What Can Massive Mobile Phone Data Tell Us about International Trade?

The Case of Spain

José María Álvarez-Pallete

Department of Research and
Chief Economist

DISCUSSION
PAPER Nº
IDB-DP-570

February 2018

What Can Massive Mobile Phone Data Tell Us about International Trade?

The Case of Spain

José María Álvarez-Pallete

Universidad Pontificia Comillas and Telefónica

February 2018

Abstract

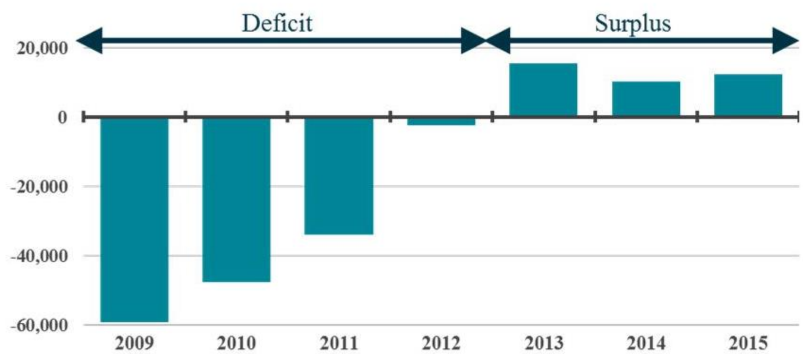
This paper studies the existing relationship between international mobile phone calls and international trade for the case of Spain, showing that it is possible to build an explanatory model with a high level of correlation between both categories of indicators. The paper further evaluates the use of “massive data” in predictive economic models and provides a review of research in this field. It has been demonstrated than when “massive data” is applied to predictive models, they can offer a higher degree of accuracy than when using traditional approaches. The application to the case of Spain shows that it is possible to anticipate international trade flows by using massive mobile phone calls data. This fact opens a wide field for applied research in international trade both for academia and policy makers.

JEL classifications: C49, F14, F47
Keywords: Massive data, Mobile phone, International trade, Predictive economic models

1. Introduction

In recent years, Spain has experienced one of the most momentous economic crises in its history (see Sauri, 2015, and OECD, 2014). The External sector has played a key role in the recovery of the country, as can be seen in Figure 1. From 2009 to 2015 Spain moved from an external deficit close to 60,000 million euros (about 6 percent of GDP) to a surplus of around 1 percent of GDP, transforming the structure of the current account balance.

Figure 1. Evolution of Current Account Balance in Spain (millions of euros)



Source: Author's compilation based on 2015 Bank of Spain data, retrieved March 2016.

When looking for explanations of this structural change in Spain's current account balance, it is noteworthy that the country experienced, in a very short period of time, a considerable improvement in its competitiveness with respect to the rest of Europe, according to data from the Consejo Empresarial para la Competitividad (2013).

This development suggests that it would be valuable to have tools that allow for a better understanding and ability to predict the foreign trade capability of an economy. To this end, this paper addresses the following research questions. First, what data sources, in addition to official statistics, provide a sound and reliable basis for understanding the size, nature and pattern of foreign trade activity?

Second, what are the scale and span of this data compared to that used by official institutions? Is there a data source that enjoys an informational advantage due to its relevance or accuracy?

Third, is it possible to provide alternative sources in order to test their reliability? Assuming a reliable source of data is available, is it possible to build a predictive model that could anticipate an economy's foreign trade activity?

Thus, the purpose of this paper is to determine whether or not it is possible to build an explanatory model of the evolution of Spain's trade balance based on a source as reliable in volume as it is in detail. With this objective in mind, we hypothesize that telecommunications networks could be a possible indicator of economic activity, due to the fact that a significant amount of business activity requires the exchange of information through communication networks.

In order to find answers to these questions, and for the purposes of this study, it is necessary to use data flows from telecommunications networks. In this sense, I show that it is possible to build an explanatory model with a high level of correlation between international mobile phone calls and international trade for the case of Spain. At the same time, I evaluate the use of "massive data" in predictive economic models and provide a review of research in this field. It has been demonstrated than when "massive data" is applied to predictive models, they can offer a higher degree of accuracy than when using traditional approaches. The application to the case of Spain shows that it is possible to anticipate international trade flows by using massive mobile phone calls data. From my point of view, the analysis contained in this paper opens a wide field for applied research in international trade, both for academia and policymakers.

This paper is structured as follows. Section 2 analyzes the literature on massive data from telecommunications and economics, finding a gap in the use of this approach for the empirical study of international trade flows. Section 3 describes the data construction process and presents some stylized facts on the relationship between international trade and international phone calls. Section 4 presents the principal model and results. Finally, Section 5 concludes.

2. Literature Review

One of the innovations provided by massive data is the introduction of economic analysis from the use of original variables that were previously difficult to observe. Among these, the geo-location variable stands out (Einav and Levin, 2013). Mobile phones—with call detail records or CDRs, where data are recorded for billing and operational purposes—provide a remarkable degree of detail of human behavior by detecting the movement of mobile terminals through the mobile infrastructure network. The mobile phone data of millions of users are obtained with high frequency and granularity.

A reference study using telecommunications network data for their application in the field of official statistics is that of Frias-Martinez et al. (2013). This paper provides details on how official statistics, principally socioeconomic variables, often suffer from important disadvantages. First, the very process of data collection is often expensive and troublesome, especially for emerging economies with financial constraints and technical deficiencies. Second, the socio-economic indicators are produced ex post when the socioeconomic changes subject to analysis have already occurred. Consequently, the study seeks to demonstrate that the use of data from telecommunications networks is more reliable, effective and immediate.

Its aim was to provide economic policy-makers with the tools to develop advanced indicators for socioeconomic variables in an efficient and reliable way. To do so, they proposed a set of indicators that could be quickly, reliably and efficiently extracted from telecommunication networks data.

One of the most recent studies carried out in this field (Coscia and Hausmann, 2015), establishes the relationship between two social networks of human interaction: one constructed based on telephone calls and another constructed from face-to-face interaction/mobile scrolling. Analysis of mobile data in Colombia concluded that relationships through social networks expressed via mobile phone calls are an aggregation of a higher order than face-to-face social relations. Both networks are isomorphic. Establishing the relationship between the two social networks opens the door to future research that could show the extent to which, for example, both types of social networks are complementary or substitutes.

This study is particularly relevant since it concludes that face-to-face relations and those carried out telephonically are almost indistinguishable. Thus, the conclusion drawn is that human behavior is practically the same whether or not it takes place through physical interactions or telecommunications.

Also, the work of Louail et al. (2014) highlights how the use of data obtained from telecoms networks can be used to draw conclusions in the field of social sciences. The authors confirm that pervasive infrastructures allow an enormous amount of data on human behavior to be obtained, thereby revealing the character and evolution of urban infrastructures.

In order to do this, the authors studied mobile network data over a period of 55 days in 31 Spanish cities. They defined a series of indices describing the average distance covered by individuals in these cities, as well as a series of hotspots based on the concentration of individuals and the movements of people towards those centers.

The authors drew several conclusions. On one hand, there is a rhythm and tempo of urban life common to all cities. On the other hand, this study opened the possibility of establishing a new quantitative classification of cities via spatio-temporal data (based on the distribution of hotspots and the activity occurring throughout the day around these centers).

Massive data offers advantages as a resource for economic analysis. Key aspects noted by Einav and Levin (2013) include real-time availability, which produces immediate predictions and identifies changes in the behavioral pattern of economic variables as they occur, or which can anticipate economic changes.

Analyzed with the appropriate tools (via machine-learning techniques), massive data increases statistical power and precision, reinforcing official statistics, identifying economic opportunities and anticipating structural economic changes. Data can facilitate predictions for many markets and, consequently, decision-making by businesses and individuals (Wu and Brynjolfsson, 2015).

Researchers have discovered the usefulness of online data. Due to large quantities of foreign trade and macroeconomic data available online, computing power and free access via the Internet, Hausmann et al. (2011) have created the reference publication *The Atlas of Economic Complexity* to measure the complexity of knowledge production of each country, based on the composition of its production. The additional development of the online data visualization tool the Observatory of Economic Complexity (Simoes and Hidalgo, 2011) offers new and innovative ways of perceiving world trade to understand better its contribution to economic growth.

In addition to Internet data use, mobile phone use—with CDRs as noted above—provides an enormous amount of data on human behavior. The mobile phone data of millions of users can be extracted in an aggregated and anonymous way to gain a better understanding of the demand for urban mobility or to offer better management of the urban environment or social welfare.

The research in this field stands out for its original use of data from mobile terminals. The literature includes studies to estimate crime (Bogomolov et al., 2014), to evaluate the social response to an earthquake (Moumni, Frias-Martinez and Frias-Martinez, 2013) or to classify cities according to their spatial structure as measured by the density of mobile connections to hotspots (Louail et al., 2014).

In addition, within the field of international trade, there are many relevant examples worthy of mention. A World Bank study undertaken with information obtained from eBay (Lendle et al., 2012) compared the impact of distance, a standard proxy for trading costs, on eBay (online) and on traditional (off-line) international trade flows, respectively. The study concluded that the impact of distance is on average 65 percent lower on the eBay online platform than offline. This difference is explained by the reduction of both information and trust issues due to online technology. The analysis estimates that wealth gains would be on average of 29 percent higher if off-line transaction problems were reduced to the level of online issues.

The possibility of exploiting massive data sets for the predictive analysis of economic activity, however, is not limited to time series obtained from data available on the Internet. The large volume of data managed by the private sector and large corporations has also been of interest within the field of economic research, in particular, the data generated by telecommunications networks.

3. Data and Stylized Facts

3.1 Data

To realize the possibilities noted above, it was necessary to find a telecommunications operator of reference in the Spanish market with a data source of sufficient magnitude and time span to provide a reliable starting point. Due to data availability, we selected Telefónica for that purpose.

The primary objective is to corroborate the hypothesis that international phone calls, available in real time and with a considerable volume of data given the total size of the universe (> 70,000 companies with international traffic), can predict the evolution of foreign trade indicators before the official publication of government statistics (DataComex). The study focuses on the specific segment of companies that make international calls, such as corporations and small and medium enterprises (SMEs).

For the purposes of this study, we constructed a dataset containing 13,207 million phone calls recorded for 49 months. As this volume of data takes up 1,190 GB of memory, important transformations were required to render the data usable for the purposes of this paper.

First, we looked into the historical series of long-duration calls, classified by country of destination. International calls placed through the fixed network were not stored for long enough nor with the detail necessarily required for this study. Thus, international voice traffic data carried through mobile networks was selected for the period between 2012 and 2015, classified by country of destination on a monthly basis.

Second, filters were applied by activity code, excluding calls from individuals, the self-employed, phone booth centers and "call centers," since these subjects are beyond the scope of this paper. For the purposes of this study, we built a sample of international mobile traffic data from large corporations and SMEs.

This dataset represents more than 70,000 corporations and SMEs that have generated international mobile voice traffic to over 200 countries, and a total volume of 1.4 million series of data and more than 67 million recorded calls during the period analyzed. This volume of data is sufficiently representative of the international call activity of a relevant sample of Spain's business sector.

Foreign trade data series from official sources, available on-line and with open access, were also extracted. Specifically, historical data series for the export and import of goods for the period between 2012 and 2015, with monthly data broken down by country, were extracted from the foreign trade database (DataComex) of the Ministry of Economy and Competitiveness of Spain (2016). This represented more than 20,000 export and import records for the period between 2012 and 2015.

3.2 Stylized Facts

By drawing parallels between the data obtained from mobile traffic calls and official external activity, it was found that, as shown in Table 1, a relationship between both time series could exist. This leads to further consideration of additional processes that allow us to determine what kind of relationship there might be (if any), in what form, and to what degree.

Table 1. International Telephone Calls and Trade Flows

Country of destination	Traffic	Country of destination	Export	Country of Origin	Import
France	100	France	38,696	Germany	35,925
United Kingdom	74.4	Germany	27,087	France	29,755
Portugal	67.6	Italy	18,669	China	23,622
Italy	64.3	United Kingdom	18,231	Italy	17,312
Germany	61.1	Portugal	17,915	United States	12,844
United States	40.8	United States	11,410	United Kingdom	12,583
Netherlands	18.6	Netherlands	7,939	Netherlands	11,446
Morocco	18.5	Belgium	6,644	Portugal	10,697
Belgium	15.1	Morocco	6,134	Belgium	7,071
Switzerland	14.6	Turkey	5,077	Algeria	6,490

Note: Top 10 countries for international traffic (outbound mobile voice) by traffic volume. Base France = 100. Top 10 export and import destinations / origins by total value (million euros) 2015.
Source: Author's compilation based on 2016 Telefónica and Ministry of Economy and Competitiveness data retrieved in March 2016.

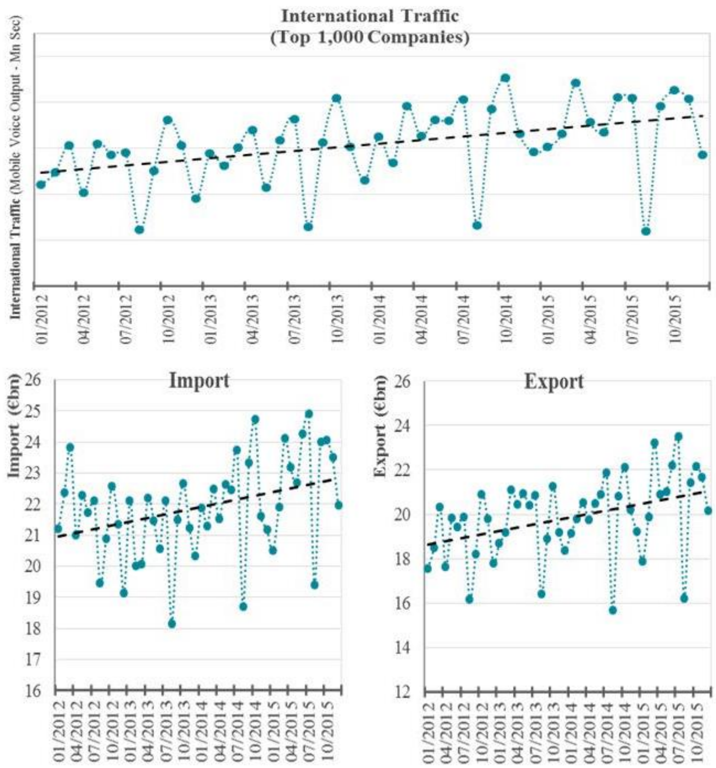
A sharp contraction in domestic demand steered the Spanish economy towards a model of economic growth based on foreign trade. The progressive re-orientation abroad led to a gradual reduction of the current account deficit and an evolution from an import-intensive foreign trade model to a more export-intensive one. During this period, not only have export volumes increased, but also the number of exporting companies.

In other words, the export base has expanded. According to the latest data available from ICEX (2016), exporting companies have gone from nearly 110,000 in 2009 to about 150,000 companies in 2015, an increase of 36 percent. However, only 31 percent of these companies were considered regular exporters, and these accounted for 93 percent of export turnover in 2015.

Drawing a parallel with what happened in the macro environment, filters were applied to the initial sample of international mobile traffic (total companies) in order to isolate "one-shot" effects or, in other words, the possible impact of international calls made by companies at a given point in time and not on a regular basis.

A sample of aggregated and anonymous data was extracted from the 1,000 companies with the greatest international traffic at the time the correlation was produced (February 2016), assuming that these top 1,000 companies constituted a representative sample of those exporting on a regular basis. A graph of the international mobile traffic time series representing the top 1,000 companies contrasted with the evolution of exports and imports—Figure 2 below—revealed something significant: behavior patterns and variations in the time series were quite similar.

Figure 2. Behavior Patterns of International Mobile Traffic:
Top 1,000 Companies by International Traffic Volume, Imports and Exports (2012-2015)



Note: Based on Telefónica's aggregated and anonymous international traffic data (outbound mobile voice) of the top 1,000 companies by international traffic volume at the time of data extraction (February 2016), and DataComex (February 2016)

It was for this reason that the international mobile traffic from the top 1,000 companies was selected as the international traffic series from which to carry out the analysis. But first, the stationarity of international traffic and imports and exports time series must be confirmed, and if it is not, a cointegration comparison must be produced. To analyze the stationarity, the Augmented Dickey-Fuller (ADF) Test was applied, showing that the traffic series and the export series are both stationary. The import series, on the other hand, is non-stationary.

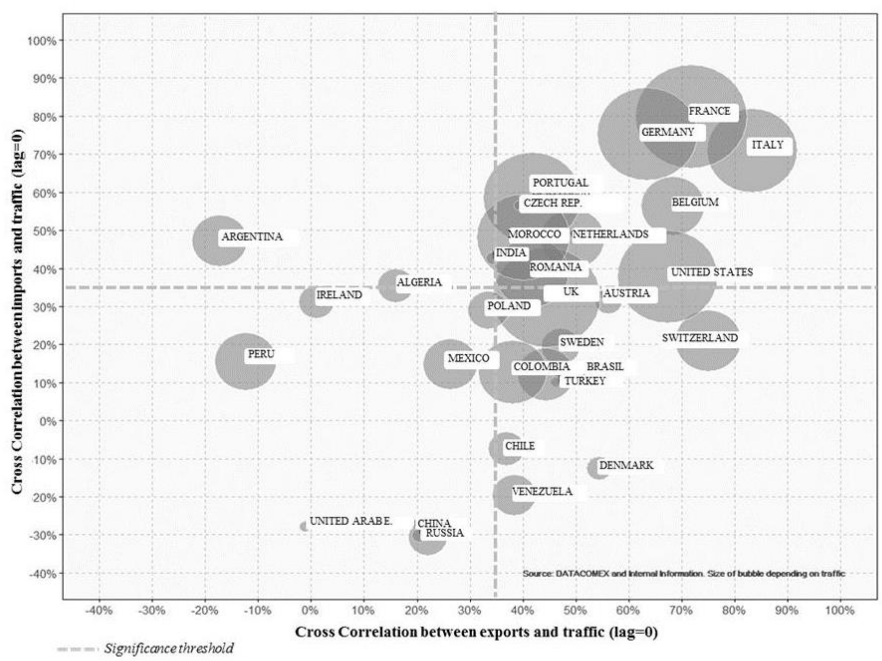
Since it is not possible to conclude the stationarity of all analyzed series, a cointegration analysis was made between the traffic series and the imports and exports series, following the Engle-Granger two-step method.

Once we were satisfied that the series were cointegrated, we applied cross-correlation analysis, the method most used to analyze the relationship between time series. Cross-correlations were produced between both time series (international traffic and trade flows) at the aggregate level by country and total aggregate level.

Following the analysis of correlations at the country level, the results were entered in a double entry graph, including cross-correlated export and import data and traffic data. This step shows that there are three groups of countries: i) those that experience higher cross-correlations between exports and imports, and traffic; ii) those that show a correlation with just one (exports or imports); and iii) countries which show a less relevant degree of correlation.

The correlation analysis conducted allows us to classify countries into four quadrants, as shown in Figure 3. The first (lower left) quadrant includes countries with a low correlation between exports and imports and international mobile traffic. A second (upper right) quadrant includes countries with a high correlation between exports and imports and international mobile traffic. A third (lower right) quadrant includes countries with a high correlation between exports and international mobile traffic (lower right quadrant). A final (upper left) quadrant includes countries with a high correlation between imports and international mobile traffic.

Figure 3. Cross-Correlations between a Country's International Traffic Level and Foreign Trade Flows



Note: Cross correlations between import-export time series and international traffic data from January 2012 to September 2014. Origin: Mobile traffic contemporaneous correlations (lag=0) (January 2012 to September 2014).
Source: Author's calculations based on 2016 data from Telefónica and Ministry of Economy and Competitiveness of Spain.

The disparity of results between countries did not allow for significant correlations for every country analyzed, so obtaining an advanced indicator from telecommunications data of imports and exports at the country level would not be feasible for all countries. This is because, while some countries, like France, Germany and Italy, have a high volume of international traffic, if significant and elevated correlations were registered for other countries, these only become important when compared to one of the trade flows (for example, Argentina and Switzerland), or are not significantly correlated with any trade flows (for example, Mexico or Peru).

After this analysis, the same exercise is conducted using the monthly series of total aggregate and anonymous international mobile traffic data (no breakdown by country of destination) of the top 1,000 companies with the greatest international traffic.

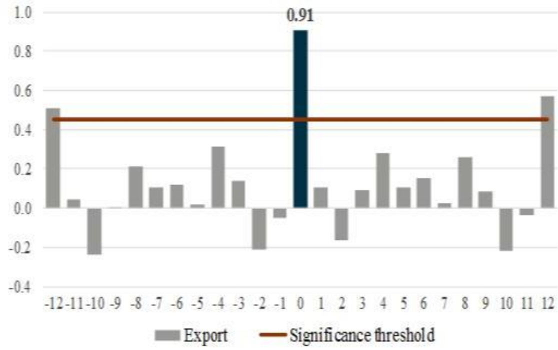
Cross-correlations were produced between the total of international traffic (outbound mobile voice) with the above criteria, and individualized import and export monthly series published by Spain's Ministry of Economy and Competitiveness, according to data compiled from the Ministry of Economy and Competitiveness of Spain (2016) and Telefónica (2016). In this last calculation, the aim is to confirm if a direct relationship exists between mobile network traffic levels and each of the trade balance calculation components (exports and imports).

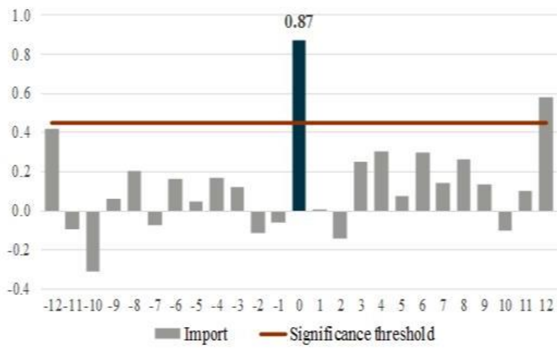
The aim is to determine whether it is possible to determine if there is a strong relationship between the variables, if the ratio is sufficiently high enough to be considered relevant and, finally, if it is greater in one case or in another. This analysis tries to demonstrate that it is possible to use data obtained from telecommunications networks / CDRs to present a fairly accurate portrait of a country's economic situation.

The results are shown in Figure 4. The correlation between international mobile traffic and exports and imports is very high and significant in both cases when there is no delay between series (lag = 0), with a higher correlation between international traffic and exports (0.91 for exports and 0.87 for imports). The analysis also shows that both series are contemporaneous.

Figure 4. Cross-Correlations between International Traffic (Outbound Mobile Voice) and Import/Export Time Series (2012-2015)

Lag (+/- 12 months)	Export	Import
-12	0.51	0.42
-11	0.04	-0.10
-10	-0.24	-0.31
-9	0.00	0.06
-8	0.21	0.20
-7	0.11	-0.07
-6	0.12	0.16
-5	0.02	0.05
-4	0.32	0.17
-3	0.14	0.12
-2	-0.21	-0.11
-1	-0.05	-0.06
0	0.91	0.87
1	0.10	0.01
2	-0.17	-0.14
3	0.09	0.25
4	0.28	0.31
5	0.10	0.08
6	0.16	0.30
7	0.03	0.14
8	0.26	0.26
9	0.09	0.14
10	-0.22	-0.10
11	-0.04	0.10
12	0.57	0.58





Note: Correlation between aggregated and anonymous international traffic data time series (outbound mobile voice) from 1,000 companies with the highest international traffic volume and external trade flows (exports and imports).

The figure also illustrates how a representative sample of the most "internationalized" companies in terms of international traffic (Top 1,000) can provide useful information on the development of foreign trade. However, the market is not static and therefore the sample should not be either. The relevance of the sample will evolve over time, adapting to market dynamics and enhancing the accuracy of the model. The first correlation between both time series is produced with a sample from the top 1,000 companies in March 2014. The correlation with trade flows (01/2012-06/2014) was equally high (90 percent for exports and 80 percent for imports).

However, maintaining a stable sample of top 1,000 companies by volume of international traffic in March 2014, and producing a correlation with the 2012-2015 import and export time series, confirmed that, although the correlations remained high and significant (0.84 for exports and 0.79 for imports), these were lower than those obtained with the current sample at the time of correlation (02/2016).

4. The Model

In this section, we discuss the model subsequently developed to include additional variables in order to corroborate the value of telecommunications variables in predictive models of economic activity.

Having demonstrated a strong relationship between the behavior of variables in international mobile traffic and export and import of goods over a certain period of time, this amplified model reinforces the initial results and demonstrates that it is possible to construct an explanatory and medium-term predictor model of foreign trade flows from telecommunications variables.

The model is based on a robust methodology that, as well as including international mobile traffic as an explanatory variable of trade flows, also introduces new explanatory variables such as the exchange rate, a calendar (working days) and ARIMA (trend and seasonality).

Forecasts are made based on an historic monthly minimum over three years. As there are limitations in the available historical monthly traffic (from 01/2012), the model focuses on predictions for 2015. It also specifies the model and the results obtained. It is completed with an analysis of the degree of forecast adjustment, contrasted with those forecasts made by the model with real foreign trade data published by the Ministry of Economy and Competitiveness of Spain (2016).

4.1 Model Specification

The dependent variables are exports and imports. By using the same independent variables for two trade flow variables, the model is specified as follows:

Log(Flow) = parameter1*Traffic+parameter2*Calendar+parameter3*Exchangerate+ARIMA
where

- Flow (dependent variable): Refers to the foreign trade flow: Exports and Imports.
- Calendar (independent variable): Refers to the number of working days per month.
- Exchange Rate (independent variable): Input into the model as the monthly average of the daily exchange rate.
- Traffic (independent variable): Refers to anonymous and aggregated mobile voice traffic from Spain taken from 1,000 companies with the highest level of

international traffic. Input into the model as monthly traffic divided by the number of working days per month: Log (Monthly traffic / Number of working days per month).

The model is based on a hierarchic structure, in that the total is calculated as the sum of defined areas, and areas as the sum of defined countries.

• Total as sum of defined areas.

• The defined areas are: European Union (UE27), Latin America, Persian Gulf, Asian Tigers, Southeast Asia, North America, Unclassified, Rest.

• A selection of 76 countries from sum of areas to which they belong.

It follows the model for imports and for exports as specified:

$$\text{Log(Imports)t} = 0.3323*\text{Traffic}_t+0{:}0488*\text{Calendar}_t+0.1544*\text{Exchangerate}_t+0.4924*a_{t-1}+0.6095*a_{t-12}$$

$$\text{Log(Exports)t} = 0.2942*\text{Traffic}_t+0.0455*\text{Calendar}_t+(-0.182)*\text{Exchangerate}_t+0.5714*a_{t-1}+0.5628*a_{t-12}$$

Error is considered by determining the degree of model fitting with Mean Absolute Percentage Error. This is calculated using two steps: i) absolute relative error is calculated for each month: [Actual-Forecast]/Actual; ii) the reported error is the average of those previously estimated.

To determine whether the proposed model is satisfactory, the autocorrelations of estimated errors are studied. A process is defined as white noise if each observation has a zero mean, the same finite variance, and the correlation between observations is zero (González and Lobato, 2003).

Box-Pierce-Llung statistic criteria are proposed in order to determine if the above is true. In this test, the null hypothesis suggests the residuals are “white noise.” If the probability of rejecting the null hypothesis is < 5 percent, it is accepted that the residuals are white noise. The result for the global model that used this applied method is that the probability of the rejection of the null hypothesis is 0 percent for imports and 2 percent for exports. That is to say, the null hypothesis is accepted with the conclusion that the proposed model is well-specified.

The assessment of foreign trade flows throughout the specified model is carried out within two periods: in May 2015 and again in January 2016. In May 2015, forecasts for 2015 are produced (with actual foreign trade data up to April 2015 and with international traffic data until May 2015).

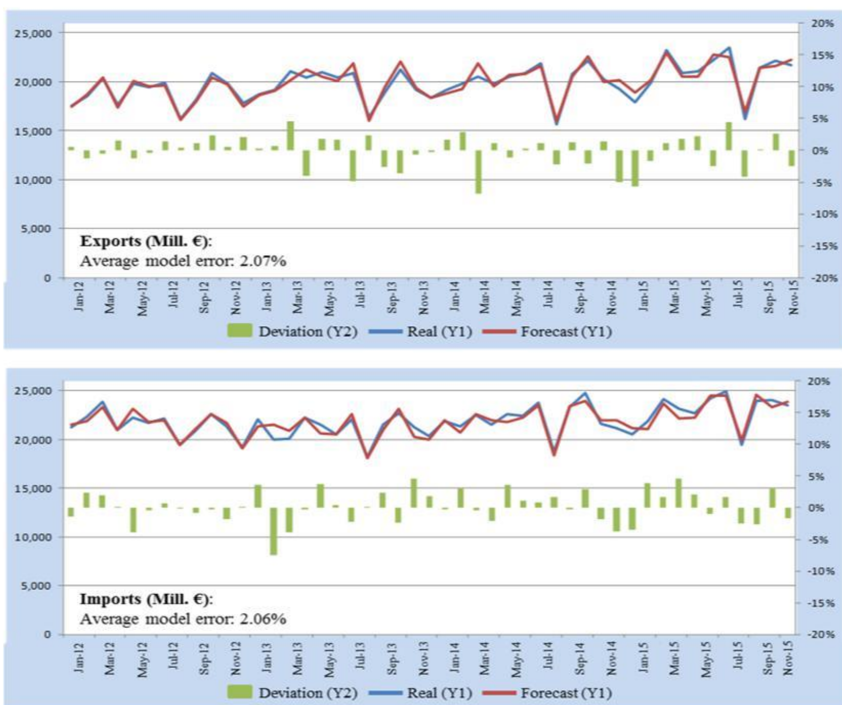
In January 2016, full-year forecasts for 2015 are produced (with actual foreign trade data to November 2015, inclusive, and from international traffic data to December 2015, inclusive).

The main results of the model will then be completed with an analysis of the degree of prediction fitting by comparing the actual performance of the export and import of goods published by Spain's Ministry of Economy and Competitiveness in DataComex with the prediction produced by the specified model. These will also be contrasted in parallel with comparable predictions made by other agencies.

4.2 Results

The model shows a goodness of fit of the expected values of trade flows. The average model error is 2.07 percent for exports and 2.06 percent for imports (Figure 5).

Figure 5. Predictive Model Fit: Average Model Error



Note: Model error as the deviation average [(Actual-Forecast)/Actual] in absolute value.
Source: Author's estimation.

The model also helps to verify the influence of international traffic variables in explaining foreign trade variables. Table 2 shows each of the variable percentages considered while explaining export or import variables. As can be observed, the influence of international mobile traffic is very high in both cases.

Table 2. Percentages Considered While Explaining Export or Import Variables

	TM International	Calendar	Exchange Rate	ARIMA
Exports	92.40%	6.00%	-1.90%	3.60%
Imports	86.00%	6.20%	4.30%	3.50%

The model also allows us to estimate that a 1 percent increase in international mobile traffic leads to a 0.4 percent increase in the export and import of goods. That is to say, the elasticity of exports and imports to changes in traffic for the overall model is around 0.4.

As mentioned, forecasts for the export and import of goods for 2015 are produced for two different points in time: in May 2015, with an update in January 2016, that produces a full set of international traffic data by the end of the year, while Spain’s Ministry of Economy and Competitiveness had yet to publish final foreign trade data for 2015 (Ministry of Economy and Competitiveness of Spain, 2016).

In May 2015, the forecast was made with actual international traffic data and exchange rates to May 2015 and foreign trade data available until April 2015. The forecast produced by the model for 2015 called for annual export and import growth rates of 5.4 percent and 4.6 percent, respectively.

By early January 2016, monthly data for international traffic for 2015 were available for the entire year, and trade flows were available through November 2015 (inclusive). The high ratio between international traffic and trade flows allows the model to conclude that the export and import of goods at the close of 2015 will have increased by 4.0 percent and 3.4 percent respectively, compared to 2014.

Table 3 shows an analysis of the goodness of fit for forecasts produced from the model (in May 2015 and at the end of 2015, before the official publication of trade data for 2015 by DataComex) compared with actual import and export trends, as published by Spain’s Ministry of Economy and Competitiveness.

Table 3. Assessing Foreign Trade Forecast Fit

Exports				
2015	%	Mil. Euros	%	Mil. Euros
Real Development	4.0	250,241	-	-
2015 Traffic Forecast Model (May 2015)	5.4	253,573	1.4	3,332
2015 Traffic Forecast Model (End 2015)	4.0	250,205	-0.01	(-36)
Imports				
2015	%	Mil. Euros	%	Mil. Euros
Real Development	3.3	274,415	-	-
2015 Traffic Forecast Model (May 2015)	4.6	277,772	1.3	3,357
2015 Traffic Forecast Model (End 2015)	3.4	274,586	0.1	170

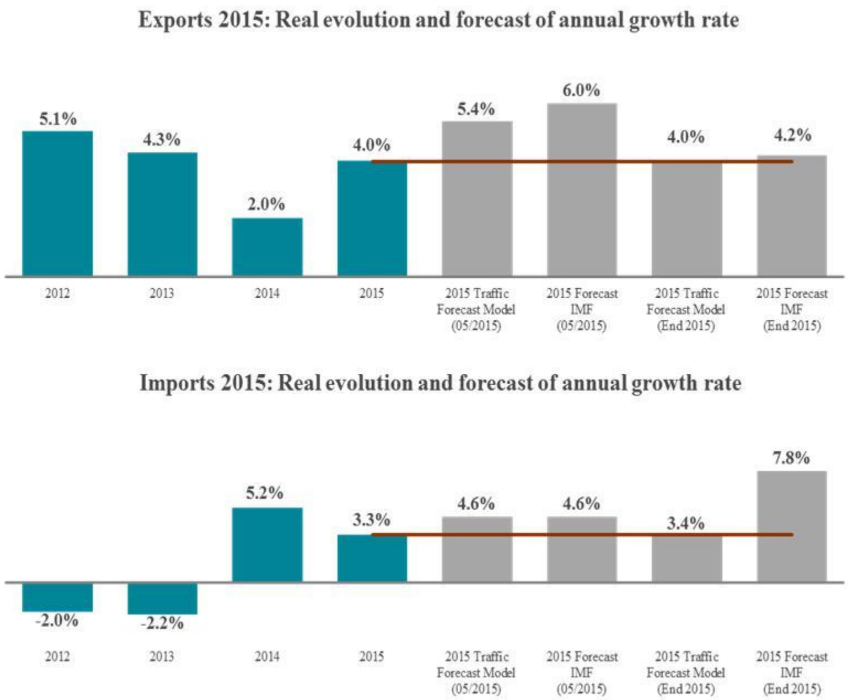
From the results, we can observe that the effectiveness of using international traffic variables in models that forecast economic activity is confirmed. Specifically, the ability of the model to predict immediately (nowcasting) based on international traffic has been confirmed. Given the availability of near-real time international traffic data and the high correlation between this variable and the foreign trade flow variable, it can predict with a considerable degree of accuracy the development of trade flows in the short term. The effectiveness of nowcasting is reflected in Table 3. It shows how the short-term deviation of the 2015 forecast is practically zero (-0.01 percent for exports and 0.1 percent for imports) and how it can estimate export and import trends before their official publication in DataComex by the Ministry of Economy and Competitiveness of Spain (2016).

The ability of international traffic variables to predict in the medium term (forecasting) is also supported. The model yields an average error level of around 2.07 percent, which provides a good fit. This is also reflected in Table 3, which demonstrates the 2015 forecast of export and import trends from traffic data up until May 2015. In medium-term predictions, a deviation of 1.4 percent (exports) and 1.3 percent (imports) with respect to real change is noted.

The forecast produced for the traffic model has also been contrasted with comparable predictions made by other agencies, taking into account that model predictions are based on trade flows in goods. For example, the International Monetary Fund (IMF) produces foreign trade forecasts in goods, as well as in goods and services, and we find that the forecast of trade in goods by the IMF (2015) is comparable with the forecast model. Bearing in mind that 2015 IMF projections were made available in May 2015, with regards to the export and import of goods, they predict the growth of Spanish exports by 6 percent and imports by 4.6 percent.

In October 2015, the IMF updated its projections for this indicator, reviewing the evolution of both flows for 2015, establishing export and import growth at 4.2 percent and 7.8 percent, respectively. Figure 6 shows the various forecasts and reveal how forecasts produced by the traffic model provide a good fit.

Figure 6. Comparison of Forecasts for Export and Import of Goods



Source: Author's calculations based on traffic model data from Telefónica (2015) and IMF (2015).

With this model, and contrasted with its degree of adjustment, the advantages of including international traffic variables in the forecast model for trade flows can be observed. First, traffic variables minimize model error, improving the forecast with respect to the same model without traffic. Model error shows an average improvement of 1.7 percent for imports and 1 percent for exports when the international mobile traffic variable is introduced.

Second, the international mobile traffic variable is available virtually in real time (the 10th of the following month). As mentioned earlier, this means the variable that is available almost two months before the official release of foreign trade data (DataComex) by the Ministry of Economy and Competitiveness of Spain (2016).[1] In conjunction with the strong relationship between international mobile traffic and trade flows, the variable provides a robust indicator for the nowcasting of economic activity.

Finally, traffic variables maximize the success of the first month's forecast. With this variable, a better fit is obtained in the first month of the forecast, and consequently an improved forecast is obtained in the medium term.

5. Conclusion

As this research shows, a strong correlation exists between international telecommunications traffic and foreign trade flows (91 percent with exports; 87 percent with imports). This demonstrated that the use of a representative sample of the most "internationalized" companies in terms of global mobile data traffic may provide useful information on the development of foreign trade. The relevance of this sample can evolve over time, adapting to market dynamics, and enhancing the accuracy of the model.

Furthermore, it is possible to obtain a very effective indicator for the immediate prediction or nowcasting of foreign trade (contemporaneous correlation and data available in near-real time), which facilitates the identification of changes in the behavior patterns of economic variables as they happen.

In fact, this relationship could be considered a "leading indicator" of economic activity, as Spain's Ministry of Economy and Competitiveness publishes monthly foreign trade data (imports and exports) with a delay of almost two months, while the international telecommunications traffic data of companies is available in near-real time (by approximately the 10th of the following month). The ability to see the evolution of indicators almost two months ahead of the publication of official data offers substantial benefits for those who need to analyze and act on import and export data, particularly in light of the new prominence of international trade as a driver of economic growth in Spain.

[1] Spain's Ministry of Economy and Competitiveness foreign trade data are published with a delay of approximately one month and three weeks

Ultimately, the hypothesis that it is possible to predict economic activity from telecommunication variables both in the short and medium term has been clearly established, and those variables' potential for setting the predictive accuracy of nowcasts has also been demonstrated. In addition, the contribution of this paper opens many avenues for future empirical research on trade using big data techniques and models emphasizing the role of networks in international trade, particularly in the context of increasing complex production and distribution networks around the world, which are data and communications intensive. The availability of high quality databases will be required for researchers to advance in this new field.

References

Bank of Spain. 2015. “Boletín Estadístico–Balance por Cuenta Corriente.” Madrid, Spain: Bank of Spain.

Bogomolov, A. et al. 2014. “Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data.” In: *Proceedings of the 16th International Conference on Multimodal Interaction*. New York, United States: Association for Computing Machinery.

Consejo Empresarial para la Competitividad. 2013. “España: Un País de Oportunidades.” Madrid, Spain: Consejo Empresarial para la Competitividad.

Coscia, M., and R. Hausmann. 2015. “Evidence That Calls-Based and Mobility Networks Are Isomorphic.” *PLoS ONE* 10(12): e0145091.

Dadhich, D., and V. Gupta. 2015. “Big Data: A Gold Mine in Telco’s Backyard.” Boston, United States: Arthur D. Little. Available at: http://www.adlittle.com/sites/default/files/viewpoints/ADL_BigDataGoldMineforTelcos.pdf

Einav, L., and J. Levin. 2013. “The Data Revolution and Economic Analysis.” In: Josh Lerner and Scott Stern, editors. *Innovation Policy and the Economy*. Volume 14. Chicago, United States: National Bureau of Economic Research and University of Chicago Press.

Frias-Martinez, V. et al. 2013. “Forecasting Socioeconomic Trends with Cell Phone Records.” In: *Proceedings of the 3rd ACM Symposium on Computing for Development*. New York, United States: Association for Computing Machinery.

González, M., and I. Lobato. 2003. “Constraste de Autocorrelación.” *Gaceta de Economía* 7(14): 41-57.

Halevy, A., P. Norvig and F. Pereira. 2009. “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems* 24: 8-12.

Hassani, H., and E.S. Silva. 2015. “Forecasting with Big Data: A Review.” *Annals of Data Science* 2: 5-19.

Hausmann, R. et al. 2011. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, United States: Puritan Press.

ICEX. 2016. “El Exportador: En Cifras.” El Exportador: Revista para la Internacionalizacion. Madrid, Spain: ICEX. Available at: www.icex.es.

International Monetary Fund (IMF). 2015. “World Economic Outlook Database.” Washington, DC, United States: IMF.

Lendle, A. et al. 2012. "There Goes Gravity: How eBay Reduces Trade Costs." Washington, DC, United States: World Bank.

Louail, T. et al. 2014. "From Mobile Phone Data to the Spatial Structure of Cities." *Scientific Reports* 4: Article 5276. Available at: https://www.nature.com/articles/srep05276.

Ministry of Economy and Competitiveness of Spain. 2016. "DataComex Database." Madrid, Spain: Secretary of State of Business.

Moumni, B., V. Frias-Martinez and E. Frias-Martinez. 2013. "Characterizing Social Response to Urban Earthquakes Using Cell-phone Network Data: The 2012 Oaxaca Earthquake." In: *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing, Adjunct Publication*, New York, United States: Association for Computing Machinery.

Organisation for Economic Co-operation and Development (OECD). 2014. "OECD Economic Surveys: Spain 2014." Paris, France: OECD Publishing.

Sauri, D. 2015. "Geografía de la Crisis Económica en España." *Economic Geography* 91: 387-388.

Simoes, A., and C.A. Hidalgo. 2011. "The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development." In: *Proceedings of the 17th AAAI Conference on Scalable Integration of Analytics and Visualization*. Palo Alto, United States: Association for the Advancement of Artificial Intelligence.

Telefónica. 2016. "Global Index of Digital Life." Madrid, Spain: Telefónica.

World Bank. 2015. *Global Economic Prospects, June 2015: The Global Economy in Transition*. Washington, DC, United States: The World Bank Group.

World Bank. 2016. "Exports of Goods and Services." Washington, DC, United States: World Bank. https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS

Wu, L., and E. Brynjolfsson. 2015. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales." In: A. Goldfarb, S.M. Greenstein and C.E. Tucker, editors. Economic Analysis of the Digital Economy. Chicago, United States: National Bureau of Economic Research and University of Chicago Press.