Service Producer Price Index for Freight Transport by Road: A model pricing approach [(+)]

Hugo Moreira [(*)] and Rui Evangelista [(*)]

Abstract

In this paper, we describe the strategy that is being followed to compile a new Service Producer Price Index (SPPI) for Freight Transport by Road (NACE 4941), which is presently not produced by Statistics Portugal. This activity represents 15.4% of the turnover provided by the service activities that fall under the scope of European short-term statistics (STS) regulations (e.g. Regulation no. 158/2005). A new survey, designed on the basis of a model pricing approach, was developed and implemented from the second quarter of 2019 onwards. Although based only on two quarters of data, preliminary results support the idea that the transmission of this new SPPI based on this approach is feasible and can begin in 2020.

Introduction

The objective of this paper is twofold. Firstly, it aims at providing an account of the work that has been carried out to compile a SPPI for freight transport services by road for Portugal. The sector dealing with this transport service (NACE 4941), ranks first in terms of turnover importance amongst all the services included in the STS Regulation (15.4%; 2015 data). As it is not presently compiled by Statistics Portugal, a SPPI for this type of service would represent an important improvement in the coverage of STS. Secondly, it presents the preliminary results stemming from 2 quarters of data taken from the new survey on prices of freight transport by road (IPPSTRM), which was designed as the source of the new SPPI.

The paper is divided into 6 sections. The first two sections provide a brief description of the road transport sector and describe the compilation strategy that was followed to build the new SPPI. The third section reports the questionnaire and sample designs of the new prices survey that had to be developed for the new SPPI. The fourth section specifically addresses the SPPI compilation process. The fifth section presents the first results stemming from the IPPTRM data. Finally, a summary and an account of the next steps in relation to the implementation of this index are given at the end of this paper.

---

1. Getting to Know the Sector Under Analysis

According to the 2016 data available at Statistics Portugal, there were 6.907 companies with NACE 4941 as their main activity and with a reported turnover and number of employed persons larger than 0. This NACE is marked by a great concentration, with the 10 and 100 biggest companies amounting to 15% and 42% of the sector's turnover. This feature, which is important to define the type of sampling design to be implemented in the construction of the new SPPI, is easily seen in the next figure.
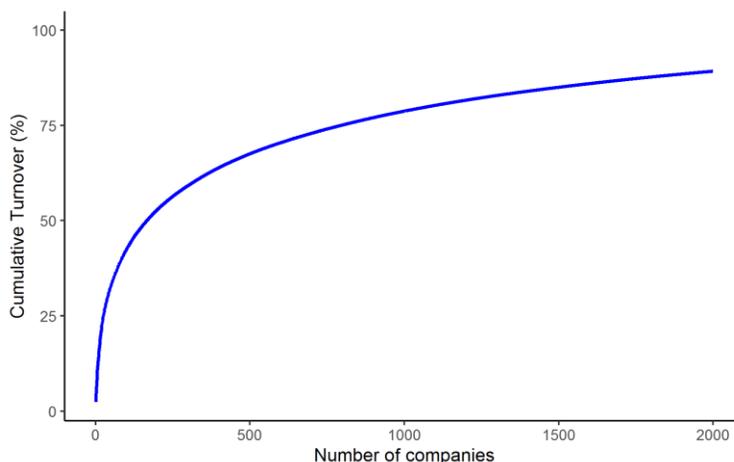
Cumulative turnover (%) by number of companies, by descending turnover order

The number of employed persons has a much skewed distribution as well. The majority of companies have less than 5 employees, as the following table shows:

Persons employed	Number of companies	%
1	1997	28.90
2-5	2852	41.30
6-10	918	13.30
11-50	971	14.10
51-100	94	1.36
100+	75	1.09

While there is no available information on service prices, there is some information on transported volumes through a Survey on Freight Transport by Road (ITRM - Statistics Portugal, 2017). This survey is the basis for the compilation of statistics on transported goods by tonnes and tonnes-kilometre, zone, type of goods (as defined by

the Standard goods classification for transport statistics or NST), and nature of freight (i.e. own account or for other companies).

The ITRM provides a clearer picture of the sector. In addition, its microdata have the potential to provide information for the design of the new SPPI. This covers not only data on individual trips, but also includes important variables (such as the number of kilometers, tonnes, goods transported, type of vehicle and origin/destination by municipality for national transportation services), that can be used in the identification of typical transportation services. For instance, the most common vehicles used to carry out a transportation service under this NACE are lorries and tractors with semi-trailers.

In addition, a round of meetings with the associations representing the sector was also conducted during the design stage of the new SPPI. More concretely, it was possible to meet the main associations of the sector (ANTRAM and ANTP), something that yielded important comments and feedback at the construction stage of the new SPPI.

2. Definition of Compilation Strategy

The literature on SPPIs constitute one of the first sources of information that could help defining the compilation strategy for the new SPPI. Unfortunately, the literature dealing with the practicalities of the compilation of a service price index for this sector is scarce. A notable exception is Goldhammer (2007) who describes the approach followed by DESTATIS to compile a SPPI for the transport and logistics sector in Germany. Some methodological guidance is also provided in Eurostat-OECD (2015), particularly in its Chapter 6, where an approach by the ONS is given as an example. In general, the approaches available in the literature tend to favor contract pricing or model pricing. An exception to this is Spain's SPPI (Ministerio de Fomento, nd), which is based on actual transaction prices collected through a "price" question in its Survey on Freight Transport by Road.

To determine the best compilation approach, it was found necessary to have the feedback of the sector's associations (ANTRAM and ANTP). In the end of this process, it was decided to follow a model pricing survey, as only larger companies would be able to report existing contracts - smaller ones do transport services on a *ad hoc* basis. Following this, it was also decided to divide the population of service freights into 3 basic markets / geographical zones: domestic, Iberian Peninsula (to/from Spain) and international. The Spanish market was detached from the remaining countries due to its relative importance for non-domestic freight transport by road (as will be seen below).

The conversations held with the operators on the market have also allowed the identification of most important price-determining variables in this type of service. Distance, weight, and type of transported goods (mainly the type of vehicle needed to transport it) were particularly emphasized by associations and operators. Finally, the

journey was defined as the price measurement unit of the survey. This implied that journeys had to be specified, something that was done in the questionnaire design stage of the survey (see the next section).

3. The New Prices Survey on Freight Transport Services

3.1. Questionnaire design

As mentioned above, after contacts with the industry, a model pricing approach was chosen as the pricing method for the survey on prices of freight transport by road (IPPSTRM). One of the drawbacks of model pricing is that it would be extremely difficult to characterize all transports in a way that all journey types would be covered by the survey, as this sector is extremely heterogeneous.

Therefore, when designing the survey and surveyed journeys, choices needed to be made as to which types of transportation would be part of the questionnaire. In doing this, there was also the need to satisfy the inherent contradiction of maximising the number of provided answers, while keeping the response burden of the survey as low as possible.

To help the specification of the journeys, we looked into ITRM's microdata (one full quarter of observations). Filtering the data only by hire or reward freights, we analysed which freights were more frequent, and decided to divide the types of transport into 9 distinct areas:

- General cargo (in curtainsider);
- Closed box;
- Stone; mining/quarry products (in tipper truck);
- Concrete (in mixer truck);
- Fuel (in fuel tanker);
- Cement (in tanker);
- Containers;
- Temperature controlled box;
- Cereals (in tipper).

Some of these types of transport do have a correspondence with the NST, but some are more general and are connected to the type of box/trailer utilised (the type of good transported is not as important in this case). There are important types left, such as wood or car transporters.

Journeys (and the questions to be asked) were also divided by geographical zone. These were the following:

- North (Portugal);
- Centre (Portugal);

- Lisbon Area (Portugal);
- South (Portugal);
- Spain (Iberian Peninsula);
- International (other European countries).

These geographical zones do not correspond to the Portuguese Nomenclature of territorial units for statistics (NUTS) areas. However, they are familiar (and commonly used) by the enterprises involved in this industry.

Questions dealing with the Spanish market only concern curtainsiders, while in international freights there are curtainsiders and controlled temperature journeys asked. For domestic trips, there is at least one question for type of transport and for zone. Some types, like curtainsiders, have more than one question per zone (which helps to capture differences in e.g. lorries/tractors with semi-trailer).

The specification of the survey's questions contained cities of origin and destination, distance between them, goods transported and their weight in tonnes, and type of vehicle (lorry/semi-trailer). This leaves the burden for companies to calculate distances from their premises to origin and from destination to premises. Additional details exist in specific questions, e.g. when there is a shuttle trip or a distribution journey. Most services included are round trips with cargo load (where each price is collected separately), unless it does not make sense (e.g. fuel transportation, cereals, etc.). Two examples of asked services can be found in the Annex.

In total, the questionnaire included 59 journeys/prices. Of these, the curtainsiders' freights have the highest weight, amounting to 21 collected prices. To lower the burden of respondents and allow a more "personalized" questionnaire, the questionnaire was divided into two sections, with the first one asking companies to select up to 3 of their most frequent services and markets/zones and the second section with the concrete price questions (which are linked to the choices made in the first section of the questionnaire).

Some companies may finish the questionnaire in the first section. An example would be a respondent that only transports goods in thermal boxes to the Spanish market, or works only with car transporters or other types of services not included in the sample of journeys.

In the second section, a justification has to be given whenever no price quote is provided (e.g. do not transport the required goods, or do not have the asked vehicle). There is also an open field where respondents are free to provide more insights. The annex provides some screen views of both parts of the questionnaire, which was sent out to selected companies using a web-based platform.

3.2. Sampling Design

Given the extreme concentration in turnover in the biggest companies, it was decided to employ cut-off sampling, using turnover as the cut-off variable. All the information on companies comes from the business register. The sample used in the new survey was divided into 4 categories, in line with the type of vehicles used to carry out transport services by road. Portugal's Institute for Mobility and Transportation compiles a register of every licensed heavy vehicle, along with the company it belongs to. Using this administrative information and cluster analysis, it was possible to find the following 4 clusters:

- Companies mainly with lorries;
- Mainly with tractors above 7 ton;
- Mainly with tractors below 7 ton;
- No registered vehicles.

The presence of the last cluster is relatively odd considering the main activity of the companies in this sector. However, it was decided to include its companies in the sampling procedure, as there was no information available that could justify excluding the group.

The only restrictions when selecting the population of companies from which a sample had to be drawn, were (i) that their main activity needed to be freight transport by road, (ii) their location was continental Portugal, (iii) had to have non-null turnover and (iv) had at least one positive turnover figure from 2015 onwards.

In practice, the cutoff sampling procedure was applied in each group, selecting units in descending order of turnover until 70% of total group turnover was covered. This procedure left a total of 666 selected companies. The sample contains more companies within the first two groups (around 30% each) than of the last two. Companies are inquired online, in the first month of the respective quarter.

Since we have no clear information about what types of transport each company provides, we cannot "direct" the survey to the ones that are able to answer the questions. Therefore, with this approach, it is expected that the number of companies with no answer/service may be relatively high. This implies that, after every instance of the survey, Statistics Portugal needs to evaluate whether a company should keep answering the questionnaire or, on the other hand, it should be removed from the sample (as it reported that it does not, for instance, cover any of the types of transport).

4. Index calculation

4.1 Weighs

The survey groups the 59 questions by type of transport and zone. While for the first variable it is not possible to derive weights (there is no information available), it is possible to calculate the relative importance of each zone/market. To this effect, we use total transported tonnes, divided by domestic and international transport. We preferred tonnes to tonnes-km since this measure would overrepresent the international freights, which usually cover longer distances.

Based on Statistics Portugal (2017), it is possible to say that international transport represents 24% of the total when measuring in tonnes. Inside it, we give transport from/to Spain the same proportion it has in terms of goods loaded/unloaded (65%). The application of this leads to the final weights:

Zone	Weight
National	0.76
Iberic (Spain)	0.16
International	0.08

This weighting scheme allows to disaggregate the index into three sub-indices (Domestic market, Iberic and International freights), a characterization that is well understood by the economic agents of this market.

4.2. Index's Compilation Steps

For the calculation of the SPPI, three sub-indices are calculated for each zone z. Each one of them is defined as a geometric mean weighted by each company's turnover:

$$I_{t,t-1}^{z} = \frac{\left(\prod_{p \in z}\left(c_{t}^{p,e,z}\right)^{\frac{VVN_e}{\Sigma VVN_e}}\right)^{\frac{1}{\Sigma \frac{VVN_e}{\Sigma VVN_e}}}}{\left(\prod_{p \in z}\left(c_{t-1}^{p,e,z}\right)^{\frac{VVN_e}{\Sigma VVN_e}}\right)^{\frac{1}{\Sigma \frac{VVN_e}{\Sigma VVN_e}}}} \quad , z = 1,2,3$$

where:

$c_{t}^{j,p,e,z}$: Price quote from company e, in zone z, for freight (question) p with transport type j, in quarter t, divided by the corresponding number of km.

$e = 1, \ldots, n$, where n is the number of companies that answered in both quarters; $z = 1,2,3$ (national, iberic, international); $p = 1, \ldots, 59$ (number of price quotations asked). For aggregation, we also need each company's turnover, referred to as VVN_e.

The final index is the aggregation using each zone's weight, as defined above.

$$I_{t,t-1} = \sum_z I^z_{t,t-1} \, w_z \, ,$$

where w_z is the zone weight.

5. Preliminary Results

5.1. Prices Survey

The first price collection started in May 2019, covering the second quarter of the present year. A total of 497 of the 666 surveyed companies replied to the questionnaire, from which 317 gave at least one price quotation, yielding a total of 2.905 prices.

The answers provided were found to be of good quality. For instance, and as expected, it was found an inverse relationship between the mean price per km and the number of km of each question. This is a consequence of the price formation mechanism, where there is a fixed "base" value, to which is summed a value per km.

Following a thorough analysis of the answers (and observations) provided by the companies which had no vehicles in the database, we concluded that this was due to one of the following situations: 1) they hire subcontractors; 2) are freight forwarders; or 3) use only light vehicles (less than 3.5t). Only 24% of the companies from this group that answered were able to supply prices.

A total of 168 companies were taken out from the sample (e.g. because they operate solely on own account or freight forwarding) and not included in the next price collection exercise. Due to the large amount of removed companies, a sample reinforcement was done, adding 177 new companies, again with a cutoff sampling procedure. In this occurrence, we decided not to select any companies with no vehicles registered: they are less reliable in terms of ability to answer, and the most important ones are already covered. For the remaining groups (e.g. companies mainly with lorries), companies are selected in accordance to the cluster's weight in the sampling frame.

The second instance of price collection covered the third quarter of 2019. We had answers from 473 companies, of which 380 answered with at least one price. 270

Categorization of companies by status of answer
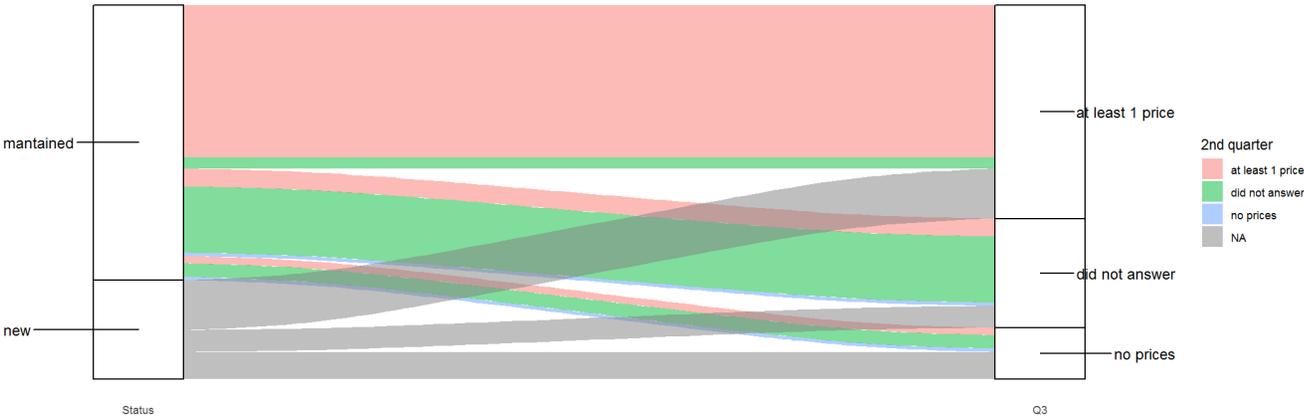
answered (with prices) in both quarters. Moreover, the 177 new companies added in the sample had a positive contribution to the number of answers. The following figure shows the flow of answers in the third quarter divided by whether they come from the second quarter or are new in the sample. The graph is coloured by the type of answer they provided in the second quarter.

As it is possible to see from the above graph, the vast majority of respondents who provided prices in the second quarter repeated this status in the third quarter of 2019.

In terms of price quotations, there are 2.905 observations in the second quarter, and 3.510 in the third. Of these, 4.976 correspond to answers of companies provided in the two quarters (a total of 2.488 pairs of price quotes). One important result is that, in around 82% of cases, the individual (pairwise) price ratio equals 1 (ie, most companies are supplying the same prices as the previous quarter).

With two quarters of data available, it was possible to show results in terms of index calculation using the pairs of price quotes and the formulas described in section 4.2. For the 2 available quarters of data, it was found a price variation of 1.21%.

5.2. Outliers and Imputation

The 1.21% result does not take into account any data cleaning or imputation procedure. In compiling this figure, no outlier removal treatment is done: it is assumed different prices for the same question happen due to competitive factors (closer distance to the cities, larger flexibility, etc.).

In the survey, companies who supply price variations of 15% or more, in absolute value for the same price quotation, are asked to provide a justification. However, sometimes this is not followed and it is not certain whether these values are data entry or other type errors. The following figure shows the pairs (answers occurring in both quarters) of prices per km given for each question on national canvas cover freights (each box is a different question). A 45º degree line (zero variation) is also plotted:
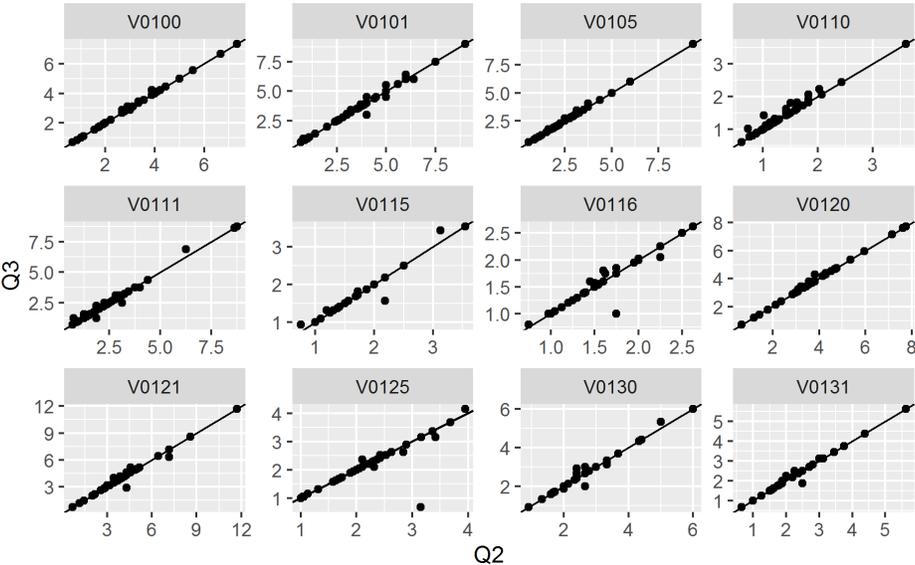
Answers on Portuguese curtainsider freights

Most answers are very close to the line, representing small variations. There are, however, some observations that show considerable deviations. At this stage, two factors need to be taken into account. First, as it was already mentioned above, some companies may provide justifications for abrupt price changes (usual responses are rises in labour costs, fuel prices or variations in the value of cargo). Second, the survey allows companies to see their previous answers, a fact that helps respondents' to insert an erroneous or inconsistent price quote.

To assess the impact of removing outlying price quotes on the Index's overall price change, we took a relaxed approach and eliminated the observations with an absolute variation of 20% or more. After this procedure, the variation dropped to 1.11%.

Moreover, since there was a non-negligible number of observations with no pairwise connection in both quarters (companies that answered in Q2 but not in Q3, and vice-versa), it was also studied the use of an imputation procedure to help increasing the number of pairs of price quotes to be used in the index. The imputation procedure replaces missing data in both quarters, thus maximising the number of pairs of data in the compilation of the SPPI.

Simple methods, such as imputation by mean or median values would, in practice, introduce a null variation in all imputed observations, and artificially lower the value of the index. A more sophisticated schema was applied to overcome this potential problem. The chosen approach is a variation of propensity score matching. It relies on the assumption that companies and questions' characteristics explain the variations, i.e. similar companies would provide similar variations in the same questions.

It works as follows. First, observations are grouped by company and question. Then an auxiliary variable is added, equaling one if the company needs imputation in that question. A Probit regression is then run, explaining the auxiliary variable in function of a company's turnover and number of workers, and question's characteristics (number of km, kg, and dummy variables for zone and type). The estimated probability is called score. With this score, we can proceed with the matching where a company-question that needs imputation is matched with one that does not need it (answered both quarters). The change rate of the second is applied to the first. All of these steps are done in R (R Core Team, 2017), with package MatchIt (Ho et. al, 2011).

The price variation given by the application of this imputation process (and outlier removal at 20%) is 1.08%, a slight decrease from the previous 1.11% figure. Since most pairs have a 0% change between quarters, this result from the previous run was expected.

The decision as whether to apply this imputation mechanism will be done in the future with the arrival of more data. As more quarters become available for analysis, it will be possible to carry out cross-validation procedures to check if propensity score matching is replacing missing data with accurate values.

Summary and the Way Forward

This paper presents the results of the development of a new SPPI for road transport freight services. As there were no prices information for this service, it was necessary to develop a completely new survey targeting enterprises in NACE 4941. After a literature review and a round of meetings with the main associations of the sector, it was decided to ask respondents for prices of pre-defined services (model prices) using a web-designed survey.

Although the new survey has only produced two quarters of data, it is already possible to see that the quality and quantity of price quotes is good enough to support the compilation of a new SPPI and its transmission to Eurostat in the first half of 2020.

References

- Eurostat-OECD (2015). Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services. Available in

https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-04-14-661

- Goldhammer, B. (2007). The new producer price indices for goods transports and logistics. Available in
https://www.destatis.de/EN/Methods/WISTAScientificJournal/Downloads/the-new-producer-price-indices-transport-goldhammer-112007.html

- Ho, D., Kosuke, I., King, G. and Stuart, E. (2011). *MatchIt: Nonparametric Preprocessing for Parametric Causal Inference*. Journal of Statistical Software, Vol. 42, No. 8, pp. 1-28. Available in: http://www.jstatsoft.org/v42/i08/

- Ministério de Fomento (n.d.). Nota Metodológica. Available (only in Spanish) in
https://www.fomento.gob.es/informacion-para-el-ciudadano/informacion-estadistica/transporte/transporte-de-mercancias-por-carretera/nota-metodologica-del-indice-de-precios-del-transporte

- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

- Statistics Portugal (2017). Inquérito ao Transporte Rodoviário de Mercadorias (Survey on Freight Transport by Road). Methodology. Available (only in Portuguese) in
http://smi.ine.pt/DocumentacaoMetodologica/Detalhes/1449?modal=1

Annex

A) Example of questions/journeys included in the questionnaire

The following service journey belongs to the curtainsider type, and the international zone. It asks for two prices (two questions total):

- Transport of 18ton of mail from Lisbon to Frankfurt (2330 km), in a tractor with curtainsider semi-trailer. Journey with empty truck to Munich (400 km).

  Return journey from Munich to Lousada (2250 km), with 13ton of general cargo.

The other example concerns a "special case". Since it is a distribution journey, we add the number of stops needed. This is a thermal controlled box question, for the Center region of Portugal:

- Distribution of 5ton of vegetables in pallets, from Bombarral to Vila do Conde, with 5 stops (total length: 300 km), in a lorry with thermal controlled box. Return with empty truck.

B) Print screens of the questionnaire

Resumo

Mercado Nacional • 1 AVISO

I   REGIÕES E TIPO DE TRANSPORTE

I.1 Indique todas as regiões onde efetua os seus serviços de transporte:

Mercado Nacional

Norte	A0100_1	☑
Centro	A0100_2	☑
Região de Lisboa	A0100_3	☑
Sul	A0100_4	☑

Mercado Ibérico

Espanha	A0100_5	☑

Mercado Internacional

Resto da Europa	A0100_6	☑

I.2 Indique os três tipos de transporte mais representativos que efetua:

Carga Geral (Lona)	A0200_1	☐
Caixa Fechada	A0200_2	☐
Inertes	A0200_3	☑
Betões	A0200_4	☐
Combustíveis	A0200_5	☐
Cimentos	A0200_6	☑
Contentores	A0200_7	☐
Frigoríficos	A0200_8	☑
Cereais	A0200_9	☐
Outros	A0200_10	☐

Confirmar

C   SERVIÇO - MERCADO INTERNACIONAL

Indique a cotação/orçamento para cada um dos serviços de transporte abaixo indicados, tendo em conta: carga transportada, distância do percurso, tipo de veículo utilizado e locais de carga e descarga. Caso não consiga fornecer uma cotação, assinale o motivo. Considere cada serviço isoladamente e suponha que este é iniciado num dia útil próximo do meio do mês de referência (ex: 15 de julho).		Cotação (sem IVA) em EUROS		Se não tem cotação, indique o(s) motivo(s):		Observações
		Anterior	Do mês	Não efetua o transporte da carga em questão	Não possui o veículo referido	
1		2	3	4	5	6
C 3. Frigorífico - Resto da Europa						
Transporte de 19 toneladas de fruta em paletes, num trator com semirreboque frigorífico, das Caldas da Rainha a Swindon, Reino Unido (2300 km). Ida em vazio para Arras (400 km).	V0410	100	100	☐	☐	
Volta com 20 toneladas de produtos hortícolas, de Arras para Coimbra (1750 km).	V0411	100		☑	☑	sem transporte