the network was, we need to compare total final consumption to the imported goods and the difference should be approximately around the imposed markup, which will be considered 30% in our case.

The analysis revealed a significant disparity between the recorded import volumes and the final consumption of milk within the network. Notably, the consumption volume surpassed the imported quantity by more than threefold, indicating that the implemented unit price filter of over 500 dram per unit was insufficient in completely excluding locally produced milk from the dataset.

In order to affirm the assertion regarding the discrepancy between imported and consumed milk amounts, a parallel experiment was conducted focusing on rice, a commodity not locally produced within the region. Unlike milk, which is subject to potential misclassification due to the presence of local production, rice serves as an ideal counterpoint as it is predominantly imported. Thus, the experiment eliminated the need for applying price or other

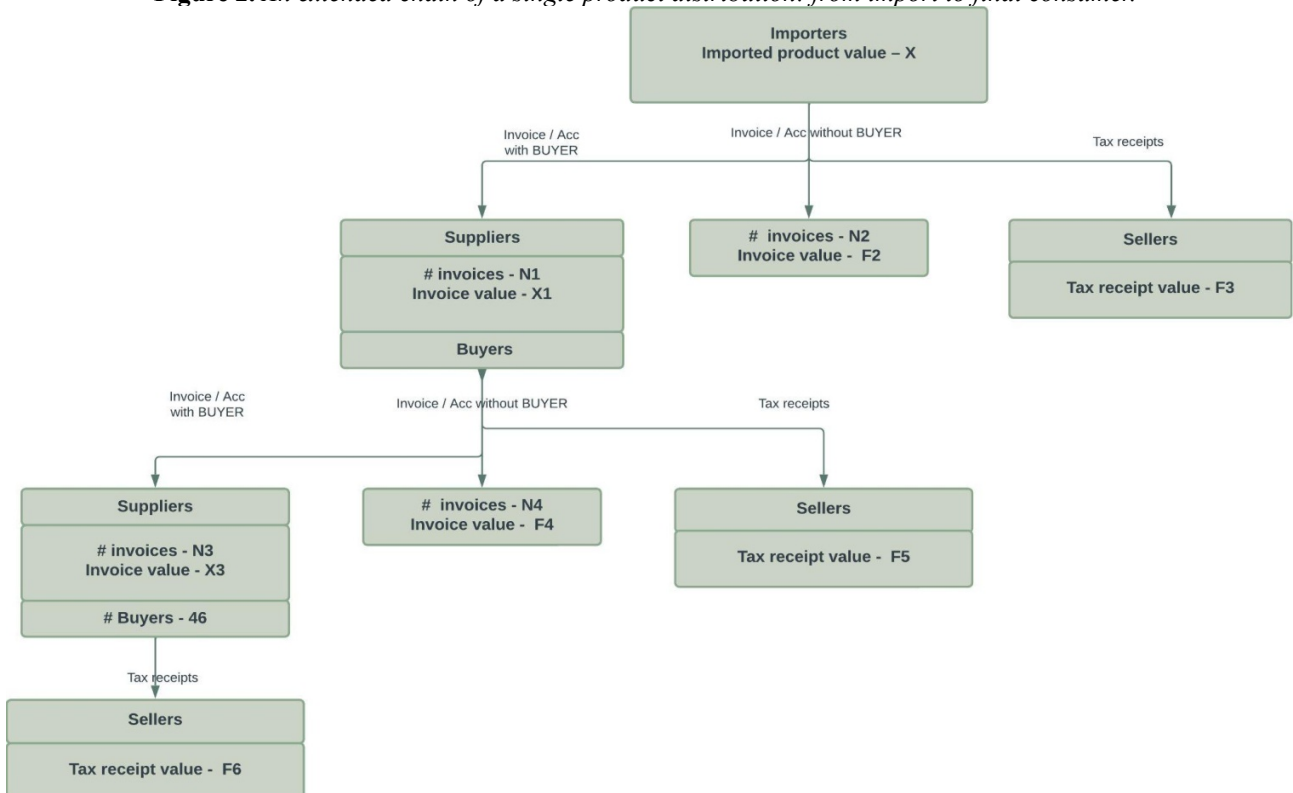
filters during the network construction phase for rice. By employing this comparative approach, any observed disparities between imported and consumed quantities of rice can be attributed solely to factors showcasing some anomalous situation, without the confounding influence of local production. This methodological extension not only bolsters the validity of the findings but also provides a comprehensive understanding of the nuances inherent in the supply chain dynamics of both milk and rice within the studied context.

As mentioned before, the final consumption can be estimated with the sum of tax receipt sales and invoice sales where the buyer is not identified in all levels.

The experiment on rice showed that the identified final consumption in the network based on automated product classification is around 1.1 times the imported amount.

Tax receipt amounts in the network are shown in absolute values but were later adjusted with 30% markup for calculations. This shows that almost all the imported rice was identified with the approach adopted by the team and proves the assumption that the reason behind the mismatch of the milk amounts was the local production.

Figure 2. An extended chain of a single product distribution: from import to final consumer.



Conclusion

In conclusion, our study underscores the efficacy of network analysis in conjunction with NLP techniques for tracking product flow from

import to final consumer. Through the integration of these analytical methodologies the networks were constructed and visualized facilitating the identification of key transactional patterns and

trends. By harnessing the power of network analysis, the interconnected relationships between importers, suppliers, and consumers were revealed, shedding light on the pathways through which goods traverse the supply chain. Furthermore, our findings emphasize the importance of incorporating NLP techniques to categorize goods accurately, thereby enhancing the traceability and reliability of supply chain analyses.

References

1. **Alexopoulos, A.N., Dellaportas, P., Gyoshev, S.B., Kotsogiannis, C., & Pavkov, T.** (2020). Detecting network anomalies in the Value Added Taxes (VAT) system.
2. **Baghdasaryan, V., Davtyan, H., Sarikyan, A., & Navasardyan, Z.** (2022). Improving tax audit efficiency using machine learning: The role of taxpayer's network data in Fraud Detection. *Applied Artificial Intelligence*, 36(1). <https://doi.org/10.1080/08839514.2021.2012002>
3. **Daley, S.**, (2010). Greeks' Wealth Is Found in Many Places, Just Not on Tax Returns. *New York Times*
4. **González-Martel, C., J. M. Hernández, and C. Manrique-de-lara-peñate.** (2021). Identifying business misreporting in VAT using network analysis. *Decision Support Systems* 141: p. 113464. doi:10.1016/j.dss.2020.113464.
5. **Mu, D., Ren, H., & Wang, C.** (2021). A literature review of taxes in cross-border supply chain modeling: Themes, tax types and new trade-offs. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 20–46. <https://doi.org/10.3390/jtaer17010002>
6. **Nelson, Ch.** (2020). Machine Learning for Detection of Trade in Strategic Goods: An Approach to Support Future Customs Enforcement and Outreach. *World Customs Journal, Volume: 13.*
7. **WCO/WTO.** *The Role of Advanced Technologies in Cross-Border Trade* (2022). <https://doi.org/10.30875/9789287071002>
8. **Wu, Y., Dong, B., Zheng, Q., Wei, R., Wang, Z., & Li, X.** (2020a). A novel tax evasion detection framework via fused transaction network representation. 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). 235-243 <https://doi.org/10.1109/compsac48688.2020.00039>

Сдана/Հանձնվել է՝ 28.02.2024

Рецензирована/Գրախոսվել է՝ 06.03.2024

Принята/Ընդունվել է՝ 11.03.2024